Advances in Intelligent Systems and Computing

George Yang (Ed.)

Proceedings of the 2012 International Conference on Communication, Electronics and Automation Engineering

Springer

Editor
Dr. George Yang
Department of Engineering Technology
Missouri Western State University
USA

Preface

Welcome to the 2012 International Conference on Communication, Electronics, and Automation Engineering and the beautiful city of Xi'an, China.

First, I like to express my sincere appreciation of the conference organizers for their hard work. Their efforts and services made this conference possible. As I understand that this is the first time for this group of organizers to work together to bring about the conference. If there are any places that need improvements, please, as I was told, provide them with your constructive suggestions.

There are tremendous amount of developments every second in the fields of Communication, Electronics, and Automation Engineering research. An academic conference is a valuable and efficient platform for us researchers to exchange ideas, to meet new people, to develop new friendship and collaboration. I am really glad to see this time we have so many world-class researchers to come to this conference to present your cutting-edge accomplishments. I firmly believe that we can make the conference a premium event in our research area. Hopefully, with the contribution of everyone, the conference will become an annual event with widespread impacts.

The conference is sponsored by Xi'an Technological University, and Shaanxi new Network and Monitoring Control Engineering laboratory. Located in Xi'an, China, Xi'an Technological University is a multi-discipline technology-focused university currently with 16,000 undergraduate and graduate students as well as 1400 faculty and staff members. Shaanxi new Network and Monitoring Control Engineering laboratory is a state funded key lab for Shaanxi province currently spearheading the research in network engineering. This conference also has the technical sponsorship by Missouri Western State University. Located in St. Joseph, Missouri, USA, Missouri Western State University is one of the Missouri State public universities with majors in liberal arts, business and professional studies. It offers over 100 undergraduate and graduate degree programs.

The papers in the proceedings were selected from over 600 high quality submissions. All papers were blind-reviewed by at least three experts in this field. The acceptance rate is about 30 percent. We were fortunate to have such a strong interest in this subject.

Finally, I hope you find the conference informative and inspirational to your research. Please enjoy the discussions with your colleagues.

Dr. George Yang
Professor and Chair
Department of Engineering Technology
Missouri Western State University, St. Joseph, USA

Contents

Track 1: Automation Control

Track 2: Data Mining and Statistics

Track 3: Simulation and Mathematical Modeling

Track 4: Human Factors and Cognitive Engineering

Track 5: Web Technology

Track 6: Optimization and Algorithm

Track 7: Network Communications

Approach towards Left-Turning Vehicles Trajectory at the Intersection Based on Video Technology

Jinglong Wu[1,2], Chao Xue[1,2], Ziran Zhao[1,2], and Bohang Liu[1,2]

[1] Shijiazhuang Tiedao University
[2] Traffic Safety and Control Lab of Hebei Province, Shijiazhuang 050043 China

Abstract. The significance of left-turning vehicles trajectory studies lies in application of intersection capacity and safety of analysis. However, data reliability plays an important role in the ability to the understand the characteristics and microscopic modeling of left-turning vehicles at the intersection. This study takes advantage of observation-based video data by using self-developed software to retrieve reliable vehicle trajectory data. In this paper, we propose a method for coordinate transformation between the video and the ground. The result shows that our method provides an effective solution to obtain traces of left-turning vehicles at the intersection.

Keywords: Intersection, Left-turning vehicles, Coordinate transformation.

1 Introduction

The study of left-turning vehicles trajectory at intersections is used for a variety of purposes: most frequently to estimate the traffic safety at intersections and also to identify geometric, environmental, and operations factors that are associated with accident[1]. Previous researches related to left-turning vehicles trajectory rely heavily upon artificial means. But in real scene, we can't get the traces that we wanted.

Lately, video detection is increasingly being deployed instead of inductive loops at intersections. We can obtain traffic data more conveniently than before by video detection. Through image processing and related arithmetic, we can extract the left-turning vehicles trajectory from video images. This paper adopted the technology of video to deal left-turning vehicles data. Based on the analysis of the coordinate transformation between the ground and the video, we can get the track of left-turning vehicles.

2 Theoretical Analysis

Targeting at the traffic feature of left-turning vehicles, it is suggested to install dedicated signal lights for left-turning vehicles. however, passing a intersection without dedicated signal lights for left-turning vehicles, there will be three ways for left-turning vehicles to pass the road: first, left-turning vehicles can pass the road at the initial green time, Second, they can pass the conflict point when the going-straight vehicles happened to pass the road. Third, they can pass the road at the rest of green signal and yellow signal after all the opposing through vehicles pass the conflict point[2].

According to the regulations of China's road traffic laws:" vehicles were allowed to go straight and turn left at green signal, furthermore, left-turning vehicles should not impede the passage of opposing through vehicles." Therefore, left-turning vehicles only turn left at the initial green signal and at the time of opposing through vehicles that have not yet reached conflict point. Vehicles can also pass the road when waiting near the conflict point until the time-highway of going-straight vehicles was allowed to across. How to organize and control the left-turning traffic flow reasonably is the key problem for management of traffic order and improving the traffic capacity at intersection. Several factors were found to affect the left-turning vehicles. These factors include vehicles types, vehicles conditions, environment of intersections, weather conditions and drivers psychology, etc.

3   Preliminary Analysis

As mentioned above, if left-turning vehicles trajectory can be measured manually through the related devices[3], it may has the positive effect on the intersection capacity and safety of analysis. Found by comparing, the data of video detection records is full and rich. Consequently, we select the video data as data source to get the track.

Data acquisition flow diagram is shown in Fig.1.
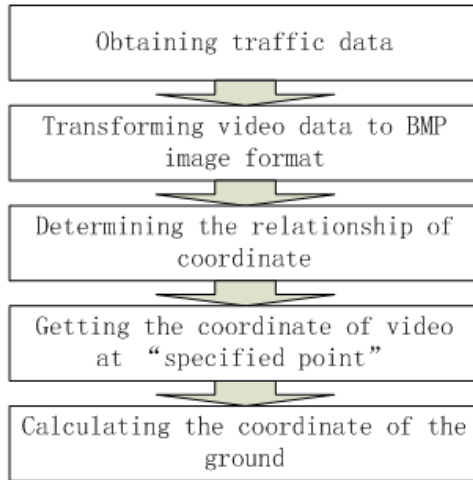
Fig. 1. Data acquisition flow diagram

This graph shows the process of data acquisition. First, through filed observation, we can get the video data, then we need to transform the video data to BMP image format[4]. It is important to ensure that data background are the same. This paper get BMP format through "Image Transfer and VirtualDubMod" software.

4   Methodological Approach

In this section, the methodological approach taken to account for coordinate transformation in the video data. It is important to note that common software could not be used to deal with coordinate transformation between the video and the ground. Our self-developed software system can record the image plane coordinates of the designated point and transform dimensional coordinates into the ground coordinates.

The relationships between the image coordinates and ground coordinates is shown in Fig.2.
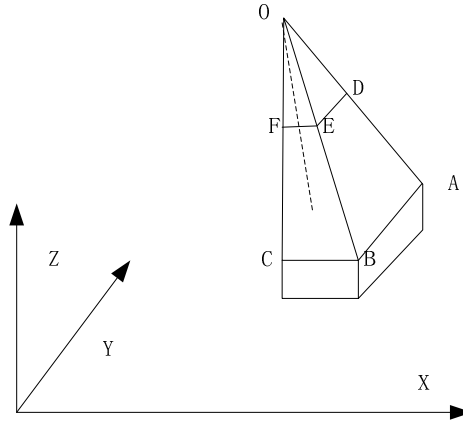
Fig. 2. The relationships between image coordinates and ground coordinates

As shown in Fig.2, O is the photography center, A is a point on the ground,. $AB = Y_A - Y_O$, $BC = X_A - X_O$, $OC = Z_A - Z_O$. According to the similar triangle calculations[5], we can get the following equation:

$$\frac{X}{X_A - X_O} = \frac{Y}{Y_A - Y_O} = \frac{Z}{Z_A - Z_O} = \frac{1}{\lambda} \tag{1}$$

Based on the related theory of coordinate transformation, the following equation is workable.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = R \begin{bmatrix} x \\ y \\ -f \end{bmatrix} = \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{bmatrix} \begin{bmatrix} x \\ y \\ -f \end{bmatrix} \tag{2}$$

According to the equations of (1) and (2), we get the other equations:

$$\begin{cases} x = -f\dfrac{a_1(X_A - X_O) + b_1(Y_A - Y_O) + c_1(Z_A - Z_O)}{a_3(X_A - X_O) + b_3(Y_A - Y_O) + c_3(Z_A - Z_O)} \\ x = -f\dfrac{a_2(X_A - X_O) + b_2(Y_A - Y_O) + c_2(Z_A - Z_O)}{a_3(X_A - X_O) + b_3(Y_A - Y_O) + c_3(Z_A - Z_O)} \end{cases} \tag{3}$$

Based on the above equations of (1), (2) and (3), we can get the following equations:

$$\begin{cases} X = \dfrac{a_1 x + a_2 y + a_3}{a_4 x + a_5 y + 1} \\ Y = \dfrac{a_6 x + a_7 y + a_8}{a_4 x + a_5 y + 1} \end{cases} \tag{4}$$

Consider the two equations (1): shows the correspond relationship between the ground coordinate and the coordinates of the points. After transfer a series of point of Vehicles coordinates in image plane to ground coordinates plane, track of left-turning vehicles on the ground can be got[6-8].

5   Application Analysis

The video data used in this paper was obtained in shijizhuang. We selected 3.2s data and got 128 pictures after the process of the Virtual DubMod software, then we selected four designated points: 1(0,0), 2(6.75,0), 3(6.75,5), 4(0,5), The designated points which we selected are presented in Fig.3.
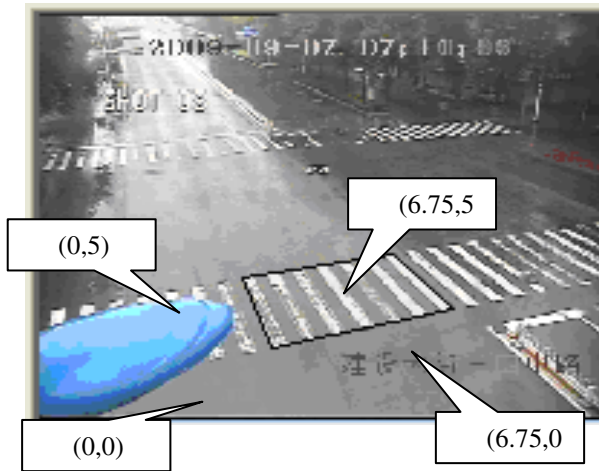
Fig. 3. Designated points

The transformation relationships between ground coordinates and image coordinates is shown in Fig.4.
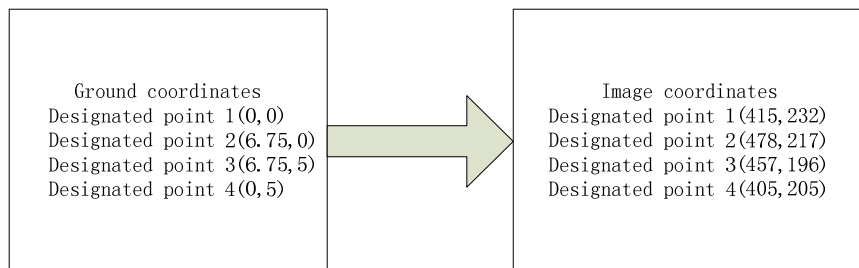
Fig. 4. The transformation relationships between ground coordinates and image coordinates

From the analysis of above, we put image coordinates into the corresponding equation, the actual left-turning vehicles trajectory can be calculated. The situation of survey is shown in Fig.5. The image coordinates of the trajectory as shown in Fig.6. The ground coordinates of the trajectory as shown in Fig.7. In order to test the accuracy of the system, the four designated points coordinate was obtained by mouse as shown in Fig.5, then put them into the conversion equation. Based on the calculation, we put the deviation between the value calculated in this paper and the actual coordinates into Tab.1. Generally, if the deviation is less than 5cm, this method proves the accuracy is reliable.

Fig. 5. The situation of survey
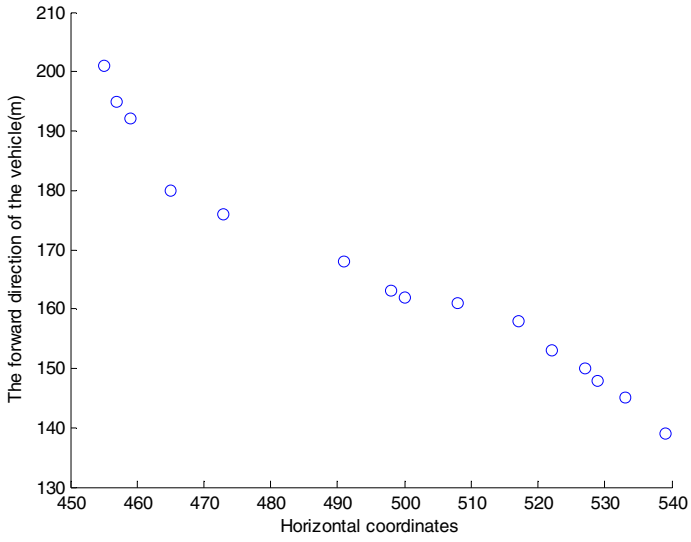
Fig. 6. The image coordinates of the trajectory
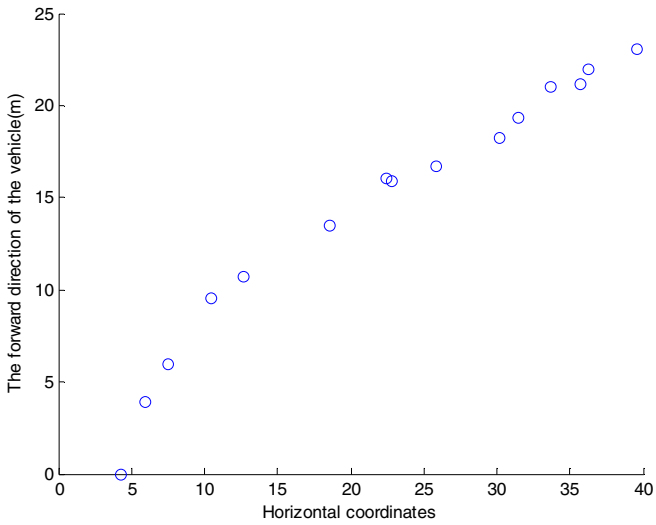
Fig. 7. The ground coordinates of the trajectory

Table 1. Error analysis of coordinates

Reference Points	Actual x(m)	Actual y(m)	Calculation x(m)	Calculation y(m)	Error of x(m)	Error of y(m)
1	6.75	0.00	6.78	0.00	0.03	0.00
2	6.75	6.00	6.75	6.00	0.00	0.00
3	6.75	10.00	6.80	10.03	0.05	0.03
4	6.75	20.00	6.79	20.12	0.04	0.12

6   Summaries

In this paper, based on the principles of photography of video technology, through self-developed system, the left-turning vehicles trajectory can be extracted from video data. Further research needs to be conducted to deal with video data in inclement weather conditions, and to seek out other possible issues. The available software and related arithmetic remains relatively difficult, and so improvements in this area also welcomed. The approach described in this paper is generalizable to other vehicles trajectory.

References

[1] CyPers, L., Lemaire, F., Vervenneetal, P.: Traffic Engineering and Control. CCATS:the Image Processing-based Traffic Senso 31(6), 371–378 (1990)

[2] Chowdhury, D., Santen, L., Schadschneider: Statistical physiscs of vehicular traffic and some related system. Phys. Rep. 329, 199–329 (2000)

[3] Once, M., Ohba, K.: Distal Image Analysis of Traffic Flow. In: Proceedings of The International Joint Conference on Pattern Recongition, Coronado, pp. 803–804 (1976)

[4] MichaloPoulos, P.G., Jaeobson, R., Anderson, C.A.: Traffic Engineering and Control. Field Deployment of AUTOSOPE in the FAST-TRACATMS/ATIS Programme 33(9), 475–483 (1992)

[5] Bando, M., Haseboe, K., Nakayama, A., Shibata, A., Sugiyama, Y.: Dynamatic model of traffic congestion and numerical simulation. Phys. Rev. E 51, 1035–1042 (1995)

[6] Ma, C.S.-Y., Chen, S.-Y., Liang-Gee: Efficient Moving Object Segmentation Algorithm Using Background Registration Technique. IEEE Trans. (2002)

[7] MichaloPoulos, P.G., Jaeobson, R.D., Anderson, C.A.: Traffic Engineering and Control. Automatic Ineident Detection Through Vieo Image Processing 34(2), 66–75 (1993)

[8] Gunilla, B.: Distance Transformations in Digital Images (1986)

Recovery Time Analysis of a Distributed Redundancy Protocol

Lu Han-Rong, Wang Wei-Jiang, Zheng Feng, and Lu Li

Air Force Radar Academy, Wuhan, China
Zhejiang University, Hangzhou, China

Abstract. Service availability is a significant feature in industry network systems that demands strict real-time constraints and stable communication interfaces. Failure detection and recovery mechanisms are indispensable to ensure the available communications. This paper analyzes a new distributed redundancy protocol for industry Ethernet. The working mechanism of the protocol is studied and the influences of the network environment, the clock synchronization error and the timing parameters on the recovery times are analyzed. Models for computing the recovery times are established.

Keywords: real-time Ethernet, distributed redundancy protocol, failure detection, recovery time.

1 Introduction

Control systems based on Ethernet are widely used in industrial plants because Ethernet is simple and cost effective. Industrial systems require stability, therefore measurement and control messages must be transmitted in an available and reliable communication facility. And industrial Ethernet systems have more strict real-time constraints compared to commercial Ethernets.

As is well known Ethernet uses Carrier Sense Multiple Access/Collision Detect (CSMA/CD) technology in the Data Link Layer (DLL) defined by ISO, where devices have to wait an uncertain time before beginning to transmit the messages. The resulting random end-to-end delay is unacceptable in an industrial plant so that the DLL must be modified or enhanced in real-time environments. There are several types of technologies solving the problem which are defined in IEC 61158 as Profinet, Ethernet/IP, EPA etc [1, 2].

Link failure and switch device failure are common in Ethernet and will bring about a series of problems such as data error, data loss, and messages being transmitted out of order. These faults will affect the plant stability seriously, thus failure detection and recovery mechanisms are mandatory in industry Ethernet environments. For a real-time constrained system, the recovery mechanism should be effective and fast enough. An important criterion to determine the strictness of the real-time constraints that a system requires is the time it can run without a working control system, and this time is defined as the grace time. If a failure occurs, the network communication system must detect the error and repair it within the grace time [3].

Using redundant links is an effective way to realize self-recovery. Fundamental topologies like ring and mesh can provide redundant links. In IEC 62439 FDIS five different solutions for redundancy are proposed: Media Redundancy Protocol (MRP) [3,4], Parallel Redundancy Protocol (PRP)[5], Cross-network Redundancy Protocol (CRP), Beacon Redundancy Protocol (BRP) and Distributed Redundancy Protocol (DRP) [6].

DRP is designed for the ring topology that is simple in structure and provides one efficient redundancy route [7]. DRP is based mainly on a distributed failure detection mode. This paper performs a timing analysis on DPR. The working mechanism is studied and the factors influencing the recovery time are analyzed. Time models to compute the recovery times are derived.

This paper is organized as 4 parts. Section 2 is an introduction of DPR. In section 3, the timing parameters are listed and analyzed and methods to estimate the recovery times are established. And section 4 presents some conclusions.

2   An Overview of DRP

DRP defines a framework for describing the operational behavior of the switch devices in a ring topology to detect a single network failure (link failure or switch failure) and to recover from it within a time limit. The switch devices in a DRP ring network are equal parties and have two ports each. Every ring port may be in one of two states: blocking state and forwarding state. All switch devices are synchronized using IEEE 1588 Precision Time Synchronization Protocol (PTP)[8].The scheduling of switch devices in DRP is done in intermittent communication cycles called macrocycles. Within a macrocycle, each switch device takes turns to send a Ring Check frame to detect a failure on the ring. Besides, each switch device sends a Link Check frame to its immediate neighbor to detect an adjacent link failure. If ether a Ring Check frame or a Link Check frame is not received in time by a certain switch device, it then multicasts a Link Alarm frame to inform the other devices of a failure. When the switch device which has one of its ports blocked receives the Link Alarm frame, it will change that port state into forwarding so that the network turns into a linear topology and the network has been thus recovered.

Failures in a DRP network are of three types: transmission medium failure, MAC layer failure or switch device failure, and DRP layer failure. In a macrocycle T_{cycle}, one certain switch device sends a Ring Check frame to be forwarded around the ring at time $t_{RingCheck}$ to test the ring state. If this frame is received by the switch device which sent it, the frame must have gone through all switch devices in the ring and the network is all right. Otherwise if this frame is not received by the originating switch device within a predefined time $T_{ex}(RC)$, the network must have gone wrong. Each switch device also sends a Link Check frame to its two neighbors at time $t_{LinkCheck}$ to test the operation state of all links and switch devices and a Link Alarm frame will be broadcasted by the switch device that has detected a failure. A MAC layer failure will be detected by ether a Ring Check frame or a Link Check frame. Some blocked switch device will be de-blocked or some non-blocked switch device will be blocked to keep the network

consistent (to recover from the failure) after receiving a Link Alarm frame. The case of a transmission medium failure is similar to switch device failure. Nonetheless, the case is quite different with a DRP layer failure. The switch device with a DRP layer failure is equivalent to a common functional switch device without failure detection. For the switch device with a DRP layer failure, all messages are stored and forwarded. It can be seen as a delay buffer between two DRP switch devices. So DRP layer failure does not affect the actual topology of the ring network. In other words, the system is still valid if the ring contains a switch device that does not support DRP.

3   Recovery Time Analysis

3.1   Recovery Time Models

When a failure is detected by a Ring Check frame the failure occurs in the interval between the Ring Check frame and the Link Check frame in a communication macrocycle. The Ring Check frame is sent after the failure in t_{RW}. The switch device which sent the Ring Check frame has to wait for t_{RW} until the failure has been confirmed. Then it has to taken t_{WA} to send the Ring Alarm frame. The Link Alarm is forwarded in the network until it is received by the switch device which has its port blocked. The state of the blocked port will be changed into forwarding and the recovery process is completed. The last step takes time t_{AR}. The whole process is shown in Fig.1. The recovery time is

$$T_{rc1} = t_{FR} + t_{RW} + t_{WA} + t_{AR}. \tag{1}$$
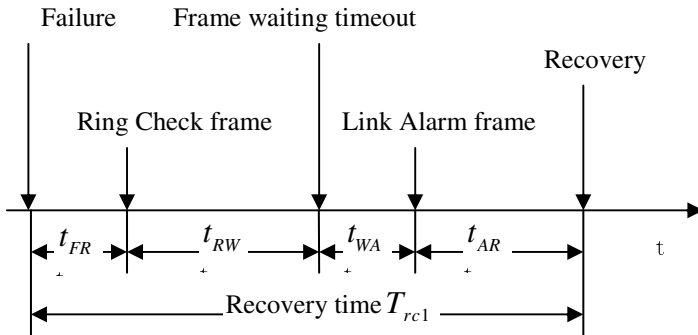
Fig. 1. Recovery time while failure is detected by Ring Check

When a failure is detected by a Link Check frame the failure occurs in the interval between the Link Check frame and the Ring Check frame. The Link Check frame is sent simultaneously by all switch devices, so the time to wait for the Link Check frame is much shorter than the time to wait for the Ring Check frame in the above case. The

Link Check frame is sent after the failure in t_{FL}. The switch device which sent the Link Check frame has to wait for t_{LW} until the failure has been confirmed. Then the process is similar to the above case and the recovery time is

$$T_{rc2} = t_{FL} + t_{LW} + t_{WA} + t_{AR}. \tag{2}$$

Failures would be detected mostly by this way so T_{rc2} is a typical parameter to estimate the recovery time of DRP.

Sometimes an inter-switch device link failure like a line being cut off can be detected by the physical layer. The Link Alarm frame is sent after the failure in t_{FA}, then it takes t_{AR} to complete the recovery process:

$$T_{rc3} = t_{FA} + t_{AR}.$$

3.2   Parameter Evaluation

To get the exact recovery time we need to compute the parameters t_{FR}, t_{RW}, t_{WA}, t_{AR}, t_{FL}, t_{LW} and t_{FA}.

Designed for industrial plants, DRP must guarantee that the real-time information, such as the sampling frame or control frame, owns the highest priority. In general, the frames are divided into 4 categories. In other words, there are four types of service flow in a DRP network: real-time frames, PTP frames for DRP, DRP frames and other frames. The priorities of frames are assigned in order from high to low in the sequence as listed above. Local service and forwarding frames of one category have the same priority. Assume that the average lengths of the frames in the 4 categories are l_R, l_{PTP}, l_{DRP} and l_{NR} respectively; the long term average rates of the corresponding service flows are ρ_R, ρ_{PTP}, ρ_{DRP}, and ρ_{NR}; the link maximum bandwidth is T.

To evaluate the parameters the forwarded delay of a DRP frame must be analyzed in advance. If a DRP frame is planned to be forwarded by a switch device, it will be stored in the switch device buffer and wait until all frames that own higher priorities to be sent out. If there are DRP frames already exiting in the buffer when a DRP frame arrives, the buffer works as an FIFO queue for the DRP frames. Besides, sending a frame is a non-preemptive mission so the forwarded delay τ_{DRP} consists of two parts: the time to send the total length of real-time, PTP and DRP frames already in the switch device buffer, and the time to send out the frame being sent, as shown in Fig.2. Let B be the backlog in the buffer of a switch device. The total length of real-time, PTP, DRP frames in the switch device buffer is $B_R + B_{PTP} + B_{DRP}$, and the length of the frame being sent is l_{bs}.
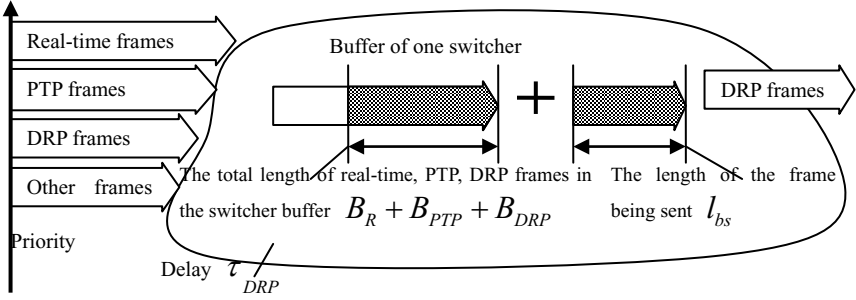
Fig. 2. Forwarded delay of DRP frames

Let the hardware delay time a switch device takes to forward a byte (usually the process of storing and forwarding) be T_t, we have the forwarded delay

$$\tau_{DRP} = (B_R + B_{PTP} + B_{DRP} + l_{bs})/T + l_{DRP}T_t. \tag{3}$$

In order to get the expectation of τ_{DRP} :

$$E(\tau_{DRP}) = (E(B_R + B_{PTP} + B_{DRP}) + E(l_{bs}))/T + l_{DRP}T_t, \tag{4}$$

We first calculate the expectations of l_{bs} and $B_R + B_{PTP} + B_{DRP}$:

$$E(l_{bs}) = (l_R\rho_R + l_{PTP}\rho_{PTP} + l_{DRP}\rho_{DRP} + l_{NR}\rho_{NR})/T. \tag{5}$$

Note that the flow of PTP frames ρ_{PTP} is small, if the service flow of real-time and DRP is relatively light, $B_R + B_{PTP} + B_{DRP} \approx 0$. In this case, τ_{DRP} can be estimated as:

$$E(\tau_{DRP}) = (l_R\rho_R + l_{PTP}\rho_{PTP} + l_{DRP}\rho_{DRP} + l_{NR}\rho_{NR})/T^2 + l_{DRP}T_t. \tag{6}$$

If the network burden is heavy $B_R + B_{PTP} + B_{DRP} \gg 0$. In this case B_R, B_{PTP}, and B_{DRP} have a much more complex relationship with l_R, l_{PTP}, l_{DRP}, ρ_R, ρ_{PTP}, and ρ_{DRP}. This relationship is analyzed by network calculus theory in another work [9].

While a switch device is down, the shortest time it will take to send the Link Alarm frame is 0 and the longest time is T_{cycle}.

$$0 \le t_{FR} \le T_{cycle}, 0 \le t_{FL} \le T_{cycle}. \tag{7}$$

$$0 \le t_{FR} \le T_{cycle} - t_{LinkCheck} + t_{RingCheck}, 0 \le t_{FL} \le t_{LinkCheck} - t_{RingCheck}. \tag{8}$$

Let $T_{ex}(LC)$ be the waiting time after a Link Check frame is sent. Note that the local clocks of different switch devices can not be synchronized perfectly. If T_{PTP} is the maximum clock deviation of two switch devices (it is also called time error), one switch device must wait at least for a time $T_{ex}(LC)+T_{PTP}$ before being prepared to send the Link Alarm frame.

$$t_{FL} = T_{ex}(LC)+T_{PTP}. \tag{9}$$

Similarly,

$$t_{FR} = T_{ex}(RC)+T_{PTP}. \tag{10}$$

When a Link Alarm frame is about to send it must wait for $\tau_{DRP(LA)}$ before the switch device port is available.

$$t_{WA} = \tau_{DRP(LA)}. \tag{11}$$

t_{FA} is determined by hardware.

Let N be the number of switch devices in the network system. The Link Alarm frame has to cost time t_{N-1} to go through (N-1) switch devices to reach the other end of the line to change the port state in t_H (dependent on switch device hardware).

$$t_{N-1} = (N-1)\tau_{DRP(LA)}. \tag{12}$$

$$t_{RW} = T_{ex}(RC)+T_{PTP} + {l_{DRP(LA)}}\big/{T} + l_{DRP(LA)}T_t. \tag{13}$$

$$t_{LW} = T_{ex}(LC)+T_{PTP} + {l_{DRP(LA)}}\big/{T} + l_{DRP(LA)}T_t. \tag{14}$$

$$t_{AR} = t_{N-1}+t_H + {(N-1)l_{DRP(LA)}}\big/{T} + (N-1)l_{DRP(LA)}T_t. \tag{15}$$

4   Summaries

In this paper, the failure detection process of DRP is modeled. The timing parameters are listed and delays in DRP are analyzed. The method to estimate the recovery times is proposed. In order to verify and validate the effectiveness of the DRP and its time model, a prototype of an industrial network system is built and experimented with. The results are reported in [9].

References

[1] IEC 61784-2. Industrial Communication networks -Part 14: Additional Fieldbus Profiles for Real-time Networks Based on ISO/IEC 8802-3. International Electrotechnical Commission, Geneva, Switzerland (2007)

[2] IEC 61784-1. Industrial Communication Networks -Part 1: Digital Data Communications for Measurement and Control - Part 1: Profile Sets for Continuous and Discrete Manufacturing Relative to Fieldbus Use in Industrial Control Systems. International Electrotechnical Commission, Geneva, Switzerland (2007)

[3] IEC 62439. High Availability Automation Networks (February 2008)

[4] Ferraria, P., Flamminia, A., Marioli, D., et al.: Experimental Evaluation of Profinet Performance. In: Proceedings of the 5th IEEE Workshop on Factory Communication Systems, Vienna, Austria, pp. 331–334 (2004)

[5] Kirrmann, H., Hansson, M., Muri, P.: IEC 62439 PRP: Bumpless Recovery for Highly Available, Hard Real-time Industrial Networks. In: Proc. of IEEE Conference on Emerging Technologies and Factory Automation, pp. 1396–1399 (2007)

[6] International Electrotechnical Commission. Industrial Communication Networks High Availability Automation Networks-Part 6: Distributed Redundancy Protocol (DRP), IEC 62439-6 (CDV). International Electrotechnical Commission, London (2009)

[7] Huynha, M., Goose, S., Mohapatra, P.: Resilience Technologies in Ethernet. Computer Networks 54(1), 57–78 (2010)

[8] IEEE 1588. IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems, New York, USA (2008)

[9] Lu, L.: Performance Analysis of Distributed Redundancy Protocol for Real-time Industrial Networks. Zhejiang University (2010)

The Infrared Video Image Acquisition and Display Design Based on FPGA

Zheng Zhixiang[1,2], Meng Xianyuan[1], Tian Yalan[3], Dong Sheng[4], and Cong Peijie[5]

[1] Tsinghua University
[2] Chongqing Communications College
[3] Chongqing College of Arts
[4] Springs Qing Company
[5] China National Software & Service Co., Ltd.

Abstract. This article introduces a hardware design method about a kind of infrared video image acquisition and display system that based on FPGA development board. This design uses VO7620 simulate video imaging sensor to collect data,SAA7113H video decoder collects video data and chang simulate video into digital video, and output ITU656 format data flow with through the I^{2}C configuration module, and Acquisition data is storaged to the FPGA development board embedded SDRAM, and then delivered to the LCD display displayed by SDRAM controller. This system is developed on the software the xilinx10.0, Modelsim6.0, and implemented in hardware, and achieve the desired results. Based on FPGA collection show is hardware real-time image data processing method in low-power, small-volume, simple economic, and has a wide range of application prospects and market value.

Keywords: FPGA, infrared, sequential control, image acquisition, LCD.

1   Introduction

The infrared camera target detection depend on the infrared radiation characteristics, it is difference in temperature between target and the surrounding environment and radiance. At night and severe weather conditions, the infrared camera has more better function than other method, therefore, it is widely used in security television monitor and military equipment, etc. Infrared imaging system is using scenery itself each part radiation difference obtain image details, chang the infrared radiation into electrical signals, then put the electrical signals into visible light, its essence is an optical - electronic systems. The basic function of the infrared system is converted infrared radiation into electrical signals, and represented the size of the electrical signals with gray level to displayed[1].

With the rapid development of microelectronics and computer science, electronic system have been developed to be a system that including ASIC, FPGA and embedded chip, etc. Also, is benning widely used in the area of video image processing apllication. Based on FPGA image processing is an ideal processing real-time image data technology. FPGA (Field Programmable Gata Array) is the further development of the product based on GAL, PAL and EP programming logical device[2], it has

high-performance, small-volume, low -power, high-reliability, design flexible and powerful, convenient to design developed process and later field maintenance, is widely used in embedded systems, and the research of image processing dedicated system will become a new hotspot of information industry. Therefore, this paper mainly introduces a hardware design method based on FPGA development board of infrared video image acquisition and display.

2   The System Design Framework and Working Principle

Infrared video image acquisition and display system of image processing circuit based on FPGA design development board shown in fig.1.
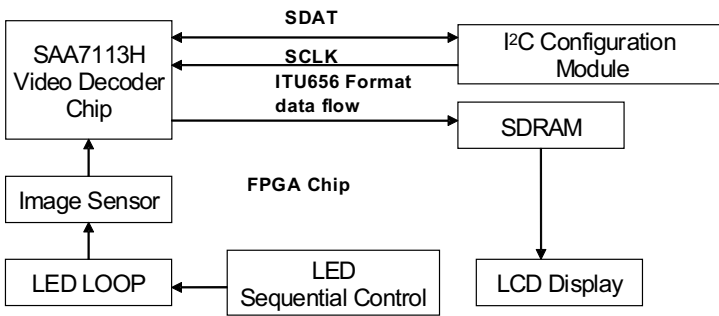
Fig. 1. System design diagram

    This system mainly is to use the FPGA chip resources, through the hardware system control infrared imaging device acquisition real-time image and output on LCD displayer, its main work is as follows: using equipped with 850nm wavelength narrowband filter simulation image sensor(VO7620) collects PAL formats video data, and then sends the data to the SAA7113H video decoding chips. Video decoder chip outputs ITU656 format data flow by I^2C configuration module[3],and stores data in SDRAM, then displays on LCD displayer.

3   Hardware Design

System hardware design includes: infrared LED sequential control, design of image acquisition unit, the I^2C configuration circuit, infrared image display unit.

3.1   Infrared LED Sequential Control

Infrared LED response time is less than 10ns.Time response refers to the time LED bright (up) and extinguishing (attenuate)[1]. The LED up characteristics relate to working current I, I increase, rise time attenuate exponentially, and decrease time does not relate to I. Therefore, using appropriate pulse based on FPGA design to control

LED lamp light and extinguish, can make LED to work by certain frequencies and time[4].The working current with pulse contact as shown in fig.2,graph (a) is one wave, which duty ratio is 1:1,pulse wave of graph (b) is each 1.0 ms launch a 100 pulse, therefore, the average current Im is only tenth of maximum.
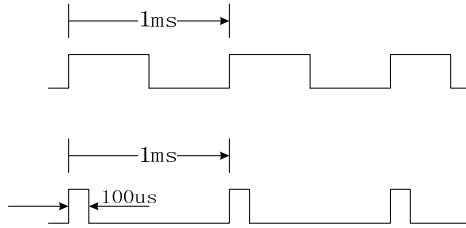
Fig. 2. Relation of working current and pulse

3.2   Image Acquisition Unit Design

Image sensor is an important unit of digital cameras, according to component different there are two models: CCD and CMOS. Compared with CCD, CMOS image sensors is inexpensive, high-integration, small-volume and saving-power etc, and can integrates time-series processing circuit and the front-end amplification and digital of image signal in a chip inside part. With the development of technology and craft, CMOS image sensors got effectively improve not only in noise, and resolution has been greatly improved, therefore, CMOS image sensors development has been hit by the height of industry importance, and has been widely used in the image acquisition and monitoring field. So here we chose OV76200 type CMOS image sensors, its internal structure as shown in fig.3 shows.
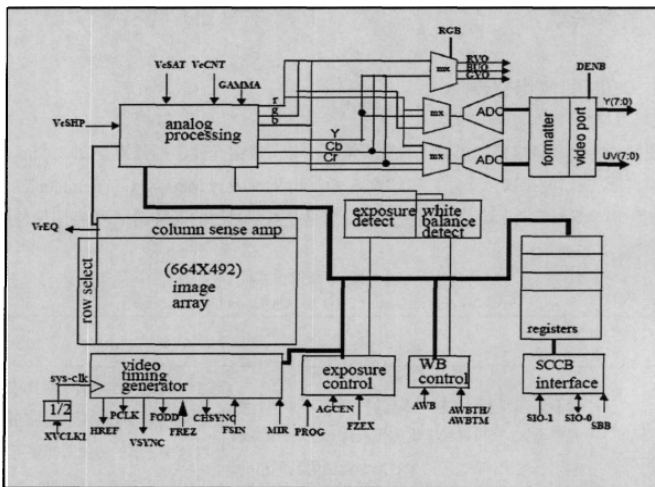
Fig. 3. OV7620 chip internal structure

OV7620 CMOS image sensors includs 664×492 photosensitive array, and integrate the frame (lines) control circuit, video sequence produce circuits, analog signal processing circuit, A/D converter circuit, digital signal output circuit and the I^{2}C programming interface. It can according to the need of the video signal output multiple standards. Video sequence produce circuit is used to produce horizontal sync, vertical sync, mixed video synchronization etc. External controller can be set by I^{2}C bus interface, read OV7620 working condition, working style and the data output format etc [5].

3.3   The I^{2}C Configuration Circuit

Configuration module is mainly to undertake Settings in SAA7113H working mode, signal timing of horizontal sync and vertical sync and the format of the output signal etc. Achieving registers configuration by I^{2}C bus, modifying register values, setting output video characteristics, completing SAA7113H initialization.

The I^{2}C controller mainly is to control the FPGA tube feet to simulate the I^{2}C serial bus SCL(serial clock line) and SDA (serial data line). Controller complete configuration process by judging this two fronts the state to. When beginning, bus in the idle state[6], When SCL's at a high-level, and SDA is hopping high-level to low-level, data begin transmitting, marked as begin, When SDA is hopping low-level to high-level, data stop transmitting, marked as end. When data is transmitting, SCL in high-level, SDA level cannot be changed, only SCL in low-level, SDA level can be changed and update data, While every byte preach finished, requires a response signal, response clock completed by the main device, and realize by releasing the data bus. This design, video decoding chips have not given response signal to come down the clock line, controller set aside time just for a response with the I^{2}C timing keep consistency. When three bytes transmission is completed, controller needs to send end signal, at this time, pulling up data first, and then the clock. Thus completing a transmission, then the next cycle, until all the data transmission configuration completed. The I^{2}C transmission process sequence shown as shown in fig.4 [5].
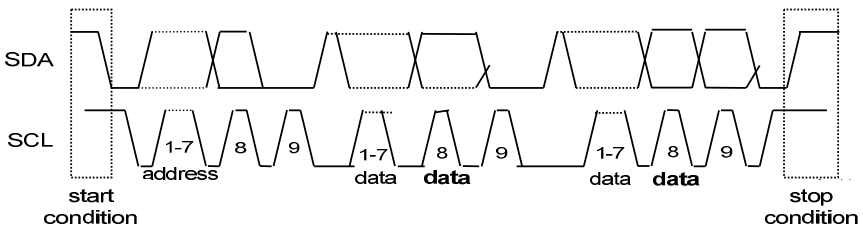
Fig. 4. The I^{2}C transmission process sequence

3.4   The Infrared Image Display Design

Choose LCD display device as infrared image, control and display by the LCD controller, fig.5 is LCD display equipment design diagram.
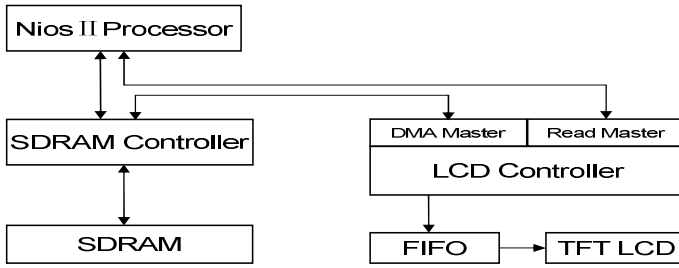
Fig. 5. LCD design diagram

LCD display system principle is: LCD controller have a primary port and a slave port, the primary port is responsible for reading the data and display from SDRAM, Nios II processor control the slave port, and make sure data base address and length of data in SDRAM. Nios II processor in SDRAM open up the Frame buffer, the Frame buffer first address is written to LCD controller and activate LCD controller. After LCD controller started, the DMA Master reads data from SDRAM control at first, and deposit in FIFO, then timing generation module reads data from FIFO based on LCD timing requirements, and send to display device to display images[7].

4   Result

Infrared video image acquisition and display system of development board based on FPGA displayed image as Fig.6.

Fig. 6. This system got the image

5   Summaries

As the FPGA device performance improvement and cost reduction, its development and application are more mature, based on FPGA in the application of infrared imaging system are discussed. Working on hardware system selection, design, realized based on FPGA infrared video image collection and display, and gives the system test results. Compared with software realization, using the FPGA design of video image collection and display system has good real-time, high- reliability, and occupies system resources rarely,high-efficiency, so it has widely application prospect in safety monitoring and military chiefs equipment.

References

[1] Yu, J.: The design and implement of embedded video image detection system based on FPGA. Jiangsu University (2008)
[2] Tan, H., Zhang, C.: EDA technology and application. Xi Dian University Publishing (2001)
[3] Zhao, Q.: Design of Fatigue Driving Detecting System Based on FPGA. Dalian University of Technology (2009)
[4] Dong, Z.: An Active Night Vision System based on IR LED Searchlight. Dalian University of Technology (2005)
[5] Xiaoxiu: Full digital wireless video communication system design and implementation. Zhejiang University of Technology (2005)
[6] Jiaojian: Design of FPGA-based digital video capturing system. Xi'an University of Technology (2010)
[7] Li, D.: Image Acquisition, Processing and Display Implementation of Infrared Imaging System. Chongqing University (2008)

Research on the Compromising Electromagnetic Emanations of PS/2 Keyboard

Litao Wang and Bin Yu

Zhengzhou Information Science and Technology Institute
Zhengzhou, P.R. China
{waltor158,byu2009}@163.com

Abstract. Computer keyboards are often used to input confidential data. Containing electronic components, keyboards will emit electromagnetic waves, which can reveal keystroke information and bring information security problems. Focus on the mainly used PS/2 keyboard, a representative control circuit is analyzed to find the sensitive signals that cause electromagnetic emanations. The PS/2 protocol is lucubrated so as to locate the potential compromising source. Further research is conducted to find how the electromagnetic emissions compromise the keystroke and the characteristics of the PS/2 keyboard's compromising Emanations. An experiment is designed and carried out to verify the research conclusion and find a way to solve the very security problem.

Keywords: PS/2 Keyboard, Compromising Electromagnetic Emanations, Information Security, Falling Edge.

1 Introduction

Electromagnetic waves are emitted while the information technique equipments are working. Some of these Electromagnetic emanations may have relationship with the information processed or transferred. As a result, the information can be recovered if emanations were received and analyzed [1-4]. This is the compromising electromagnetic emanations, and it can degrade the information security level of the device. Many researches have been done on the video display units, both CRT and LCD displays. Images shown on the display of the computer are reproduced [1, 3]. As the researches go on, more and more attention has been drawn to the electromagnetic information security. Besides video display unit, hard disk, mouse, CD/DVD driver and many other peripheral devices of the computer have been studied concerning the compromising emanations [5]. Success information recoveries of electronic device, such as RS-232 device [2], laser printer [4] and keyboard [6, 7], have been carried out with the development of signal acquisition and processing technology.

Keyboard is the main input device of computer. And the emission security of keyboard was researched as early as the 1990s, when Han Fang did some measurement and analysis on the emanations of the keyboard [8]. Han's research found the not so strong electromagnetic emission of the keyboard, and didn't get any obvious relationship between the emission and the keystrokes. In resent years, a breakthrough

has been made in the research of keyboard information leakage with the help of advanced measuring technique and instruments. That is in 2008, Vuagnoux et al. found an effective method to deal with the keyboard's emissions [6], and their best practical attack fully recovered 95% of the keystrokes of a PS/2 keyboard at a distance up to 20 meters, even through walls. They concluded that most of modern computer keyboards (no matter PS/2, USB, wireless or laptop) generate compromising emanations. Graduate student Zhang Jingqin has been doing her research on compromising electromagnetic emanations of keyboard by 2010 [7]. Zhang analyzed the compromising mechanism of the keyboard and compared the compromising emanations of different distances and keystrokes. An algorithm to recover the keystroke information is designed in her research. Both Vuagnoux and Zhang have successfully recovered the low speed keystrokes of keyboard not far away. However, the compromising electromagnetic emanation source and the reason for the distinctive emission signal received by the antenna were not covered in their papers. And thus they didn't give any available information to improve the emission security of the computer keyboard.

Compared with the USB and wireless, PS/2 interface is better supported by the low level hardware of the modern computer. PS/2 keyboard is still the mostly used keyboard due to its compatibility even when the computer breaks down. What is more, the PS/2 protocol is comparatively regular and uncomplicated. Therefore PS/2 keyboard is chosen as the object of our compromising electromagnetic emanations research.

In this study, a representative PS/2 keyboard control circuit is analyzed. Keystroke-related sensitive signals that might emit electromagnetic waves are researched. PS/2 protocol is studied in-depth and the potential compromising source is located. It is affirmed that the falling edge of PS/2 keyboard's data and clock signals compromise the keystrokes. A measurement method and the experiment setup are introduced to verify the conclusion given by the analysis. The experiment results are studied and discussed, and some ideas that will help to improve the information security of the PS/2 keyboard are got.

2   Signal Analysis of the PS/2 Keyboard

PS/2 computer keyboard consists of shell, keycaps, switch matrix, control circuit and transmission line. The shell and keycaps that make up the external appearance of the keyboard are generally plastic. So they have nothing to do with the signals' electromagnetic emission.

Under each keycap there is a keyboard contact, which is actually a tiny switch. All the keyboard contacts comprise the switch matrix, which is arranged in rows and columns. The control circuit with microcontroller sends column selected low level signal to the switch matrix one after another. When a key is pressed the corresponding row data signal is pulled to low, and this message is sent to the control circuit. The column and row indicate which key is pressed and such information is encoded conforming to PS/2 protocol and transmit to the host through the transmission line.

Now the working process of PS/2 keyboard is clear, signals runs inside can be analyzed. As far as the information security of keyboard is concerned, any signal related to the keystroke should be classified as RED signal. The other signals such as power signal and ground signal are BLACK signals apparently.

Fig.1 shows a typical keyboard control circuit, with the encoder chip UR5HC418 [9]. Low level column select outputs CS0~CS15 are sent successively and loop around to detect the keystroke. When a key is pressed, one of the row data inputs RDAT0~RDAT7 will turn to low. The microcontroller inside the encoder will notice the low level pair of column and row. This information is traced and translated into a scan code, which is transferred to the computer through the data line KBDAT in serial mode, synchronized by the clock signal KBCLK.
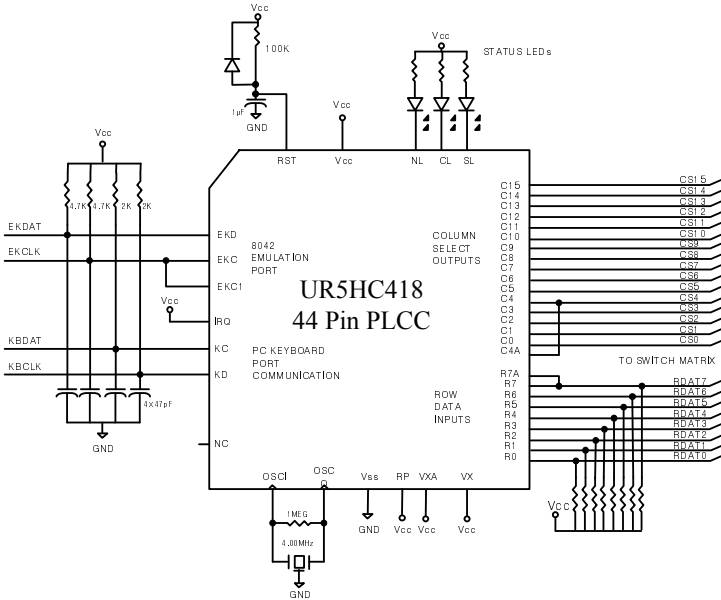
Fig. 1. Example control circuit of PS/2 keyboard

As is analyzed, the signals CS0~CS15, RDAT0~RDAT7, KBDAT and KBCLK may be RED signals that could leak the keystroke information. But whether any key is pressed or not, column select signals are sent one-by-one with an equal time interval. Row data signal is detected unless related key is pressed and corresponding column is selected (low level signal). And what's more, there won't be any noticeable current change. So the electromagnetic emission of the row and column is not remarkable enough to be captured. Even this kind of emanations are received, there won't enough difference to identify which key is pressed.

The port KBDAT in the control circuit connects to PC keyboard port clock line and KBCLK connects to PC keyboard port data line. They transmit the coded keystroke information in serial mode and can be recovered comparably easier if the electromagnetic emission is received.

3   Keystroke Information and Its Emanation

As analyzed above, the major sources of compromising emanations are the data and clock signals that are closely related to PS/2 protocol [10]. PS/2 protocol is a unified

open industry standard. This standard has lots of benefits with the implementation and operation of the keyboard, but also makes it easier to recover the keystrokes from the emitted signals.

PS/2 keyboard implements a bidirectional synchronous serial protocol. Data is sent one bit at a time on the data line and is read on the falling edge of the clock signal each time it's pulsed. Clock is valid only in data transition period. If the bit string of sending data is recovered from the electromagnetic emanations, the scan code which reflects the keystroke would be compromised.

According to PS/2 protocol, Data and Clock are both open collector, which means they are normally held at a high logic level but can easily be pulled down to ground (logic 0). Data and Clock are read on the microcontroller's port A and B, respectively. Both lines are normally held at +5V, which can be pulled to ground by asserting logic 1 on D and C, shown in Fig.2. As a result, Data equals D, inverted; and Clock equals C, inverted. Although this circuit module may be implemented into the modern ICs, the basic structure and its feature maintain.
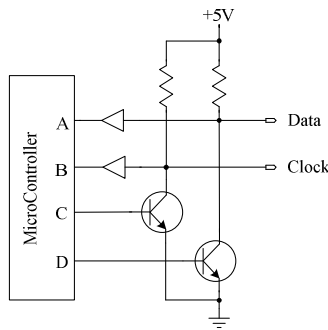
Fig. 2. Open-collector interface to data and clock of PS/2 keyboard

Every time the data or clock transit from logic 1 to logic 0, a '1' is written to the pin D or C. When a breakover voltage enters the base of triode transistor, NPN-type transistor switches into conduction between collector and emitter. A sharp increase in collector current will be generated. This means transient impulse current will happen in the falling edge of the data and clock signals. With reference to the Maxwell Equations and Fourier Transform, the impulse current will emit remarkable electromagnetic waves to the space. This emission might be done by any antenna-effect components of the keyboard.
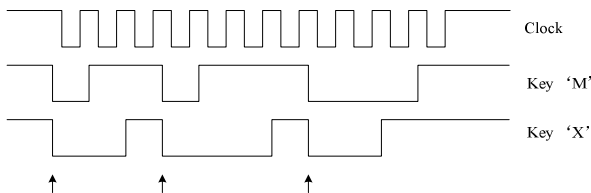
Fig. 3. Demonstration of different keys that have identical falling edges

If the electromagnetic emission of data and clock signals is received by antenna, the falling edges of the signals will be located in time-domain of the received information. Now that the falling edge of the data and clock signal is recovered, the clock signal can be separated based on the characteristic of minimum equal time interval. The PS/2 data transition starts with '0' and have a stop bit '1', so there will be a falling edge of the data before the first one of the clock. Only the falling edges of the data won't uniquely determines the scan code. For example key 'M' (make code is 0x3B) and 'X' (make code is 0x22) have the same serial of falling edges, as shown in Fig.3. Even though, it's still an effect attack of the keyboard for the greatly reduced subset of possible transmitted scan codes.

4   Experiment and Analysis

In order to verify the foregoing analysis, it is needed to get the time-domain information of the data and clock signals, which will help to find the difference between their rising edges and falling edges. The electromagnetic emanations of the keyboard shall be measured to see whether it is at the falling edges that leak the electromagnetic energy most.

Experiment Setup: The experiment is carried out in a semi-anechoic chamber. Oscilloscope (Tektronix TDS5054B) is used to show the time-domain signal of the data and clock, as well as the electromagnetic emanations received by a biconical antenna (30MHz-200MHz). The antenna is placed 3 meters from the PS/2 keyboard (SK-9270), which is connected to a laptop computer powered by battery.

To begin with, the data (labeled '1') and clock (labeled '2') is measured. The detail of data and clock signals is shown in Fig.4, from which we see it is much sharper of the falling edges compared with the rising edges. It is believed that there will be more electromagnetic emission at the falling edge with the same transmitting antenna.
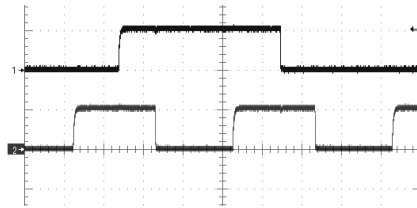
Fig. 4. Sharper falling edges of data and clock signal

When the electromagnetic emission of the keyboard is measured, three probes are connected to the data, clock and antenna respectively. When the key 'V' is pressed, the oscilloscope is triggered and the signals are displayed, shown in Fig.5. In this figure, the data signal is marked with '1', clock with '2' and antenna with '3'. The frequency of the clock is 12.43 kHz. The level of the data and clock is 5 volts, and the signal received by antenna is about 4 millivolts.

From the experiment result (Fig.5), the following characteristic will be found: Firstly, there is a noticeable impulse when and only when it's at the falling edge of data

or clock signal. Secondly, of the data and clock signals, corresponding impulse of one signal's falling edge is a little higher when the other signal is at its high level.

The first characteristic is consistent with the analysis conclusion in part 3 and will help reduce the subset of possible transmitted scan codes. If the second characteristic holds, the level of the other signal will be determined at the falling edge of one signal. It's provable that from this distinctive signals received by antenna, the scan code of key can be uniquely determined, which means that the keystroke of the PS/2 keyboard is compromised. The falling edge emanation ought to be weakened or eliminated to avoid the information compromising of the keyboard and to ensure the information security of system.
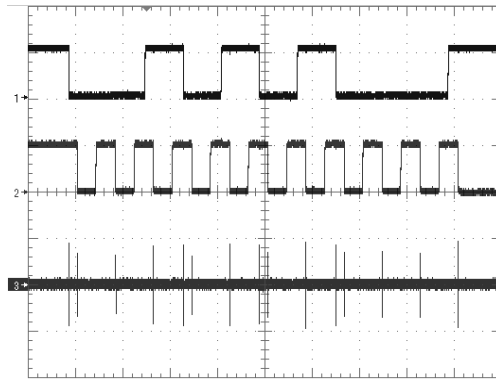
Fig. 5. Time-domain signals when key 'V' is pressed

5   Summaries

The emission source and characteristics of PS/2 keyboard's electromagnetic compromising emanations are studied in our research. Efficient measure to solve this security problem radically is to change the way that falling edges of data and clock take shape in the PS/2 keyboard's control circuit. Further work should be done to find the reason why the second characteristic exists, whether it holds for all PS/2 keyboard and to find a way to implement the protective measure.

References

[1]   van Eck, W.I.M.: Electromagnetic Radiation from Video Display Units: An Eavesdropping Risk? Computers & Security, 269–286 (April 1985)
[2]   Smulders, P.: The Threat of Information Theft by Reception of Electromagnetic Radiation from RS-232 Cables. Computers & Security, 53–58 (September 1990)
[3]   Kuhn, M.G.: Compromising Emanations: Eavesdropping risks of Computer Displays. Technical Report UCAM-CL-TR-577 (2003),
       http://www.cl.cam.ac.uk/TechReports/

[4] Tosaka, T., Taira, K., Yamanaka, Y., et al.: Feasibility study for reconstruction of information from near field observations of the magnetic field of a laser printer. In: Proceeding of the 17th Int. Symp. on EMC, Zurich, pp. 630–633 (2006)

[5] Zhou, C., Hou, D., Chang, Q., et al.: Testing and Analysis of EM Leakiness Spectrum from Digital ITE. Computer Engineering 31(17), 189–190 (2005)

[6] Vuagnoux M., Pasini S.: Compromising Electromagnetic Emanations of Wired and Wireless Keyboards (2008), `http://www.usenix.org/events/sec09/tech/full_papers/vuagnoux.pdf`

[7] Zhang, J.: Information Recover Based on Compromising Electromagnetic Emanations of Keyboard. Beijing Jiaotong University, Beijing (2010)

[8] Han, F.: Electromagnetic Information Leakage and its Protection of Computer. Science Press, Beijing (1993)

[9] SEMTECH Corporation. DeskCoder UR5HC418 datasheet (EB/OL), `http://www.semtech.com/`

[10] Chapweske, A.: The PS/2 Mouse/Keyboard Protocol (EB/OL), `http://www.computer-engineering.org/ps2protocol`

Pulse Wave Velocity Measurement
with Velocity Vector Imaging

Cun Liu[1,2], Qigang Zhu[2,3], Yanling Zheng[4], Yuanliu He[1], Hongxia Xu[1], Juan Su[1],
Lili Zhang[1], Xiaohong Zhou[1], Zhe Ma[5], and Changchun Liu[6,*]

[1] Jinan Central Hospital, Shandong University, Jinan
[2] School of Control Science and Engineering,
Shandong University
[3] College of Information and Electrical Engineering,
Shandong University of Science and Technology
[4] School of Mathematical Sciences,
University of Jinan, Jinan
[5] Qilu Hospital, Shandong University,
Jinan 250012, Shandong
[6] School of Control Science and Engineering
Shandong University, Jinan 250061
changchunliu@sdu.edu.cn

Abstract. Pulse-wave velocity (PWV) is typically used for estimating the stiffness of arteries This study was to implement a new method and rapid two-dimensional ultrasound method for the measurement of PWV in the carotid artery. The longitudinal dynamic image of CCA was analyzed using syngo Velocity Vector Imaging technology. Due to the propagation of pulse wave, nearby different sampling point will have same curves after the transit time between those sampling point. The regional PWV was calculated according to the formula and be applied for the measurement of the other vascular arteries. The noninvasive imaging method for calculating the regional PWV may contribute to a good alternative screening test in the evaluation of cardiovascular diseases and reduce the overall morbidity and mortality of cardiovascular diseases. In summary, our results demonstrate that velocity vector imaging can be an effective and promising approach for evaluate pulse wave velocity of the carotid artery.

Keywords: Pulse-wave velocity, Ultrasound, Velocity vector imaging, Common carotid, Atherosclerosis.

1   Introduction

Arteries are not simple tube conduction structures. The repeated process of pumping blood by the heart produces pressure and flow waves. These waves travel throughout the circulatory system, undergoing transformation in shape and amplitude as they travel from central to peripheral arteries. They moderate systolic pressure increases

---

[*] Corresponding author.

and maintain sufficient diastolic level to guarantee myocardial perfusion. While blood can be considered incompressible, vessel walls have elastic properties that result in finite pulse wave velocities[1]. One of the most reliable indices of the aortic stiffness is the pulse wave velocity (PWV); the speed of propagation of the flow wave in the aorta and an important indicator of arterial compliance. PWV is higher on a stiffer wall and its increase leads to cardiac pressure overload, myocardial hypertrophy [2], and enhance the risk of coronary heart disease [3]. A reliable means to measure aortic wave velocity would therefore be of substantial clinical benefit. Several studies have shown that this parameter is an independent predictor of cardiovascular mortality in the elderly, hypertensive, diabetics, and patients with chronic renal failure as well as in the general population [4,5,6,7].

Typically, the most common method to determine wave velocity is obtaining pressure or flow curves at two locations along an artery and dividing the known distance between measurement sites by the transit time between the foot of each waveform. Hence, PWV = D (meters) / ΔT(seconds) [8,9]. Despite the simple definition of PWV, some problems still remain, which limit the interpretation of available findings and the general applicability of the PWV measurement [10,11,12]. The accuracy of PWV measured from two separate points suffers from errors of distance measurements and/or time-delay measurements [12]. A method with higher temporal and spatial resolution is thus needed in order to calculate the regional PWV.

Ultrasound is a non-invasive imaging technique, which has been successfully used for the morphological and functional assessment of cardiovascular disease. It also offers a reliable platform for their biomechanical assessment. Various ultrasound techniques have been used to detect and track the vessel wall motion. Recently, velocity vector imaging (VVI) has been obtained using ultrasounds [13,14,15].

The purpose of this study was to implement a new method and rapid two-dimensional ultrasound method for the measurement of PWV in the carotid artery.

2   Methods

For this study, A total of 20 healthy volunteers were recruited from hospital staff who wanted a CCA ultrasound and agreed to participate. The volunteers had no history of cardiovascular event or hypertension, and with normal physical examination, blood pressure (BP), ECG and Echocardiogram-Doppler ultrasound values. All volunteers gave their informed and written consent to study participation and the human part of the study was also approved by the Local Ethics Committee. All volunteers underwent CCA ultrasound examination. The examined subjects were in supine position, breathe calmly. Neck was fully exposed with face toward opposite side. Ultrasound was performed through transverse and longitudinal directions from up to down according to a standardized protocol over the CCA; images in long axis view of CCA were stored and transferred to a computer for off-line analysis. Intima-media thickness was measured at the point 1cm proximal to the carotid artery bifurcation from the lumen-initial interface to the medial adventitial border [10]. Two dimensional dynamic image for was collected when volunteer was told to hold breath

for several seconds. The longitudinal dynamic image of CCA was stored to be analyzed using the new off-line software (syngo Velocity Vector Imaging technology [VVI], Siemens). The 2D velocity was used to derive vessel wall displacement off-line. When the photograph was frozen, twenty-four points were marked on the endomembrane of the anterior of carotid uniformly by hand. Three sampling points would be added automatically between every two marked points, and the wall of carotid was divided into ninety-six small segments. The reference point was put in the proximally to the heart and anteriorly to the anterior wallcenter of CCA. The velocity of each sampling point was simultaneously calculated as instantaneous motion divided by the time interval. Due to the propagation of pulse wave, nearby different sampling point will have same curves after the transit time between those sampling point. The measurement the length of a straight line between two sampling point and the transit time between two sampling point was calculated using the new off-line software. The PWV, assessed between two sampling point was calculated according to the following formula: The estimated wall velocities were PWV = D (meters) / ΔT(seconds) (Fig1, Fig2).

3   Statistical Analysis

The PWV of 20 subjects were assessed by 2 expert echocardiologists blinded to the results of each other. Data are expressed as mean $\pm$ standard deviation. Intra- and inter- observer variability were reported as the correlation coefficient between measurements as well as the mean difference between respective measurements. All data analysis was performed by SPSS version 13.0 (SPSS Inc., Chicago, USA). A p value <0.05 was considered Statistical significance.

4   Results

Mean age was 50.4± 17.6 years in healthy volunteers. The PWV could be determined in all healthy volunteers, with good quality velocity-encoded images and corresponding velocity-time curves. Figure 1 shows that the PWV, assessed between two sampling point was calculated according to the following formula: The estimated wall velocities were PWV = D (meters) / ΔT(seconds). The regional PWV of CCA ranged from 1.13 m/s to 20.9 m/s. Using the Bland-Altman method, intraobserver mean differences for 2 measurements of PWV were 0.03 ± 0.48 (p>0.64), and interobserver mean differences were 0.19 ± 1.10 (p>0.136), respectively. The inter-observer variation coefficients for PWV of CCA were $6.9 \pm 1.8\%$.The intra-observer variation coefficients were $5.7 \pm 2.2\%$. There was good intraobserver and interobserver reproducibility for the PWV measurements.
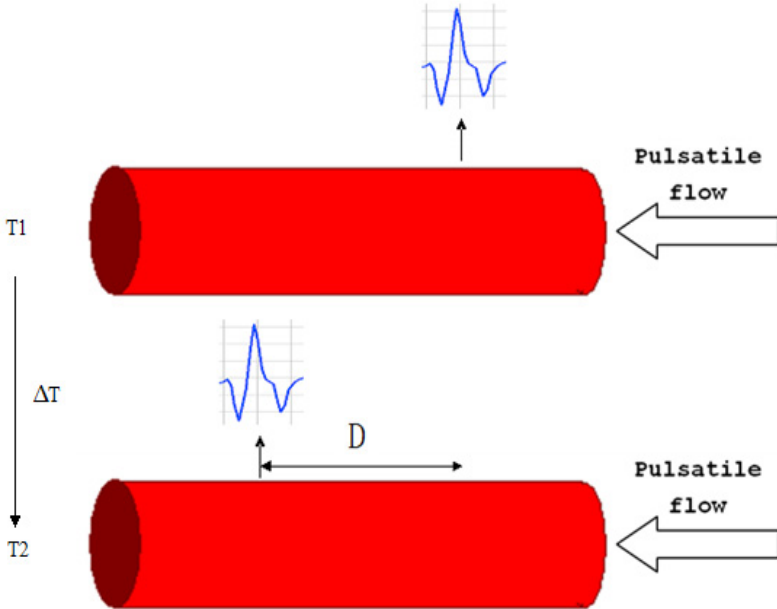
Fig. 1. Illustration of the pulse wave propagation along the CCA and the pulse wave imaging field of view

The PWV, assessed between two sampling point was calculated according to the following formula: The estimated wall velocities were PWV = D (meters) / ΔT (seconds).
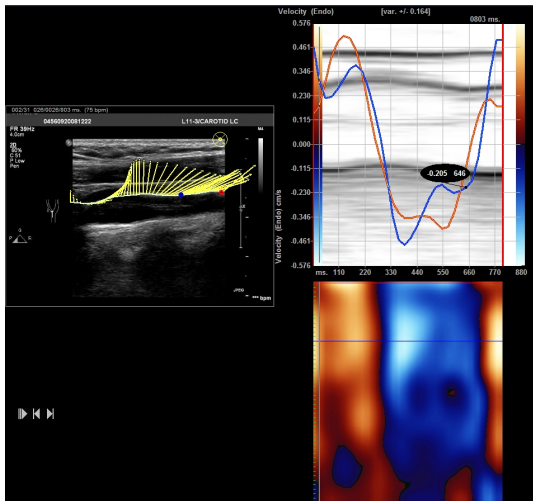
Fig. 2. Example of a velocity vector imaging of the CCA

Due to the propagation of pulse wave, nearby different sampling point will have same curves after the transit time between those sampling point. The measurement the length of a straight line between two sampling point and the transit time between two sampling point was calculated using the new off-line software. The PWV, assessed between two sampling point was calculated according to the following formula: The estimated wall velocities were PWV = D (meters) / ΔT(seconds).

5   Discussion

In recent years, increased emphasis has been placed on the association of aortic stiffness with aging, as well as a variety of cardiovascular diseases, including atherosclerosis, heart failure, hypertension, diabetes, and aortopathies[16,17]. Changes in the composition and structure of the wall will alter the stiffness of the arterial wall[18]. Pulse-wave velocity (PWV) is typically used for estimating the stiffness of arteries [19,20,21,22].

Ultrasound allows noninvasive measurements of PWV, and has been an area of recent research interest[23,24,25,26]. During recent years, VVI, a novel echocardiographic imaging technique based on routine two-dimensional grayscale echocardiographic images that is independent of the angle of the transducer have emerged as valuable echocardiographic tools for more comprehensive and reliable assessment of myocardial function[27,28]. It also offers a reliable platform for their biomechanical assessment.

This study is a novel, noninvasive measurements of PWV using VVI. The main advantages of our method are simplicity of the measurement and the short acquisition time. All data were easily acquired within several seconds by ultrasound without requiring specific devices. This method can provide of the wall velocity of CCA at the time of pulse wave propagation, the curves of sampling point provided a more accurate characterization of the beginning of the systolic phase, and more reproducible results. The measurement the length of a straight line between two sampling point and the transit time between two sampling point was simplicity calculated using the new off-line software. The regional PWV was calculated according to the formula and be applied for the measurement of the other vascular arteries. The noninvasive imaging method for calculating the regional PWV may contribute to a good alternative screening test in the evaluation of cardiovascular diseases and reduce the overall morbidity and mortality of cardiovascular diseases.

Several limitations of this method warrant mention. First, although our imaging frame rates were higher (typically 38-50 frames/s), compareing with PWV, the frame rate is not high enough to precisely measure small time shifts. So the more higher frame rate is able to obtain more images and more detailed spatio-temporal information of the propagation for depicting the pulse wave[26]. Second, two-dimensional speckle tracking algorithms are inherently dependent on image quality and endocardial border definition, any technical factors affecting gray scale may in theory have an impact on the accuracy of VVI measurements, so VVI may be considered semi quantitative at the present time.

6   Summaries

In summary, our results demonstrate that velocity vector imaging can be an effective and promising approach for evaluate pulse wave velocity of the carotid artery. However, large-scale randomized clinical trials are needed to determine the future role of this method.

References

[1]   Milnor, W.R.: Wave propagation. In: Hemodynamics, 2nd edn., pp. 225–259. Williams & Wilkins, Baltimore (1989)

[2]   Nussbacher, A., et al.: Hemodynamic effects of unloading the old heart. Am. J. Physiol. 277, H1863–H1871 (1999)

[3]   Franklin, S.S., et al.: Is pulse pressure useful in predicting risk for coronary heart Disease? The Framingham heart study. Circulation 100, 354–360 (1999)

[4]   Safar, M.E., London, G.M., Plante, G.E.: Arterial stiffness and kidney function. Hypertension 43(2), 163–168 (2004)

[5]   Laurent, S., Boutouyrie, P.: Arterial stiffness: a new surrogate end point for cardiovascular disease? J. Nephrol. 20(suppl. 12), 45–50 (2007)

[6]   De Loach, S.S., Townsend, R.R.: Vascular stiffness: its measurement and significance for epidemiologic and outcome studies. Clin. J. Am. Soc. Nephrol. 3(1), 184–192 (2008)

[7]   Gosse, P., Roche, F., Dauphinot, V., Maudoux, D., Pichot, V., Barthelemy, J.C.: Components of arterial stiffness in a population of 65-year-old subjects: PROOF study. Journal of Hypertension 26(6), 1138–1146 (2008)

[8]   Laurent, S., Cockcroft, J., Van Bortel, L., Boutouyrie, P., Giannattasio, C., Hayoz, D., Pannier, B., Vlachopoulos, C., Wilkinson, I., Struijker-Boudier, H.: European Network for Non-invasive Investigation of Large Arteries. Expert consensus document on arterial stiffness: methodological issues and clinical applications. Eur. Heart J. 27(21), 2588–2605 (2006)

[9]   Weber, T., Ammer, M., Rammer, M., Adji, A., O'Rourke, M.F., Wassertheurer, S., Rosenkranz, S., Eber, B.: Noninvasive determination of carotid-femoral pulse wave velocity depends critically on assessment of travel distance: a comparison with invasive measurement. Journal of Hypertension 27(8), 1624–1630 (2009)

[10] Xu, J.P.: Do we need a better approach for measuring pulse-wave velocity? Ultrasound Med. Biol. 29, 1373–1373 (2003)

[11] Karamanoglu, M.: Errors in estimating propagation distances in pulse wave velocity. Hypertension 41, E8–E8 (2003)

[12] Zhang, X.M., Kinnick, R.R., Fatemi, M., Greenleaf, J.F.: Noninvasive method for estimation of complex elastic modulus of arterial vessels. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 52(4), 642–652 (2005)

[13] Cho, I.J., Chi, Y.S., Yang, W.-I., Kim, S.-A., Chang, H.-J., Jang, Y., Chung, N., Ha, J.-W.: Assessment of mechanical properties of common carotid artery in Takayasu's arteritis using velocity vector imaging. Circulation Journal 74, 1465–1470 (2010)

[14] Valocik, G., Druzbacka, L., Valocikova, I., Mitro, P.: Velocity vector imaging to quantify left atrial function. International Journal of Cardiovascular Imaging 26, 641–649 (2010)

[15] Lei, Z., Yan, L., Fei, Z.P., Xia, Z.Y., Ping, J.X., Ting, L.X., Qiang, C.W., Xi, L.C., Cheng, Z., Yun, Z.: Peak radial and circumferential strain measured by velocity vector imaging is a novel index for detecting vulnerable plaques in a rabbit model of atherosclerosis. Atherosclerosis 211, 146–152 (2010)

[16] Metafratzi, Z.M., Efremidis, S.C., Skopelitou, A.S., de Roos, A.: The clinical significance of aortic compliance and its assessment with magnetic resonance imaging. J. Cardiovasc. Magn. Reson. 4, 481–491 (2002)

[17] London, G.M., Cohn, J.N.: Prognostic application of arterial stiffness: task forces. Am. J. Hypertens. 15, 754–758 (2002)

[18] Wilson, K., Bradbury, A., Whyman, M., Hoskins, P., Lee, A., Fowkes, G., McCollum, P., Ruckley, C.V.: Relationship between abdominal aortic aneurysm wall compliance and clinical outcome: A preliminary analysis. Eur. J. Vasc. Endovasc. Surg. 15, 472–477 (1998)

[19] Nichols, W.W., O'Rourke, M.F.: McDonald's Blood Flow in Arteries, 5th edn. Hodder Arnold, New York (2005)

[20] Davies, J.I., Struthers, A.D.: Pulse wave analysis and pulse wave velocity: A critical review of their strengths and weaknesses. J. Hypertens. 21, 463–472 (2003)

[21] Hirata, K., Kawakami, M., O'Rourke, M.F.: Pulse wave analysis and pulse wave velocity—A review of blood pressure interpretation 100 years after Korotkov. Circ. J. 70, 1231–1239 (2006)

[22] Laurent, S., Cockcroft, J., Van Bortel, L., Boutouyrie, P., Giannattasio, C., Hayoz, D., Pannier, B., Vlachopoulos, C., Wilkinson, I., Struijker-Boudier, H.: Expert consensus document on arterial stiffness:Methodological issues and clinical applications. Eur. Heart J. 27, 2588–2605 (2006)

[23] Eriksson, A., Greiff, E., Loupas, T., Persson, M., Pesque, P.: Arterial pulse wave velocity with tissue Doppler imaging. Ultrasound Med. Biol. 28, 571–580 (2002)

[24] Hermeling, E., Reesink, K.D., Reneman, R.S., Hoeks, A.P.G.: Measurement of local pulse wave velocity: Effects of signal processing on precision. Ultrasound Med. Biol. 33, 774–781 (2007)

[25] Hoctor, R.T., Dentinger, A.M., Thomenius, K.E.: Array signal processing for local arterial pulse wave velocity measurement using ultrasound. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 54(5), 1018–1027 (2007)

[26] Luo, J., Fujikura, K., Tyrie, L.S., Tilson, M.D., Konofagou, E.E.: Pulse wave imaging of normal and aneurysmal abdominal aortas in vivo. IEEE Trans. Med. Imaging 28(4), 477–486 (2009)

[27] Chen, J., Cao, T., Duan, Y., et al.: Velocity vector imaging in assessing myocardial systolic function of hypertensive patients with left ventricular hypertrophy. Can. J. Cardiol. 23, 957–961 (2007)

[28] Canesson, M., Tanabe, M., Suffoletto, M.S., et al.: Velocity vector imaging to quantify ventricular dyssynchrony and predict response to cardiac resynchronization therapy. Am. J. Cardiol. 98, 949–953 (2006)

Finite Element Analysis on Biomass Pelletizing Process

Tie-Li Ye[1], He Li[1], Liang Wang[1], and Xiao-Fei Fan[2]

[1] College of Mechanical and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266510, China
Terryye99@yahoo.com.cn,
{xinfangwina,6057079}@163.com
[2] Jinan Pressure Vessel Factory,
Jinan 250022, China
378431294@qq.com

Abstract. Biomass pelletizing is an important branch of biomass energy conversion technology. In this article, using parametrical design module of three-dimensional solid modeling software Unigraphics NX(UG) constructs geometric model of plane-die belonging to the pivotal part of plane-die pelletizing machine. Using finite element simulation software ANSYS simulates biomass pelletizing process so as to obtain stress field of plane-die and biomass, which provide an important reference for the design and improvement of plane-die pelletizing machine.

Keywords: Biomass, Plane-die, Simulation.

1   Introduction

Biomass compression molding is to put agricultural and forestry waste, such as the crushed crop stalks, into the molding machine, then compressed into the required shape and size by exerting certain external pressure [1]. According to the different molding mechanism, compression molding machine can be divided into four categories: piston press, extruder press, matrix pellet press, twist press [2]. Among them, matrix pellet press can be divided into plane-die pelletizing machine and ring-die pelletizing machine.

Plane-die is a pivotal part of plane-die pelletizing machine. Fig.1 shows the molding parts of plane-die pelletizing machine.
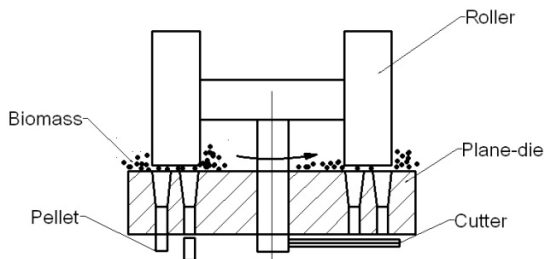
Fig. 1. Structure schematic drawing of molding parts

2   Parametric Design and Modeling

UG is a large-scale application software associated with the parametric technology, which has very powerful parametric design capabilities. In the finite element simulation results analysis of the plane-die, the geometric model need probably be repeated the call and modified. To great extent, UG   parametric design method provides the convenience for fast modification of the model parameter.

Preliminary Determination of the Plane-Die Parameters. Structure parameters of plane-die determine the fuel quality of compression molding. Its parameters mainly include the plane-die opening area, arrangement of model hole, dimensions of model hole, length-diameter ratio of model hole, etc. According to the mechanical industry related design standards, preliminarily determine that the outer diameter of plane-die (PM_OD1) is equal to 255mm, the thickness of plane-die (PM_HDE) is equal to 50mm, the diameter of plane-die (PM_HD) is equal to 10mm, other parameters settings are shown in Fig.2.
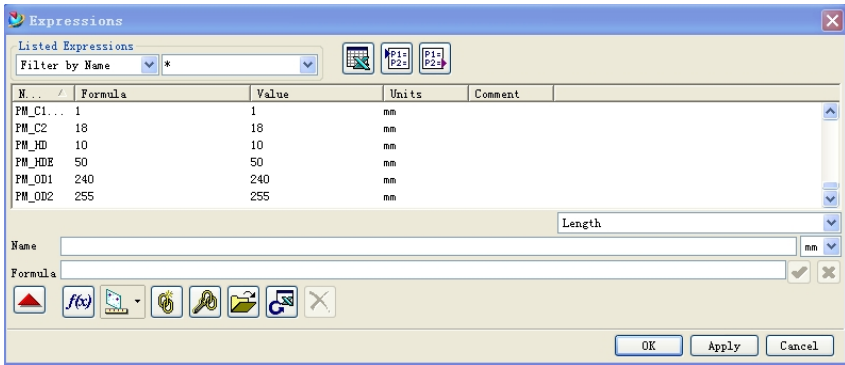
Fig. 2. Parameters dialog box of plane-die

Parametric Design of the Plane-Die. Through creating features module, characteristic operation module, editing features module and dimension driven technology of UG, constructing plane-die model and producing design variables, then through the
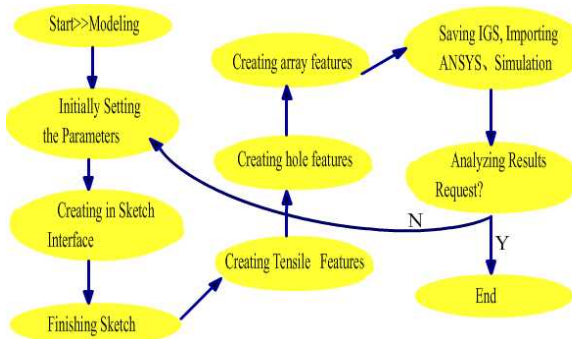
Fig. 3. The flow chart of modeling of plane-die
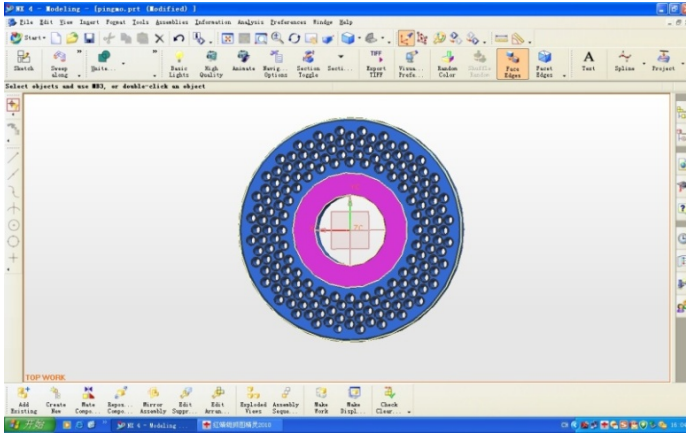
Fig. 4. Plane-die

modification of design variables, to realize the parametric design of plane-die interaction levels. Fig. 2, Fig. 3, Fig. 4 show the parametric modeling process of plane-die.

3   Finite Element Simulation of Biomass Molding

Using finite element analysis software ASNYS simulates the compression molding process of biomass in order to obtain the condition of contact stress and strain of plane-die and changes of the extrusion process of biomass.

The Contact Type Analysis belonging to ANSYS Structural Analysis. Contact problem is a highly nonlinear behavior and requires large computing resources. Contact problem is divided into two kinds: one kind is the rigid body - soft body contact, another kind is the soft body - the soft body contact. In the ordinary circumstances, when soft materials and hard materials are at contact, problems can be assumed to be the rigid body- the soft body contact [3]. So this article is based on the surface-the surface contact analysis belonging to the rigid body-the soft body problem. The rigid body is plane-die, and the soft body is biomass materials.

Building of Finite Element Model. We assume that the biomass material is homogeneous and isotropic compressible continuum. Because of involving large plastic deformation in the course of analysis, solid182 unit type is selected to build both geometric model. The unit is set to be axisymmetric behavior, and its real constant do not need to be defined. The contact problem belongs to surface-surface contact .The target surface is the hole wall of plane-die and the contact surface is the soft biomass material face. Therefore we will use the contact unit CONTA172 and target unit TARGE169 to simulate the contact surface. And the friction coefficient is zero point one. To assume that biomass molding process meets equivalent pressure principle, that is, the built model has equivalent pressure in hole axial direction. So the small part of model can be extracted for analysis [4] (Fig.5). The overall model for analysis is shown in Fig.6.

Fig. 5. Plane-die area of extraction   **Fig. 6.** Overall analysis model
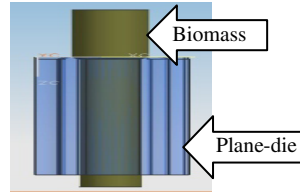
Use of biomass materials in this article is sawdust. 12Cr1MoV is selected as the material of plane-die, because the material has high strength and high wear resistance. Material properties are shown in Tab.1 and Tab.2 [5].

Table 1. Property parameters of sawdust

EI(MPa)	μ	ρ(kg/m^3)
4.4MPa	0.38	365

Table 2. Property parameters of 12Cr1MoV

EI(MPa)	μ	ρ(kg/m^3)	σs (MPa)
2.1x10^{5}MPa	0.3	7.85x10^3	225

Because the analysis model and load are axisymmetric, it is used axisymmetric methods to analyze. In order to save computing costs and observe the results more directly in post-treatment, a quarter of the entire model is adopted for modeling analysis.

Division of the Mesh. According to the information shown in the Tab.1 and Tab.2, the physical properties of materials are set. Then based on the physical properties of the material biomass, we respectively make divided grid on biomass and plane-die. In order to ensure the accuracy and convergence, save computer resources, the distribution of the unit and the node on the unit are effectively controlled. Because sawdust in the extrusion process has a greater plastic deformation and displacement, using the manual meshing generates a mapped grid. The mesh of the plane-die adopts intelligent meshing combined with manual meshing.

To Establish Contact, Exert Boundary Condition and Load, Solving. Note that set the same real constant number to match Target element TAGRE169 to contact element CONTA172. When boundary conditions are exerted, the bottom of plane-die is applied the fixed constraints along Y-axis direction and the fixed constraints along the X axis direction, because the plane-die is connected to the machine body. All the nodes in the left line of biomass model are applied symmetry boundary conditions along the X axis direction and upper surface of biomass model is applied the displacement loading along the Y-axis negative direction. Taking into account the nonlinear and friction factors of the biomass in the extrusion process and combining with

elastic-plastic mechanics and theoretical knowledge of finite element method, ANSYS solving set is carried out to ensure the convergence and the accuracy of results.

Finite Element Results and Analysis. Fig.7 and Fig.8 show the deformation of the biomass materials in the extrusion process when respectively t = 1s, t = 2s. From the two figures, we can see clearly that the biomass flow is more uniform, and no dead zone in the extrusion process. In the near exit surface and near the die hole wall, there is a greater line bending along X-axis direction. This is because between the surface of plane-die hole and biomass has a greater friction, when biomass materials flow. The outer layer displacement is behind the middle layer.
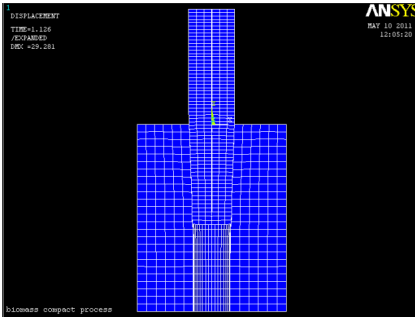
Fig. 7. Deformation when t = 1s   **Fig. 8.** Deformation when t = 2s
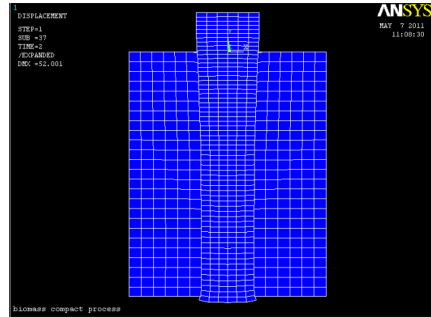
Fig.9 and Fig.10 show the contours of stress of compression modeling of the biomass materials along Y-axis when respectively t = 1s and t = 2s. It can be shown from the figures that along the Y direction, because of the biomass molecular interactions, intermediate place of the biomass flows faster, compared to both sides, resulting in uneven stress distribution. In this case, it is easy to make the surface of biomass fuel cracked, even broken. But it can also be shown from the two figures that the edge and middle part of the biomass have a little flow rate difference, thus the modeling characteristics of the fuel is better. This also proves that each parameter of the plane-die is basically reasonable, which is determined in parametric design of plane-die in the 2.2 section.
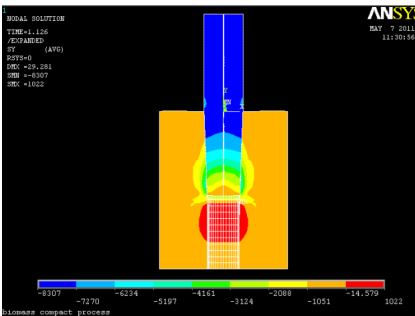
Fig. 9. Stress contours along Y-axis when t=1s   **Fig. 10.** Stress contours along Y-axis when t=2s
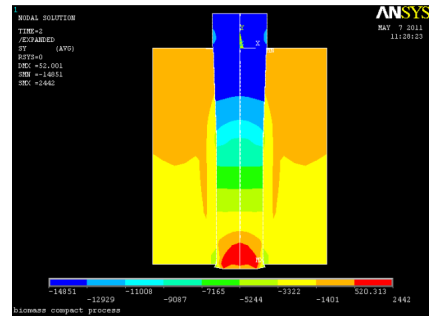
4   Summaries

In the article, parametrical design module of three-dimensional solid modeling software Unigraphics NX is used to build plane-die geometric model, which belongs to the key part in the plane-die pelletizing machine. And analysis model is extracted correspondingly according to the equivalent principle of pressure. Through ANSYS finite element analysis software, corresponding finite element model is established and the compression process of biomass is simulated. Ideas of parameterized modeling and analysis of simulation results provide important theoretical basis for the design and improvement of plane-die machine and furnish important reference value for the optimization of modeling process.

References

[1] Zhao, L.-Y.: Development and Utilization of Biomass and Biomass Energy. Agricultural Environment and Development, 30–33 (June 2007)
[2] Cheng, K.-C., Jiag, C.-Q., Zhong, J.-L.: Research Review of Biomass Compression Molding Fuel Technology. Energy Engineering, 8–11 (1996)
[3] Wang, Q.-W., Zuo, F., Hu, R.-X.: ANSYS 10.0 Senior Application Example about Mechanical Design, pp. 203–215. Machinery Industry Press (August 2006)
[4] Jing, G.-X.: Biomass Fuel Plane-die Pelletizing Machine Design Theory and Simulation Research (June 2010)
[5] Cheng, D.-X.: Mechanical Design Manual, pp. 311–319. Chemical Industry Press (November 2007)

Based on PLC Control System in the Application of Wastewater Treatment Plant

Niu Qinzhou, Jia Wei, and Yan Fei

School of Information Science and Engineering, Guilin University of Technology,
Guilin, China

Abstract. Based on the Siemens PLC S7-300 modified type CASS process as the main body of the activated sludge level 2 biochemical treatments, and aimed at medium-sized and small enterprises sewage treatment to realize the good sewage treatment of automatic control system with high reliability, convenient and flexible control and scalability.

Keywords: CASS, PLC, Automatic control, Sewage treatment.

1 Introduction

With the development of economy, lives of people have been further improved, but it also brought with polluting of the environment. Sewage is one of the causes of the environmental pollution, in the long-term, our country put more and more vigor in sewage treatment and the control level requirements are increasingly high, after reusing the wastewater to effectively reduce the environment pollution, and conforms to our country sustainable development of scientific strategy[1]. Therefore, establishing stable and reliable sewage treatment system is imperative. Taking the sewage treatment plant of a county as an example, the author introduces sewage treatment of automatic control system.

2 Sewage Treatment Process Selection and Process Flow

2.1 Wastewater Treatment Options

CASS process is in sequencing batch type activated sludge (SBR), and on the basis of the reaction tank for long direction along the pool two parts, front design for biological choose area (pre reaction district), mainly in the reaction zone, the main reaction district have installed the elevating draining water device, aeration and precipitation within in the same pool and tell the cycle of the two conventional activated sludge pond and return sludge system, COD removal rate above 90%, BOD5 95%, and removal has very good to nitrogen phosphorus removal.

The characteristics of CASS process are low construction cost, small floor space, low operating cost, high degree of automation, convenient management, decarburization phosphorus don't need plus another potion, high the treatment effect, good effluent and stable and reliable operation.

2.2   Sewage Treatment Processes

The altered CASS pool of 4 cases within arrangement, according to time by pool water, precipitation and aeration and idle stage constitute a cycle. Process flow diagram is shown in fig.1 shows:
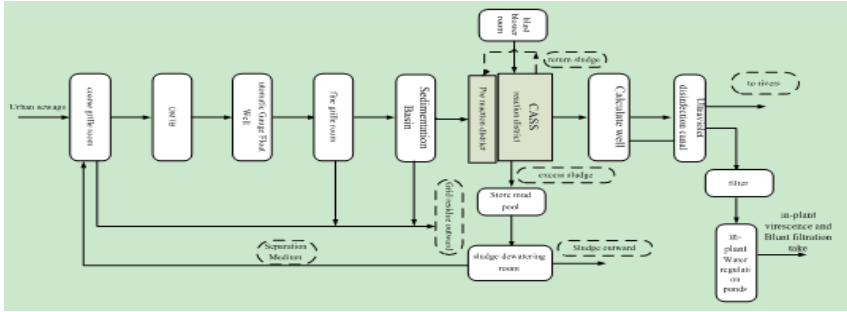
Fig. 1. Graph one sewage treatment plant process flow diagram

The sewage treatment plant treats 10,000 tons everyday, sewage treatment of the altered CASS biochemical pool as the main body of the activated sludge level 2 biochemical treatments, the sludge treatment part is the surplus within the sludge biochemical pool by sludge concentration after dehydration shipped to landfills. The processed meets emission standard, and the water in the river through a small amount to the water has been treated inside the plant after water filters into, one adjusting pool for flushing green belt, the other part of road watering greening for plant.

3   Automation System

3.1   System Structure

This control system USES adopts structure of "the centralized management, decentralized control, data sharing", according with current industrial automation
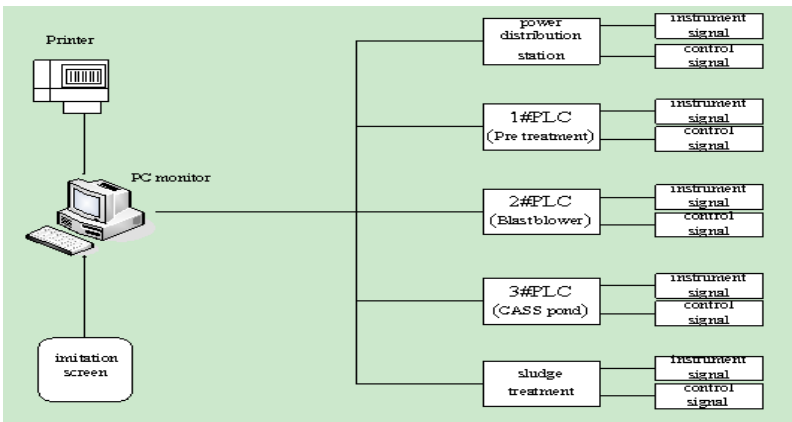
Fig. 2. Self-control system diagrams

detection system development trend and eventually achieve technological parameters and equipment plant concentrated check and process automatic control. Its control system graph as shown in fig.2 shows:

3.2   The Basic Composition Automation System

Whole factory automation system consists of four parts:

(1) Production process auto-control system: mainly is the realization of the production field unattended and computerization management, plant provides a production control information exchange platform.

(2) On-line measuring instrument: realize real-time monitoring of process parameters according to the examination, and the parameters of the detected promptly to control adjustment.

(3) Production management of computer network system: to achieve production reliable and efficient operation, and realized the paperless office.

(4) Integrated wiring system: plant production, management structures for information exchange and information processing platform.

3.3   Operating Mode of the Automatic Control System

In sewage treatment plant, the design process for the automatic control system provides three operating mode:

(1) Dynamic operation mode, the design of wastewater treatment plant by automatic control process

(2) Computer manual operation mode, PC display system status, the equipment room does not exist chain relations;

(3) Mechanical manual operation mode, in automatic control system malfunctions or maintenance period can run this model. In this mode, the operation of equipment does not suffer PLC control and the switch by the control cabinet control.

4   The Application of Siemens s7-300 in Wastewater Treatment Plant

Siemens is one of the world's major PLC producers, its product are widely used in home early and PLC products of S7 - S7-200, 300, 400 s7-300 are has high market share and has gradually occupy leading position in China. This system is medium S7-300 modular with PLC, meeting the application of medium performance requirements, whose main advantage is easy to realize the distribution between each separate module and can be widely combination and expand and convenient user master [2]

4.1   Central Control Rooms

Have two configuration of the real time monitoring software as the central control monitoring server (hot backup), statements printer and print setup accident alarm in the console and central control monitoring server, industrial Ethernet switches

interaction interfaces cabinet are installed on the network. The central operation station computer as the man-machine switches mouth, server through network adapter and industrial control systems and plant management nets system connection this system including the user login, real-time process flow diagram and alarm display, real-time curves and historical curve, parameter setting, events login, statements processing.

User login: system for different user endowed with different operating privileges, divided into administrator privileges and operator permissions.

Real-time process flow diagram and alarm display: real-time dynamic reflects the operation process, including the equipment running status and real-time process parameters. Alarm show there are four main states of that have not been confirmed not disappear police has not been confirmed, alarming disappear, alarming not disappear is confirmed, alarm disappear is confirmed, the system according to the state police handling.

Real-time curves and historical curve: check process real-time curves and history of the curve, the larger or smaller, can view any time parameters curve.

Parameter setting and the event log: all the parameters can be adjusted on-line adjusting, when the operator for parameters adjustment of events will be recorded. Statements treatment: the system automatically records of the process operation data, data rolled into statements form, convenient print.

4.2   Pretreatment Control Stations (1 # PLC)

Primary monitor is coarse grille (spiral lose slag machine), improve pump room, fine grille (spiral lose slag machine), sink the main electric equipment sand pool and on-line instrument measuring.

(1) Thick grille: mainly enters the sewage after grille filters out some larger floater and sundry. Mainly uses the level different control and timing control. According to actual condition, when grille is greater than before, when 50cm level difference after spiral lose slag machine is running, when grille starting the machine before and after less than 30cm, level difference coarse grille machine stop running and then spiral lose slag machine stop. Time control mode has two kinds of cases, one is when liquid meter fail coarse grille machine adopts time control mode, running time according to the operation condition of actual wastewater treatment plant set; two is the level difference and time control manner, while running, when the level difference in a set period of time, fails to reach the set value, has been monitoring system will automatically start PLC spiral lose and coarse grille machine slag machine, lest coarse grille machine is sundry winding and level difference plan fault from sewage jam.

(2) Ascending pump: mainly from coarse grille will flow of water ascending to a height, and then into fine grille. PLC is mainly acquisition ascending field instruments and pumps the signals, and through the automatic control program realization of automatic operation control pumps equipment. Ascending pump room three submersible sewage pumps, frequency-inverter regulation.

First the well water room set to ultra-low water level is low, medium and above normal, high, higher. General water level in water to maintain simply start frequency pumps can. When the water level rises, increasing frequency conversion speed, high water level and maintain noisy over 10S, then start a fixed pump, if water levels

continue to rise, and achieve high water level, is in start a water pump, when the water reaches high water level and maintain 10S, the alarm starts. When the water level drops to lower levels, the first operation of the first stop running, if pump water continued to drop, then stop the other a water pump, if water levels continue to decline of variable frequency pump rotation speed, reduce a little low water level and maintain 10S to when the police. Three submersible sewage pumps amphibious a provision, two pumps according to level height and their respective pump total runtime. Ensuring the basic total runtime is equal and the best running status of the equipments.

(3) Fine grille: through ascension, still exist some the tiny fiber sewage after pump removing. In order to reduce the tiny material must use the fine load to treat the intercept tiny floater. Like the coarse grille, Fine grille control uses the liquid level different control and timing control.

(4) Sand setting pool: after in fine grille go out after the residue in water sands, guaranteeing the follow-up processing structures and the equipment running and prevent the biological treatment structures of interference. Here by using two groups of eddy current sink sand pool, each group of installing a pump and mixer, blender analyzed to practice long-term operation. The bottom of the debris is analyzed and practice pump water separator sends to sand, according to start automatically set operation cycle every operation, and could pump 10 minutes and 20 minutes setback.

4.3   Variable Power Distribution Control Station (2 # PLC)

Primary monitor are variable power distribution, the main electric fan wind on-line detection instrument equipment and realizing the function:

(1) According to the detection of biochemical pool of PLC qualitative parameters (DO value and SS value) control of blower, adjust the frequency, oxygen and examine the working state of the blower;

(2) Transferring ensemble the power system, direct through industrial field bus and controlled communication complete data acquisition work to guarantee the automation system becomes a complete and whole system.

4.4   CASS Pool Control Station (# 3 PLC)

CASS pool is the sewage treatment plant, an important part of the automatic control system, # 3 PLC primary monitors biochemical pool, off-work water main electric equipment, on-line measurement instrument and off-work water quality parameters measurement instrument.

(1) CASS biochemical pool has four groups; each frame biochemical pool is divided into preprocessing reaction district, oxygen area and the main reaction area. PLC opens inlet valve and the inlet valve aeration (aeration time according to the field instruments), according to set parameters DO value, adjust the frequency of the blast blower, reached the control requirements of the oxygen. After the water aeration into precipitation phase (precipitation time according to the site condition regulation), the precipitation phase ended, into right-handed water stage, and finally into idle stage. Water stage, in abandoned the surface of the water with decreasing speed and slow

down, on this stage will adjust the drainage, water decanter frequencies that decrease time for half an hour, so when not in the drainage of the sludge sedimentation disturbance lower after treatment, will the upper clarify liquid discharge 3. CASS pool operation control table as shown in Tab.1 shows:

Table 1. CASS pool operation control list

Pool	Situation	0-1	1-2	2-3	3-4	4-5	5-6	6-7	7-8
1#CASS	Inflowing water	√	√	√		√	√	√	
	aeration	√	√			√	√		
	Sediment			√				√	
	Drain water				√				√
	Unloading mud				√				√
2#CASS	Inflowing water	√		√	√	√		√	√
	Aeration			√	√			√	√
	Sediment	√				√			
	Drain water		√				√		
	Unloading mud		√				√		
3#CASS	Inflowing water	√	√		√	√	√		√
	Aeration	√			√	√			√
	Sediment		√				√		
	Drain water			√				√	
	Unloading mud			√				√	
4#CASS	Inflowing water		√	√	√		√	√	√
	Aeration		√	√			√	√	
	Sediment				√				√
	Drain water	√				√			
	Unloading mud	√				√			

(2) PLC collection MLSS, liquid level values DO, according to the parameters such as DO value timely adjust collected the frequency of blower, oxygen to make changes in the pool to control water quality requirements. Testing backflow sludge pump, excess sludge pump, mixer, blower, the wind valve adjusting weir door, water decanter operation state, and the wind valve, water fan door, the water decanter adjust weir limit switches operating and fault signal.

(3) Detecting the ammonia nitrogen, COD (COD), MLSS (sludge concentration), PH (PH), a water quality parameters such as water, for timely adjust control parameters.

(4) Changing with transferring &transforming ensemble with power directly, the system by computer backstage Ethernet and controlling completed acquisition data communication.

5   Summaries

The sewage treatment system is based on Siemens s7-300 PLC control system design in sewage treatment plant running more than half a year, always in the stable condition, greatly reducing the operating personnel labor strength and improving sewage treatment plant. According to the operation efficiency of on-line detection parameters, and then adjusted control parameters and finally make completely achievement to the A level standard.

References

[1] Li, J.: The application of PLC in urban sewage. PLC&FA, 69–69 (January 2005)
[2] Ren, S., Bian, C., Sun, Y., S7-300/400, P.L.C.: S7-300/400 PLC Principle and Practical development guidelines, pp. 25–30. Machinery Industry Press (2007)
[3] Wang, L., Wu, M.: Sewage treatment technology and engineering example, pp. 217–218. China Petrochemical Press (2006)
[4] Guha, D., Chakraborty, D.: A New Approach to Fuzzy Distance Measure and Similarity Measure between Two Generalized Fuzzy Numbers. Applied Soft Computing (2009)
[5] Lemmi, A., Betti, G.: Fuzzy Set Approach to Multidimensional Poverty Measurement (October 2006)

Effects of Even-Order Nonlinear Terms and Oversampling Rate on Digital Baseband Predistortion Linearization

Xiaoning Feng, Peng Zhang, and Tianxiao Cui

Communication University of China

Abstract. Power amplifier (PA) is widely used in communication systems, however, it has nonlinear characteristic. Digital baseband predistortion is an effective approach to compensate for nonlinear distortion of the PA. Accurate models are of significant importance to study PA nonlinear characteristics and design predistorters. Polynomials have been used extensively for modeling the behavior of the PA or the predistorter (PD). For bandpass signals, attention has been paid mainly to odd-order nonlinear terms. In this paper, we reveal the performance of introducing even-order nonlinear terms in baseband predistorter models. Moreover, we analyze the predistorted spectral characteristics of predistorter models. Theoretical analysis and simulation results show that, when the oversampling rates of source signals are designed properly, it is beneficial to include even-order nonlinear terms in baseband predistorter models.

Keywords: Even-order, Power amplifier, Digital baseband predistortion, Oversampling rate.

1 Introduction

It is well known that a RF power amplifier is inherently nonlinear. Among all linearization techniques, digital baseband predistortion is a highly cost-effective approach. A predistorter (PD) is added in the final stage of the baseband processing, prior to upconverter and Power amplifier (PA). The PD creates a complementary nonlinearity, which ideally is the exact inverse of the PA nonlinearity. As a result, the overall PD-PA chain is approximately linear.

Polynomial models have been used extensively for memoryless nonlinear PA and PD. Only odd-order polynomial terms in the PA nonlinearity impact the bandpass signals. For this reason, people take it for granted that when constructing a polynomial-based predistorter, only odd-order nonlinear terms should be considered. However, in this paper, we will show that it is beneficial to include even-order terms in baseband predistorter models [2].

Predistortion means that spectrum spread happens to the source signal due to nonlinear distortion of the PD. Spectrum spread may bring on undersampling and aliasing of the predistorted signal. A radical solution to the problem is oversampling the digital baseband signal and increasing indirectly the sampling rate of the predistorted signal. In this paper, we will show that when the oversampling rate is chosen properly, it is beneficial to include even-order nonlinear terms in predistorter models.

The rest of this paper is organized as follows: Section 2 describes the polynomial-based predistortion system. Section 3 proposes the theoritical basis of including even-order terms in PD. The spectrum characteristic of predistorted signal will be studied in section 4. Section 5 gives and dicusses the results of computer simulation. The last section is a summary of this paper.

2   The Polynomial-Based Predistortion Method

The characteristic curve of the PA is commonly smooth, and its nonlinearity can be described by a polynomial with complex coefficients. The intermodulation products (IMP) of the polynomial even-order terms are far away from the passband, while the IMP of the odd-order terms may lie in or near the passband. So, normalized baseband model of the PA can be depicted as the following odd-order polynomial

$$\tilde{a}(t) = \tilde{d}(t) \sum_{k=1 \text{ odd}}^{\infty} \tilde{\beta}_k \left| \tilde{d}(t) \right|^{k-1} \tag{1}$$

where $\tilde{d}(t)$ and $\tilde{a}(t)$ are the input and output signals of PA respectively, $\tilde{\beta}_k \ (k=1,3,5,\cdots)$ is the complex coefficient of the kth term, and $\tilde{\beta}_1 = 1$ on account of normalization. $\tilde{\beta}_k = \tilde{\beta}_{Ik} + \tilde{\beta}_{Qk}$ is made up of the in-phase ($\tilde{\beta}_{Ik}$) and quadrature ($\tilde{\beta}_{Qk}$) parts, meaning that the PA has both AM/AM amplitude distortion and AM/PM phase distortion. Generally speaking, $\tilde{\beta}_k$ decreases quickly with k increasing. When $k \geq 7$, the effect of the nonlinear term corresponding to $\tilde{\beta}_k$ can be omitted. Equation (1) is therefore approximated as

$$\tilde{a}(t) = \tilde{d}(t) \sum_{k=1 \text{ odd}}^{5} \tilde{\beta}_k \left| \tilde{d}(t) \right|^{k-1} \tag{2}$$

Using the same method, the output of the PD (also input of the PA), $\tilde{d}(t)$ is expressed in term of the input signal $\tilde{m}(t)$ as

$$\tilde{d}(t) = \tilde{m}(t) \sum_{k=1 \text{ odd}}^{K} \tilde{\alpha}_k \left| \tilde{m}(t) \right|^{k-1} \tag{3}$$

where K is the maximum polynomial order, and k can be even or odd.

Since the nonlinear characteristic of the ideal PD is the inverse of the PA, the PA output $\tilde{a}(t)$ and the normalized PD input $\tilde{m}(t)$ must satisfy $\tilde{a}(t) = \tilde{m}(t)$. It can be seen from the resulting expression of substituting (3) into (2) that, the joint characteristic of the PD and the PA is not ideally linear but a polynomial with order higher than both of them.
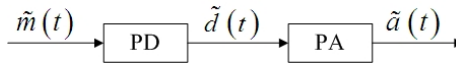
Fig. 1. The functional block diagram of linearized transmitter

3   The Impact of Even-Order Terms on Baseband Signals

The basic method has been reviewed in last section. There is no need to consider even-order nonlinear terms when bandpass signals go through nonlinear systems. Now, what we want to know is whether it is helpful to consider the effect of even-order terms in baseband system.

Let $\tilde{m}(t)$ be a signal whose frequency components lie in the band

$$-\omega_c \leq \omega \leq \omega_c$$

Suppose that $\tilde{m}(t)$ goes through a nonlinear PD to produce $\tilde{d}(t)$. Apparently, $\tilde{d}(t)$ contains certain frequency components that are not present in $\tilde{m}(t)$. We are interested in knowing which of those components fall in the band of $[-\omega_c, \omega_c]$, referred to as in-band.

When (3) is expanded, we get

$$\tilde{d}(t) = \tilde{\alpha}_1 \tilde{m}(t) + \tilde{\alpha}_2 \tilde{m}(t)^2 + \cdots + \tilde{\alpha}_k \tilde{m}(t)^k + \cdots + \tilde{\alpha}_K \tilde{m}(t)^K \qquad (4)$$

To facilitate analysis, we simplify $\tilde{m}(t)$ as a dual-tone signal which has two frequency components ($e^{j\omega_1 t}$ and $e^{j\omega_2 t}$), then investigate their interactions in the presence of kth-order nonlinearity. Recall that

$$\begin{aligned}
\tilde{\alpha}_k \tilde{m}(t)^k &= \tilde{\alpha}_k \left(e^{j\omega_1 t} + e^{j\omega_2 t} \right)^k \\
&= \tilde{\alpha}_k \sum_{i=0}^{k} \binom{k}{i} \left(e^{j\omega_1 t} \right)^i \left(e^{j\omega_2 t} \right)^{k-i} \\
&= \tilde{\alpha}_k \sum_{i=0}^{k} \binom{k}{i} e^{j[i\omega_1 + (k-i)\omega_2]t}
\end{aligned} \qquad (5)$$

The interesting question is: "Under what condition will the new frequency component $i\omega_1 + (k-i)\omega_2$ lie in the band between $-\omega_c$ and ω_c?"

Apparently, ω_1 and ω_2 are both in the interval of $[-\omega_c, \omega_c]$, that means

$$-\omega_c \leq \omega_1 \leq \omega_c \quad , -\omega_c \leq \omega_2 \leq \omega_c$$

It follows easily that $-k\omega_c \leq i\omega_1 + (k-i)\omega_2 \leq k\omega_c$.

Therefore, $i\omega_1 + (k-i)\omega_2$ falls in a band between $-k\omega_c$ and $k\omega_c$, which contains $[-\omega_c, \omega_c]$, hence, no matter k is odd or even there is possibility that $i\omega_1 + (k-i)\omega_2$ falls in band. This explains why it is necessary to include even-order nonlinear terms in the baseband PD. This conclusion can also be applied to the case in which $\tilde{m}(t)$ is an OFDM signal.

4   The Spectrum of the Predistorted Signal

As investigated in last section, predistortion causes spectrum spread. Reference [2] had proved in detail that when OFDM multi-carrier signal predistorted by K-odd-order polynomial PD, the output signal becomes K times as wide as the source signal owing to spectrum spread. It is easy to prove that when K-order PD polynomial contains even-order terms, the predistorted signal will also extends to K times of the source signal, but its center frequncy remains unchanged.

The kth order of the predistorter polynomial creates k-order IMP. Since the kth IMP is proportional to the amplitude of kth sub-carrier, and amplitude of sub-carrier in normalization model is far less than 1, high-order IMP amplitude is lower than low one. Furthermore, the farther IMP is away from the center frequency, the smaller the number and amplitude of IMP will be.

Fig.2 and Fig.3 show the spectrum characteristics of different PD polynomial orders, where ω_c is center frequency of carrier and B is bandwidth of source signal.
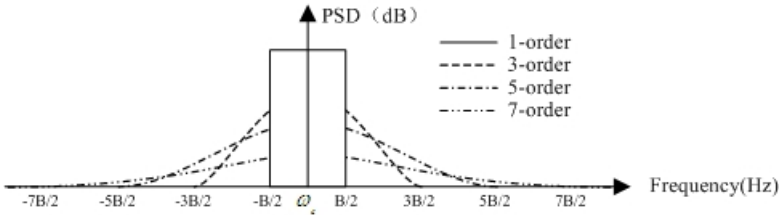
Fig. 2. The spectrum characteristics of different PD polynomial orders（odd）
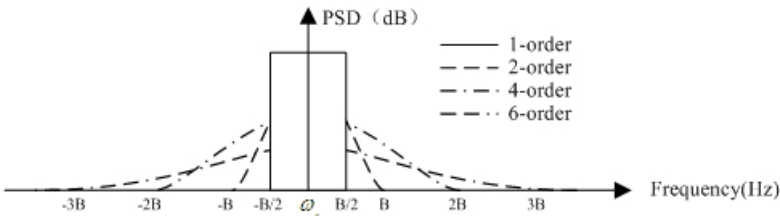
Fig. 3. The spectrum characteristics of different PD polynomial orders（even）

5   Simulation Analysis

Taking example of the OFDM signal adopted by China Mobile Multimedia Broadcasting (CMMB) standard, the spectral characteristics of the output signal of polynomial PD are analyzed by MATLAB simulation.

We first constructed a PA polynomial model with coefficients $\tilde{\beta}_1 = 1$, $\tilde{\beta}_3 = -0.3 - 0.1j$, $\tilde{\beta}_5 = -0.3 - 0.1j$ respectively, and then measured its

AM/AM and AM/PM characteristics. Next, we fitted PA linear characteristic according to (3) using three sets of polynomial orders: $K_1 = \{1,3,5\}$, $K_2 = \{1,3,5,7,9\}$, $K_3 = \{1,2,3,4,5\}$. Note that K_1 and K_3 have the same maximum nonlinearity order whereas K_2 and K_3 contain the same number of terms. It should be pointed out that, the PD polynomial coefficients were calculated by the least mean square (LMS) algorithm.

5.1   3 Times Oversampling

When the source signal is 3x oversampled, the predistortion effectiveness of different PD polynomial coefficients is shown in Fig.4. Obviously, the performance of PD polynomial with coefficients K_3 is better than that with K_1 and K_2. The performances of 4x and 5x oversampling are close to that of 3x oversampling. Their good spectrum satisfies the specification of CMMB standard. Note that since PD polynomials with coefficients K_1 and K_2 perform almost the same, the curve of polynomial with coefficient K_2 is not displayed in Fig.4.
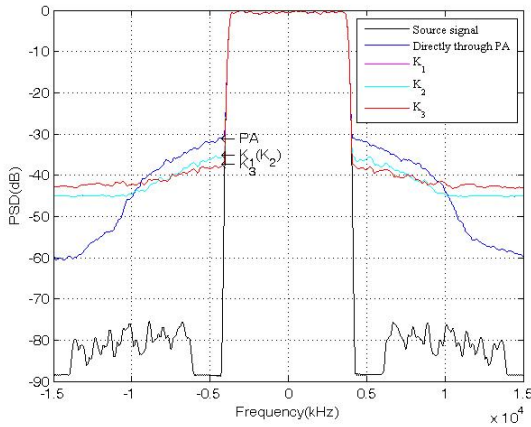
Fig. 4. The effectiveness of different orders (3x)

5.2   2 Times Oversample

When the source signal is 2x oversampled, the predistortion effectiveness of different PD polynomial coefficients is shown in Fig.5. The serious spectral aliasing of the predistorted signal results in intolerable linearizetion effectiveness, no matter which kind of order K is. So, to avoid serious spectral aliasing, the oversampling rate must be 3x or higher.
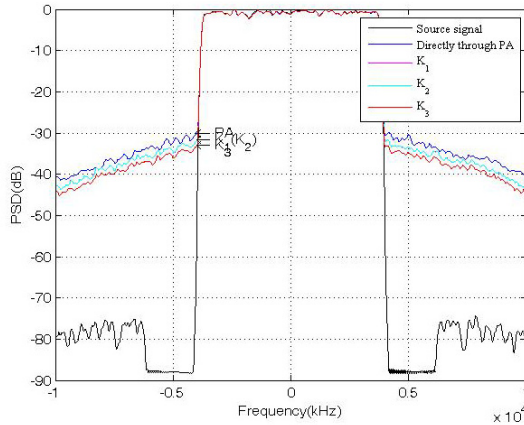
Fig. 5. The effectiveness of different orders (2x)

6   Summaries

We first theoretically analyzed the reason why it is beneficial to introduce even-order nonlinear terms in predistorter models. Then, by taking an example of the OFDM signal, we proved the correctness of the hypothesis. By introducing even-order terms, the accuracy of transmission system could be increased obviously. In addition, oversampling rate is very important in predistortion system. To avoid spectral aliasing, the oversampling rate must be designed properly first. Simulation results show that, when the oversampling rate of the source signals equals to 3x or higher, including even-order nonlinear terms in predistorter models can enhance predistortion performance, that is why this method is recommended.

References

[1]   Ding, L., Zhou, G.T.: Effects of even-order nonlinear terms on predistortion. In: IEEE10th Digital Signal Processing Workshop and the 2nd Signal Processing Education Workshop, pp. 1–6 (October 2002)

[2]   Zhang, P., Yang, F., Yang, G., Liu, C.: Effects of Digital Baseband Predistortion on Linearized Transmitter Design. In: IEEE 2nd International Conference on Power Electronics and Intelligent Transportation System, pp. 60–63 (2009)

[3]   GY/T 220.1-2006. Mobile multimedia broadcasting part 1: framing structure, channel coding and modulation for broadcasting channel (2006)

[4]   Westesson, E., Sundstrom, L.: A complex polynomial predistorter chip in CMOS for baseband or IF linearization of RF power amplifiers. In: Proc. IEEE Int. Symp. Circuits Syst., pp. 206–209 (May 1999)

[5]   Stapleton, S.P., Cavers, J.K.: A new technique for adaptation of linearizing predistorters. In: Proc. 41st IEEE Veh. Technol. Conf., pp. 753–758 (May 1991)

[6]   Ding, L., Zhou, G.T., Morgan, D.R., et al.: A robust digital baseband predistorter constructed using memory polynomials. IEEE Transaction on Communications 52(1), 159–165 (2004)

[7]   Kim, J., Knostantinou, K.: Digital predistortion of wide band signals based on power amplifier model with memory. Electron. Lett. 37(23), 1417–1418 (2001)

[8]   Stapleton, S.P., Costescu, C.: An adaptive predistorter for a power amplifier based no adjacent channel emissions. IEEE Trans. Veh. Technol. 41, 49–56

[9]   Zhang, P., Wu, S.L., Zhang, Q.: Adaptive digital predistortion schemes to linearize RF power amplifiers with memory effects. Journal of Beijing Institute of Technology 17(2), 217–221 (2008)

[10] Jian, W., Yu, C.X., Wang, J.X., et al.: OFDM adaptive digital predistortion method combines RLS and LMS algorithm. In: 4th IEEE Conference on Industrial Electronics and Applications, pp. 3900–3903 (May 2009)

Design and Implementation of High-Precision Multi-function Time Calibrator

Li Yongjun[1], Xu Xiaorong[2], Jiang Pingge[1], and Tian Yafang[3]

[1] School of Physics and Electronics,
Henan University,
Kaifeng, Henan, China
`lyjustc@163.com`
[2] College of Computer Science and Technology,
Hunan University of Arts and Science,
Changde, Hunan, China
[3] Zhengzhou Railway Vocational & Technical College,
Zhengzhou, Henan, China

Abstract. In order In order to test Multi-Rate, timing reference frequency and switching period error of the multi-function electricity meter, the system of high-precision multi-function time calibrator is designed and implemented. In this system, W77E58 is adopted as the center of the control and computation, Intel 8254 and the multi-channel selector are applied as the measuring kernel, the high-precision quartz crystal is used as frequency standard, the time provided by GPS is used as time standard. This system is proved to be stable, simple high precision, multi-channel measuring.

Keywords: Time Calibrator, Frequency Standard, Intel 8254, GPS.

1   Introduction

As of the increasing popularization of multi-rated and multi-function electric energy meter in electric power industry, related measuring equipments are also developing rapidly.   Multi-rated electric energy meter tables differentiate sections by time[1], so a time standard is needed besides an electric power standard when checking multi-rated and multi-function electric energy meter. The function of time calibrator is testing multi-rated and multi-function electric energy meter the errors of standard frequency, switching period signal of multi-function time meter. Therefore, a calibrator must have high precision of frequency and time standard. This paper designs and implements a high precision multi-function time calibrator system with the quality of simple design and cost low, CPU W77E58 is acting as the core of control and calculation, multi-channel switch selector CD4051 and D flip-flop compose the select network for input signal, counter Intel 8254 and measuring control circuit constitute the measurement circuit of standard timing frequency, 8 input D flip-flop、NOR gates CD4078 and flip-latch 74LS273 consist the measurement circuit of switching period signal. High precision quartz crystal provides 10MHZ standard frequency, GPS GSU36 provides time standard and PPS second pulse in this

system. It's a high-speed measuring, and has good stability, high accuracy and multi-channel surveys.

2   Measuring Principle

Comparison method is used in this paper in order to have frequent measurement, i.e., periodicity of standard frequency signal f_s is N, frequency of measured signal f_x is f_s / N. Relative error r_e and Date error d_e are used to indicate frequency error, as shown below in （1）and（2）. f_b in （1）is frequency nominal value of measured signal. （3）is switching period error, measured from measurable switching time recorded by CPU[2].

$$r_e = [(f_x - f_b)/f_b] \times 10^6 \quad (PPM) \tag{1}$$

$$d_e = r_e \times 24 \times 60 \times 60 \tag{2}$$

$$s_e = t_{RM} - t_{BD} \tag{3}$$

3   Hardware Structure

The system hardware structure is shown in Fig.1, CPU W77E58 is the core of calculation and control, Intel 8254 and multi-way selector compose frequency measurement circuit to measure sampling signal, GPS GSU36 module provides time standards and PPS second pulse when measuring and constant temperature quartz crystal provides 10MHZ standard frequency. PC synchronize time to the checked meter by getting measuring results through communication module. The keyboard and display receive user operation as man-machine interface, display specific status information.
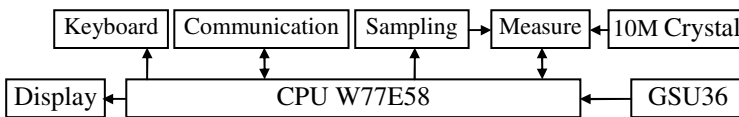
Fig. 1. Hardware structure of the system

4   Design and Implement

Input signal sampling is shown in Fig.2.Input switching period signal and standard frequency of multi-rate and multi-function electric energy meter into the system. Through 8 8-to-1 switches 4051, the system measures 64 input standard frequency signals F[1…64]. The 64 signals are divided into 8 groups, each 8-to-1 switches 4051 is one group. D0, D1, and D2 in CPU select one group of signal among the 8 groups and transmit to U10, which is also an 8-to-1 switch 4051. D4, D5, and D6 in CPU select one group of signal FXIN among the 8 groups and transmit to automatic measuring gate-controlled circuit.

Fig.3 shows gate-controlled circuit used to measure the timing reference frequency. Signal FXIN gets multiplication through 9014, then with PRE and CLR to CLK of D trigger 74F74. When U1A is set to 1 and U2A is cleared to 0, the system starts to work. The first rise of FXIN makes the two triggers to set 1 at the same time. Open gate AND2 and start GATE1 on 8254 counter T1 to order Intel 8254 to count 10MHz standard frequency signal. The second rise of FXIN makes the two triggers to set 0 at the same time, and U2A 0. Close gate AND2 and GATE1 to order Intel 8254 stop counting; set U1A to 0, W77E58 pauses, reads the counting results from 8254, and make out FXIN frequency, frequency error, and date error according to (1) and (2).
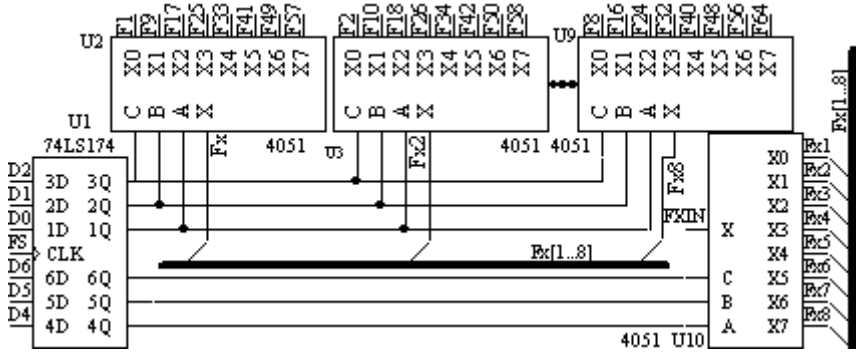
Fig. 2. Input signal sampling
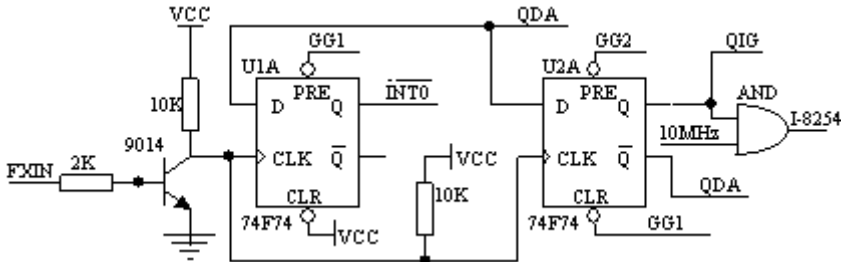
Fig. 3. Gate-controlled circuit used to measure the timing reference frequency

Fig.4 shows the circuit between CPU and Intel 8254, the core of the system. As the core of the controlling system and functioning as a counter, CPU W77E58 is a fast programming microprocessor compatible with MCS51 series monolithic processor[5]. In Fig.4, P0 is 8-digit data bus, through 74LS373, low 8-digit address bus is obtained, while P2 gets high 8-digit address bus[3]. P1.0 and P1.1 in gate P1 assist D trigger in Fig.3 to complete frequency test. P1.4,P1.5, and P1.6 make up keyboard matrix. P3.5 is working condition indicator. RXD0 and TXD0 serve as RS232 interface and communication with host computer. RXD1 receives time standard sent by GPS. INT0 frequency test input pauses, INT1 switching period signal input pauses, INT2 receives standard second pulse PPS sent by GPS. Meanwhile, W77E58 controls all chip selecting signals of the system from 74ls138. W77E58 operates on the basis of these chip selecting signals as well as data and address bus.
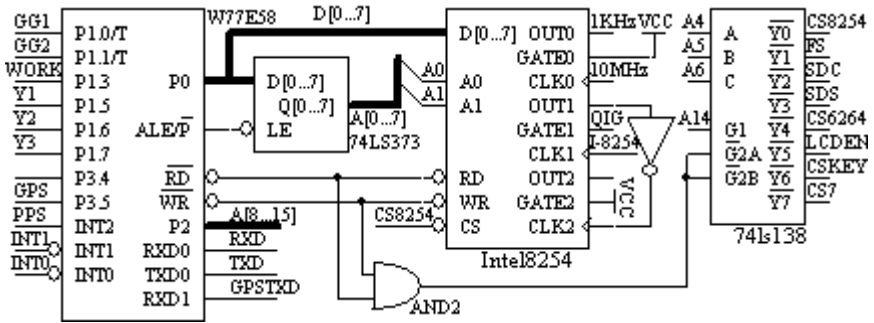
Fig. 4. Circuit between CPU and Intel 8254

Intel 8254 is a programming timer/counter with 3 separate 16-digit down counters, each counter having 6 working modes[4]. Counter 0 is used to divide 10MHz frequency by 10^4 so as to obtain standard 1KHz signal. Under the control of CPU W77E58, Counter 1 and 2 make frequency measurement on input signal FXIN. As is shown in Fig.4, chip selecting signals obtained by Intel 8254 are all from 74ls138. From Fig.4, we can easily find that address of Intel 8254 is 4000H~4003H, i.e., control register port address is 4003H, port address of 3 counters is 4000H, 4001H, and 4002H respectively. Operations such as initializing and read-write to Intel 8254 are realized through these addresses. Standard timing frequency and error measurement programming is as bellows:

```
MOV    DPTR, #FS
MOV    A,#00H
MOV    @DPTR,A; measure F1 in Fig 2
SETB   GG1
SETB   GG2; begin to meaure;
```

Waiting interrupt INT0, ending of measurement, reading the initial value of counter 1 and counter 2, calculating standard frequency and error.

```
MOV    A,#40H
MOV    DPTR,#4003H
MOVX   @DPTR,A; Latch count value of counter 1
MOV    DPTR,#4001H;
MOVX   A,@DPTR
MOV    @R0,A; Read low 8 bits
INC    R0
MOVX   A,@DPTR
MOV    @R0,A; Read low 8 bits;
```

Latch and read count value of counter 2

Sampling of switching period signal and frequency signal is similar. But each measurement of switching period error is 8 lines as a group, as shown in Fig.5. Data of SD[1…8] is sent to 8 D triggers' CLK. After SDC clears all 8 D triggers to 0, switching pulse signals from any line will make its corresponding D trigger to Q
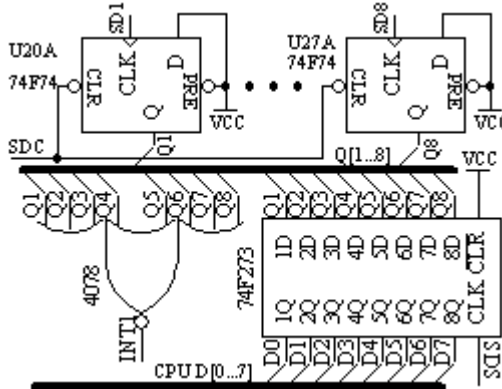
Fig. 5. Chart of measuring Switching period error

setting. 8 input negater CD4078 output low level. This low level is sent to INT1 of CPU, CPU pauses, reads 74LS273 data in Fig.5, and counts switching period time and switching period error. Programming of switching period time and switching period error is as follows:

```
MOV     DPTR ,#SDC
MOVX    @DPTR,A
MOV     DPTR ,#CS7
MOVX    @DPTR,A; Produce a low-level pulse which makes the trigger U1A to
U8A zeros in Fig.3;
Waiting interrupt INT1
MOV    DPTR ,#SDS
MOVX    A,@DPTR; Read switching period signal to the CPU, calculate the
error of switching period signal
```

5   Software Design

There are four processing programs in this software: key handling processing program, address display processing program, one second handling processing program, and communications handler, each with a marker site. Main program searches for these four marker sites. When the detected marker site is 1, corresponding program works. counter 1 and counter 2 in series are used in Intel 8254 to achieve measurement extension. But counter initialization in Intel 8254 is set after one pulse is received, so there should be at least one pulse output on counter 1 before measurement in order to make sure that counter 2 can work properly. In this system, Intel 8254 pre works for more than 10ms. Fig.6 is Intel 8254 Software flow chart.
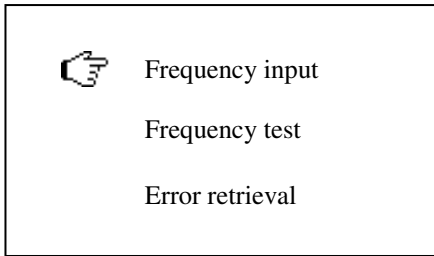
Fig. 7. Main menu
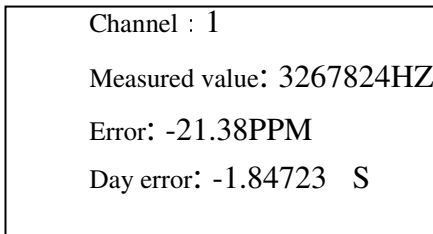
Fig. 8. Result of the experiment
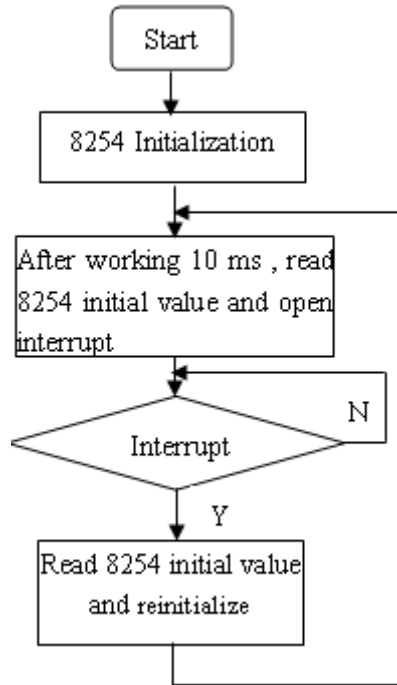
Fig. 6. Software flow chart

6   Test Results

This system has realized and been used in multi-function electric energy meter checking device. Frequency measurement range is 0.1Hz～50K Hz；f frequency measurement accuracy is ±0.2×10-6; Date error accuracy ≤20ms; GPS time setting accuracy is 10-6s; internal crystal oscillator stability ≤5×10-8/s. After test, this system is in compliance with the checking requirements of electric energy meter in JJG596－1999, DL/T614-1997, and JJG307－88[6]. This system is applied in the checking of multi-rate and multi-function electric energy meter. Fig.7 is the main menu. Fig.8 is measurement result on line 1. This system is suitable for the checking of multi-rate and multi-function electric energy meter. But this method is only suitable for the measurement of low frequency signal（below 50KHz）and measurement speed is slower than FPGA technology because Intel 8254 is under cycling measurement and is affected by the counting speed of W77E58. Frequency input measured value error Day error.

References

[1]   Li, Q.-B.: Design of High-precision Multi-function Time Calibrator. Zhigong: Journal of Sichuan University of Science&Engineering (NaturalScience Edition) 22(5), 89–92 (2009)
[2]   Gui, C., Wu, L.: Design and Implementation of Self-Control Clock Based on GPS Time Source. Shiyan: Journal of Hubei Automotive Industries Institute 18(2), 49–53 (2004)

[3] He, L.: SCM Advanced Tutorial: Application and Design, 2nd edn., pp. 132–175. Beijing University Press, Beijing (2007)
[4] Zhou, T.: Research & Apply on SCM Low_frequency Impulsator. Electronic Engineer (6), 29–30 (2000)
[5] WINBOND Electronics Corp.8-bit Micro-controller W77E58 (EB/OL) (2003),
      `http://www.winbond.com`
[6] Measurement of Quality and Technical Supervision Division. Meter Technical Manual, pp. 149–171. Publishing House of China Metrology, Beijing (2000)

Research on Defect Detection in Rubber Rings

Zhu Hong-Li[1], Wang Xian-Rong[1], and Zhu Hong-Yan[2]

[1] College of Information Science and Engineering,
Henan University of Technology, Zhengzhou, 450001, China
[2] School of Electrical and Automation Engineering,
Hefei University of Technology, Hefei, 230009, China
zhl780221@163.com

Abstract. Defect detection is a largely critical step in the production of rubber rings. This paper presents a algorithm of edge detection based on digital image processing. Firstly, the target image should be preprocessed to eliminate noise; secondly, the Sobel operator is used for edge detection, and thirdly these images are dilated by an algorithm of mathematical morphology, and are emulated by the MATLAB software. The result indicates that this algorithm functions well in repairing fissures between fine defects, enhances the brightness of images, and has a better detection precision.

Keywords: Digital Image Processing, Rubber Ring, Adaptive median filter, Sobel operator, Dilation.

1   Introduction

The edge of image is a collection of points with disconnected or sharply-various gray level values of image, the edge detection of rubber rings is to draw the border line of defects and background. In the course of producing rubber rings on the industrial scale, various forms of defects will appear inevitably, such as rough cutting, uneven thickness, rough edges, bubbles, and soon. Taking bubbles---one kind of defects in rubber rings for example, in this paper we provide a method of defect detection algorithm based on digital image processing, and then analyses its course and list its result[1].

2   Algorithm Research

The target image generally includes noises, so the image preprocessing is necessary to eliminate noises before edge detection is carried out by the Sobel operator. Finally the mathematical morphology can be used to dilate the disconnected images and link the breaks between defects, for obtaining continuous edges, i.e. a better detection effect. The algorithm block diagram is drawn in Fig.1.
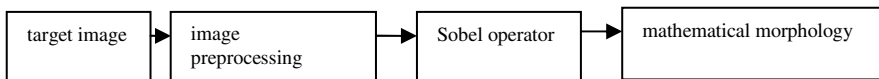
Fig. 1. Algorithm block diagram

The noise pollution produced by shooting, saving, and transmitting images of rubber rings can result in degraded phenomenon of image, such as image distortion, image blurring. It can explained by many factors, for example, the motion blur that is produced by relative movement between cameras and rubber rings in the course of detection; noise existing in the imaging system; the surrounding effects of the detection system. Images preprocessing of rubber rings with defects is aimed at removing noise as much as possible for the convenience of image segmentation.

Mean filter, median filter and adaptive median filter are the three mostly-common algorithms to handle noise. Each of them has its typical features: mean filter is also called linear filtering and can effectively handle additive noise by replacing a previous pixel with the average of several pixels around it. However, it easily produces the image blurring.

Median filter replaces the value of a pixel by the median of the gray levels in the neighborhood of that pixel: $f^{'}(x, y) = \underset{(s,t) \in S_{xy}}{median}\{g(s,t)\}$ It can effectively handle impulse noise. Adaptive median filter, an upgrading median filter, compared with the traditional median filter, can handle impulse noise with probabilities even larger, and preserves details while smoothing non-impulse noise. So the latter is adopted to handle noise in the paper [2].

Adaptive median filtering algorithm works in two levels, denoted level A and level B, as follows:

level A : A1=Zmed—Zmin
            A2=Zmed—Zmax
If A1>0 AND A2<0, Go to level B
Else increase the window size of Sxy
If window size ≤Smax, repeat level A
Else output Zxy
level B : B1=Zxy—Zmin
            B2=Zxy—Zmax
If B1>0 AND B2<0, output Zxy
Else output Zmed
Here, Zmin=minimum gray level value in Sxy
Zmax=maximum gray level value in Sxy
Zmed=median of gray levels in Sxy
Zxy=gray level at coordinates(x, y)
Smax=maximum allowed size of Sxy

We can conclude from the algorithm that the purpose of level A is to determine if the median filter output, Zmed, is an impulse or not. If not, we go to level B to judge whether the coordinates (x, y) waiting for being processed is a noise or not. If B1>0 and B2<0 is true, then the coordinates (x, y) cannot be an impulse, the algorithm outputs the unchanged pixel value, Zxy. Distortion is reduced in the image. That means the coordinates isn't processed at all. If B1>0 and B2<0 is false, then the coordinates (x, y) is a noise, and the algorithm outputs Zmed.

If A1>0 and A2<0 is false, then the median value is a noise. The algorithm then increases the size of the window until a non-impulse is found. If the maximum window

size is reached, and A1>0 and A2<0 is false, the algorithm returns the value of coordinates(x, y), Zxy.

The test indicates that adaptive median filtering can effectively eliminate the noise of rubber rings, and more importantly, it can also protect the image details[3]. Therefor, this paper uses adaptive median filtering to do the image preprocessing. The result of tests is indicated in the Fig.2, in which the Fig.2(a) is an image with salt and pepper noise, Fig.2(b) is the resulting image of adaptive median filtering.

(a) salt and pepper noise   (b) adaptive median filtering
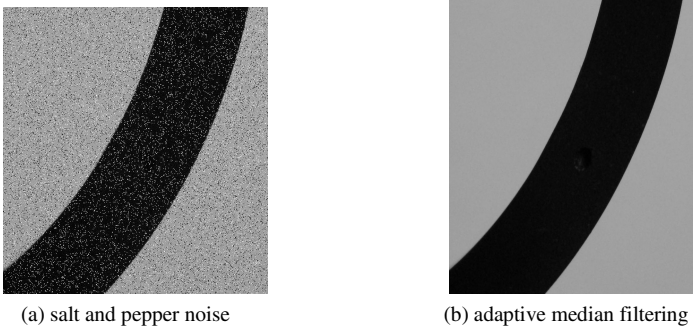
Fig. 2. Adaptive median filter

Edge detection is one significant part of digital image processing. The final aim of defect detection in rubber rings is to separate those edges with defects from all edges that are boundary lines where targets and background are to be extracted. The operators of edge detection based on the differential method include Roberts operator, Prewitt operator, Sobel operator, Candy operator, Laplacian operator, LoG operator and so on. The first four are operators based on first derivative, while the last two belong to those based on second derivative. In the process of algorithm, the appropriate threshold value can be selected to extract the edges through the convolution and operations of the template and every pixel in the image. What the first derivative differs from the second derivative is that the former corresponds the place of edge with its maximum, while the later does it with its zero crossing point.

Among edge detection operators based on differential method, Roberts operator adopts the difference of two adjacent pixels on the diagonal direction as the approximate gradient amplitude, to detect edges, with a higher positioning accuracy and the better detection effects on the vertical and horizontal edges than oblique edges. However it is sensitive to noise. Prewitt operator and Sobel operator, the two most popular in practice, both use the gray weighting algorithm of all adjacent points surrounding the pixel point, and detect edges according to the principle of reaching the extreme value on the edge point. Prewitt operator work much easier, but is inferior to Sobel operator in removing noise. Sobel operator can detect edge points more accurately and smooth noise, but the range of edges it detects is wider. Canny operator is the best one to detect step edges by relying on the traditional differentiation, with a higher capacity of de-noising. But it easily smoothes some information of edges away and has a slow calculation speed [4].

Laplacian operator, a linear second order differential operator, is very sensitive to noise, and generally not directly used for edge detection. LoG operator is an improved

Laplacian operator, adopts the Gaussian function to undergo the smoothing filter, and carries out the Laplacian operation to the smoothed images. The use of Laplacian operator is to provide a picture that can determine the location of the edge.

Edge detection is one significant part of digital image processing. The final aim of defect detection in rubber rings is to separate those edges with defects from all edges that are boundary lines where targets and background are to be extracted.

Sobel operator is a group of directional operator, detects edges from different directions. The way it works is not to simply calculate the average value and then the difference, but to enhance weight of the central pixel's four directional pixels. [4].

Directional operator uses a group of template to do convolution with the same pixel in the image, and selects the biggest value as the edge intensity, and chooses the corresponding direction as the edge direction.

Its gradient:

$$Gx = (z_7 + 2z_8 + z_9) - (z_1 + 2z_2 + z_3), Gy = (z_3 + 2z_6 + z_9) - (z_1 + 2z_4 + z_7)$$

Prewitt operator is the average filtering, while Sobel operator is the weighted average filtering. In Prewitt operator the pixel neighborhood has the equal effect on the current image pixels, while in Sobel operator, there are different weights according to the distance between the current image pixel and the neighborhood pixels, that is to say, the smaller the distance, the bigger the weight. The weight 2 of Sobel operator achieves some smoothing effects by increasing the importance of the central point. Due to the introduction of average factors, it has some certain function of smoothing the random noise in images. Because Sobel operator is based on the calculus of two rows or two columns apart, edge elements on the both sides are enhanced, and the edge appears thick and bright [4]. The experimental result is shown in the Fig.3.

Fig. 3. Sobel operator's edge detection

Mathematical morphology is a subject established on the mathematical theory, its mathematical basis is the set theory, putting an image as a set. The fundamental idea is to use structural elements with certain shape to measure and extract the corresponding shape in the image, in order to achieve the purpose of digital image processing. The

basic operations of the mathematical morphology are: dilation, corrosion, open and close.

Dilation is the most basic morphological transform of mathematical morphology. Assume A and B is a set of the Z^2 , then the dilation of A by B can be defined by:

$$A \oplus B = \{z \mid (B')_z \bigcap A \neq \varnothing\}$$

in which B' is the mapping of the B, with its definition : $B' = \{x \mid x = -b, b \in B\}$. The dilation of A by B is first to get the mapping B' of B, and to use z to displace on the mapping B'. The dilation of A by B is the collection of the displacement z. Meanwhile, the intersection of B'and A rewarding can not be empty. So it can be rewritten as: $A \oplus B = \{z \mid [(B')_z \bigcap A] \subseteq A\}$, in which B'can also be called as the structural element of dilation. The schematic diagram is shown in Fig.4.
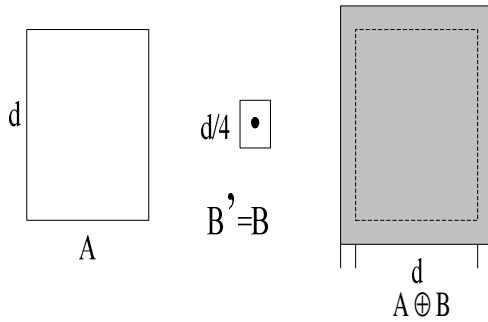
Fig. 4. Schematic diagram

In the above diagram, A is a collection, B is the structural element, the black spot is the original point of B. Because B is symmetrical around its original point, the mapping of B is the same as B. The method of dilation is to compare the original point of B respectively with the points in A, and if there is one point in B falling in the range of A, then the corresponding point of B's original point is the dilated point. Fig.5 is the resulting picture of the dilation, which includes all ranges of A (dashed part). The results of dilation vary according to different origin coordinates of structural elements.

As for the binary image, if the pixel point of B is represented by 1, then the dilation is to change the 0 value of B's neighboring pixel point as the 1 value of pixel point, expanding its ranges to all directions.

To repair the fissure is the most commonly-used application of dilation, such as the defect detection of black rubber rings this topic focuses. Although Sobel operator can better detect the edges of defects in image, these edges are neither continuous nor completely integrate, the morphological dilation can amend very well the fissures between defects.

3   Result

Compared with those before dilation, defects after dilation have apparently greater brightness, and fissures between fine edges are connected. A better detection is accomplished. The result is shown in Fig.5.

Fig. 5. Result

4   Summaries

This paper has discussed the algorithm of defect detection in rubber rings, analyzed advantages and disadvantages of several common algorithms, and finally presented a new defect detection algorithm, and carried out experimental emulation with MATLAB, whose result demonstrates its detection precision is higher than those traditional edge detection algorithms.

References

[1] Liu, C., Li, X., Lu, C., Hu, C.: Research on Detection Method for Convex Dot Defect of Rubber Ring. Journal of Electronic Measurement and Instrument (supplement), 178–181 (2008)
[2] Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 2nd edn., pp. 262–264, 241–242. Publishing House of Electronics Industry, Beijing (2007)
[3] Zhao, F., Luan, X., Sun, Y.: Edge Detection Operators in Digital Image Processing. Communication and Information Processing 28(3), 68–72 (2009)
[4] Castleman, K.R.: Digital Image Processing, pp. 402–403. Publishing House of Electronics Industry, Beijing (1998)
[5] Chang, C.-L., Juan, C.-L.: The Compare Research on Defect Detection. New Technology New Product (1), 3–4 (2009)

Research on the Theory and Application of Touchscreen Tactile Feedback Based on Soft Vibration

Lu Xiang[1] and Cao Yue[2]

[1] Logistics Information Engineering Department,
Logistics Engineering University,
Chongqing, China
nuaaphoenix@163.com
[2] Information Engineering Department,
Chongqing City Vocational College,
Chongqing, China
Cymoon1982@qq.com

Abstract. Users can't feel the true tactile feedback through the virtual keys of touchscreen, and the feeling will lead to input leakage or wrong input easily. Soft vibration, which is used to simulate force feedback of mechanical keys, is a technology of tactile feedback, and this technology can solve the problem of tactile feedback in touchscreen. Aiming at the touchscreen, several ways of tactile feedback are analyzed. The soft vibration is proposed, and the process of soft vibrotactile feedback by some input ways is researched. Compared with other techniques of tactile feedback, the soft vibratactile feedback has many advantages, such as higher realistic degree, simple structure, friendly HMI, and so on.

Keywords: Soft vibration, Touchscreen, Tactile feedback, Human-machine interface.

1 Introduction

The rapid development of touch-sensing technology has subverted the traditional mechanical buttons and potentiometers concept, which greatly improved the friendly human-machine interface, allowing users to operate electronic device directly without the help of the buttons and trackball[1][2]. However, we note that the force feedback of mechanical buttons makes the user feel each input behavior without our vision, and this is a good feature that touchscreen does not have. Although we can use the screen buttons trigger the sound or vibration of the way to remind the user's input operation, there is inadequate to tell the the input is correct or not. In addition, each key sounds a single form of feedback or vibration feedback, and in some cases, it may cause problems to the users[3][4].

Based on the analysis above, this paper takes touch screen input as the research object, identifies the user input with soft vibration in order to enhance the friendly touch-screen input, and reduces the input leakage and error input.

2   Implementation of Tactile Feedback

Despite the touch screens' touch capabilities bring a friendly interface to users, it also creats some new problems: users typically get the result of feedback effects through sound or variational virtual button icon, rather than the force feedback effects of mechanical buttons or slider potentiometer. All that would result in actual use efficiency decreased[5][6][7].

Nowadays, a concrete realization of tactile feedback for the main mode are as followed: vibration of a single-stage, touch screen with haptic vibration feedback module integration and soft vibration.

2.1   Vibration of a Single-Stage

Vibration of a single-stage is a basic way.to achieve tactile feedback in vast majority of touch screen devices. It produces vibration when touch screen's virtual keys are touched, which hints the act effectively. This approach has the advantage of simple, low cost, high penetration rate. But the disadvantage is obvious: vibration is felt stiff, so that frequent vibration could lead to rapid decrease in user satisfaction rate, and it is only for ordinary compact handheld devices[8]. For small or lager devices, especially for large equipment, such a partial or whole components of vibration can cause loose, short circuit, damaged or even other more serious accidents.

2.2   Touchscreen with Haptic Vibration Feedback Module Integration

There is a substantial difference between single-stage vibration effect and the experience feeling of the physical buttons. The ultimate solution for eliminating this fundamental difference is integrating touch screen with tactile feedback module [9][10]. Nokia, for example, recently developed a new technology, which allows a place on the touch screen vibration region[11]; Apple Inc. also applied a soft touch screen patent a few years ago, which leads a real touch feeling to users as touching a true keyboard by planting a transparent touch pad point under the screen[12].

In addition, Microsoft has applied for a patent, and it uses a new material to change its surface shape according to the wavelength of ultraviolet radiation. In this patent, Microsoft would make such a shape memory and touch-screen pixel-sized, cover the surface of the touch screen, and then determine the position of the fingers about to touch the screen with the camera installed under the screen. The user's fingers feel like stone, wood, glass and other surfaces touch through to the appropriate location of injection specific wavelengths of ultraviolet light[13].

These new materials and new technology even though produce a more satisfactory tactile feedback, but it also has some of its inherent defects. Based on consideration of cost and versatility, the current philosophy of touch screen with haptic vibration feedback module integration still remains in the laboratory stage.

2.3   Soft Vibration

Tactile feedback of Single-stage is mechanical, but simple and low cost. The touch screen with haptic vibration feedback module integration works well, similar to

physical buttons. However it is not available in the current operability related to new materials, new technology and high-cost, low- benefits.

Based on the analysis above, this paper presents a flexible way of tactile feedback named soft vibration: achieves vibration by the function of touch screen's virtual keys and slider. The soft vibration is characterized by changing the vibration frequency, waveform, amplitude and period, so that users feel tactile feedback similar to the physical effects.

3   Implementation of Soft Vibration

The implement of soft vibration is as shown in Fig.1. The module of soft vibration is composed of a number of micro-vibrator, which distribute in the points below the touch screen. When the user's finger make a move on the touch screen, such as clicking or sliding, Touch Screen Controller A analyses user's operation intent, and sends the command which should be executed to the corresponding MCU applications mandating the next step, according with the intensity, and direction of touch point. At the same time, the module achieves tactile feedback with different frequency, waveform, amplitude and period of vibration by several vibrators on the bottom of the screen.
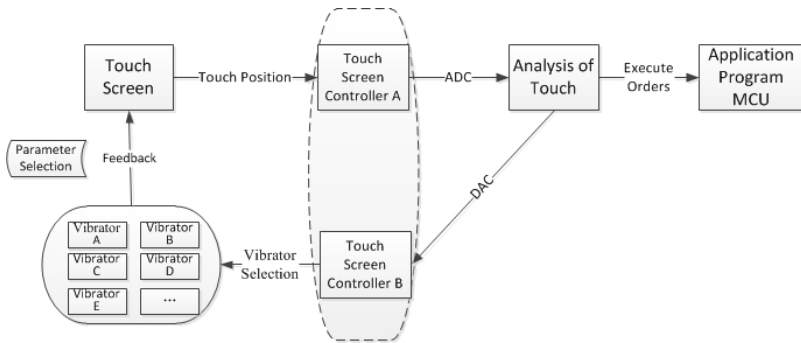
Fig. 1. Process of soft vibrotactile feedback

3.1   Implement of Soft Vibration by Clicks

The touch screen with five micro-vibration sensor, for example, as shown in Fig.2. Let the vibrator A vibration feedback force define as N when X point of action which user clicks is on the top of the vibrator A. Then when the user clicks the X point as shown in Fig.2, the feedback force of vibrator A is αN, the feedback force of vibrator B is βN, the feedback force of vibrator C is γN, the feedback force of vibrator D is δN, and the feedback force of vibrator E is εN. Where $\alpha, \beta, \gamma, \delta$ and ε are the feedback coefficients, which behave a non-linear downward trend according to the distance between the vibrator and the X point: the greater the distance, the coefficient is smaller. When the X point is infinitely close to point A, the feedback force of point B, point C, point D and point E is close to zero.
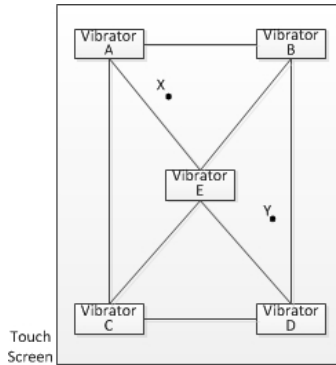
Fig. 2. Position distribution of vibration sensors

3.2   Implement of Soft Vibration by Sliding

Sliding action shows slightly more complex than the clicks. In general mechanical slider, the slider from the initial state, is to overcome static friction resistance; once it starts sliding, dynamic friction decreases as the acceleration; there is a strong force feedback before the end of the sliding when speed is reduced to zero. Therefore, the sliding action of the touch feedback needs three stages to reflect actual sliding.

As shown in Fig.2, when the user's finger slides from X point to Y point on the touch screen, the first step is to determine whether the action is continuous clicks or continuous sliding. It is bounded by a fixed time τ. When the time t between two clicks $\prec \tau$, the action is judged as a sliding. When the time t between two clicks $\geq \tau$ the action is judged as two clicks. As the sliding action, the feedback force from X point to Y point is $\alpha'\alpha$ N, where the α' is a dynamic parameter: it is a larger number at the beginning of sliding, and then maintains a constant value, and finally increases again when stops.

3.3   Some Problems of Soft Vibration

The technical solutions of Soft vibration need to introduce multiple micro-vibrator, and a number of   simultaneously working micro-vibrator will produce greater impact on each judgment of touch action, and then increase the production of false judgments.

In addition, the touch screen controller itself would produce some internal noises for other multiple sensors such as accelerometers, gravity sensing sensors, distance sensors, light sensors and other sensors. The sweat, oil, cream, stains of user fingers' would disturb the judgment of touch position and touch action.

These aspects are needed to be studied carefully before soft vibration being used actually.

4   Summaries

Tactile feedback is one of the most hot application researches, and its realizability directly affects the degree of friendly in HMI of touch screen.

Compared to the simple single-stage vibration tactile feedback and tactile feedback touchscreen with tactile feedback module integration ideal, the soft vibration overcomes inherent deficiencies of the first two, but also takes some advantages from both into account. In the optimized design, it is the integration of the touch screen with tactile feedback module, and then uses the bottom of the interactive system design, simulates to approximate the original entity key operating experience to improve the HMI friendly.

Although the tactile feedback of soft vibration will lead to increased hardware costs and increased volume of touch screen devices, but with an increasing number of variety of hand-held device or fixed device touch panel is increasing steadily, increasing the proportion of the virtual keyboard application, this problem will continue to be resolved.

References

[1] Jangwoon, K., Jaewan, P., HyungKwan, K., et al.: HCI (Human computer interaction) using multi-touch tabletop display. In: PACRIM 2007, pp. 391–394. IEEE (2007)

[2] Jefferson, H.: Low-cost multi-touch sensing through frustrated total internal reflection. In: Proceedings of the 18th Annual ACM Symposium on User Interface Software and Technology, Seattle, pp. 315–319 (2005)

[3] Forlines, C., Wigdor, D., Shen, C., et al.: Direct-touch vs. mouse input for tabletop displays. In: CHI 2007, pp. 647–656. ACM (2007)

[4] Weng, X.-P.: Touch sensing technology and applications, pp. 2–3. Beijing University of Aeronautics &Astronautics Press, Beijing (2010)

[5] Wilson, A.D.: PlayAnywhere: A compact interactive tabletop projection-vision system. In: Proc. UIST 2005, pp. 83–92. ACM (2005)

[6] Cheng, Z.-H., Yang, S.-F.: Touchscreen Applications, pp. 1–6. Posts & Telecom Press, Beijing (2010)

[7] Ankur, A., Shahram, I., Manmohan, C., et al.: High precision multi-touch sensing on sur faces using overhead cameras. In: Proceedings of The Second IEEE International Workshop on Horizontal Interactive Human-Computer Systems, Rhode Island, pp. 197–200 (2007)

[8] Ye, Y.-W., Qian, J.-W.: Principles and advantages of Piezoelectric multi-touch technology. Advanced Display, 47–52 (2009)

[9] Benko, H., Wilson, A.D., Baudisch, P.: Precise selection techniques for multi-touch screens. In: CHI 2006, pp. 1263–1272. ACM (2006)

[10] Wu, M., Shen, C., Ryall, K., et al.: Gesture registration, relaxation, and reuse for multi-point direct-touch surfaces. In: Proc. of the 1st IEEE International Conference on Horizontal International Interactive Human Computer System, pp. 183–190. IEEE Computer Society, Washington DC (2006)

[11] Application development trends of touch technology–Multi-couch solution diversity (EB/OL) (2011), http://www.touco.cn

[12] Dietz, P., Leigh, D.: Diamond touch: a multi-user touch technology. In: Proc. of the 14th Annual ACM Symposium on User Interface Software and Technology, UIST, pp. 219–226. ACM Press, New York (2001)

[13] Nestler, S., Echtler, Dippon, A.: Collaborative problem solving on mobile hand-held devices and stationary multi-touch interfaces. In: Proc. of Workshop on Designing Multi-touch Interaction Techniques for Coupled Public and Private Displays, pp. 36–40 (2008)

Interior Ballistic Research on Gas and Steam Launch Power System

Chen Qinggui, Qi Qiang, Zhou Yuan, Zhao Ruyan, and Wang Bin

Naval Aeronautical and Astronautical University, China

Abstract. To guarantee the missile leaves launch unit smoothly, launch interior ballistic computation model based on gas and steam launch power is improved. One missile is taken for example to compute its interior ballistic. And changing laws of interior ballistic parameters are obtained. Results indicate that the modified model can decrease gas pressure in the unit effectively when missile leaves the launch unit. This is of some reference value for the design of gas and steam launch power system.

Keywords: gas and steam launch power system, submarine-launched missile, interior ballistic.

1   Introduction

Gas and steam launch power system is usually adopted to launch Submarine-launched missile. It is composed of ignition-assurance framework, gas generator, cooling vehicle and syphon. Mixture of gas and vapor is used to lauch missile. In the missle-launch process, change of interior ballistic parameters has much to do with the launch of missile. While the working time of gas generator affects interior ballistic parameters. To get satisfying interior ballistic parameters, interior ballistic model[1] is modified and computed.

2   Launch Interior Ballistic Model

Flux of gas and cooling water that enter into the launch unit is expressed as

$$m_g = \frac{gA_t\mu\sigma_f}{C^*}\int_0^t p_c \mathrm{d}t + m_{g0} \tag{1}$$

$$m_l = \alpha\mu_l n_l s_l \sqrt{2\rho_l\lambda}\int_0^t \sqrt{P_c}\, \mathrm{d}t + m_{l0} \tag{2}$$

Where A_t is gular area of nozzle; C^* is characteristic velocity of powder; σ_f is pressure-resuming coefficient; μ is flux coefficient of nozzle; m_{g0} is mass of ignition powder; u_l is flux coefficient of water-spurting hole; n_l is number of water-spurting hole; λ is water-spurting pressure difference coefficient; m_{l0} is previous water mass.

The motion equation of missile is calculated as

$$Ma = (1 + x_k)p_t s_t - F \tag{3}$$

Where F is resisting force; x_k is kinetic energy coefficient; P_t is pressure in the launch unit; S_t is sectional area of launch unit.

Interior ballistic equations[1-2]of different processes are as follows:

(1) Pyrogenation Process.

$$
\begin{cases}
Ma = (1 + x_k)P_t S_t - F \\[2mm]
t_t = \dfrac{x_e m_g c_{vg} t_v + m_l c_l t_l + m_a c_{va} t_a - (\frac{1}{2}MV^2 + \int_0^l F dl)}{m_g c_{vg} + m_l c_l + m_a c_{va}} \\[4mm]
P_t = \dfrac{x_p (R_g m_g + R_a m_a) T_t}{S_t (l_0 + l)}
\end{cases}
\tag{4}
$$

(2) Boiloff Process.

$$
\begin{cases}
Ma = (1 + x_k)P_t S_t - F \\[2mm]
\dfrac{P_t}{x_p} = \dfrac{(R_g m_g + R_a m_a) T_s}{S_t (l_0 + l)} + \dfrac{R_l T_s}{v - b} + \dfrac{A_2 + B_2 T_s + C_2 e^{-5.475 T_s / T_k}}{(v - b)^2} + \\[4mm]
\dfrac{A_3 + B_3 T_s + C_3 e^{-5.475 T_s / T_k}}{(v - b)^3} + \dfrac{A_4}{(v - b)^4} + \dfrac{B_5 T_s}{(v - b)^5} \\[4mm]
x = \dfrac{x_e m_g c_{vg} t_v + m_l c_l t_l + m_a c_{va} t_a - (m_g c_{vg} + m_l c_l + m_a c_{va}) t_s - (\frac{1}{2}MV^2 + \int_0^l F dl)}{m_l (\Delta H - p_s v_s'')}
\end{cases}
\tag{5}
$$

(3) Superheat Process.

$$
\begin{cases}
Ma = (1 + x_k)P_t S_t - F \\[2mm]
\dfrac{P_t}{x_p} = \dfrac{(R_g m_g + R_a m_a) T_t}{S_t (l_0 + l)} + \dfrac{R_l T_t}{v - b} + \dfrac{A_2 + B_2 T_t + C_2 e^{-5.475 T_t / T_K}}{(v - b)^2} \\[4mm]
+ \dfrac{A_3 + B_3 T_t + C_3 e^{-5.475 T_t / T_K}}{(v - b)^3} + \dfrac{A_4}{(v - b)^4} + \dfrac{B_5 T_t}{(v - b)^5} \\[4mm]
t_t = \dfrac{x_e m_g c_{vg} t_v + m_l c_l t_l + m_a c_{va} t_a - [(c_l - c_{pl}) t_s + \Delta H - pv] m_l - (\frac{1}{2}MV^2 + \int_0^l F dl)}{m_g c_{vg} + m_l c_{pl} + m_a c_{va}}
\end{cases}
\tag{6}
$$

Where x_e is energy coefficient; x_p is pressure coefficient; c_{vg}, c_{va}, c_l are specific heat at constant volume of gas, air and specific heat of water; c_{pl} is specific heat at constant pressure; v is specific volume; ΔH is latent heat; A_2, A_3, A_4, B_2, B_3, B_5, C_2, C_3 are coefficients of M-H equation of state[3].

3   Modification of Launch Interior Ballistic Model

Above equations are concluded in the situation where gas and vapor in the launch unit change constantly before gas generator stops work[4]. To guarantee the missile leaves the launch unit smoothly, gas pressure in the launch unit must be reduced when the missile leaves the launch unit. It is demanded that gas generator stops work before the missile leaves the launch unit. Gas in the unit don't change in mass during the time from gas generator's stopping work to missile's leaving unit. Fixed-mass gas does swelling work. In this process, equations are same to that of variational-mass gas except in the energy equations. Interior ballistic equations of fixed-mass gas doing work are as follows [5].

(1) Boiloff Process.

$$
\begin{cases}
Ma = (1 + x_k) P_t S_t - F \\[2mm]
\dfrac{P_t}{x_p} = \dfrac{(R_g m_g + R_a m_a) T_s}{S_t (l_0 + l)} + \dfrac{R_l T_s}{v - b} + \dfrac{A_2 + B_2 T_s + C_2 e^{-5.475 T_s / T_k}}{(v - b)^2} + \\[4mm]
\qquad \dfrac{A_3 + B_3 T_s + C_3 e^{-5.475 T_s / T_k}}{(v - b)^3} + \dfrac{A_4}{(v - b)^4} + \dfrac{B_5 T_s}{(v - b)^5} \\[4mm]
x = \dfrac{U_{g_{i-1}} + U_{l_{i-1}} + U_{a_{i-1}} - (m_g c_{vg} + m_l c_l + m_a c_{va}) t_s - \frac{1}{2} M(V_i^2 - V_{i-1}^2) + \bar{F}(l_i - l_{i-1})}{m_l (\Delta H - p_s v_s'')}
\end{cases}
\tag{7}
$$

(2) Superheat Process.

$$
\begin{cases}
Ma = (1 + x_k) P_t S_t - F \\[2mm]
\dfrac{P_t}{x_p} = \dfrac{(R_g m_g + R_a m_a) T_t}{S_t (l_0 + l)} + \dfrac{R_l T_t}{v - b} + \dfrac{A_2 + B_2 T_t + C_2 e^{-5.475 T_t / T_K}}{(v - b)^2} \\[4mm]
+ \dfrac{A_3 + B_3 T_t + C_3 e^{-5.475 T_t / T_K}}{(v - b)^3} + \dfrac{A_4}{(v - b)^4} + \dfrac{B_5 T_t}{(v - b)^5} \\[4mm]
t_t = \dfrac{U_{g_{i-1}} + U_{l_{i-1}} + U_{a_{i-1}} - [(c_l - c_{pl}) t_s + \Delta H - pv] m_l - \frac{1}{2} M(V_i^2 - V_{i-1}^2) + \bar{F}(l_i - l_{i-1})}{m_g c_{vg} + m_l c_{pl} + m_a c_{va}}
\end{cases}
\tag{8}
$$

Where $U_{g_{i-1}}$, $U_{l_{i-1}}$, $U_{a_{i-1}}$ are energy of gas, vapor and air separately at time $i - 1$; V_i, l_i and V_{i-1}, l_{i-1} are velocity and displacement of missile at time i and $i - 1$; $\bar{F}$ is average resistanting force of time i and $i - 1$.

According to Thaler progressional expanding formula, the displacement and velocity of missile are

$$
l_n = l_{n-1} + \Delta t v_{n-1} + \frac{1}{2} \Delta t^2 a_{n-1} + \frac{1}{6} \Delta t^3 \dot{a}_{n-1}.
\tag{9}
$$

$$
v_n = v_{n-1} + \Delta t a_{n-1} + \frac{1}{2} \Delta t^2 \dot{a}_{n-1}.
\tag{10}
$$

4   Examples and Analysis

Interior ballistic models are computed with Matlab language, changing laws of interior ballistic parameters ($l \sim t$, $v \sim t$, $a \sim t$, $P_t \sim t$, $T_t \sim t$) are obtained by using iterative method. For the convenience of comparison, computation conditions are that mass of missile, velocity of leaving unit and launch depth are same. Changing laws of interior ballisitc parameters are shown in figure (1)-(5).
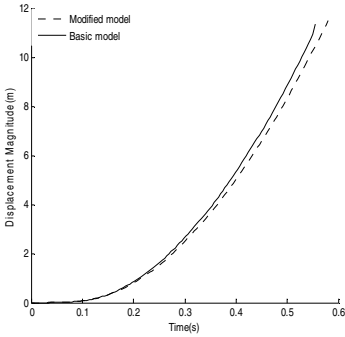
Fig. 1. Displacement-time curve
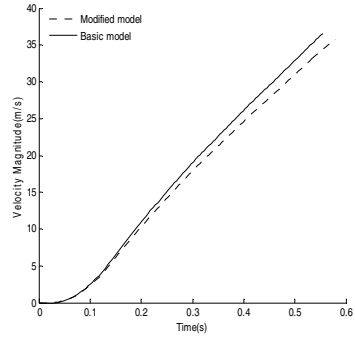
Fig. 2. Velocity-time curve
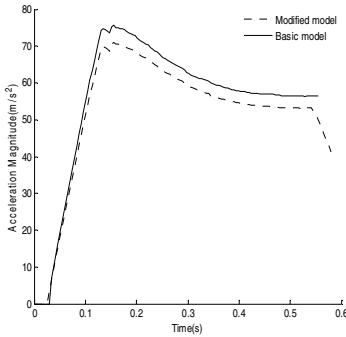
Fig. 3. Acceleration-time curve
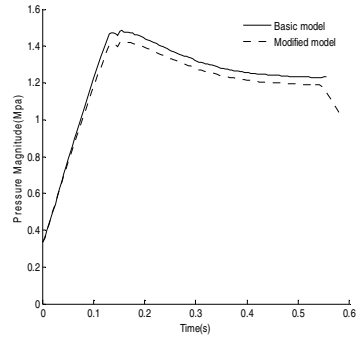
Fig. 4. Pressure-time curve
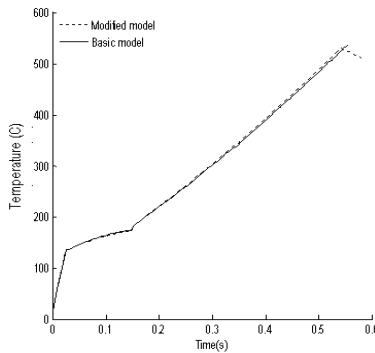
Fig. 5. Temperature-time curve

From the simulation results, following conclusions can be drawn.

(1) The missile moves at an even acceleration in the launch unit approximately so that launch interior ballistic is balanced.

(2) If gas generator stops work before the missile leaves the launch unit, gas pressure in launch unit can be decreased effectively. Effects of unit meatus pressure on missile can be diminished greatly.

(3) The changing law of acceleration is similar to that of pressure. The missile moves slowly in initial motion stage, while acceleration increases rapidly. With the pick up of missile's velocity, the increase of launch unit's cubage exceeds that of vapor's supply, which makes the pressure decrease and the acceleration reduce.

5   Summaries

Interior ballistic model based on gas and steam launch power is modified and computed in this paper. Results show that gas pressure in the launch unit can be decreased effectively if gas generator stops work before missile leaves the launch unit. The missile can leave the launch unit smoothly in the situation that targets of launch interior ballistic are achieved. This is of great importance for the design of gas and steam launch power system.

References

[1] Zhao, X.-F., Wang, J.-J.: Interior Ballistics for Submarine-to-ground Ballistic Missile Launch System, pp. 33–43. Harbin Engineering University Press (2001)
[2] Li, X.-H., Wang, J.-J.: Launch Power System for Submarine-to-ground Missile, pp. 79–88. Harbin Engineering University Press (2000)
[3] Martin, J.J., Hou, Y.C.: Development of An Equation of State for Gases. AIChE J. 1, 142–149 (1955)
[4] Zhang, R.-J., Bao, F.-T.: The CAD Development for Interior Ballistic Design of Gas and steam Launch Power System with Water Injection in Block, pp. 11–39. Westnorth Engineering University (2005)
[5] Zhang, R.-J., Bao, F.-T.: Computation of Internal Ballistic Properties Between Gas and Steam Launching Systems in Two Different Modes of Water Injection. Journal of Solid Rocket Technology, 5–9 (2005)

Hierarchical Image Analysis Based on Fuzzy Control

Zhiyong Zhang[1], Xiaoning Li[1], Yueqi Liu[2], and Xiaofeng Li[3]

[1] Changchun Normal University, The Computer Science and Technology Institute,
Changchun, Jilin Province 130032, China
[2] Changchun Normal University,
Changchun, Jilin Province 130032, China
[3] Jilin University,
Chang Chun, China

Abstract. Based on fuzzy mathematics gained considerable development in the theory and application have achieved fruitful results impressive. Applications of fuzzy mathematics has been related to automatic control, image and text recognition, artificial intelligence, geology, seismology, medical diagnostics, weather analysis, aviation, aerospace, ship train car to drive, traffic management, decision evaluation, business management and socio-economic many aspects.

Keywords: Fuzzy mathematics, Image analysis.

1   Introduction

The application of automation technology is very active and fuzzy as a fruitful field. The automatic control of well-known authority Austrom once pointed out: fuzzy logic control, neural network control and analysis of expert control are the three typical intelligent control methods[1].

1974, British scholar EHMamdani fuzzy set theory the first time successfully applied in the control of boiler and steam engine, the first in the field of automatic control of fuzzy control to open applications in practical engineering precedent.

In only 30 years, fuzzy mathematics gained considerable development in the theory and application have achieved fruitful results impressive. Applications of fuzzy mathematics has been related to automatic control, image and text recognition, artificial intelligence, geology, seismology, medical diagnostics, weather analysis, aviation, aerospace, ship train car to drive, traffic management, decision evaluation, business management and socio-economic many aspects.

Assuming rigid body motion binocular device for general E, the image is under the camera from the eyes of the metric calculation of reconstruction of the scene [2]. The problems with the above sports scene reconstruction problem are equivalent. Because binocular device for general rigid motion E of the images obtained with the eyes stationary object device on the scene in a fixed position relative to the rigid body motion for the resulting image is the same.

The early 90s, fuzzy appliances popular in Japan, to Japanese companies brought great commercial profits, but also to promote the U.S., Europe and other countries to further promote the development of fuzzy technology. In 1985 the world's first fuzzy

logic chips in the United States came the famous Bell Labs, which is fuzzy technology marks another milestone toward practical. Japan, the United States, Germany and many other well-known companies[3] are actively engaged in research in this area, have developed a number of commercial fuzzy logic chips in 1986, Japan established a fuzzy controller hardware system (fuzzy control specific devices). The late 80s of last century to the mid-90s has a fuzzy approximate reasoning, fuzzy adaptive control, fuzzy adaptive neural networks and fuzzy inference systems. The application of fuzzy technology to inject new vitality opened up very attractive bright prospects.

Our research in the field of fuzzy theory in the world advanced level, has published dozens of writings of the fuzzy area. In engineering applications is weak, continuous monitoring system has been proposed to facilitate engineering design methods and fuzzy integrated control methods. Since the late 90s of last century there was fuzzy appliance control[4].

2   The Basic Model of Visual Images

Satisfy the following properties: point corresponding to the spatial point sets and the difference between sets of points in space an affine transformation, which makes the existence of affine transformation, denoted Claimed two points correspond[5], for the corresponding affine point. For example, the vertices of two six-sided composition of two images corresponding to two points is the affine point correspondence, as between any two rectangular exists an affine transformation to transform one set of vertices to another vertex set of knives. Stated in the following series of questions exist in both the corresponding affine points, so they can apply the method given in this section to solve.

Wavelet transform is a space (time) and frequency of the local transformation, which can effectively extract information from the signal as fuzzy mathematics. By dilation and translation operations functions such as function or signal can multi-scale refinement analysis, Fourier transform to solve many difficult problems can not be solved. Contact the application of wavelet mathematics, physics, computer science, signal and information processing, image processing, seismic exploration, and other disciplines. Mathematicians believe that wavelet analysis is a new branch of mathematics, it is the functional analysis, Fourier analysis, like coherence analysis, numerical analysis of the perfect crystal; signal and information processing experts believe that wavelet analysis is a time - scale analysis and multirsolution analysis a new technique, which signal analysis, speech synthesis, image recognition, computer vision, data compression, seismic exploration, atmospheric and ocean wave analysis and other aspects of studies have meaning and application of scientific results.

The main purpose of signal analysis is to find a simple and effective method of signal conversion, the signal contains important information can be revealed. Wavelet analysis is a kind of signal analysis, before the advent of the wavelet analysis, Fourier transform is the most widely used signal processing, one of the best means of Fourier transform is a time domain to frequency domain conversion tool to each other, from the physical sense, the real Fourier transform is to decompose the waveform into a superposition of sine waves of different frequencies and. Fourier transform is the physical meaning of this important decision the Fourier transform in signal analysis and signal processing in a unique position.

Fourier transform used in both directions infinitely extended sinusoidal wave as the orthogonal basis functions, the development of a periodic function into a Fourier series, the development of non-periodic function into a Fourier integral, Fourier transform functions for spectrum analysis, reflection time of the signal spectrum, better reveal the characteristics of a stationary signal. Wavelet analysis is a new transformation method, which inherited and developed the short time Fourier transform thinking localized, while the window size does not overcome the disadvantages with the frequency changes, providing a frequency change with a frequency window of time is the signal processing time-frequency analysis and the ideal tool. Its main feature is the prominent problems by changing some aspects of the full features, therefore, wavelet transform in many areas have been successfully applied, in particular digital algorithm of discrete wavelet transform has been widely used in many problems in the change of study. Since then, the introductions of wavelet transform more and more people's attention, and its applications to more and more widely.

Coding for the low frequency coefficients as fuzzy mathematics, the DCT-based dual-CVQ system is used. Because the images of the energy are concentrated in low frequency, so the encoding properties of low-frequency coefficients will greatly affect the entire coding system image quality of recovery. A large number of experiments show that a simple Level CVQ can not get satisfactory results, in the case of high compression ratio to restore the image quality appears blocking artifacts. In this paper, two CVQ system, namely the results obtained with a CVQ image and the original coefficient compared to get worse value image, and then on the difference image CVQ coding. Since most of the difference between the values of the image values concentrated near 0, the individual value of the jump occurs, so changes to a lesser extent the formation of the vector, and then gets better quantified through CVQ effect.

Stratified Reconstruction from affine point of principle is the corresponding reconstruction method for calculated measures, first of all calculated from the point of the corresponding projective reconstruction [4'5]; affine points according to a corresponding projective reconstruction projective transformation, projective transformation by this Real feature vector to determine the plane at infinity in the projective reconstruction of space coordinates, which are affine reconstruction; reconstruction calculated from the affine camera intrinsic parameter matrix, and ultimately get metric reconstruction.

After the robot visual image recognition and image preprocessing, you can identify the object and its orientation [6]. Therefore, after processing using the corresponding algorithm, the robot can make a difference on matters related to reflection. Solving the robot recognition processing problems, improving the level of awareness of the robot to reduce the identification error is about a very key issue. Visual robot on image pre-processing capability advantages and disadvantages requires a reasonable image quality evaluation algorithm rapidly as the support, but to build a robot model of the visual image information data base, can improve the robot ability to identify, solve the image Identification of Error.

Image is decomposed into a series of wavelet sub-image, because each level of decomposition of the low-frequency part of the coefficient distribution and the distribution of image pixels, the same as the original, so go through the same low-pass and high pass filter formed by the phase image convolution in the same direction must have a similarity. Assuming the image is L-level decomposition, the definition of i-level decomposed into sub-image resolution level for the 1. Some of the first

resolution level m-1 m-resolution sub-image and image-level sub- the FFT, and found that the direction of the sub-image with the same energy distribution in the same pixel. In order to prove the same directional sub images with different resolution level similarity, we will be verified in the experimental section. take advantage of this similarity must will reduce the number of bits needed to encode.

Now discuss the application of the reconstruction resolution level m m-1 sub-image resolution generated sub-class image classification information. We define the resolution level m-vector sub-image block size is m × m, choose the first1 resolution rate-level sub-block size image vector 2 m× 2 m, the same directional sub-level m-resolution images and the resolution level m-1 the same number of directional sub image with the same vector. Since the same directional sub-level images with different resolution similar, we apply the reconstruction of the first-level sub-m resolution images to classified information vector s m-1 instead of the same sub-image resolution level corresponding position vector of the classified information, thus reducing the transmission of bits required for classified information the number in the entire encoding process to send only the minimum level sub-image resolution of classified information. In order to verify the reliability of this method, this article will be classified in the experimental part of the correctness of the test.

3   Reconstructing Surfaces

Reconstructing surfaces is a crucial technology in reverse engineering; the quality of reconstruction directly affects subsequent design and manufacture [7]. Although Surfacer software is capable of fitting curve rapidly and easily, curved surface modeling is less than the other CAD. In general, surface acquired by surfacer need to be modified farther in other CAD, for example, imageware. Consequently combine the surfacer with Pro/E and have reconstructed the curve and surface in this paper. It is too difficult to create a whole car mode once, curved surfaces is separated into several piece and be fitted and merged and trimmed, finished car body surface is shown.

$$T = diag(A, A, 1)(F, 10) \tag{1}$$

Encoding the high frequency coefficients needs the following strategy with Formula (1): for each resolution level of each sub-image using classified vector quantization, vector image vector into the edge of the transition vector, flat vector three categories, each category with their respective codebook vector quantify the resolution of level between the use of the similarity of the image the same direction, using the first transmission-level low-resolution images[8], then transmit high-resolution level image, high-resolution images of the vector-level classified information from the previous level sub-image reconstruction production, which re-m resolution level 1 sub-image resolution generated sub-class image classification information. As the resolution level of the sample sub-image sizes, but at the same resolution level the image of the sub-image that different information Therefore, the statistical distribution coefficient is not the same, so if the vector using the same code book to quantify, it is impossible to obtain good results, and codebook generation is quite difficult, because it is difficult to find a wide range of, for a variety resolution images and a variety of levels, a variety of directional sub image of a unified code book. To solve this problem, we use multi-resolution codebook, that is, different resolution levels in different types of sub-images with different codebook vector[9].

4 Surface CNC Machining

Under normal circumstances, the robot can use the code, and inertial positioning plate method, but in the classical three-dimensional visual, the need to estimate the Euclidean coordinate system under visual matrices, be able to complete the stereo visual system calibration[10]. Chang Yong method is based on some spatial point coordinate system in Europe under the coordinate Yu Qi correspondence between image coordinates, Jian Li visual Zhongxinjuzhen the Yueshufangcheng, Conger-queding visual Zhongxin matrix. In practice, in order to get the Euclidean coordinates, need to make a calibration reference material, reference material on the calibration feature points through the accurate measurement of visual as the estimated center of the space required when matrix point[11].

5 Euclidean Space and the Visual Center of the Space Transformation

Remember the first i-view camera image absolute conic (IAC) is, by the affine reconstruction can be one of the constraint equations IAC is the first i-1 with the first view of the infinite homograph is 1 and the camera viewpoint the IAC. This method can be used to solve the intrinsic parameters.

References

[1] Wei, T., Li, H.: The new stereo visual model based on velocity vector obstacle avoidance planning. Computer (2009)
[2] Zhuo, Z., Zhao, Y., Yan, L.: Wu too summer, phase clouds: Granite surface bidirectional specular component and diffuse component of comparative study. Infrared and Millimeter Waves (2007)
[3] Wu, J., Wang, T., cable Zhiyong, Bao, Z.: DPSS projection matrix based on single-channel moving target detection method. Data Acquisition and Processing (2009)
[4] Chen, M., Liu, Y.: A moving object based on motion vector extraction method. Electronic Design Engineering (2010)
[5] Zheng, Z.Y., Yao, G.-Z., Jin, G., Hai, C., Zhang, H., Dai, J.: Binocular stereo vision information processing II. Three-dimensional spatial filtering effect on the eyes. Psychology (1983)
[6] Chi, T.: Remote sensing technology: strong support for the space calculation. Computer World (2006)
[7] Sun, F., Wu, F.-C., Hu, Z.-Y.: Determined by the projection plane parallel to the plane at infinity homography. Journal of Software (2003)
[8] Ye, M.: Of standard definition TV and HDTV clarity. TV Works (2000)
[9] A song East. Computer control. Computer World (2006)
[10] Zhi, S., Fu roots, Z.: A multi-scale mathematical morphology classification of similar objects invariant moments. Beijing Aeronautics and Astronautics (2009)
[11] Zhu, H.-Y.: Transformation method using vectors derived velocity and acceleration in cylindrical coordinates, the natural coordinates and spherical coordinates in the expression. Neijiang Teachers College (1998)

Automatic Speech Embedded Word Method

Xiaoning Li[1], Xiaofeng Li[2], Zhiyong Zhang[1], and Zhuo Zhang[3]

[1] College of Computer Science and Technology of Chang Chun Normal University,
Chang Chun 130032, China
xiaonin4523@163.com
[2] Jilin University, Chang Chun 130022, China
li_xf@jlu.edu.cn
[3] Changchun City Experimental High School, Chang Chun 130000, China

Abstract. This article has been key issues in the automatic transcription of speech automatic review ways to make a simple introduction. In the letter, the language of text to speech (text to speech, TTS) and automatic speech recognition systems (automatic speech recognition, speech recognition) system, the practical application, due to the emergence of new words and many proper nouns collection, storage difficulties pronunciation dictionary does not exist is often encountered, need these words automatic transcription. On this basis, gives an automated voice system implementation, impact on performance, some technical details of the analysis system and method are given hand-written on the comparison of experimental results show that under the rules, significantly better than the latter.

Keywords: Information processing, Speech processing, Embedded systems, Automatic speech.

1 Introduction

Automobile industry is one of the mainstays of the global manufacturing. With the market changing, regeneration velocity of the automobile is becoming more and more quickly. Automobile manufacturers develop new automobile continuously in order to occupy market. Model and sample of the car are capital element in the development of the car body, and traditional car model is made by manufacturing 1:4 or 1:5 oil sludge model by hand, as show in Fig.1.

In recent years, electronic technology of transcription is real smart move in more and more widely practical use to more efficiently the urgent needs of the convenience and small machine, but the traditional touch screen aspects can not be satisfactorily answered. At the same time, automatic speech recognition (ASR, Automatic Speech Recognition) technology has been rapid development of some simple speech recognition system can be used such as telephone voice dialing, smart toys, robot control and other embedded platforms. As the language itself is the most common human form of communication, therefore, embedded speech recognition technology will be the future of intelligent human-computer interaction for mobile devices is an important choice.

Letter, the language, written language is phonetic text to speech (text to speech, TTS) and automatic speech recognition (automatic speech recognition, automatic

speech recognition) system, an essential link. Speech text generated in accordance with the symbols of the string corresponds to the pronunciation of the phonemes, then convert the phoneme string acoustic output; in the text speech recognition and voice dictionary a bridge between the acoustic models. In general, the pronunciation dictionary to store the pronunciation of words connected, but can not always find the word in the dictionary is the root cause, not including the letters, and more of the language, pronunciation is very stable, all of the words forming the word; and letter writing, may be relatively unstable, on the one hand with the progress of time, often inventing new in spirit, on the other hand, the letters in other languages will gradually introduce the same words, in order to increase the number of terms. English, linguists generally believed that, in addition to proper nouns, vocabulary increased by about 10 billion dollars. In addition, such as name, biological, chemical and other terms of the number of proper nouns are also significant. All of these make it almost impossible to cover all of the words in the dictionary pronunciation, especially in the current hand-held equipment, storage space is limited, and can not store very large dictionary. Therefore, people need to have a "backup" solution, automatically, without a dictionary word transcription. In the literature, which of course have a variety of titles, because of its prime phoneme conversion (shaped to phoneme conversion, G2P), letter sound conversion (letter to sound conversion), etc. In this paper, the most commonly used means the G2P.

Automatic spelling words to learn, is based on the language, orthography (spelling), between the words and pronunciation is a regular.

On the other hand, the law can not cover all of the sounds, there are always exceptions. G2P is how to identify, express and use of this rule. Degree of complex legal, as well as in different languages with some exceptions, the situation is very different relationship; therefore, difficult G2P different languages are very different. For decades, scholars have proposed many solutions to the problem, in essence, G2P, can be divided into rule based handwriting (written rules) and machine learning (machine learning category 2). This paper reviews these two methods, including some of the more G2P-based machine learning methods typical of the decision tree of the prototype system, based on analysis of the performance of some technical details, and gives the methods and methods of Comparing the results of the handwriting on the rules.

Handwritten G2P rules

G2P is based on the first hand-written rules. Linguists on the basis of language understanding, artificial voice began to sum rules, these rules are usually expressed as

2   There Are Three Methods of Machine Learning in G2P

2.1   Decision Tree

Decision tree is a widely used classification method, in which cars (classification and regression tree) approach is widely used G2P. Shopping Cart is a binary decision tree; each non-leaf node corresponds to a yes, no problem, and all the leaves Node corresponds to a category. For an example of classification, from the tree's root node, the node according to its answer questions Figure 1 shows an example.
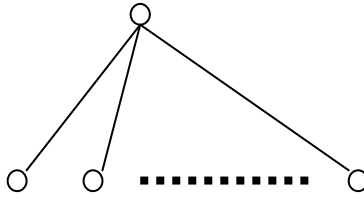
Fig. 1. Traditionally making the car model

Second child node 1, and so on [5], until you reach a leaf node, then we get examples of classification. Specific to the G2P problem, an example of a phonetic alphabet which includes the letters, the problem usually contains letters and words such as "whether the correct e" the first letter, "the first phoneme left side is a vowel", etc.; assigned to the child nodes of non-leaf nodes, and improving the purity of the highest principles; leaf node corresponds to the arrival of this classification is assigned to the phoneme nodes for all training data. Shopping cart many show details of the training, including design issues (including the length of the window), the node selection.

$$\Delta(t) = \sum_{i=1}^{m} (t_i - g_i)$$

The gram model is the standard speech recognition language model approach. [5] Proposed N - Gram's method to solve the problem G2P news, theory is playing a word w, assuming that it is spelled Wei is the first phoneme jag , then the joint probability can be defined.

Where A is all possible routes, ti=f(nodei),gi=k(nodei) for the realization of any one route. f(nodei) aligned according to letters and sounds, the joint probability. Reference language model of the N - gram model, k(nodei) to retain the previous N - 1 had only historical probability, A reduced to f(nodei), not only can reduce the model, rather than a serious impact on the scale of the model accuracy. Training model parameters for the k(nodei) to spell the word, its pronunciation by the equation obtain In addition to this method can do G2P, using the same set of models, you can also turn to do P2G, just type k(nodei) was revised to them. To improve the speed, you can enter k(nodei) remove, just use the Viterbi search method the maximum path.

As not one letter corresponds to a voice, as well as a few letters correspond to phonemes, a number of phonemes corresponding to more than one letter, the letter is not pronounced, it is first of all to align letters and phonemes. For a number of letter sounds and letters corresponding to the situation of non-pronunciation. [1]

Two types of solutions, as an empty phoneme, the corresponding non-pronunciation of the letters a, b, the definition of the word's first prime number, a key word by one or several letters correspond to sounds. The former is the more common approach, which can refer to the text. More similar letter to the appropriate solution "empty" the letter announced that it will define the location, you can also define a new phoneme combinations, such as often occurs in English pronunciation x page [Ohio] and [Shnghai] The new definition can be two phonemes, namely, the two combinations, which deal easier. Letters and good air, the air the voice of the definition of general use after several iterations (expectation maximization) algorithm, the results of the best route.

2.2   HMM Model

Hidden Markov Model (HMM, HMM) speech recognition is the most commonly used methods of acoustic modeling. The G2P problem HMM [6] is used in the proposed solution. The starting point is basically represents a word, the pronunciation of 564 Li et al: Automatic Speech spelling the word the word is just an outward manifestation of this performance will be the impact of changes in the existing noise situation. Thus, HMM modeling, pronunciation is a hidden variable, but the spelling is the observation output, G2P is hidden in the search process variables known state HMM observations. Still assume that the spelling and pronunciation of the word w, respectively, a(W) and m (w), best sound.[6]

Where: node x is the prior probability of pronunciation, you can use the N - gram model, and HMM models. Since each phoneme, including the four states, HMM model for each output state observer is a letter, four countries, a phoneme can correspond to the largest, such as four letters, "country" the "ountry", each country can skip the rest of the country the end of the model may also arise whether phoneme letters, including four-letter string, this assumption is appropriate in English.

Although different implementations[2], but G2P machine learning methods based nature of the contents of the letter is determined by the sound of the letter (to be described later PBA's slightly different). Is generally believed that only the letters from the target closer to the letter sound effects, so you can add a word of the sliding window on each side of the central objective of the letter, retain a certain number of letters, the letters will not be considered out of the window. Consider the letter "o" sound, that only "c" left "tag NGR" and right "ntray" can affect the pronunciation. How to choose the length of the window is a need to consider, too short a skeleton does not contain sufficient information, the second long skeleton noise during the training period, reduce the model's generalization ability. In fact, there can be only the length of the right side of the skeleton, such as the word "Photo" and "photographic" spelling, but the latter one more letter, in three areas, corresponding to different vowels pronunciation, the first "o" of the "C "8-bit word pronunciation is the distance[3], usually not long unilateral window. In general, appropriate skeleton length needs to be tested and experience to be determined.

2.3   Non-parametric Method

Non-parametric method does not require the training of model parameters, but with phonetic word dictionary matches all words do, according to results of the competition and some of the selection criteria, points out the word pronunciation direct splicing. IB is also known as nearest neighbor (neighbor) method, assuming that X and Y are strings of letters before the window, to determine their similarity

$$X = \int f(t(y))dy$$

Where: n is alphanumeric [7], X, and Y is my first letter, nodei is a joy, meaning similarity, can be defined as the two letters the same as 1, otherwise 0. Each letter string X skeleton, you can find in the dictionary most similar, the two are the corresponding letter for the X in the middle of the letter sounds in the pronunciation. Improve the similarity calculation, the letter of the string sounds of the letters in different positions similar to the function of the size of.

Where X weight gain known as the information (information gain weight) can calculate the entropy of the training data. This improved method is called IB1 period,

about law. Analog voice (pronounced analogy, PBA) Dedina and Nussbaum first to propose and implement their sound system [8], the pronunciation dictionary system, matching module, voice grid, decision-making function of Part Y. It is first phonetic characters, each word in the dictionary to do the maximum matching, matching method, first two are left-aligned, then record all of the locations and characters are letters and corresponding sounds the same dictionary string, and then transferred to a short letter word right, re-match, until the two words so far in the right alignment. Match results and the corresponding forms of organized sound grid, each grid node, a letter word that arc connecting these two nodes record of correspondence between sounds and letters in the dictionary matches the number that matches the string. Each grid corresponds to the path from start to finish the pronunciation of the word, last step is to choose a "best" path. Understandably, a small arc of the path number, that each period of longer the arc, that there are in the dictionary or more consecutive letters to find a match for the corresponding pronunciation.

The possibility of more correct, therefore, chooses the path of minimal arc, or the number of shortest paths, the corresponding pronunciation. When the same path length ratio, we need a decision function to decide which to be chosen. Given the simplest method to choose the path of the corresponding arc to tie the maximum number of times, made a strategic investment portfolio options proposed several methods [9] and better results. Non-parametric method has the advantage of these two is simple, intuitive; drawback is to use a dictionary to complete, for the current hand-held device, the dictionary a few megabytes of memory costs too much. But also need to traverse the dictionary each word several times to do some matching, processing speed should be slower than other methods. In addition to these methods, neural networks, finite-state jump.

3   Shopping Cart Based G2P System

This choice Webster's Pocket Dictionary (1974 edition of the test data, Sejnowski and Rosenberg to their NETtalk system training) [10], will be a manual adjustment of words and pronunciation, commonly known as NETtalk dictionary. The dictionary contains 20 008 words, the different positions of the same phonemes, stress removal, but only to enter, left 19,940 words, more than the number of words to sound more than one word. This will provide information directly to the line using the dictionary does not automatically pre-aligned. If not specified, the following results, these words do 10 cross-validation results. In G2P experiment, the overall performance of the two systems using numerical calculation, that is the correct conversion rate and accuracy of text conversion morpheme. This experiment only shows the word correct rate conversion, which is the total number of words and the right to change the proportion of the total number of words the results.

3.1   Reference System

Shopping cart G2P-based baseline system, the main two issues need to be considered, namely, feature selection problems and design issues. Issues have the background and 2 classes, such as the phoneme "L 2 O" means "(the current letter) the right of the first two letters, 738 letters generated by Tsinghua University (Natural Science) 2008. The current phoneme is left of the first phoneme among baseline system, all the questions," single-issue", that it is a problem to implement the "one" specific letters or

sounds, rather than whether to the left of the first letter vowel, this classification problem. The problem of the correspondence, which contains 5 letters and left the current three-phoneme, sounds current location.

From the tree, two conclusions can be drawn: First, from the target letters and letter sounds further small current pronunciation, the second largest in the right voice over to the left of the target letter writing. Shopping Cart generation algorithm is an important step in how to select a node corresponding to the problem. Split the target node of each child node contains a sample of training data belonging to the same category as far as possible, number of samples is as little as possible, paraphrase, "other types of impurities " low, the better. Training samples belonging to the root of all the beginning, and then split each node are not in a lower purity of the local optimal solution search node. Based on the questions asked of the two child nodes of node splitting to answer tons of T1 and T2, does not reduce the function is defined as the purity.

Location: node1 and node2 with f(node1) and f(node2), respectively, the sample proportion in the total sample of Y, f(nodei) is a node, the function is not pure, generally have two options, they are the Gini coefficient (Gini index) and entropy (entropy), whose role is expressed as.

Where f(nodei) is the proportion of nodei in the same level. Experimental results show that the performance of the two systems were not significantly different functions, the system chosen entropy function.

3.2   Reverse Conversion

Section A pointed out that the English pronunciation of the letters left than the right to present a great impact, you can guess the correct phoneme phonological effects should be more on the left. Therefore, you can try from the right letters of a word, turn left, make full use of that right has been converted phoneme information. Training of scale is more than section A.

System performance tests are inseparable from the data, G2P system, the so-called data, one or more of the pronunciation dictionary. The rules-based approach can be used as a dictionary word test set; and machine learning in general requires the training set and test set dictionary-based approach is divided into two parts, or cross-validation method. Cross-validation, each copy of the data set to be the test set, training set, the rest of the election, so N times cycle test (N-fold cross validation) and other means to summarize all the results. [4] Two kinds of data classification method, which can be more effective use of various data, the results obtained, because the data is chosen to avoid uncertainty factors. In G2P literature, test data between the different languages used in different courses, even if there is a language, select the dictionary and thesaurus of the basic classification method is also different. Test data Di selection and test suite design is not completely consistent, although there are kinds of methods of G2P Yi gradually increase, but did not admit the obvious method advantage.

Through training (or more appropriate), the model over-fitting the training data, including noise, resulting in the training data is not visible, to promote ability. Specific to the use of this method of shopping cart, you can solve the following training.

First, the tree nodes split and stop criteria. Until the baseline system can not continue to divide in order to improve the system, using Chi-square test method, when the child nodes and parent nodes of the phoneme distribution was no significant difference, it is no longer split.

Secondly, according to the first letter and sound conclusions from the target closer to section a letters influenced current pronunciation. Therefore, this problem can be sorted according to target letters from the lower value of the same purity, the priority issues from the past.

Finally, the issue of setting lean in the baseline system, use-related problems as a candidate[11], the current five letters of the letter, in fact, far more likely to believe the training set of noise. Through the experiment, each letter of the three is the best system performance use. Table 1 shows the conversion through the system to improve results than the baseline word accuracy of about 5% growth.

Table 1. Experimental results

scale	origin	HMM	CDHMM
15 872	27.2%	28.1%	28.6%
17 850	29.1%	29.7%	30.6%
21 310	30.4%	30.8%	31.9%

While the initial process of Fourier analysis as the analytical thermal analysis tool, but it still has the typical way of thinking and analysis of the reduction of the characteristics of doctrine. "Arbitrary" function by some decomposition can be expressed as a linear combination of sine function in the form, and the sine function in physics is relatively simple to fully research the function class, the idea of atomism with the idea of chemically much alike! What's amazing is that the Fourier transform of modern mathematics has very good properties, making it so easy to use and useful, people have lamented the magic of creation: 1. Fourier transform is a linear operator, if given the appropriate norm, It is unitary operator.

Fourier transform easily find the inverse transform, and change form and are very similar; 3. sine basis functions are eigenfunctions of differential operation, which makes the linear differential equation can be transformed into ordinary coefficient of algebraic equation. linear time-invariant physical system, the nature of the frequency is unchanged, so the system can respond to incentives for the complex by combining its response to sinusoidal signals of different frequencies to obtain; 4. the famous volume product theorem states that: Fourier transform of the complex can be a simple convolution product operation, which provides a simple means of calculating the convolution; 5. discrete Fourier transform can be calculated using a fast digital computer (the algorithm is called Fast Fourier Transform algorithm (FFT)).

It is because of the good nature of the Fourier transform in physics, number theory, combinatorial mathematics, signal processing, probability, statistics, cryptography, acoustics, optics and other fields have a wide range of applications . Fourier Transform on FPGA implementation of digital signal processing, Fourier transform of the basic operation, widely used in presentation and analysis of discrete time-domain signal areas. However, due to its computational and transform points is proportional to the square of N, therefore, N is large, direct application of DFT spectral transformation algorithm is not realistic. However, the emergence of fast Fourier transform to make fundamental changes have taken place. This paper describes the use of FPGA to realize 2k/4k/8k point FFT design approach.

4   Summaries

This article describes the rules and on the two types of machine learning methods based on hand-written G2P, and gives the shopping cart using the systematic approach to implementation. Although in recent years, machine learning methods based on more and more attention, but some scholars believe that, in the handwritten rules-based approach can do more than before and no worse. Relatively small, and therefore inconclusive. Therefore, this paper proposes a simple comparative experiment. This downloaded code is written using the Elovitz rules because it uses the phoneme set and NETtalk inconsistent The results generated need to do post-processing, mapping to NETtalk phoneme set, in the process, and not introduce new errors. 17 850 word test set, the word accuracy was 27.16%, significantly lower than this car system. 654 Li, et al: Automatic phonetic alphabet word method, summing up, a simple review of the methods to hand-written rules and made two categories based on machine learning methods, and gives a prototype system implementation, system design analysis showed that some problems do. Comparison of results gives the two class method. In future work, there are several areas to further improve the prototype system, including those aimed at addressing the global optimal decision tree, on the N - gram model phonemes presented detailed questions.

References

[1]   Daelemans, W., Van den Bosch, A., Weijters, T.: IGTree: Using trees for compression and classification in lazy learning algorithms. Artificial Intelligence Review 11, 407–423 (1997)

[2]   Marchand, Y., Damper, R.I.: A multi-strategy approach to improving pronunciation by analogy. Computational Linguistics 26(2), 195–219 (2000)

[3]   Dedina, M.J., Nusbaum, H.C.: Pronouce: A program for pronunciation by analogy. Computer Speech and Language 5, 55–64 (1991)

[4]   Elovitz, H.S., Johnson, R., McHugh, A., et al.: Letter-to-sound rules for automatic translation of English text to phonetics. IEEE Transactions on Acoustics, Speech and Signal Processing 24, 446–459 (1976)

[5]   Wang, G., Liang, W., Liu, J., et al.: Moderate vocabulary English speech recognition system embedded on a chip. Qinghua Daxue Xuebao/Journal of Tsinghua University 45(10), 1393–1396 (2005)

[6]   Ververidis, D., Kotropoulos, C.: Emotional speech recognition: resources, features, and methods. Speech Communication 48(9), 1162–1181 (2006)

[7]   Li, H.-X., Li, L.-X., Wang, J.-Y., et al.: Fuzzy Decision Making Based on Variable Weights. Mathematical and Computer Modelling, 163–179 (2004)

[8]   Ramamohan, S., Dandapat, S.: Sinusoidal model-based analysis and classification of stressed speech. IEEE Trans. on Audio, Speech, and Language Processing 14(3), 737–746 (2006)

[9]   Nwe, T.L., Foo, S.W., De Silva, L.C.: Speech emotion recognition using hidden Markov Models. Speech Communication 41(4), 603–623 (2003)

[10]   Ramamohan, S., Dandapat, S.: Sinusoidal model-based analysis and classification of stressed speech. IEEE Trans. on Audio, Speech, and Language Processing 14(3), 737–746 (2006)

[11]   Morrison, D., Wang, R.-L., De Silva, L.C.: Ensemble methods for spoken emotion recognition in call-centres. Speech Communication 49(2), 98–112 (2007)

A QoS_Based Self Adaptive Control Strategy and Implementation for BPEL Process

Yu Nan[1], Guangsheng He[1], Zhuo Zhao[2], and Cheng Liu[1]

[1] Department of Electronic & Information Engineering,
Information and Science Technology, College of Zhengzhou, China
ddApple2011@163.com, {48894098,377829688}@qq.com
[2] Shengda Economics, Trade & Management,
College of ZhengZhou University,
ZhengZhou, China
zzzhaozhuo@126.com

Abstract. The traditional BPEL language statically describes the binding relationship of business process and the required web services, and it can't adapt to the dynamic operating SOA environment. This paper presents a late binding and dynamic service based on the selected BPEL process orchestration and implementation mechanisms. It designs and implements a dynamic service invoking framework based on this design. In the process of BPEL execution, the service selection and service call based on QoS operated interactions, so that it has some self adaptive features to improve the robustness of the system, also the efficiency of the system is ensured by the static filtering and ranking mechanisms.

Keywords: Business Process Execution Language (BPEL), Service-oriented Architecture (SOA), Self Adaptive, Service Selection, Late Binding.

1 Introduction

BPEL (Business process execution language) is a XML-based standardization of a process description language in SOA (service-oriented architecture) architecture. It is used to operate the available web service of the current SOA environment in a process of implementation, in order to complete the needs of specific business logic. The prepared BPEL processes can be run in any standards-compliant layout engine (orchestration engine), and the BPEL process is analyzed and implemented by orchestration engine. The orchestration engine will manage and organize every required operation service during the visit.

In a traditional BPEL process, designers need to give the details of each service information (e.g. location, interface descriptions, etc.), that is, the service used in the process needs to be specified (binding), the relationship of processes and the corresponding services is fixed in the design. During the operation, the orchestration engine only calls the determined web service In accordance with the BPEL process description and definition. However, SOA environment is a dynamic environment. During the process of web services operation, although the function description is

unchanged, the non-functional attributes (QoS properties) is dynamic changed. For example, some services of BPEL process become unavailable due to a network reasons, or some services QoS properties no longer meet the requirements process (such as the workload is too large and the time of implementation is too long). Therefore, BPEL processes running should also be a dynamic process that requires some kind of adaptive mechanism to dynamic change the binding relationship between process and the required services according to environment changes during the operation. And when needed, it can reselect the useful services in the current available services list in order to increase the robustness of process execution.

In response to these problems, a new architecture and a control strategy are proposed. This paper improves the standard service orchestration engine, and adds a service selection mechanism into the analysis and implementation of the BPEL process so that the operation of BPEL process can be self-adaptable.

The self-adaptable features are described as following:

(1) Through the late binding, the binding of BPEL process and the specific services are postponed at the runtime, which makes implementation of the adaptive behavior at runtime become possible.
(2) A service selection mechanism is implemented, the services of BPEL processes is preselected based on the UDDI server to get a list of available services in smaller range, to ensure the efficiency of service optimization at runtime.
(3) For providing the basis for service optimization at runtime, dynamic data of QoS attribute values of each candidate services is collected by QoS monitor broker.
(4) Based on the above mechanism, a self-adaptive service selection engine is realized. Before the service is called, the local optimized services are selected to bind with BPEL to ensure the robustness of the implementation process, according to the current available services list and the services' dynamic QoS attribute values.

This paper firstly provides an appropriate solution for the problems mentioned above, and then proposes a general framework and the implementation of the framework; at last a conclusion is made about the framework.

2   Late Binding and Self Adaptive Process

SOA architecture provides a loosely coupled, distributed operating environment, with the development of SOA technology, a large number of services appear in the network, and the services with similar functions are provided in different forms of implementation by different service. Because of the dynamics and uncertainty of web services, the available services of network in the runtime dynamic changes.

As a workflow language, BPEL is one of the main application forms of the web services. But in a traditional BPEL processes, the binding of business processes and specific services required must be specified and established in the design stage and this association is not changed in the deployment and operation time. It is clear that the binding of business processes and specific services required can't adapt to the dynamic operating environment provided by SOA. If a specified service becomes unavailable, then the whole business implementation process will be interrupted (Additional, the

network will provides some same or similar services for users). Thus, an adaptive mechanism is needed that it could make BPEL processes deployed dynamically binding with the corresponding service at the runtime. On the other hand, there are a large number of services with the same or similar functions in the network, so it requires an evaluation and selection mechanism based on the needs of the business process to select the appropriate service. The evaluation and selection not only refer to the web service function, but also closely relate to its non-functional attributes (QoS properties). For specific services, the change of operation status will affect the value of non-functional properties. For example, if the workload is too large, the running time will obviously become longer than before. Therefore, it needs to monitor the QoS services running real-timely at the runtime. Finally, to ensure the efficiency of BPEL process service selection, the anthology at the runtime should be reduced.

Now, the issues above will be discussed, and the corresponding techniques and solutions are provided as following.

2.1 Late Binding

Late binding is a very important technology in software systems design. To the code and the corresponding resources, it makes the system more flexibility and scalability, also makes it possible to configure the system at the runtime. For example, Dynamic Link Library (DDL) technology is a typical late binding technique. It can dynamically change the behavior of the system through the dynamic configure a new DDL file. Open Database Link (ODBC) is also an implementation of late binding ideology. It delays the binding time between database application and database services by defining the common access interface which is abided by both the application side and database providers, so that it can configure different database resource for same database application codes. In fact, more broadly, the history of software technology development is a history of late binding.

Similarly, deferring the time of binding the business processes and services can provide more flexibility for the execution of the BPEL process so that the dynamic selection of services can be realized. For BPEL process, the binding time can be roughly divided into three stages: designing phase, deploying phase and operating phase. Although the boundaries between the three phases are not clear, and the three phases may overlap, it is clear that the more close binding is to runtime, the more flexible the execution of BPEL processes is. In this solution, the binding time is deferred before the calling of every service. At the same time, the candidate services are optimized based on the description of BPEL process before the binding time to get the required services and local optimization services with the non-functional characteristics. Thus, a dynamical BPEL execution process is constructed. Because of the dynamically interacting of BPEL process and service selection, the efficiency of service selection becomes a key factor to be considered.

2.2 Static Screening and Pre-service

To ensure the efficiency of service selection, the candidate sets of the selected service at runtime need to be minimized. At the same time, the process should be brought forward that are necessary for the service call and time-consuming at the design period or the loading process to complete statically. Such as the business process mapping data

among the various services can present this. Therefore a service and pre-screening mechanism is needed for the operation of the business process of dynamic service selection to make full and effective preparations. This mainly includes the following aspects:

Firstly, a description of web services includes functional features and characteristics of non-functional. The functional features includes a description of input and output parameters, the premise of the service call and the result of service calls (WDSL), when a web service is deployed and it provides services for users, the functional characteristics often will not be changed. At the side of BPEL, when the process is deployed and loaded, the implementation of the required services and operations are no longer changed. In the execution of BPEL process, the set of functionality required for all stages of the process is relatively stable. Furthermore, the service selection itself is a time-consuming operation; the service functional selection need to be brought forward before the period of deployment or load in the BPEL process, because the bottleneck is caused by computational complexity and network latency.

Secondly, a BPEL process can be seen as a process with numbers of different functional components. At the runtime, each functional component is implemented by a specific service. A list of candidate services will be obtained by the selection of the available services for the required functional. Regardless of the services' functional are same or similar, it will meet the functional requirements of the process. Because all the services list are coming from different service providers, the interface provided by service selection may not match the interface required by BPEL process. Thus, the data adaptation is a very important factor. However, the data adaptation is a time-consuming process so that it must be finished early.

Finally, the process of service selection generals some candidate lists of services. The dynamic service selection at the runtime is operated in the lists. And the selection is based on the current QoS attribute values. In order to improve the efficiency of the selection, the pre-selected list can be ranked and obtain the first K entries of the each list to general a dynamic candidate list. The deployed BPEL process will dynamically bind with the first of the new candidates list to complete the process.

2.3   The Detection of QoS Attribute

When the web service is deployed and running, although the characteristics and description of its function is relatively stable, one of the deployed services' non-functional characteristics is dynamic changing, such as it is unavailable due to network reasons or a new service is deployed. Thus, in the BPEL side, the candidate lists and each service's QoS attributes need to timely update at the runtime so that the selected service is the best choice for requirement.

The properties of each service QoS in the candidate list are provided by QoS probe, the agents provide the BPEL process orchestration engine a series of interfaces defined. At the runtime, the BPEL process orchestration engine obtains each QoS's attribute value through the interface provided by the QoS probe. However, the detection of QoS's attribute values is a very time-consuming process, so this paper proposes a scheme using asynchronous propelling movement mechanism. QoS probe updates the candidate list according to certain BPEL process. This mechanism ensures that the dynamic selection at runtime is the latest data.

2.4   Service Selection and Self Adaptive at Runtime

In support of the above mechanisms, BPEL business process is executed interactively with the service selection. And the selection of services at runtime based primarily on two aspects of information, one is demanded for the BPEL process itself, and the other is demanded for the candidate services' dynamic QoS state. BPEL orchestration engine maintains a candidate set of services (the set is obtained via static selection) to update the QoS states of every service by appropriate frequency, and re-rank the different functions of each candidate service in the list according to the latest set of data values. Since the quality of each service is described by a number of different QoS attributes dimensions, such as commonly used execution time, reputation, price and so on. The composite score weighted sum of the various dimensions is used to describe the utility of a service, and the weight is given by BPEL process designer. The service in the candidate list is ranked by the value of utility. In the implementation of BPEL process, before the relevant service is called, the orchestration engine selects the appropriate list from candidate set base on the required functional properties, and service with the highest rank will be called.
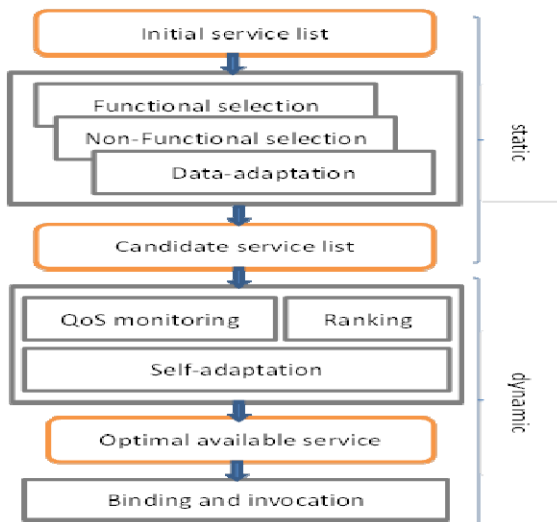
Fig. 1. BPEL process operating sequence

In the framework described above, the BPEL process flow is shown in Fig.1, including static state and dynamic state. When the BPEL process is deployed and loaded, an initial candidates set of services is obtained by static selection. And the service of the set is adapted to the data according to BPEL process services' requirement. After configuration QoS monitoring agent, the required service is dynamic selected and called in the implementation.
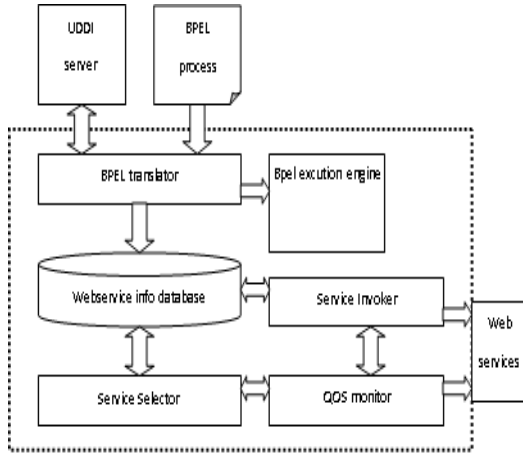
Fig. 2. The framework of BPEL orchestration engine and services invocation

3   The Design and Implementation

According to the description above, this paper proposes and implements an adaptive dynamic BPEL orchestration engine and services invocation framework, as shown in Fig.2.

The whole system is developed base on open source BPEL engine Orchestra, including four parts: BT (BPEL translator), SS (Service selector), SI (Service invoker), and QM (QoS monitor).

BT translates the abstract BPEL processes deployed in the loading process, and obtains the description of the services required, then translates it to SAWSDL description document and submits it to the UDDI server. UDDI server obtains a service list which meets the functional requirements based on the each function in the BPEL process. All of the information is contained in the list. The candidate set of services is processed and maintained by SS. SS preprocesses the candidate list, including data adapter and QM initialization, ranking the candidate list based on QoS property values and requirements of BPEL process. Then SS maintains a candidate list for each processes deployed, and the size of the list (the value of K) is defined according to the requirements.

QM is implemented on a middleware systems M4ABP (Monitoring for Adaptive Business Process) for adaptive business processes. After initialing a set of candidate services, each QoS property values of current service will be collected and stored. SS accesses QM according the defined frequency and updates the QoS property values, while the re-rank mechanism is started. The higher the frequency is, the stronger of the adaptability of BPEL processes is, and more computational resources are required. Thus, the freshness of the data and system resources are a pair of contradictory factors, the update frequency setting needs to be weighed. At the runtime, SI is responsible for the implementation of BPEL processes. SI selects and calls each service in the specified process. Before an abstract service is called, SI accesses SS to select the highest ranked service in the candidates list. At the runtime, SS and QM will be implemented

alternated according to the frequency defined. Thus, under the framework proposed in this paper, the service calling and selection implement alternated in BPEL process to adapt to the dynamic changes of SOA environment, and the robustness of the implementation of BPEL is improved.

4 Summaries

For the dynamic characteristics of SOA environment, this paper presents a BPEL orchestration and invoking mechanism based on late binding and dynamic service selection, and based on this mechanism, the dynamic service invocation framework is proposed aiming to BPEL business processes, the design details and implementation is also discussed. The future research will focus on how to add a services retrieval mechanism based on semantic and more efficient service selection algorithm to support a more comprehensive adaptive behavior.

References

[1] Halpern, M.: Binding. In: Encyclopedia of Computer Science, 3rd edn., p. 125. Chapman and Hall (1993)
[2] Kopeckỳ, J., Vitvar, T., Bournez, C., Farrell, J.: Semantic annotations for WSDL and XML schema. IEEE Internet Computing 11(6), 60–67 (2007)
[3] Le Duc, B., et al.: Non-functional Data Collection for Adaptive Business Process and Decision Making. In: Proc. of MW4SOC 2009 Workshop, pp. 7–12 (2009)
[4] Truck, I., Akdag, H.: A Tool for Aggregation with Words. Inf. Sci., Special Issue: Linguistic Decision Making: Tools and Applications 179(14), 2317–2324 (2009)
[5] Zhang, W.: Design and Implementation of Process Establishment in Web Service Composition Tool Based on BPEL. Computer Soft and Theory (2009)
[6] Zhao, G.J, Shen, L.M.: The Research to Bpel Workflow Supported for Dynamic Invocation to web Services. Computer Soft and Theory (2010)
[7] Dai, J.P., Dong, Y.Q.: E-Business Service System for Manufacture Based on BPEL. Computer Engineering (2009)
[8] Bo, S., Li, S.Y.: Research and implementation of BPEL for web services. Computer Engineering and Design (2007)

Dynamic Workflow of Clinical Pathway System

Lin Tongchuan and Qi Deyu

Research Institute of Computer Systems
South China University of Technology
Guangzhou 51006, China
`{linckham,qideyu}@gmail.com`

Abstract. Workflow is one of the important technologies in enterprise business process automation. It has many advantages such as raising efficiency of business operation, improving resource utilization, increasing flexibility and adaptability of business execution and so on. However, traditional workflow systems have shortcomings of hard coding and flexibility. Based on the analysis of these shortcomings, a dynamic workflow system based on rule engine is put forward to enhance the flexibility of the system. This system has greatly improved the maintenance and flexibility.

Keywords: Work flow, Rule engine, Clinical pathway.

1 Introduction

For last ten years, many researchers have given different definition for workflow. So far, there is not completely unified definition for workflow. For example, the definition of Workflow Management Coalition WFMC[1], IBM Almaden Research Center[2], Amit Sheth[3] and W.M.P. Van der Aalst [4]. According to these definitions and myself understanding of workflow, in this paper I give a definition: Workflow is a computing model of working process, that is saying how the work of the workflow before and after the organization together in the computer logic and rules of the appropriate model to represent and calculate its implementation, in short, business process automation or semi-automated implementation. Workflow mainly includes the concept of business process, definition of process, activity, workflow management system and the instance of process [5].

Rules engine is a component that is embedded in the application. Its task is to test and compare the current submitted to the engine's data object and the loading business rules in the engine, activating the business rules which are consistent with the current data status, according to the execution logic declared in the business rules, trigger the operation of the corresponding application. The rules engine contains the functional modules of context, agenda, working memory and rules container [6].

Clinical pathway is a means to establish a standard for a particular disease, treatment modalities and therapeutic procedures, is a comprehensive model on clinical treatment. It can promote the treatment organization and the methods of disease management, which is based on medical evidence and guidelines; finally it can achieve the purpose to regulate medical practice, reduce variation, reduce costs and

improve the quality. Relative of the guide, the content of clinic pathway is more concise, readable, and adapting for multi-disciplinary, multi-section action. It is for diagnosis and treatment of specific disease process, it focuses on synergy, result and time during the process of treatment [7].

The existing operating model of workflow system has some flaws and limitations when it applies for actual application [8]. Firstly, the rules of business process are fixed in the code by the form of "If-Then-Else"; secondly, most workflow systems lack of central rules information repository, centralized management mechanism of rules and strategy; finally, in response to complex business rules and record variations, the workflow system can't handle these complex rules and rapidly respond to changes in the rules. However, the introduction of the rules engine technology can bring great improvement in this situation. The rules engine technology separates the business logic code from the application and defines these codes as rules that computers can identify. Rules engine is charging for rules management, maintenance and computing, providing rules service for workflow engine and providing operating pages for business operator. The function of rules service is to reason and compute the rules in the context, and return the result for the process instance. Workflow engine is particularly charging for process operation, sometimes calling the rules of business decisions when necessary. When business rules have changed, there is no need for modifying relative source code; it only need update the rules repository in the edit pages by business operator. This system has greatly improved the maintenance and flexibility.

2   Overall Design

2.1   System Framework

Dynamic Workflow of Clinical Pathway System adopts the classic MVC framework, Struts is used for presentation layer, Spring is used for business layer, Hibernate is used for persistence layer. The development environment of system is MyEclipse6.5 +Tomcat6.0+JBoss Rule+Oracle9i; The framework of system as show in Fig.1.
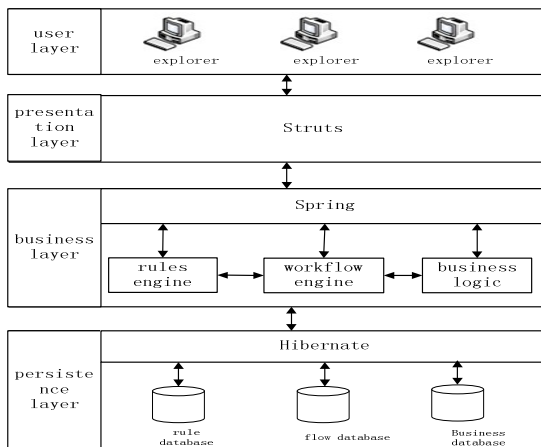
Fig. 1. The framework of clinical pathway system

The presentation layer is the window of clinical pathway system, it provides the interface for user to access and operate; the business layer is charging for the business functions of the whole system, including patient management, management of clinical pathway execution and so on; management of clinical pathway execution is a core part of this system, it mainly charges for the judgment of pathway in or out; the persistent layer is charging for storing or loading verify data in system, like business data, workflow data, rules engine data and so on. This framework can greatly improve the development efficiency.

2.2   Design of Work Flow

This system uses Fire Workflow tools to help create system flow models. Fire Workflow uses XML to define flow files, at the same time it combines with state, activity diagram and Petri net. This system mainly contains two workflow: Register-Diagnosis flow; clinical pathway flow.

2.3   Design of Business Logic Rules

In this system, control the circulation of workflow and the execution of activity need rules. There are two main rules. First rule is fracture of shaft of femur:

   If "symptoms contain deformity" and "check contains fracture of shaft of femur"
   Then "validation success" and "enter clinical pathway"
   Else   "Validation failure" and "can't enter clinical pathway"
   Second rule is consultation activity.
   If "doctor think need"   Then "set flag equals true" and "enter consultation"
   Else   "Set flag equals false" and "end consultation"

3   Implement of System

3.1   Algorithm of Workflow

The algorithm flow chart of fracture of shaft of femur clinical pathway is shown as show in Fig.2.

3.2   Key Code of Workflow

The key technology is how to combine with workflow engine and rules engine. As shown in Fig.3, doctor diagnosis task has applied this key technology.

After the doctor logins on workspace, he or she will find a list of to-do tasks; then the doctor select one of to-do tasks to deal with. The doctor diagnosis the patient and enter the relative page for operation, as show in Fig.3.

After entering the diagnosis page, when the symptom is selected as deformity, check is selected as fracture of shaft of femur, disease is selected as fracture of shaft of femur, then click the button "validate" to validate, the return result will be success, otherwise

Fig. 2. Algorithm flow chart

Fig. 3. Doctor diagnosis

will be failure and end the diagnosis flow. The "validate button" will call JBOSS Rule Engine to judge whether pass. The key code of validation function is as below:

```
this.setAccessFlag(false);
testsRulesEngine=new TestsRulesEngine();
testsRulesEngine.assignTests(this);
```

When validation is passed, then click the button "enter clinical pathway" and do clinical pathway diagnosis for the selected patient. First, the diagnosis function will call Fire Workflow engine to initialize the clinical pathway flow, which is mentioned in 2.2. the button of "enter clinical pathway" executes code as follow:

```
pi=(ProcessInstance)wflsession.createProcessInstance("gugugantest",account==
null?"":account.getUserID());
pi.setProcessInstanceVariable("sn",Utils.getFormatedDateStringForSn(new
Date()));
pi.setProcessInstanceVariable("patientID",patient.getPatientID());
pi.setProcessInstanceVariable("doctor",patient.getDoctor());
pi.setProcessInstanceVariable("disease","");
pi.run();
```

After the workflow instance is initialized, call JBOSS Rule Engine to judge the condition, in order to dynamically change the executed direction. Key code as follows:

```
testsRulesEngine= new TestsRulesEngine();
testsRulesEngine.assignTests(this);
pi.setProcessInstanceVariable("flag", this.huiZhengFlag);
```
The next step is followed the chart in Fig.2.

4   Summaries

After analyzing the advantages and disadvantages of the workflow engine and rules engine, I propose the combination of workflow engine and rules engine and use it in hospital clinical pathway system, and for the sample with femoral shaft fracture clinical path dynamic workflow system.This paper describes the technology framework of clinical path of dynamic workflow system and the concrete realization of each module. The dynamic workflow system of clinical pathway in this paper can effectively improve hospital's efficiency and improve resource utilization in the hospital, it has broad application prospect.

References

[1] Mohan, C.: Recent trends in workflow management products, standards, and research (1997)
[2] Alonso, G., Agrawal, D., Abbadi, E.: A Functionality and limitations of current workflow management systems. IEEE Expert 1(9), 12(5) (1997)
[3] AmitSheth, M.R.: Modern Database Systems: The Object Model, Interoperability, and Beyond. ACM Press (1995)
[4] van der Aalst, W.M.P.: Three good reasons for using a Petri-net-based workflow management system. Engineering and Computer Science 11(14), 179–201 (1996)
[5] Wu, C., Deng, S.: Workflow System design and its realistic. Zhejiang University Press (2006)
[6] Peng, L.: Rule engine principle analysis. FuJian Computer (9), 42–45 (2006)
[7] Pierkakalo. Clinical Pathway (OL),
     http://baike.baidu.com/view/2109245.html
[8] Cheng, M.: Research and the design of the dynamically workflow model based on rules engine. Wuhan Technology University (2009)
[9] Li, G.: Integration Spting+Structs+Hibernate Application developing. Tsinghua University Press (2007)

Camera Calibration
Based on Extreme Learning Machine

Chai Zhaohu, Ren Xuemei, and Chen Qiang

School of Automation, Beijing Institute of Technology, Beijing, China
Chai-Zhaohu@163.com

Abstract. An extreme learning machine (ELM) based camera calibration method is proposed for monocular vision system in this paper. Extreme learning machine is used to depict the relationship between the image space and the world space, in which any prior knowledge on camera model or parameters is not needed, and faster training speed and higher precision are gotten. The influence of environmental noise to the calibration precision can be effectively attenuated. The validity of the proposed method is proved by comparison with camera calibration based on BP algorithm.

Keywords: Extreme learning machine, Camera calibration, Computer vision.

1   Introduction

Camera calibration is a necessary step in computer vision in order to extract information from 2-D images. The most widely used calibration techniques require the development of elaborate mathematical models and prior knowledge of many parameters. It is difficult to construct a suitable model and obtain reasonable values for unknow calibration parameters. To overcome this problem, neural network is used in camera calibration. In [1], a non-linear optimization based on feedforward neural network to camera calibration was proposed and utilized in stereoscopic vision of industrial robots. Zhao [2] proposed a calibration approach based on BP neural networks for binocular vision, and used it in robot curve tracking successfully. Among the aforementioned work, the learning speed of the networks is in general far slower than required, which is a major bottleneck in their applications. In order to solve these problems, Huang [3] [4] proposed a new learning algorithm called extreme learning machine (ELM) for single-hidden layer feedforward neural networks (SLFNs).

Unlike those traditional implementations, ELM randomly chooses the input weights and analytically determines the output weights of SLFNs. This algorithm tends to provide the best generalization performance at extremely fast learning speed, and is already used in complex nonlinear approximation and pattern classification [5]. However, the application of extreme learning machine in camera calibration is still seldom reported.

In this paper, extreme learning machine is utilized to calibrate the camera. Compared with calibration method based on BP neural network, the proposed technique is with simpler structure, faster training speed and higher precision. Computer simulations are

provided to test the proposed technique, and good results are obtained. The rest of the paper is organized as follows. Section 2 describes the basic principle of extreme learning machine and the calibration procedure. Section 3 provides three simulations to validate the proposed technique. Summaries are given in Section 4.

2   Camera Calibration Based on Extreme Learning Machine

2.1   Extreme Learning Machine

For N arbitrary distinct samples (x_i, t_i), where $x_i = [x_{i1}, x_{i2}, .., x_{in}]^T \in \mathbf{R}^n$ and $t_i = [t_{i1}, t_{i2}, \cdots, t_{im}]^T \in \mathbf{R}^m$. The output vector is:

$$o_j = \sum_{i=1}^{\tilde{N}} \beta_i g\left(w_i^T \cdot x_j + b_i\right), \quad j = 1, \cdots, N, \tag{1}$$

where $\mathbf{w_i} = [w_{i1}, w_{i2}, \cdots, w_{in}]^T$ is the weight vector connecting the *ith* hidden neuron and the input neurons, $\beta_i = [\beta_{i1}, \beta_{i2}, \cdots, \beta_{im}]^T$ is the weight vector connecting the *ith* hidden neuron and the output neurons, and b_i is the threshold of the *ith* hidden neuron.

That single-hidden layer feedforward neural networks (SLFNs) including $\tilde{N}$ hidden neurons with activation function $g(x)$ can approximate N samples with zero error, which means $\sum_{j=1}^{N} \|o_j - t_j\| = 0$, *i.e.*, there exist β_i, $\mathbf{w_i}$ and b_i such that

$$\sum_{i=1}^{\tilde{N}} \beta_i g\left(\mathbf{w}_i^T \cdot x_j + b_i\right) = t_j, \quad j = 1, \cdots, N, \tag{2}$$

Then, (2) can be rewritten as

$$\mathbf{H\beta = T} \tag{3}$$

where

$$\mathbf{H}\left(w_1, \cdots, w_{\tilde{N}}, b_1, \cdots, b_{\tilde{N}}, x_1, \cdots, x_N\right)$$
$$= \begin{pmatrix} g\left(w_1^T \cdot x_1 + b_1\right) & \cdots & g\left(w_{\tilde{N}}^T \cdot x_1 + b_{\tilde{N}}\right) \\ \vdots & \ddots & \vdots \\ g\left(w_1^T \cdot x_N + b_1\right) & \cdots & g\left(w_{\tilde{N}}^T \cdot x_N + b_{\tilde{N}}\right) \end{pmatrix}_{N \times \tilde{N}} \tag{4}$$

$$\beta = \begin{bmatrix} \beta_1^T \\ \vdots \\ \beta_{\tilde{N}}^T \end{bmatrix}_{\tilde{N} \times m} \quad \text{and} \quad \mathbf{T} = \begin{bmatrix} t_1^T \\ \vdots \\ t_{\tilde{N}}^T \end{bmatrix}_{N \times m} \tag{5}$$

$\mathbf{H}$ is called the hidden layer output matrix of the neural network; the column of $\mathbf{H}$ is the hidden neuron's output vector with respect to inputs $\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_N$.

Huang [4] has given a strict proof that instead of using the iterative algorithm, the random assignment for values of the network input weight and bias can be utilized when the activation function of the hidden layer is infinitely differentiable. Thus, the training of SLFNs can be realized by finding the minimum norm least squares solution $\tilde{\boldsymbol{\beta}}$ of the corresponding linear system $\mathbf{H}\boldsymbol{\beta} = \mathbf{T}$. Thus, $\tilde{\boldsymbol{\beta}}$ can be obtained by solving (3) as

$$\tilde{\boldsymbol{\beta}} = \mathbf{H}^{\dagger}\mathbf{T} \tag{6}$$

where $\mathbf{H}^{\dagger}$ is the Moore-Penrose generalized inverse of $\mathbf{H}$.

The orthogonal project method can be used when $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ is nonsingular and

$$\mathbf{H}^{\dagger} = \left(\mathbf{H}^{\mathrm{T}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{T}} \tag{7}$$

However, $\mathbf{H}^{\mathrm{T}}\mathbf{H}$ may not always be nonsingular or may tend to be singular in some applications and thus orthogonal projection method may not perform well. The singular value decomposition can be generally used to calculate the Moore-Penrose generalized inverse of H in all cases.

2.2 Design of Camera Calibration Method Based on ELM

Aiming at monocular vision calibration, we construct 2-30-2 type SLFNs. The input and the output of the networks are image coordinates and world coordinates, respectively.

The process of camera calibration method based on ELM is summarized as follows:

1) *Randomly generate input weights and hidden nodes bias;*
2) *From the overall samples, randomly select n groups and m groups of samples for training and testing, respectively;*
3) *Calculate the hidden layer output matrix H using n groups of training samples*
4) *Calculate output weights matrix $\tilde{\boldsymbol{\beta}}$ by using (6).*

For evaluation, we define the root mean square (RMS) error of single sample as:

$$E_{p,rms} = \sqrt{\frac{\Delta x^2 + \Delta y^2}{2}} \tag{8}$$

And the RMS error of all the testing samples is defined as:

$$E_{rms} = \sqrt{\frac{E_{1,rms}^2 + E_{2,rms}^2 + \cdots + E_{s,rms}^2}{s}} \tag{9}$$

where Δx and Δy are testing errors of X coordinate and Y coordinate, respectively, and s is the number of testing samples.

3   Simulations

The data provided by Zhang [6] is utilized in our simulations. The image resolution is 640×480 and the size of the pattern is 17cm×17cm. 8×8 squares are included in the model plane means that there are 256 corners. Due to monocular vision calibration, we assume the model plane is located at Z=0 for simplicity.

To show the superiority of the proposed method based on ELM, BP neural network is used for comparing, which is 2-12-2 type. We use the hyperbolic tangent function given by Tan-Sigmoid as the activation function in hidden layer while all neurons in other layers have linear activation functions. The calibration precision, speed and the dependence on training sample numbers of the two methods are respectively discussed in Section Ⅲ-A and Section Ⅲ-B, and the parameter sensitivity of extreme leaning machine is discussed in Section Ⅲ-C.

3.1   Comparition of the Calibration Precision and Speed

In the first simulation, 60 and 10 groups of samples are randomly chosen as training samples and testing samples respectively, the testing results are shown in Table 1, where T_{train} is the training time.

From table 1, it is obvious that the proposed method can calibrate more accurately and hundreds of times faster than the solution based on BP neural network.

Table 1. Composition of the two methods on calibration precision and calibration speed（mm）

real coordinates		output of BP			output of ELM		
x_w	y_w	x_w	y_w	$E_{p,rms}$	x_w	y_w	$E_{p,rms}$
0	-12.7000	-0.2003	-12.2422	0.3533	0.0005	-12.5766	0.0873
80.4334	-12.7000	80.4299	-12.7856	0.0606	80.4153	-12.5495	0.1072
148.1666	0	148.4074	-0.4943	0.3888	148.1947	-0.0352	0.0318
22.5778	-22.5778	22.7595	-22.7512	0.1776	22.5825	-22.5089	0.0488
112.8888	-35.2778	112.8519	-35.2886	0.0271	112.8696	-35.1957	0.0597
12.7000	-57.8556	12.6250	-57.9218	0.0707	12.6601	-57.7686	0.0677
80.4334	-45.1556	80.4036	-45.2029	0.0395	80.4242	-45.1059	0.0358
135.4666	-45.1556	135.4259	-45.1830	0.0347	135.4640	-45.1496	0.0047
45.1556	-80.4334	45.2067	-80.4211	0.0372	45.1152	-80.3199	0.0852
125.5888	-80.4334	125.4819	-80.4435	0.0759	125.5879	-80.3786	0.0388
E_{rms}		0.2681			0.0629		
T_{train}		2.813s			0.016s		

3.2   Comparition of the Dependence on Training Sample Numbers

In order to compare the dependence on parameters of the two methods, 220, 150, 100, 60, 40, 20 groups samples are randomly chosen as training samples for the two methods, and 20 groups samples are chosen for test. The results are shown in Table 2.

Table 2. Comparison of the dependence on training sample numbers

Sample number	E_{rms}	
	BP solution	ELM solution
220	0.0788	0.0583
150	0.0795	0.0564
100	0.0916	0.0581
60	0.2701	0.0625
40	0.3728	0.0695
20	6.9723	0.0986

It can be seen from the Table 2, compared with the method based on BP neural networks, the calibration precision of the proposed method is less influenced by the number of training samples, and can reach good performance when the training sample number is reduced to 20.

3.3   Parameter Sensitivity of Extreme Leaning Machine

For the calibration based on ELM, only the parameter μ of activation function needs to be adjusted in training process of the networks. The influence of μ to calibration precision is shown in Fig.1.
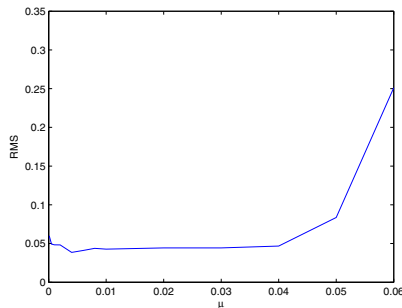
Fig. 1. The influence of parameter μ to calibration precision

From Figure 1, it can be observed that the method based on ELM is insensitive to the parameter μ, when $\mu \in (0, 0.04)$.

4   Summaries

This paper has proposed a camera calibration method based on extreme learning machine, which is used to project image space to the world space directly. Neither camera model nor prior knowledge about the parameters is required. The algorithm of ELM has higher efficiency, and it can effectively simplify the calibration process. The simulations show that the proposed method based on ELM can provide more accurate machine vision calibration than BP neural networks.

References

[1] Lynch, M.B., Dagli, C.H., Vallenki, M.: The use of feed forward neural networks for machine vision calibration. International Journal of Production Economics 60-61, 479–489 (1999)
[2] Zhao, Q., Sun, Z., Lan, L.: Neural network technique in camera calibration. Control and Decision 7, 336–338 (2002) (in Chinese)
[3] Huang, G.B., Zhu, Q.Y., Siew, C.K.: Extreme learning machine: A new learning scheme of feedforward neural networks. In: Proc. International Joint Conference on Neural Networks, JCNN 2004, Budapest, Hungary, pp. 25–29 (2004)
[4] Huang, G.B., Siew, C.K.: Extreme learning machine with randomly assigned RBF kernels. International Journal of Information Technology 11, 16–24 (2005)
[5] Zhu, Q.Y., Qin, A.K., Suganthan, P.N., et al.: Evolutionary Extreme Learning Machine. Pattern Recognition 38, 1759–1763 (2005)
[6] Zhang, Z.Y.: A flexible new technique for camera calibration. IEEE Transaction on Pattern Analysis and Machine Intelligence 22, 1330–1334 (2002)

Analysis of Electromagnetism Disturbing Wireless Fuzes by Fault Tree Method

Zhang Yan-Pu, Wei Xian-Jie, and Tan Chao-Bin

Department 3, Artillery Commanding Institute, Langfang, 065000, China

Abstract. Modeling of electromagnetism disturbing wireless fuzes logically by Fault Tree Method firstly in the paper, and then Magnitude of Critical Concernment(CIg(i)) and Frame Concernment(Ig(i)) to basic faults of this disturbing has been researched by analyzing every basic fault qualitatively and quantitatively, so weightiness to every basic fault can be evaluated. This paper will be of some importance to enhance the ability of resisting to electromagnetism disturbing for wireless fuzes.

Keywords: Wireless Fuzes, Electromagnetic Disturbing, Fault Tree Method, Critical Concernment, frame Concernment.

1 Introduction

Wireless fuzes are important electron devices in some control and guided system. Analysis of the effect of electromagnetism disturbing wireless fuzes by fault tree method and studying weightiness to every basic fault, will enhance and reform the ability of resisting to electromagnetism disturbing in a degree.

2 Modeling of Electromagnetism Disturbing Wireless Fuzes by Fault Tree Method

Traditional fault tree analysis method is an effective method to analyze system reliability. Accoding to this method, now analyzing those disturbing as follows: wireless fuzes disturbing by electromagnetism can be divided into two basic forms, disturbing inside and disturbing outside showing in Fig.1.
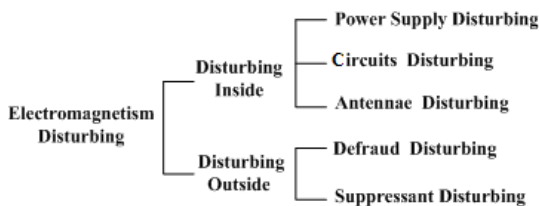
Fig. 1. The wireless fuzes disturbing by electromagnetism

Disturbing inside generally includes: power supply disturbing, circuits disturbing and antennae disturbing. power supply disturbing contains power supply abnormity, high voltage, low voltage, and other power supply disturbing by noises. circuits disturbing contains design bugs, high frequency impedance dismatching, circuits disturbing by positive feedbacks themselves, and so on. antennae disturbing contains antennae setting unsuitable, electromagnetism leakage into fuzes. The errors can be tested by experiment in disturbing inside and avoided by mends of bettering devise, so disturbing inside is not dominant in all this disturbings[1][2].

Disturbing outside mostly includes: defraud disturbing and suppressant disturbing. defraud disturbing is an important method of electronic countermeasure, with a spurious signal created from simulation of perceptive frequency to a signal of a wireless fuse, leading to explode ahead of schedule. Suppressant disturbing is another method of generating mighty electromagnetic impulse to interfere electronic devices of wireless fuzes, regardless of their frequencies, such as electromagnetism impulse bombs, nucleus impulse bombs, and high-power microwave weapons. Certainly, thunder disturbing is a sort of   disturbing outside too[3][4]. The fault tree structure of electromagnetism disturbing wireless fuzes is shown in Fig.2.
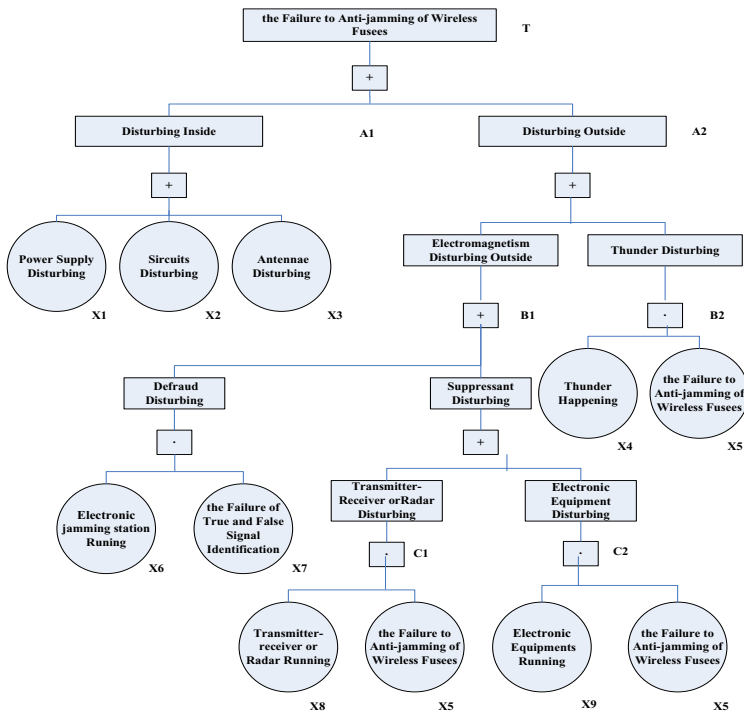
Fig. 2. The fault tree structure of electromagnetism disturbing wireless fuzes

3   The Fundamental Theory of Qualitative Fault Tree

3.1   The Basic Fault Tree Structural Function

It is $T=A_1+A_2=X_1+X_2+X_3+B_1+B_2=X_1+X_2+X_3+X_4 \cdot X_5+X_6 \cdot X_7+C_1+C_2$

$= X_1+X_2+X_3+X_4 \cdot X_5+X_6 \cdot X_7+ X_5 \cdot X_8 +X_5 \cdot X_9$.

So, the Least Detached Setting can be deduced as follows.

$K_1=\{X_1\}, K_2=\{X_2\}, K_3=\{X_3\}, K_4=\{X_4 \cdot X_5\}, K_5=\{X_6 \cdot X_7\}, K_6=\{X_5 \cdot X_8\}, K_7=\{X_5 \cdot X_9\}$.

3.2   Analysis of Weightiness to Every Basic Fault

Accordding to the approximate formulation of weightiness to every basic fault:

$$I(j) = \sum_{X_j \in K_r} \frac{1}{2^{n_j - 1}}$$

In the algorithm, $I(j)$ is the approximate calculated value of structural coefficients of basic faults X_j. $X_j \in K_r$, every basic fault X_j is attributed to the least detached setting K_R. Here, nj is the sum of every basic fault X_J embodied every least detached setting. calculate approximate calculated values of Structural coefficients as tab. 1.

Table 1. Approximate calculated values of structural coefficients

I(1)	I(2)	I(3)	I(4)	I(5)	I(6)	I(7)	I(8)	I(9)
1	1	1	1/2	3/2	1/2	1/2	1/2	1/2

As table 1 showing, I(5)> I(1)= I(2)= I(3)> I(4)= I(6)= I(7)= I(8)= I(9), it is concluded that I ø (5) > I ø (1) = I ø (2) = I ø (3) > I ø (4) = I ø (6) = I ø (7) = I ø (8) = I ø (9) .

In the structure, the basic fault X_5(the Failure to anti-jamming of wireless fuzes)and disturbing inside(X_1, X_2, X_3) is in the highest flight, and that disturbing outside (X_4, X_6, X_8, X_9) is subordinate. In fact, disturbing Inside can be eliminated by experiment, But disturbing outside is a great probability of occurrences in modern wars, dominant in all this disturbings.

4   The Electromagnetism Disturbing Analysis Based on Qualitative Fault Tree

To calculate weightiness to every basic fault more veracious, the quantitative fault tree analysis follows.

4.1 Calculation to Probability of Up - Affair Using Algorithm of Independent Approximator

$$g = 1 - \prod (1 - \prod_{x_i \in k_r} q_i)$$

the Basic Fault Tree Structural Function above(T= $X_1+X_2+X_3+X_4 \cdot X_5+X_6 \cdot X_7+X_5 \cdot X_8 +X_5 \cdot X_9$), And the data from table 2. (probability of every basic fault is from References[1])

Taking probability of every basic fault happening qi into Algorithm (3), probability of up- affair g = 1-(1-q_1) (1-q_2) (1-q_3) (1-$q_4 \cdot q_5$) (1-$q_6 \cdot q_7$) (1-$q_5 \cdot q_8$) (1-$q_5 \cdot q_9$) = 0.017412961 can be obtained. That is to say, probability of electromagnetism disturbing wireless fuzes is 1.74% around. So the effect of electromagnetism disturbing must be taked into account in modern wars.

4.2 Calculation to Concernment Coefficient to Probability of Every Basic Fault Happening

Table 2. BASIC fault probability

code name	Basic fault name	q_i	$1-q_i$
X1	Power Supply Disturbing	10^{-6}	0.999999
X2	Circuits Disturbing	10^{-6}	0.999999
X3	Antennae Disturbing	10^{-6}	0.999999
X4	Thunder Happening	10^{-3}	0.999
X5	the Failure to Anti-jamming of Wireless Fuzes	10^{-2}	0.99
X6	Electronic Jamming Bugs Running	0.5	0.5
X7	the Failure of True and False Signal Identification	10^{-2}	0.99
X8	transmitter-receiver or radar disturbing	0.5	0.5
X9	Electronic Equipments Running	0.75	0.25

According to algorithm of $I_g(i) = \dfrac{\partial g}{\partial q_i}$, concernment coefficient to probability of every basic fault happening has been obtained. (Tab. 3)

Table 3. Concernment coefficient to probability of every basic fault happening

Ig (1)	Ig (2)	Ig (3)	Ig (4)	Ig (5)
0.98258802	0.98258802	0.98258802	0.009825968	1.23726145
Ig (6)	Ig (7)	Ig (8)	Ig (9)	
0.009875246	0.493762331	0.00994987	0.009900121	

4.3   Analysis of Critical Concernment to Every Basic Fault Happening

Using algorithm of $CIg(i) = \dfrac{q_i}{g} I_g(i)$, critical concernment to every basic fault happening has been obtained (Tab. 4).

Table 4. Critical concernment to every basic fault happening

CIg(1)	CIg(2)	CIg(3)	CIg(4)	CIg(5)
$5.6428e^{-5}$	$5.6428e^{-5}$	$5.6428e^{-5}$	$5.6429e^{-4}$	0.71054
CIg(6)	CIg(7)	CIg(8)	CIg(9)	
0.28356	0.28356	0.28570	0.42641	

As Tab4 showing, critical concernment to every basic fault happening sequenced as follows.

CIg (5)> CIg (9)> CIg (8)> CIg (6)= CIg (7)> CIg (4)= CIg (3)= CIg (2)>=CIg (1)

5   Summaries

(1) The fault X_5 is the greatest value in all critical concernments $CIg(i)$(the capability to anti-jamming of wireless fuzes). It is testified that to ameliorate designs to boost the anti-jamming capability is key to prevent disturbing.

(2) The fault X_8(transmitter-receiver or radar disturbing), the fault X_7(the failure of true and false signal identification), the fault X_6(electronic jamming stations runing) each accounts for a great proportion in all critical concernments $CIg(i)$. Especially in modern electronic-counter battlefield, electronic jamming transmitter, radar, transmitter-receiver would radiate massive electromagnetic waves, and finally engender a complex electro-magnetic disturbing circumstance. To cut down this influence in greatest possibility, firepower-destroyed and electromagnetism suppressant anti-jamming must be applied. At the same time, it is quite urgent to enhance the ability of resisting to electromagnetism disturbing, to reform target information identifications and to countercheck defraud disturbing.

(3) The fault X_1(power supply disturbing), X_2(circuits   disturbing) and X_3 (antennae disturbing), among inside disturbing are a little probability of happenings, but this need to test fuzes repeatedly and select furthest optimized schemes in finalizing the design of wireless fuzes. If not, inside disturbing would go up to a high degree. Yet the effect of the fault X_4(thunder happening) can be greatly reduced by optimizing lightning guard design of wireless fuzes.

References

[1] Zhang, J.-L.: Safety Systems Engineering, pp. 12–13. North China Engineering Academy, Shijiazhuang (2000)
[2] Yi, J.-Z.: Safety Technology and Administer. Ordnance Engineering Academy, Shijiazhuang (1992)

[3] Du, H.-Q.: Wireless Fuzes Anti-jamming Principles. Weapon Industry Publishing Company, Beijing (1985)
[4] Chen, Y.-Q.: Electronic Fuzes Electromagnetism Effect and Reinforced Technology Research Ordnance Engineering Academy (1999)

Research on Function Block Application in Process Control System

Zhuang Xia

Civil Aviation Flight University of China, Sichuan, 618307

Abstract. Present and future digital process control systems need to many requirements, for example, increase security and safety, reduce time to market, minimize training costs, support integrated methodology for implementation, and so on. Process control systems are required to fulfil these requirements in terms of their architecture and their operation during all the phases of the life cycle. The accepted basic concept for the design process control system is to describe all necessary implementation-specific functions with function block. IEC 61804 defines the function block for process control. In this paper, we discuss how a consistent function block capability may be provided for all fieldbus technology utilized in a control system. Examples will be given of how this standard has been applied in modern control systems to give a consistent interface to Foundation Fieldbus and PROFIBUS. Some detail will be presented on the standard means that is defined for manufacturers to describe function block capability of a field device.

Keywords: Process control system, Function block, Device, Standard.

1 Introduction

A variety of digital fieldbus devices and fieldbus technologies have been introduced within the process industry. There has been a gradually acceptance of the fact that a variety of communication technologies are needed to fully address the application requirements of a manufacturing facility. However, engineers responsible for the specification, engineer, implement of control system require that a common interface and functionality be proved in the control system. This capability should be independent of the underlying fieldbus technology or manufacturer of the fieldbus device. The draft IEC 61804 standard defines how a control system can be structured to provide this flexibility in the utilization of fieldbus technology[1].

The draft IEC 61804 standard is an end user driven specification of the requirements of distributed process control systems based on Function Blocks. This requirement specification (which is defined as part 1) and its associated function block standard (part 2) originate from the power plant industrial sector. It is validated by applications in oil and gas, petrochemicals, pharma ceuticals and fine chemicals, pulp and paper, food and beverage, waste water treatment plants, steel milling and others. The specification defines the requirements for function blocks to provide control, and to facilitate the maintenance and the technical management as

applications, which interact with actuators and measurement devices[2]. These parts were prepared by IEC SC65C WG7 and are available as draft IEC 61804.

2   Fonction Block

IEC 61804 defines a function block model which graphical representation is shown in Fig.1. It consists of the components FB Head and FB Body. The body carries the data flow (Data inputs and data outputs, algorithms, internal data and contained data, which are not involve in the data flow, but adjust the algorithms) and the head of the event of flow (event inputs, event outputs and execution control).
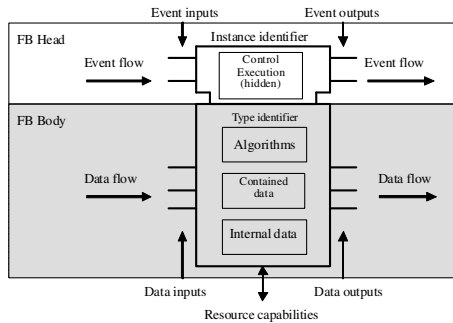
Fig. 1. Structure of Function Block

The execution order of function blocks in a centralized program is determined by the order of the function block call in the program, or by the scheduling of the task system of one resource[3]. In the distributed environment in terms of IEC 61804 the execution order of function blocks is determined by the event flow of the FB heads. Because of the event connection between the execution control head of a function block and the local operating system which is seen as scheduling function. The head of the function block is the configurable part of the distributed operation system. In other words, the distributed operation system is build by the local operating system of each resource and the connections between function block execution control heads.

3   Function Block Application Components

3.1   From Control System to Function Block

Function Blocks are encapsulations of variables, parameters and their processing algorithms. The variables, parameters and algorithms are those required by the design of the process and its control system. They can be derived from the Pipe and Instrumentation Diagram (P&ID)[4]. These are presented in 2 different clauses. One deals with "rich" FBs covering complex but common functions suchas control loop (for example proportional, integral, differential PID) required by the majority of the

process industries. Another covers a set of elementary FBs (EFB) such as Boolean functions required to compose very specific and unique functionality.

The application can be distributed among several devices. The devices are connected via a communication network or a hierarchy of communication networks. There are different function block types, which encapsulate specific functionality of devices performing an automation application. The Technology Block represents the process attachment of a device. It contains the measurement or actuation principles of a device. The Application Function Block contains application related signal processing, such as scaling, alarm detection or control and calculation. Elementary functions and elementary function blocks contains mathematical and logical functions with specific additional exception handling procedures. The Device Block represent the resource of the device, which contains information and function about the device itself, the operation system of the device and the device hardware.
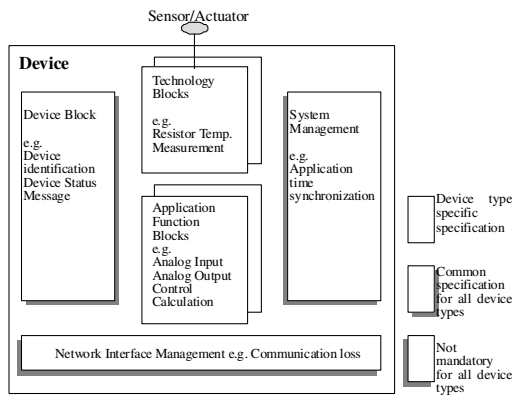
Fig. 2. Device Structure

All devices in the scope of IEC 61804 are expected to have the same logical device structure. The number and types of blocks, which are instantiated in a device, are device and manufacturer specific. There is a data flow chain from signal detection through the Technology Block and Function Blocks and vice versa. The signals between the parts of the chain can be internal within the blocks or visible as linkages between blocks. The logical chain of technology and function block is called a channel.

3.2   IEC 61804 Function Block Overview

The description of an IEC 61804 function block is a list of algorithms that are encapsulated in the block together with the related input and output variables and parameters. There are algorithms that are related to the process signal flow and algorithms that are related to other block specific functions. The parameter table shows all the needed accessible parameters of the block.
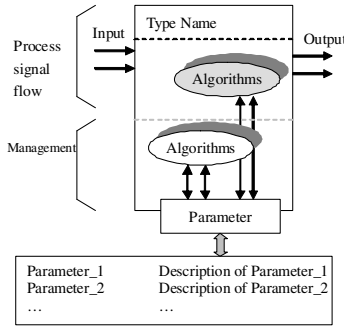
Fig. 3. IEC 61804 function block overview

The algorithm description is done individual for each algorithm by a device designer/manufacturer in the appropriate language, e.g. plain English, Harel State Diagram, IEC 61131 FBD (function block diagram) or IEC 61131 ST (structured text). The block variables, parameters and algorithms included in a block will be those that are significant for the algorithm and device. As a minimum, function blocks will include the variables and parameters defined in the P&ID.

4   Function Block Application

The technology and application blocks built a function chain along with the process signal flow. Together they built a measurement or actuation channel. Measurement and actuation channels perform together with control and calculation function blocks in the application (Fig.4). The technology blocks are technology dependent and the function blocks are technology independent. There are different implementations of an application, depending from the used technology of the devices. The application may be performed by implementing the application in measurement and actuation devices only (i.e. a complex device may perform measurement, control and actuation.) or in measurement and actuation devices together with controllers and other components of the system. A controller may be for instance integrated in the application as one calculation function block or an actuation device may take parts of programmable functions from controller devices in terms of calculation function blocks.
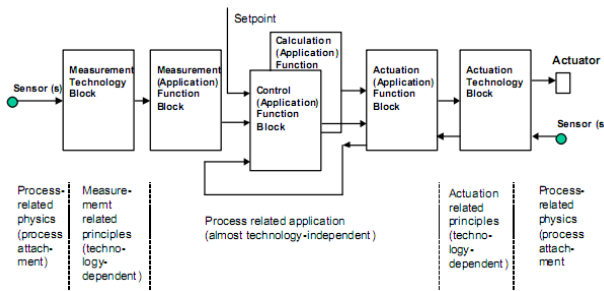
Fig. 4. Application process signal flow

There are different device characteristics, called device types. The main characteristic aspect is the execution control methods. Execution control of function block algorithms is a feature of each device. There are many possible execution policies within devices and in a distributed system:

- Free running
- Device internal time schedule (time synchronization)
- System wide time synchronization
- Communication service triggered
- Device internal event triggered
- Distributed execution control

Which execution control is used depends on the capability of the devices. Therefore the execution control is defined by concrete fieldbus systems[5].

5   Example Application

The function blocks resulting from the design of a process control system are abstract representations and may be implemented in different ways in different device types (Fig.5). Function blocks can be implemented e.g. in field devices, PLC, visualization stations an d device descriptions. Additionally other applications such as system engineering and the supervisory system have to handle or interact with the function blocks. Functions defined for a function block in the conceptual model are not necessarily mapped one-to-one to the device; they can be mapped to the device and a proxy if the current technology doesn't solve it in the device.
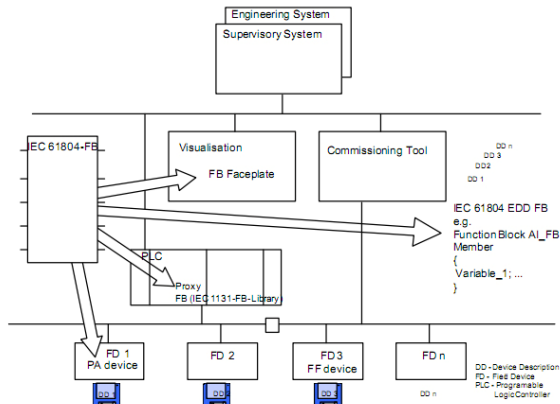
Fig. 5. IEC 61804 function blocks can be implemented in different devices

One example of how the IEC 61804 specification may be applied to give a consistent function block interface to different technologies is the DeltaV Control system from Fisher-Rosemount Systems, Inc. The system architecture is based on Foundation Fieldbus function block specification and thus allows control strategies to

be defined independent of whether the function block executes in the fieldbus device or in the DeltaV controller. Fieldbus devices that do not support function block such as those based on AS-Interface, Profibus DP, and DeviceNet are brought into the system using the concept of a proxy in the controller to add function block capability to the measurement or actuator value provided by the device.

To the end user, these proxy function blocks may be configured for monitoring and calculation application in the same manner as Foundation Fieldbus function blocks. The advanced signal processing and alarming that would normally be done by function blocks in the fieldbus device are done instead in the controller. Though the same functionality is provided through the proxy, there are differences in the dynamic associated with the signal processing. In many cases, these timing differences will have not impact on the application. An example configuration application that utilizes inputs and output from Foundation Fieldbus, AS-Interface, and Profibus DP is shown in Fig.6.
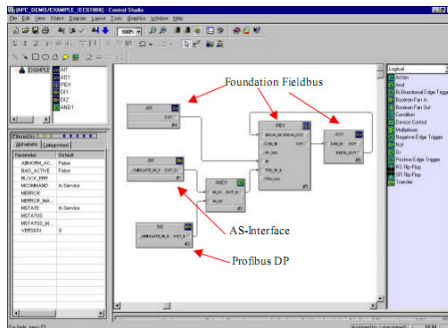
Fig. 6. Function block example for Foundation Fieldbus, AS-Interface, and Profibus DP

6   Summaries

Process control systems that use fieldbus, function blocks and device descriptive languages already exist in the market. It is clear that the majority of these systems will continue to use function blocks into the foreseeable future, as they are modular and re-usable. However control system manufacturers, implementers and users are faced with:

- Multiple technology platforms that are changing rapidly.
- Multiple communication standards.

The approaches are similar, but the technologies are different.

As for the mapping on the several technologies, an important role is played by the proxy approach. It is expected that process plants users use the standard as a point of reference in specifying their control system requirements and in evaluating technology to meet these requirements in formal manner. Provided that some function blocks are not completely solved in the existing devices, the proxy approach allows the system integrator to complement these devices by solving the missing functionalities with some additional software installed in the controller. System

integrators are able to define the proxies needed for the different applications and for the different industrial sectors. Thus IEC 61804 allows the users to carry a top-level view of their control systems forward into the future and the manufacturers to answer with evolving solutions.

References

[1] IEC 61804-Function blocks (FB) for process control Part 1: General Requirements Committee Draft, International Electric Commission, Draft, Tech. Comm. 65, Working group 6, Geneva

[2] IEC 61804-Function blocks (FB) for process control Part 2: Specification Committee Draft, International Electric Commission, Draft, Tech. Comm. 65, Working group 6, Geneva

[3] Brennan, R., Zhang, X., Norrie, H.: A Reconfigurable Concurrent Function Block Model and its Implementation in Real-Time Java. Journal of Integrated Computer-Aided Engineering, 263–279 (2002)

[4] Chen, H.-C., Chang, J.-F.: EP-based PID control design for chaotic synchronization with application in secure communication. Expert Systems with Applications, 1169–1177 (February 2008)

[5] Mahalik, N.P., Yen, M.: Extending fieldbus standards to food processing and packaging industry. Computer Standards & Interfaces, 586–598 (March 2009)

Uniform Quasi-synchronization of Fractional-Order Chaotic Systems in the Presence of Parameter Mismatches Using Active Control

Ying Shi, Jinde Cao, and Shun Chen

Department of Mathematics, Southeast University, Nanjing 210096, China

Abstract. This paper investigates the uniform quasi-synchronization of a class of fractional-order chaotic systems in the presence of parameter mismatches by utilizing the active control. In practical situations, parameter mismatches exactly exist. Based on the Laplace transform theory and Gronwall-Bellman lemma, a suitable condition for uniform quasi-synchronization with a bound on the synchronization error is derived. Finally, numerical simulations are given to show the effectiveness of the theoretical result.

Keywords: Fractional-order systems, Uniform quasi-synchronization, Parameter mismatches, Active control.

1   Introduction

In recent years, fractional-order systems attract more and more physicists and engineers' attention [1], which is mainly due to the fact that many real world physical systems are well characterized by fractional-order systems. Meanwhile, synchronization of fractional-order systems is also receiving increasing attention [2], some good results of which are derived.

In order to achieve the synchronization, many control techniques including continuous feedback and discontinuous feedback have been developed, such as nonlinear feed-back control [3], adaptive control [4], back-stepping control [5], impulsive control [6], active control [7], etc. In this paper, active control is used to guarantee the synchronization of fractional-order systems basted on the Laplace transform theory. In fact, it is an effective method to synchronize two different chaotic systems.

In many papers, synchronization schemes assume that the drive-response systems are identical [8] However, there are always some mismatches between drive system and response system in practical applications. In certain cases, parameter mismatches are detrimental to the synchronization quality. Mismatches robustness and its effects on synchronization are studied in [9]. Despite their importance for practical applications, studies on the effect of parameter mismatches in chaotic synchronization are still quite scarce in the literature. Particularly in fractional-order system, few people had investigated the parameter mismatches problem.

In this paper, uniform quasi-synchronization of fractional-order systems in the presence of parameter mismatches is investigated. By means of the Laplace transform

theory and Gronwall-Bellman lemma, a suitable condition for the uniform quasi-synchronization is given. We show that the parameter mismatches influence the synchronization of fractional-order systems, that is, fluctuations of drive system can affect the synchronization of fractional-order systems. Via using the active control, the synchronization error bound is derived under suitable conditions.

The rest of this paper is organized as follows. In Sec.2, the definitions of fractional-order Caputo derivative and the Mittag-Leffler function are presented. In Sec.3, a sufficient condition for the fractional-order non-identical systems in the presence of parameter mismatches is obtained. In Sec.4, numerical examples are given to show the effectiveness of the result, and in Sec.5, conclusions are finally drawn.

2   Preliminaries

This section gives some necessary preliminaries to derive the main results of this paper.

2.1   Notations

Through the paper, R^n and $R^{n \times n}$ denote $n-$dimensional real space and the set of all $n \times n$ Real matrices, respectively. The superscript T denotes matrix transposition. I_n is the $n \times n$ identity matrix. $\|\cdot\|$ is the Euclidean norm in R^n. If A is a matrix

$\|A\| = \sup\{\|Ax\| : \|x\| = 1\} = \sqrt{\lambda_{\max}(A^T A)}$, where $\lambda_{\max}(\cdot)$ is the largest eigenvalue of A

2.2   Some Supporting Definitions and Lemmas

There are many definitions of fractional derivatives, such as the Riemann-Liouville and the Caputo fractional derivatives. Fractional differential equations in terms of the Riemann-Liouville derivatives require initial conditions expressed in terms of initial values of fractional derivatives of the unknown function, but initial conditions for the Caputo derivatives are expressed in terms of initial values of integer-order derivatives. Furthermore, the Caputo derivative of a constant is 0. From the physical point of view, the Caputo definition is better than the Riemann-Liouville definition. So we choose the Caputo derivative and not the Riemann-Liouville derivative.

Definition 1 [10]. The Caputo fractional-order derivative is

$$D^\alpha f(t) = \frac{1}{\Gamma(N-\alpha)} \int_{t_0}^t \frac{f^{(N)}(\tau)}{(t-\tau)^{\alpha+1-N}} d\tau, \tag{1}$$

Where D^α is generally called the Caputo differential operator of order $\alpha > 0$, and $N = \lceil \alpha \rceil$, i.e.,

N is the first integer which is not less than α, $\Gamma(\cdot)$ is the Gamma function.

Definition 2 [11]. A function frequently used in the solution of fractional-order system is the Mittag-Laffler function, defined as

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}, \alpha > 0, \beta > 0. \tag{2}$$

The Laplace transform of Mittag-Leffler function with two parameters is

$$\int_0^\infty e^{-st} t^{\alpha k+\beta-1} E_{\alpha,\beta}^{(k)}(at^\alpha) dt = \frac{k! s^{\alpha-\beta}}{(s^\alpha - a)^{k+1}}, \tag{3}$$

where $\mathrm{Re}(s) > |a|^{1/\alpha}$.

Lemma 1 [12]. If $A \in C^{n \times n}$ and $\alpha < 2$, β is an arbitrary real number, η is a constant such that $\alpha\pi/2 < \eta < \min\{\pi, \pi\alpha\}$ and $C > 0$ is a real constant, then

$$\left\| E_{\alpha,\beta}(A) \right\| \le \frac{C}{1 + \|A\|}, \eta \le \left| \arg(\lambda_i(A)) \right| \le \pi, i = 1, 2, \cdots, n, \tag{4}$$

Where $\lambda_i(A)$ $(i = 1, 2, \cdots, n)$ denote the eigenvalues of matrix A.

Lemma 2 (Gronwall-Bellman Lemma) [13]. Assume that: 1) $u(t)$ and $f(t)$ are real-valued piecewise-continuous functions defined on the real interval $[a,b]$; 2) $K(t)$ is also real-valued and $K(t) \in L(a,b)$; 3) $u(t)$ and $K(t)$ are nonnegative on this interval. If for all $t \in [a,b]$ such that $u(t) \le f(t) + \int_a^t K(\tau)u(\tau)d\tau, \forall t \in [a,b]$, then

$$u(t) \le f(t) + \int_a^t f(\tau)K(\tau)\exp\{\int_\tau^t K(r)dr\}d\tau, \forall t \in [a,b].$$

3   Problem Formulation

Consider the fractional-order drive system

$$D^\alpha x(t) = A_1 x(t) + f(x(t)), t \ge 0, \tag{5}$$

where $x(t) \in R^n$ is the state vector, A_1 is $n \times n$ constant matrix, $f(x(t)) : R^n \to R^n$ is nonlinear continuous vector-value function and satisfying the Lipschiz condition: there exists constant K such that

$$\|f(s_1) - f(s_2)\| \le K \|s_1 - s_2\|.$$

The corresponding response system is designed as

$$D^\alpha y(t) = A_2 y(t) + f(y(t)) + U(t), t \ge 0, \tag{6}$$

where $y(t) \in R^{n \times n}$ denotes the state vector of system (6), A_2 is $n \times n$ constant matrix, $U(t)$ is active control of system (6). The purpose is to design a suitable $U(t)$ to synchronize the state of drive system (5) and response system (6).

Remark 1. The paper assumes that there exist parameter mismatches between the drive system and response system, that is, $A_2 \ne A_1$. Let $\Delta A = A_2 - A_1$ denotes the parameter mismatches error.

Let $e(t) = y(t) - x(t)$ be the synchronization error between the states of drive system (5) and response system (6). The active control function is constructed as follows

$$U(t) = (2A_1 - A_2)x(t) - A_1 y(t) + U(y(t) - x(t)) \quad,$$

where $U \in R^{n \times n}$ is gain matrix. So the error system can be expressed as
$$D^\alpha e(t) = (\Delta A + U)e(t) + f(y(t)) - f(x(t)). \tag{7}$$

Remark 2. Evidently, the synchronization error $e(t)$ can never converge to zero, i.e., the complete synchronization is impossible. Yet we can find an error bound and make the error small as possible by choosing a suitable control. Uniform quasi-synchronization with error bound is adopted to the non-identical chaotic systems (8) and (10).

Definition 3 [14]. Let Ω denotes a region of interest in the phase space that contains the chaotic attractor of system (5). The synchronization schemes (5) and (6) are said to be uniformly quasi-synchronized with error bound $\varepsilon > 0$ if there exist a $T > 0$ such that $\|y(t) - x(t)\| \le \varepsilon$ for all $t \ge T$ starting from any initial values $x(0) \in \Omega$ and $y(0) \in \Omega$.

The quasi-synchronization concept given in Definition 1 is global since it dose not depend on the initial values. In this paper, synchronization results are in terms of uniform quasi-synchronization.

4   Main Results

Theorem 1. Suppose that $\Psi = \{x \in R^n : \|x\| \le \varphi\}$ is the phase range of interested chaotic systems, fractional-order drive system (5) and response (6) are uniformly quasi-synchronization, if $\Delta A + U$ satisfies $\alpha\pi/2 < \left|\arg(\lambda_i(\Delta A + U))\right| < \pi$ and the parameter mismatches satisfy $\|\Delta A + U\| > \dfrac{CK}{\alpha}$ $(0 < \alpha < 1)$.

Proof: The Laplace transform of fractional-order differential equation (7) with the order $0 < \alpha < 1$ is

$$E(s) = L\{e(t); s\} = (s^\alpha - (\Delta A + U))^{-1}\{s^{\alpha-1}e(0) + L\{f(y(t)) - f(x(t)); s\}\}, \quad (8)$$

where $L\{g(t); s\}$ denotes the Laplace transform of $g(t)$.

Then taking Laplace inverse transform for Eq. (8) by using the Laplace transform formula of Mittag-Leffler function in two parameters Eq. (3) yields that

$$e(t) = E_{\alpha,1}((\Delta A + U)t^\alpha)e(0) + \int_0^t (t-\tau)^{\alpha-1}E_{\alpha,\alpha}((\Delta A + U)(t-\tau)^\alpha)$$
$$\{f(y(t)) - f(x(t))\}d\tau.$$

Taking the norm on both sides of the above equation by inequality (4), one has

$$\|e(t)\| \le \frac{C\|e(0)\|}{1 + \|\Delta A + U\|t^\alpha} + \int_0^t \frac{CK(t-\tau)^{\alpha-1}}{1 + \|\Delta A + U\|(t-\tau)^\alpha}\|e(\tau)\|d\tau. \quad (9)$$

In Lemma 3, set

$$u(t) = \|e(t)\|, \ f(t) = \frac{C\|e(0)\|}{1 + \|\Delta A + U\|t^\alpha}, \ K(\tau) = \frac{CK(t-\tau)^{\alpha-1}}{1 + \|\Delta A + U\|(t-\tau)^\alpha}.$$

Then by the inequality in Lemma 2, one has

$$\|e(t)\| \le \frac{C\|e(0)\|}{1 + \|\Delta A + U\|t^\alpha} + \int_0^t \frac{C\|e(0)\|}{1 + \|\Delta A + U\|t^\alpha} \frac{CK(t-\tau)^{\alpha-1}}{1 + \|\Delta A + U\|(t-\tau)^\alpha}$$

$$\exp\{\int_\tau^t \frac{CK(t-r)^{\alpha-1}}{1 + \|\Delta A + U\|(t-r)^\alpha}dr\}d\tau$$

$$= \frac{C\|e(0)\|}{1+\|\Delta A+U\|t^{\alpha}} + C^2 K\|e(0)\|$$

$$\int_0^t \frac{1}{1+\|\Delta A+U\|\tau^{\alpha}} \frac{(t-\tau)^{\alpha-1}}{(1+\|\Delta A+U\|(t-\tau)^{\alpha})^{1-\frac{CK}{\alpha\|\Delta A+U\|}}} d\tau$$

$$= \frac{C\|e(0)\|}{1+\|\Delta A+U\|t^{\alpha}} + 2C^2 K\|e(0)\| \int_0^{\frac{t}{2}} \frac{\tau^{\alpha-1}}{(1+\|\Delta A+U\|\tau^{\alpha})^{2-\frac{CK}{\alpha\|\Delta A+U\|}}} d\tau$$

$$= \frac{C\|e(0)\|}{1+\|\Delta A+U\|t^{\alpha}} + 2C^2 K\|e(0)\| [(1+\|\Delta A+U\|(\frac{t}{2}))^{-1+\frac{CK}{\alpha\|\Delta A+U\|}} - 1].$$

Let $t \to \infty$, we have

$$\|e(t)\| \le \frac{2C^2 K\|e(0)\|}{\alpha\|\Delta A+U\| - CK} \tag{10}$$

Remark 2. From Eq. (10), we know that parameter mismatches error ΔA, the control gain matrix U and the order α are important factors which determine the synchronization error bound and thus influence the effect of synchronization.

5   Summaries

In this paper, we investigate uniform quasi-synchronization of chaotic fractional-order systems by active control. In drive-response systems, parameter mismatch are assumed to be existent. Since the Lyapunov method has some restrictions in dealing with fractional-order system, the Laplace transform theory is utilized instead of the Lyapunov method. The rigorous theoretical analysis show that the parameter mismatches and the order can affect the error bound and thus affect the effect of synchronization. In proof, Gronwall-Bellman inequality plays a vital important role. For the past decade, this inequality has generated a lot of different forms, which have been applied to analysis existence, uniqueness and stability of differential equations.

References

[1]   Igor, P.: Fractional Differential Equation, vol. 198. Academic, New York (1999)
[2]   Li, C.P., Deng, W.H., Xu, D.: Chaos Synchronization of the Chua System with a Fractional Order. Physica A 360(2), 171–185 (2006)
[3]   Huang, L., Wang, M., Feng, R.: Parameters Identification and Adaptive Synchronization of Chaos Systems with Unknown Parameters. Physics Letters A 342(4), 299–304 (2005)

[4] Yassen, M.T.: Adaptive Control and Synchronization of a Modified Chua's Circuit System. Applied Mathematics and Computation 135(1), 113–128 (2003)

[5] Guo, H., Shi, P., Kui, Y.: Generalized Projective Synchronization between Two Different Chaotic Systems Using Active Backstepping Control 355(4-5), 326–330 (2006)

[6] Yang, Y., Cao, J.: Exponential Lag Synchronization of a Class of Chaotic Delayed Neural Networks with Impulsive Effects 386(1), 492–502 (2007)

[7] Wang, X., Song, J.: Synchronization of the Fractional Order Hyperchaos Lorenz Systems with Activation Feedback Control 14(8), 3351–3357 (2009)

[8] Lu, J., Cao, J.: Adaptive complete Synchronization of Two Identical or Different Chaotic (Hyperchaotic) Systems with fully Unknown Parameters. Chaos 15(043901) (2005)

[9] Johnson, G.A., Mar, D.J., Carroll, T.L., Pecora, L.M.: Synchronization and Imposed Bifurcation in the Presence of Large Parameter Mismatch 80(18), 3956–3959 (1998)

[10] Michele, C.: Linear Models of Dissipation Whose Q is Almost Frequency Independent-II. Geophysical Journal of the Royal Astronomical Society 13(5), 529–539 (1967)

[11] Erdeilyi, A.: Higher Transcendental Function, vol. 13. McGraw-Hill, New York (1955)

[12] Wen, X., Wu, Z., Lu, J.: Stability Analysis of a Class of Nonlinear Fractional-Order Systems. IEEE Transactions on Circuits and Systems II 55(11), 1178–1182 (2008)

[13] Gronwall, T.H.: Note on the Derivatives with Respect to a Parameter of the Solutions of a System of Differential Equations. The Annals of Mathematics 20(4), 292–296 (1919)

[14] Li, C., Chen, G., Liao, X., Fan, Z.: Chaos Quasisynchronization induced by Impulses with Parameter Mismatches. Chaos 16(023102) (2006)

Adaptive Reclosure Technology for High-Voltage Overhead Lines Combined with Underground Power Cables Based on Travelling Wave Principle

Zheng Haitao[1], Chen Ping[2], and Wang Li-Ping[1]

[1] Gansu Electric Power Design Institute, Lanzhou, Gansu Province, China
ezht_dq_gsedi@163.com, leepingwang@tom.com
[2] Shandong University of Technology, Zibo, Shandong Province, China
pingchen1969@263.net

Abstract. For the 110kV and above high-voltage overhead lines combined with underground power cables, a set of adaptive reclosure technology is developed based on travelling wave principle. Firstly, the fault section is determined on line by comparing the time difference between the arrival time of the fault induced initial travelling wave detected at two ends of the combined line with a setting value. Then, according to the different fault section, the corresponding reclosure blocking logic is determined, i.e. the reclosure blocking logic is activated if and only if the fault occurs inside the underground cable section of the combined line. In order to test the validity of the presented adaptive reclosure technology, an adaptive reclosure control system is developed based on double-ended travelling wave principle for an actual combined line. In the system, one reclosure control unit is installed at each terminal of the combined line and synchronized by a GPS based power system synchronous clock. The two reclosure control units can communicate with each other through 2M multiplexing optical-fiber channel. At each terminal of the combined line, the reclosure control unit has a reclosure blocking contact signal output (open normally) which is parallel connected to the two ends of the blocking autorecloser hard strap on the protective relaying cabinet for the combined line. Simulation test and actual operation experience show that the adaptive reclosure control technique for HV combined lines based on double-ended travelling wave principle is feasible.

Keywords: autorecloser, data acquisition, fault location, Global Positioning System, HV combined lines, propagation, protective relaying, signal analysis, surges, synchronization, transducers.

1 Introduction

The high-voltage underground power cables are becoming more and more common with the development of urban power grid. As a result, a large amount of HV overhead lines combined with underground power cables will appear in HV power grid with 110 kV and above voltage level.

In most cases, the underground cable fault is permanent, so the reclosure function of the protective relays had to be out of service for pure cables. For the HV combined lines, because the existing protective relays don't have the ability to identify cable fault, so their reclosure functions are also disabled. This will obviously affect the power supply reliability of the HV combined lines.

The key problem of combined line autorecloser is how to identify the fault location. If the fault is located at the overhead line section, then the autorecloser operation is permitted. Otherwise, the autorecloser must be blocked. In this sense, accurate fault location technology is the basis to realize combined line autorecloser.

The traveling wave fault location technology used for transmission lines has been the concern of relay protection professionals [1], owing to its advantages of high location precision and wide application range, etc. Since 1990s, with the developments of modern micro-electronics, communication and digital signal processing techniques, the traveling wave based fault location techniques have been endowed with new vitality [2], and get more and more on-line applications on overhead transmission lines [3]-[7]. The practical operation experiences show that the location error of modern traveling wave based fault location systems for AC transmission lines could be within ±200 m [8]-[9].

In this paper, a set of adaptive reclosure technology for HV overhead lines combined with underground power cables based on travelling wave principle is developed and applied to an actual 110kV combined line.

2   Basic Principle of Adaptive Reclosure for HV Combined Lines Using Traveling Waves

A combined line is composed of a series of overhead line sections and cable sections. The fault section can be determined on line by comparing the time difference between the arrival time of the fault induced initial travelling wave detected at both ends of the combined line with a setting value.

A fault on a combined line generates voltage and current surges traveling towards two terminals of the line at a speed close to the light, as shown in Fig. 1.

Define ΔT_P and Δt_{MN} as follows:

$$\Delta T_P = \frac{L_C}{v_C} - \frac{L_O}{v_O}$$

$$\Delta t_{MN} = t_M - t_N$$

Where L_C and v_C are the cable section length and the velocity of traveling waves in the cable respectively, L_O and v_O are the overhead line section length and the velocity of traveling waves in the overhead line, t_M and t_N are the absolute arrival times of the fault induced initial surges measured at terminal M and N respectively.
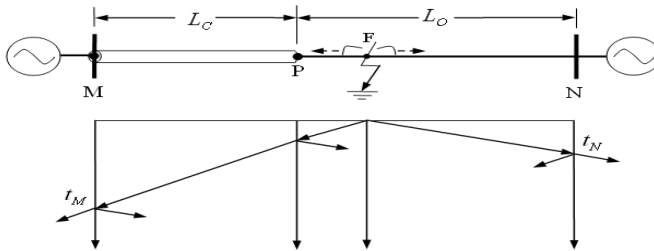
Fig. 1. Lattice diagram for propagation of fault induced traveling waves in a simple combined line

The adaptive reclosure logic for combined line MN can be designed as follows:

1) If $\Delta t_{MN} = \Delta T_P$, then it is derived that the fault is just the connnetion point P, and the reclosure logic needs to be blocked;

2) If $\Delta t_{MN} < \Delta T_P$, then it is derived that the fault is inside the underground cable section MP, and the reclosure logic needs to be   blocked;

3) If $\Delta t_{MN} > \Delta T_P$, then it is derived that the fault is inside the overhead line section PN, and the reclosure logic needs to be permitted.

3   Key Techniques for Realization of Adaptive Reclosure for HV Combined Lines

3.1   Inductiong of Traveling Wave Signals

For a long time people assumed that conventional voltage and current transformers which are designed for measuring power frequency components could not reproduce high frequency travelling waves effectively. A specially designed voltage coupler was employed in previously available travelling wave system. It requires modification of tune circuit of power line carrier (PLC), and is inconvenient and expensive. Some designs proposed application of optical transformers. It is not technically and economically practical to install additional optical transformers to existing primary systems.

Extensive simulation and test results show that capacitor voltage transformer (CVT) can not transform transient signal effectively, while TA can reproduce current transient with surprisingly fast risetime. It has been proved by large amount of fault recordings that better than 300 meters fault location accuracy can be achieved based on secondary output of TA.

Using conventional TA's output, the adaptive reclosure system for HV combined lines can be easily wired into secondary circuit like conventional protection equipments. This scheme is more cost effective and acceptable to field application.

3.2   High Speed Data Acquisiton

Surges from fault are detected by a simple triggering circuit in previously available fault locators based on traveling wave principles. There usually exists significant detection delay to its actual arrival time since it may takes many microseconds for an actual surge to rise to a preset threshold. Being subject to influence of noise signals, such as lightings, this detection method is not reliable.

Using modern microelectronic techniques transient travelling waves can be digitally recorded at very high sampling rate. The recorded transients then can be processed using sophisticated digital signal processing techniques, such as correlation, wavelet transform. This makes surge detection more accurate and reliable.

To get better than 500 meters distance calculation resolution, data acquisition sampling frequency should not be less than 500 kHz. Such high speed data acquisition could not be accomplished by normal data acquisition technique, which control signal sampling, converting, reading and storing processes using a microprocessor. A special data acquisition circuit in which all control signals are generated by a hard logical circuit is designed. It continuously monitors and samples input signal, and stores data in a preset time interval when triggered by a transient. The recorded transient data are then transferred to central processing unit (CPU) for further storing and processing.

3.3   Precise Time Synchronization

The adaptive reclosure method for HV combined lines is based on double-ended travelling wave principle, so it requires equipments at both ends have better than 3 microseconds time synchronization accuracy to get better than 500 meters fault location accuracy. For a long time, there was no public source providing such high time synchronization accuracy. It is obviously expensive and impractical to set up a dedicated synchronization system in power network.

The Global Positioning System (GPS) can provide better than 1 microsecond time synchronization accuracy, and is open to public uses. Availability of GPS technique promoted development of modern travelling wave system. Time synchronization of the field equipments are obtained by combining the serial time of day message and 1 pulse per second (1pps) strobe from a GPS based power system synchronous clock.

3.4   Remote Communication

The adaptive reclosure method for HV combined lines needs a means to exchange the two terminals' measurements. It must be "on line" and therefore can use dedicated or multiplexing optical-fiber (2M) channel.

4   Adaptive Reclosure System for HV Combined Lines

The structure of the developed adaptive reclosure system for a HV combined line is shown in Fig. 2.
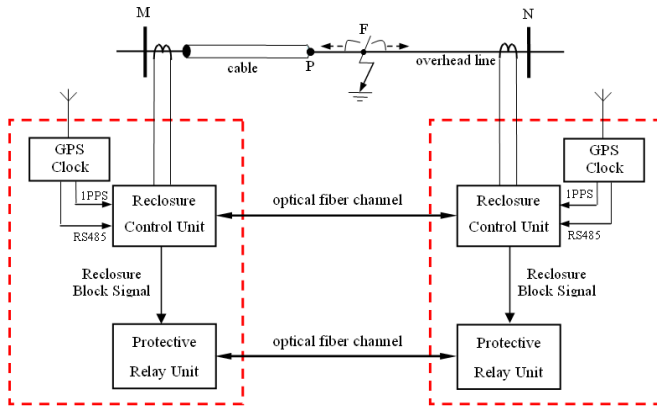
Fig. 2. Structure of the presented adaptive reclosure system for a HV combined line

At each terminal of the combined line, there are three devices including a GPS based power system synchronous clock, a reclosure control unit based on double-ended travelling wave principle and a conventional microprocessor based protective relaying unit.

Each reclosure control unit has a reclosure blocking contact signal output (open normally) which is parallel connected to the two ends of the blocking autorecloser hard strap on the protective relaying cabinet.

The reclosure control unit should send reclosure block signal to the corresponding protective relay unit only when the fault point is inside the cable section or the dead band near the connection point of cable and overhead line section.

5   Operation Experiences of the Adaptive Reclosure System

In March 2010, a developed adaptive reclosure control system for combined lines was installed in Yantai Power Supply Company, Shandong, China, monitoring the combined line between Fushan substaion and Songjiang substaion. The primary connection diagram is shown in Fig. 3.

The two reclosure control unit communicates with each other through 2M multiplexing optical-fiber channel.

Fushan-Songjiang combined line is a 110kV single power supply line, so only the protective relay unit for the combined line is equipped only at Fushan side.

The combined line is composed of one overhead line section whose length is 8.975 km, and one underground cable section whose length is 1.695 km.
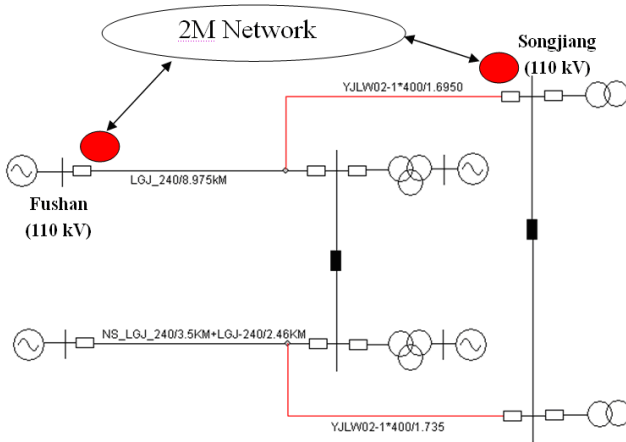
Fig. 3. Primary connection diagram for the combined line between Fushan and Songjiang

At 14:57:23, May 13th, 2010, a fault occurred in Fushan-Songjiang combined line, and the   reclosure action was successful.

Table 1 gives the travelling wave triggering time recorded by the reclosure control units.

Table 1. Travelling wave triggering time recorded by the reclosure control units

Fushan substation—M (overhead line terminal)	Songjiang substation—N (underground cable terminal)
2010-05-13/14:57:23	2010-05-13/14:57:23
356555μs	356570μs

The time difference was $\Delta t_{MN} = 12$ μs, and the setting value was $\Delta T_P = 20.6$ μs. Obviously, the fault was inside the overhead line section ($\Delta t_{MN} < \Delta T_P$). So, the reclosure control units of both substation couldn't send a reclosure block signal. Consequentially, the protective relay unit at Fushan side would send a reclosure order.

According to the above analysis, it can be concluded that the action of the reclosure control units based on double-ended travelling waves principle is correct.

The fault generated transient waveforms recorded by the two reclosure control units at both ends of Fushan-Songjiang combined line are shown in Fig. 4 and Fig. 5.
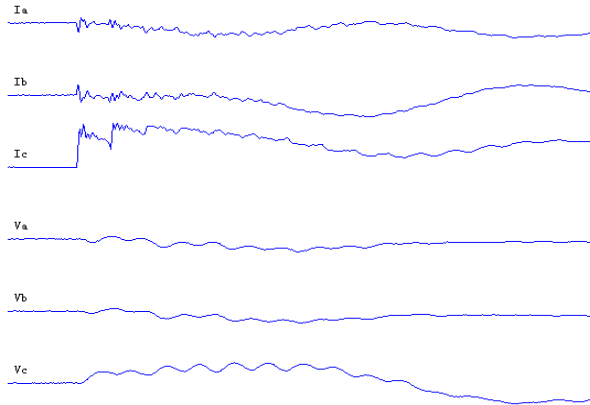
Fig. 4. The fault generated transient waveforms recorded at Fushan substation
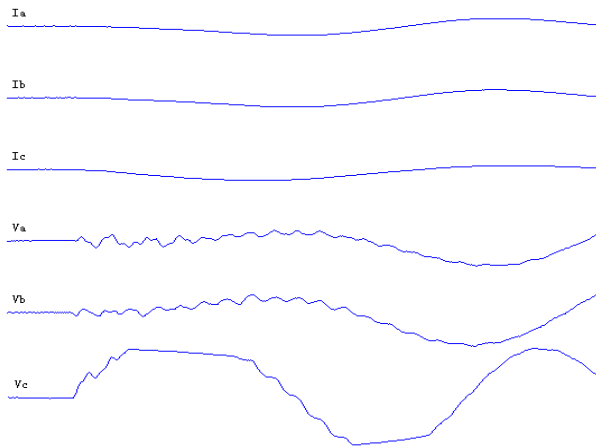
Fig. 5. The fault generated transient waveforms recorded at Songjiang substation

6   Summaries

For high-voltage overhead lines combined with underground power cables, the fault section can be identified on line based on double-ended travelling wave principle.

An Adaptive reclosure control scheme for HV combined lines using double-ended travelling wave based fault sectioning principle is presented, and the corresponding reclosure control system is also developed. The system can identify the fault section of combined lines on line, and send a reclosure block signal to local protective ralay unit if and only if a unferground cable fault occurs.

Actual operation experience shows that the adaptive reclosure control technology for HV combined lines based on double-ended travelling wave principle is feasible and it has a wide application perspective with the rapid development of urban power grid.

References

[1] Ge, Y.: New Types of Protective Relaying and Fault Location Theory and Techniques, 2nd edn., pp. 301–335. Xi'an Jiaotong University Press, Xi'an (2007)

[2] Xu, B., Li, J., Chen, P., et al.: Modern traveling wave based fault location technology and its application. Automation of Electric Power Systems 25, 62–65 (2001)

[3] Lee, H., Mousa, A.M.: GPS traveling wave fault locator systems: Investigation into the anomalous measurements related to lightning strikes. IEEE Trans. Power Delivery 11, 1214–1223 (1996)

[4] Gale, P.F., Taylor, P.V., Naidoo, P., et al.: Traveling wave fault locator experience on Eskom's transmission network. In: Proc. 2001 Seventh International Conference on Developments in Power System Protection, pp. 327–330 (2001)

[5] Jian, Q., Zhen, H., Hua, Y., et al.: Study on traveling wave propagation in double-circuit parallel transmission line. Proceedings of the Chinese Society of Electrical Engineering 24, 30–34 (2004)

[6] Ji, T., Xue, Y.-D., Sun, T.-J., et al.: Fault location for distribution feeders based on traveling waves. Automation of Electric Power Systems 29, 66–71 (2005)

[7] Xue, Y.-D., Xu, B.-Y., Li, J., et al.: Traveling waves based fault location in 10 kV automatic blocking and continuous railway power lines. Automation of Electric Power Systems 30, 68–73 (2006)

[8] Chen, P., Xu, B., Li, J., et al.: Development of modern traveling wave based fault location system. Automation of Electric Power Systems 27, 81–85 (2003)

[9] Chen, P., Xu, B., Li, J.: The Optimized Combination of Fault Location Technology Based on Traveling Wave Principle. In: Proceedings of Asia-Pacific Power and Energy Engineering Conference, APPEEC 2009, Wuhan, China, March 28-31, pp. 28–31 (2009)

Fault Detection Based on Granular Delaminating Reasoning

Yu Hang, Xiao Ming-Qing, and Chen Li-An

ATS Lab. Engineering College,
Air Force Engineering University,
Xi'an, 710038, China

Abstract. According to the requirements of self-repair and reconfiguration to weapon system, and the characters of distribute systems, the distribute weapon system has lots of function nodes and complex inner structure. Study on advanced method of fault detection of distribute weapon system becomes the focus of weapon maintain and logistic. The paper use the complex networks theory to model the distribute system; use the community structure of complex networks to analyze the structure of system, the method of granular delaminating reasoning has been used to predigest the question space. At last use the distribute test system as an example to test the method, which shows new trains of thought for distribute weapon system fault detection.

Keywords: complex networks theory, community structure, granular delaminating. fault detection.

1 Introduction

According to the requirements of self-repair and reconfiguration for weapon system, and the characters of distribute systems, the distribute weapon system has lots of function nodes and complex inner interaction relation. The method of fault detection of distribute weapon system becomes the focus of weapon maintain and logistic. The main methods of fault detection are based on fault spread graph [1, 2] and qualitative model. According to the distribute system always working in the unstable environment, the working period and system status always changing, the traditional fault detection method cannot fit the requirement. In this paper, the system structure was analyzed according to the complex networks theory. The topological model of system is built. And detected the fault use the method of granular delaminating. The method can predigest the complexity of system analyze and promoted a new view on complex system fault detection.

2 System Model Based on Complex Networks Theory

Use the complex networks theory to construct the system model. Use the node to express the component, equipment or module, and use the edge to express the communication between these components in the system or the module and

arrangement structure. When use the complex networks theory to construct the fault detection model, node express the atom node of system and the edge express the data field or the function communication, the community structure express the module or the arrangement structure.

Definition 1. The system model $S = \{V, E\}$,

Nodes set $V = \{v_i \mid v_i$ is the atom component of system$\}$

Edge set $E = \{e_{i,j} \mid$ the edge from S_i to $S_j\}$

Definition 2. The adjacency matrix is defined $P = \{P_{i,j}\}$

$$P_{i,j} = \begin{cases} 1 & \text{has direct edge from } v_i \text{ to } v_j \\ 0 & \text{no direct edge from } v_i \text{ to } v_j \end{cases}$$

Distribute system constructed by atom components which actualize the system's function, and the atom component is drive by the input information from outside or other atom component, the dispose result of this atom component will transmit to other atom component. All the atom components in system dispose according to this method in turn. The input may contain multi-input such as the AND or OR logic etc. Harmonized these atom components can achieve the function of the whole system [3].

A distribute system constructed by atom components $(v_1, v_2, \ldots, v_N)$. Let each V_i has the function f_i , the input is u_i , and use y_i formulate the output: $y_i = f_i(u_i, y_1, y_2, \cdots, y_N)$, $i = 1, 2, \cdots, N$.

Take a distribute system which contains 4 atom components as an example, the model as figure 1(a) shows, the drive communication between atom components shows as the edge in graph according to the input or output, figure 1(b) shows the adjacency matrix of model.
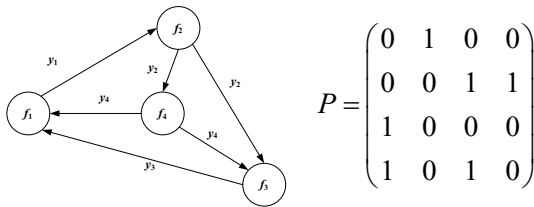
a. system structure model   b. adjacency matrix of model

Fig. 1. System model

If there is fault atom component in system, the system reconfigurated, and the topological structure changed. The system structure $\overline{S}$ turn to $\overline{S'}$, and the subsystem turn to $(S_i, S_{k1'}, S_{k2'}, \ldots, S_{kn'})$, the function of system also reconfigurated, as figure 2 shows, when lost the atom component s_2 ,the system model shows in figure 3(a), and the adjacency matrix shows in figure 3(b).
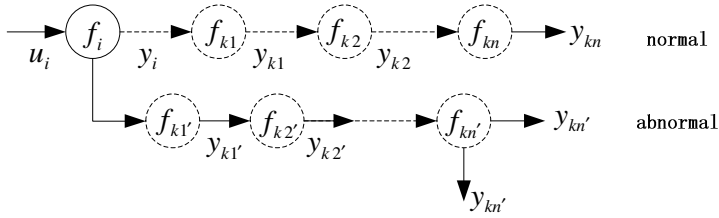
Fig. 2. System function reconfigrated in abnormal status

$$P' = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 1 & 0 \end{pmatrix}$$
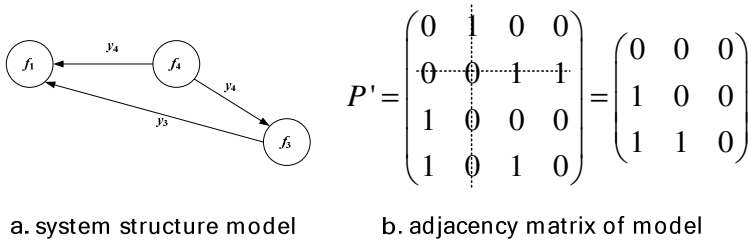
a. system structure model   b. adjacency matrix of model

Fig. 3. Reconfigrated model

3   Granular Delaminating Based on Community Digging

Granular transform use the method of granular to predigest the complex question, which make people can analyze, disposal and get knowledge. Granular is a method which disposed the complex question, form a higher level granule and keep the certain analyze function. The method predigests the complexity of question, get some basic character of question and make a primary answer to the question. This primary knowledge can direct the farther study on the complex question [4]. Quotient space theory also predigests the complexity by granular transform. The complexity predigests through the process of granular transform mainly depend on the fidelity principle, the disloyalty principle and quotient space chain, which means description of structure is helpful for understanding the principle. The structure question can be solved during the process of quotient topological structure transform between different levels of granular degree according to the topological structure. Research the answer for question in different levels of granular according to the principles, and combine these answers to get the answer of the complex question.[5]

Community structure is a classification of nodes in graph according to the graph theory. Character of this kind of classification is the communication among nodes in the community is tight but is sparsity among communities. Mapping to the real system, communities can show the module or layer structure of system, and these kind of structure character just can combine with the granular delaminating to find the answer of complex question and analyze the structure and function of real system when mapping to the real system.

4   Hierarchical Fault Reasoning Based on Quotient Space

Hierarchical quotient space chain method can simply described as: Face to the complex and cosmically question, mapping the maturity information of original to different levels of granular spaces according to the topological relation of elements. All the granular question we get are higher level than the original space, and only contain the information about certain target; analyze the structure and character of original question on these granular spaces which build the different granular levels of quotient space; according to the level sequence put these quotient spaces in line is definite as the hierarchical quotient space chain. Get the certain information of question from the higher level quotient space according to the fidelity principle and use these information to guide the certain target information analyze in the lower level quotient space. This process operation starts at the highest quotient space and finished in the lowest quotient space to get the answer of question.[6]

Use the community detect algorithm to get the community structure of network in lower hierarchical is predigest the question once. If still cannot find answer when question is predigested, make the community as a node, and take the connection among the community as the edge, this make a new network and use the detect algorithm on the new network can get the community structure in higher granularity. According to the process above, make the network recursive predigest of network and can output the multi-granularity network view, and find the predigest network structure which meet the demands.

The multi-granularity predigests algorithm of network structure is as follow:

Step 1: Use community detection algorithm gets the network community structure information;
Step 2: Make the community as node, and the relationship among communities as edge, constructed the new network topological graph;
Step 3: Check the network topological structure, if the granularity meet the demands of question solve, turn to step 4, else turn to Step 1;
Step 4: Break, the capable predigest community network structure view has been find.

5   Experiment Validation

Take the distribute test system as an example to test the fault reasoning method. In the distribute test system, atom components are test equipments, the datum field is test network. Atom components communicated with each other with communication link and construct the test network. TUA contains TUA controller, power source, signal conditioning module, interface and so on, they constructed by the configurable unit (CU), the configurable units cooperated with each other to collect and dispose the UUT signal.

The output of complex network model is the distribute system's network topological view, show in figure 4. The nodes means the atom components, because of the complexity of distribute system; the figure 4 is not show the lowest granular topological structure. The nodes and function parts of distribute system in this hierarchical show in blank 1.

Blank 1. The atom components of distribute test system

NO.	Name	NO.	Name	NO.	Name
1	Data signals interface	7	Buck-boost gas path	13	Stress test control gas path
2	Analog signals interface	8	Currency DC monitor	14	Currency AC monitor
3	Normal signal level interface	9	Human-computer interaction interface	15	Signal conditioning card 1
4	Air operated chuck gas path	10	Test control computer power source	16	Signal conditioning card 2
5	Wireless RF interface	11	Programmable AC power source	17	Signal conditioning card 3
6	Programmable DC electrical source	12	Refrigeration gas path	18	Datum backup storage

According to the community detection algorithm to detect the network and predigest the structure. Construct the network structure according the function relationship of system. And predigest the network structure according to the multi-granularity predigests Algorithm, record the result in the network hierarchical tree which is show in figure 5.

Make the fault recall Hierarchical fault reasoning according to the hierarchical tree, and find the fault source parts in the hierarchical quotient space step by step. The process of reasoning shows in figure 6.

The gray communities and nodes is the space which needn't search. The reasoning process shows that the reasoning space is obviously reduced and reasoning time is shorten. When reasoning process finished, the fault source can be orientated in nodes 3, 7 and 13. And then, make the possibility sequence according to the fault probability. The fault rate of atom components are (0.02,0.06,0.04,0.03,0.09,0.08,0.07,0.10, 0.05,0.08,0.07,0.05,0.06,0.04,0.07,0.04,0.02,0.03).
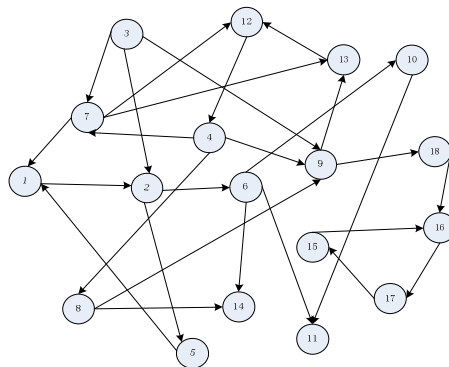
Fig. 4. Network topological structure of system
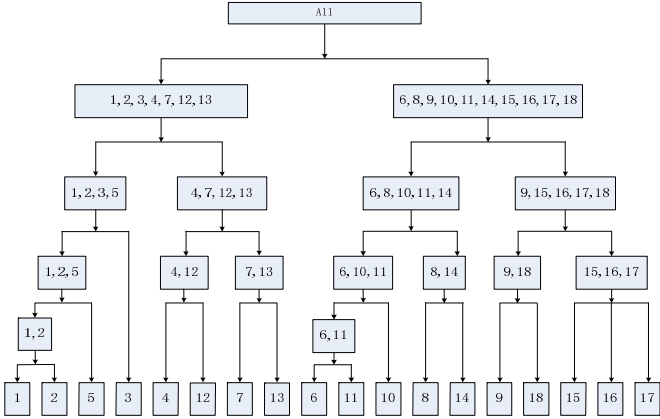
Fig. 5. Network hierarchical tree
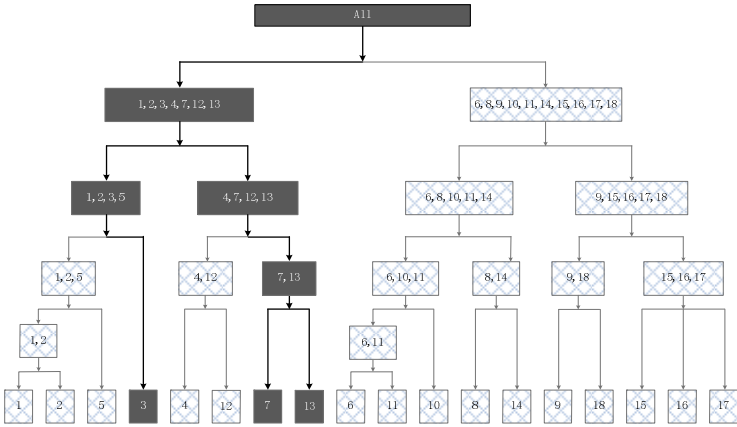
Fig. 6. Hierarchical fault reasoning process

The fault possibility of node 3,7,13 are: $P_3 = 0.04 \times 3 = 0.12$, $P_7 = 0.07 \times 4 = 0.28$, $P_{13} = 0.06 \times 4 = 0.24$.

6   Concluding Remarks

The complex network theory has been used in the distribute system structure analyze, and combine the quotient space granularity transform with community detect method, promoted the distribute system fault detect method. And applied the method in the distribute test system, the result of experiment shows, the method can orientated the fault source, and give the fault possibility sequence when there have the experience knowledge.

References

[1] Iri, M.: A graphical approach to the problem of locating the origin of the system failure. Journal of the Operational Research Society 23, 295–311 (1980)

[2] Kokawa, Miyazaki, S., Shingai, S.: Fault location using digraph and inverse direction search with application. Automatica 19, 729–735 (1983)

[3] Mori, S.: Autonomous Decentralized Systems–From Concept to Application. Science Press, Beijing (2008)

[4] Bargiela, A., Pedrycz, W.: Granular Computing: An Introduction. Kluwer Academic Publishers (2003)

[5] Zhang, L., Zhang, B.: The theory and application of tolerance relations. International Journal of Granular Computing, Rough Sets and Intelligent Systems 1(2), 179–189 (2009)

[6] Zhang, Y., Luo, B., Yao, Y.: Quotient Space and Granular Computing–Structure Problem Solving Theory and Method. Science Press, Beijing (2010)

Parameter Identification of Synchronous Machine Based on TLS-ESPRIT

Jiang Jian, An Wei-zhong, and Liu Hui

Gansu Electric Power Design Institute
Lanzhou, Gansu Province, China
JiangJian_XL@163.com, Liuhui9582@gmail.com

Abstract. An array signal processing method TLS-ESPRIT (total least squares-estimation of signal parameters via rotational invariance technique) is applied in parameter identification of synchronous machine in this paper. This method decomposes original signal into signal subspace in sigulai value decomposition and noise subspace and is processed by TLS(total least square), so it is a high precision method and improve the capacity of anti-noise. It can extract the transient and sub-transient parameters of synchronous machine accurately. The simulation example demonstrate that this method is capable of higher precision, stronger anti-noise capacity and easier than other identification methods. The dynamic simulation test proves that method is available and feasible.

Keywords: Synchronous machine, Parameter identification, Total least squares-estimation, signal parameters, rotational invariance technique, Subspac.

1   Introduction

Synchronous machine is the important equiment in power system, it's parameters determines its action, and the accurate parameters of synchronous machine make important effect on researching and analysing some projects about power system working and controlling system design[1]. Many scholars concentrate their attention on how to measure the parameters of synchronous machine exactly[2].

Suddenly three-phase short-circuit test is a good way to measure the parameters of synchronous machine, which also is the way Chinese standard GB/T-1029 and GB577-89 recommending[3]. Usually, parameter in short circuit test may be test after overhauling stator winding or in the acceptance test[4]. The traditional way of calculating the parameters of synchronous machine is to extract aperiodic component and periodic component by means of calculating fluctuation envelope to add and subtract aiming at short-circuit current, then a method based on least-squares curve fitting is used to calculating transient parameters. While this algorithm has low precision and big errors. In reference 5 Prony method is used to identify synchrounous machine parameters, however, this method has big errors under the influence of noise. Paper 6 used Meryer wavelet basis to extract component of the short-circuit correctly and reduce influence of noise, so the accuracy was enhanced. Using Prony algorithm

on the basis of extracting wavelet is limited because the wavelet basis exactly matching the signal can't be found.

ESPRIT is a method about array signal processing, with which to solve the DOA estimation[7], the signal is divided into signal subspace and noise subspace in order to reduce the noise effect. When the signal parameter is calculating, it will improve the precision of the unknown parameters that the TLS(total least squares) is introduced-it equivalents to another removing noise and processing disturb. Now some research scholars use this method in the electrical power system and this method meets the expectation[8]. In this paper TLS－ESPRIT was used to identify the Transient and Super transient parameters, the simulation examples and dynamic model test proof that this method is practicable and effective.

2   TLS-ESPRIT Arithmetic and Deviding Subspace

2.1   ESPRIT Arithmetic Thought

ESPRIT was firstly advanced by ROY and others in 1989[9], which has been an important method for moden signal processing and applied widely.

Suppose oscillating signal x(n) can be expressed as acombination of a series of sine signals changing by the insex rule and white noise, the expression at   sampling time n is:

$$x(n) = \sum_{k=1}^{p} c_k z_k^{\ n} + w(n) \tag{1}$$

In which, z_k is called signal apex. $c_k = a_k e^{j\theta_k}, z_k = e^{(-\sigma_k + j\omega_k)T_s}$, T_s is sampling cycle. $a_k, \theta_k, \omega_k, \sigma_k$ stands for respectively amplitude、initialization phasic、angular frequency、degradation factor of the Kth attenuation component, of which f_k is the frequency, $f_k = \omega_k / (2 * \pi)$; and w is white noise whose average value is 0.As sampling signal usually is real ones, model rank p doubles the number of signal's actual sine real component. define Q snapshot data vectors:

$X = [x(n)\ , x(n+1), \cdots x(n+Q-1)]^T$, thereinto $Q > p$, so defining

$$X = \Gamma_Q \phi^n c + W = S + W \tag{2}$$

in the formula:

$$c = [c_1, c_2, \cdots c_p]^T, \phi = diag(z_1,\ z_2, \cdots z_P), \tau_Q(z_i) = [1,\ z_i, \cdots z_i^{Q-1}]^T$$

$$\Gamma_Q = [\tau_Q(z_1), \tau_Q(z_2), \cdots, \tau_Q(z_p)]W = [w(n), w(n+1), \cdots w(n+Q-1)]^T$$

ϕ is also called spin operator, obtained which, the frequency and damp coefficient of every signal component will be found. In order to express expediently, S_1, S_2 is stipulated to represent the new matrix comes from S deleting the first and last row respectively, so:

$$S_1 = \Gamma_{Q-1} \phi^{n+1} c = J_1 c \tag{3}$$

$$S_2 = \Gamma_{Q-1} \phi^{n} c = J_2 c \tag{4}$$

$$J_1 = J_2 \phi \tag{5}$$

In the formula, J_2 is considered to be changed into J_1 as a result of being circumrotated. Working-out ϕ will make for all the parameters of the signal.

2.2   TLS-ESPRIT Arithmetic

At present there are many different ways to implement TLS-ESPRIT, for example, in the 9 reference, the matrix bundle is formed by the autocorrelation matrix and cross-correlation matrix of data sequence, and then the signal parameters will be obtained by using generalized eigenvalue decomposition. Correlative matixes should be calculated heavy if this way was used. So in this paper the TLS-ESPRIT in the 10 reference will be adopted. The collected snapshot data series $x(0), x(1), \cdots x(Q-1)$ will compose $X_{N \times M}$ HANKEL matrix, in which, $N > p, M > p$ .We perform SVD(singular value decomposition) on matrix X , $X = U \Sigma V^{T}$. The diagonal elements of Σ is $\xi_1 \geq \xi_2 \cdots \geq \xi_p > \xi_{p+1} \approx \cdots \xi_{\min(N,M)} \approx 0$. At this time V may be devided into signal subspace V_1 and noise subspace V_2 according to the order of singular value. Let $V = [V_1 \quad V_2]$. The right singular vectors associated with $\xi_1, \xi_2 \cdots \xi_p$ correspond to the column-vectors of signal subspace V_1 . Let V_3 and V_4 represent the new matrices that come from V_1 deleting the first row and the last row respectively. As the noise and other disturbance for the signal are considered, the following expression is given:

$$(V_3 + e_1) = (V_4 + e_2)\phi \tag{6}$$

Extract the solution of equation (6), namely get the minimum and optimal ϕ from Frobenius norm of perturbation matrix $D = [-e_1, e_2]$, thus TLS[11](total least squares) will be introduced, and then we perform SVD on $[V_4 \quad V_3]$,

$$[V_4 \quad V_3] = R \Lambda P^{T} \tag{7}$$

Deviding P into $P_{2p \times 2p} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix}$, after calculating the eigenvalue $\lambda_k (k = 1, 2, ...p)$ of $P_{11} P_{21}^{-1}$, we will find the frequency and attenuation factor of each component in the original signal. The formula for calculation is as follows:

$$f_k = \frac{\arg(\lambda_k)}{2\pi T_s} \tag{8}$$

$$\sigma_k = \frac{\ln|\lambda_k|}{T_s} \tag{9}$$

For Q sampling signals, $X = \lambda S$,

$$\lambda = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \cdots & \lambda_p \\ \vdots & \vdots & & \vdots \\ \lambda_1^{Q-1} & \lambda_2^{Q-1} & \lambda_p^{Q-1} \end{bmatrix} .$$

Base on least squars technique, there is $S = (\lambda^T \lambda)^{-1} \lambda^T Y$, thereby the amplitude and the phase of each component can be estimated in the signal:

$$a_k = 2|S| \tag{10}$$

$$\theta_k = angle(S) \tag{11}$$

2.3   Signal Subspace V_1 and Noise Subspace V_2

It is very important to the veracity of estimating signal's parameters that the signal subspace V_1 should be partitioned accurately. The partitioning into signal and noise subspace is equaled with that how to calculate the effective rank p of HANKEL. Towards the signals whose components are unknown, the partition can be carried though according to the AIC[12], MDL rule and judging the eigenvalue[7] while if the signal's components are known artificial effective rank can be available. At the time of stator three phase short circuit the current of synchronous machine (salient pole generator and non salient pole generator) can be expressed by the following formula:

$$i_a = -\frac{E}{2}e^{\frac{-t}{T_a}}[(\frac{1}{x_d''}+\frac{1}{x_q''})\cos\phi_0 + (\frac{1}{x_d''}-\frac{1}{x_q''})\cos(2\omega t + \phi_0)]$$

$$+ [(\frac{1}{x_d''}-\frac{1}{x_d'})e^{\frac{-t}{T_d''}} + (\frac{1}{x_d'}-\frac{1}{x_d})e^{\frac{-t}{T_d'}} + \frac{1}{x_d}]E\cos(\omega t + \phi_0) + e(t) \tag{12}$$

$$= i_0 + i_{2\omega} + i_\omega + e(t)$$

So the fault current is considered to contain aperiodic component, fundamental frequency component, double frequency component and gaussian white noise, therefore we may take the value of 9 as the effective rank of HANKEL. Only the current data of the DC component and fundamental frequency component need to be analysed.

3   Simulation Example

An simulated analysis will be performed as a perfect synchronous machine serves as an illustration $T_a = 0.4348s$, $T_d^{'} = 0.8350s$, $T_d^{''} = 0.017s$, (per unit value) $x_d^{'} = 0.2290$, $x_d^{''} = 0.1830$, $x_q^{''} = 0.1855$, the short circuit epoch angle $\phi_0 = \dfrac{\pi}{3}$, $E = 1$. When the gauss white noise is injected into the perfect current of this machine with sudden three phase short circuit, the waveform of the current is illustrated in Fig.1.
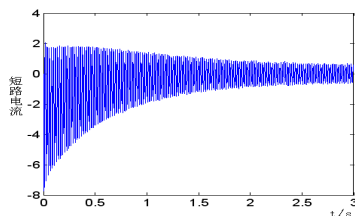
Fig. 1. The waveform of fault current in noise

3.1   TLS-ESPRIT Extracting Current Componet

The fault current will be extracted respectively using TLS-ESPRIT, Wavelet and traditional method as follows. Using TLS-ESPRIT method we get the HANKEL matrix by means of gathering 3000 points of fault current, which will be calculated. Wavelet method uses Meryer wavelet basis mentioned in the 6 reference, while traditional method uses linear interpolation to get the upper and lower envelopes of the short currents. Fig 2 shows the waveforms of DC component and perfect DC without noise extracted respectively by the three ways, in which the real line stands for perfect DC waveform while the broken line stands for extracted DC waveform.
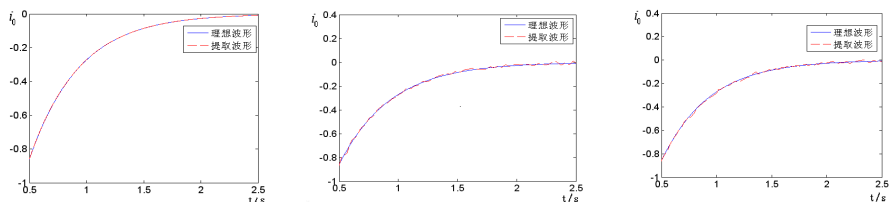
（a）DC component extracted by TLS-ESPRIT
(b) DC component extracted by wavelet
(c) DC component extracted by Traditional method

Fig. 2. Compare of actual and obtained DC current component

From Fig 2, it is obvious that the DC component extracted by wavelet or traditional method fluctuates a little, while the DC component extracted by TLS-ESPRIT approximates to the real signal, and the parameters(amplitude、attenuation factor) of

DC component can be obtained directly. Fig 3 shows the contrast between the waveform of fundamental component obtained by TLS-ESPRIT and wavelet with the waveform of actual fundamental component.

（a）Fundamental component extracted by　TLS-ESPRIT　　　　（b）Fundamental component extracted by wavelet
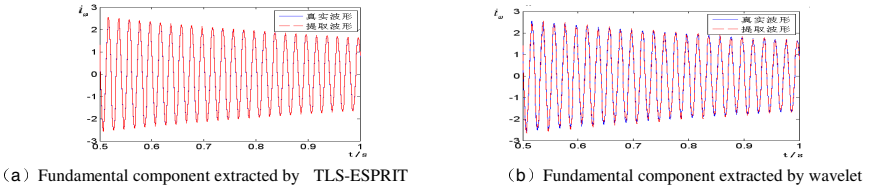
Fig. 3. Contrat of actual and obtained fundamental component

In order to eliminate the fluctuating of the signal extracted by wavelet on both sides and embody the advantages of wavelet, the data of DC component i_0 and fundamental component i_ω both obtained using each method at the time slice from $0.5s$ to $2.5s$ compares the root mean square deviation with the actual fundamental component data, the comparison result is as shown in Tab 1.

Table 1. The comparison of root mean squarecurrent component

	i_0	i_ω
Traditional method	19.03	*
Meryer wavelet	0.77	5.24
TLS-ESPRIT	0.03	0.09

Tab 1 shows that the mean square of each current component obtained by TLS-ESPRIT is of the order of magnitude of 10^{-4}, as the effect is much better than the wavelet and traditional method does.

3.2　TLS-ESPRIT Parameters Extraction

The method analyzing the exponential form transformed from the signal by TLS-ESPRIT is similar to the prony algorithm, so the frequency f、attenuation factor σ, amplitude a, initial phase θ of the signal will be worked out according to fomula (13), (14),(15),(16), and then we can calculate the transient and subtransient parameters of synchronous machine. Parameters of fault current various component extracted by TLS-ESPRIT (DC component i_0 and fundamental component i_ω only need) are as shown in Tab 2:

Table 2. The parameters of current component extracting by TLS-ESPRIT

Current component	Component parameters			
	f	a	σ	θ
i_ω	50.0006	3.8754	-1.2056	1.0500
	50.0377	1.0972	-57.9430	1.0518
	50.0011	0.4917	-0.0242	1.0251
i_0	0	5.4310	-2.2993	3.1416

When the parameters of synchronous machine are being calculated using Tab.2, the following three points should be pay attention to:

(1) Synchronous reactance x_d is considered to be known quantity because it can be measured accurately by no-load and short-circuit characteristic rules.
(2) The generator's parameters associated with each component can be estimated according to the various parameters of machine, for example, the time constant corresponding to the maximal attenuation factor of fundamental component -57.9430 should be subtransient decay time constant.

According to the two above point, we can calculate the transient and subtransient parameters of synchronous machine.

Tab.3 shows identification result of T_a, $x_d^{'}$, $T_d^{'}$, $\frac{1}{x^{"}_d} + \frac{1}{x_q^{"}}$ and comparison of relative error, in which $x_2 = \frac{1}{x^{"}_d} + \frac{1}{x^{"}_q}$.

Table 3. Part of identification result of machine parameters and comparison of relative error

parameters	traditional method	relative rror (10^{-2})	wavelet	relative error (10^{-2})	TLS-ESPRIT	relative error (10^{-2})
T_a	0.5843	34.36	0.4377	0.6700	0.4349	0.0230
$x_d^{'}$	0.2120	7.42	0.232	1.31	0.2306	0.70
$T_d^{'}$	0.7248	13.2	0.811	2.874	0.8295	0.66
x_2	9.262	14.68	10.773	0.76	10.862	0.0617

Tab.4 shows the identification result of machine subtransient parameters, in analysis result, we can find out that precision of general algorithm is low because of not considering noise effect. It's relative ideal for analysing with combination of prony and wavelet with selection problem of wavelet base and poor self-adaptive. The proposed identification method with high identical accuracy and strong anti-noise ability in this paper overcomes shortcomings of wavelet.

Table 4. Identified results of sub-transient parameters comparison of relative error

param eters	traditional method	relative rror (10^{-2})	wavelet $+$ PRONY	relative rror (10^{-2})	TLS-ESPRIT	relative rror (10^{-2})
$T_d^{''}$	0.0274	59.41	0.0174	2.3530	0.0173	1.7341
$x_d^{''}$	0.2204	20.44	0.2115	15.57	0.184	0.55
$x_q^{''}$	0.2117	14.12	0.1654	10.835	0.1843	0.65

4   Dynamic Former Examination Analysis

In order to validate the availability of analysing actual problem by the method put forward in this paper, sudden 3-phase short circuit test is done to a stimulant synchronous generator in dynamic former lab. Dealing with the collected fault current obtained obtained by the method in this paper, we are capable of getting the generator's parameters as following:

$$T_a = 0.4378s, T_d^{'} = 0.1561s, T_d^{''} = 0.0370s, \quad x_d^{'} = 0.1041, x_d^{''} = 0.0546, x_q^{''} = 0.0586$$

with these parameters we will work out the waveform of fault current further, shown as figure 4
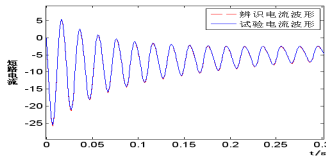
Fig. 4. The waveform of identification and experiment

After comparing waveform of identification current and test waveform (filterd), we can calculate out that the root mean square deviation of the two waveforms is 0.37%, so the result shows a good agreement, as shows that the measured parameters are comparatively accurate and verifies validity and feasibility of this method.

5   Summaries

(1) TLS-ESPRIT is a method for extracting signal parameters with high precision, which has strong robustness. Firstly we devide HANKEL matrix formed from gathering current signals into signal subspace and noise subspace for reducing the influence of noise, secondly, we reduce the noise and other disturbance usig TLS(total least squares).

(2) Here the fault current signal is converted into exponential form for the calculation and this method used in this paper may be applied to extract the parameters in the current signal directly without use in combination with

other methods. The suggested method has the merits such as simple algorithm, high precision, easy realization.

(3) The suggested method can be used not only in parameters identification on the salient-pole synchronous generator, but also in parameters identification on the non-salient pole machine.

References

[1]   Ma, H., Hu, Q.: Transient parameter measurement of synchronous machine. Proceedings of EPSA 12(2), 8–12 (2000)

[2]   Wang, L., Wang, G., Ma, W., et al.: Parameter identification of synchronous machine based on wavelet transform and Neural Network 27(3), 1–6 (2007)

[3]   Li, T., Gao, L., Nie, Y., et al.: Parameter identification of synchronous machine based on HHT. Proceedings of CSEE 26(8), 153–158 (2006)

[4]   Su, X., Zhou, S.: Application of Prony method to parameter identification of synchronous generators. Electric Power Automation Equipment 26(9), 1–4 (2006)

[5]   Wu, X., Ma, W., Wang, G., et al.: Parameter identification of synchronous machine based on wavelet transform and prony algorithm 27(19), 38–42 (2003)

[6]   Zhang, X.: Modern Signal Processing. Tsinghua University Press (2002)

[7]   Zhang, J., Xu, Z., Wang, F., et al.: TLS-ESPRIT based method for low frequency oscillation analysis in power system. Automation of Electric Power Systems 31(20), 84–88 (2007)

[8]   Roy, R., Kailath, T.: ESPRIT—Estimation of signal parameters via rotational invariance techniques. IEEE Trans. Acoust., Speech, Signal Processing 37, 297–301 (1989)

[9]   Zhou, Y.-Z., Chen, T.-Q.: An adaptive ESPRIT algorithm. Systems Engineering and Electronics 23(9), 5–11 (2001)

[10] Golub, G.H., Van Loan, C.F.: An analysis of the total least squares problem. SIAM J. Number Anal. 17, 883–893 (1980)

[11] Wax, M., Kailath, T.: Determining the Number of Signals by Information Theoretic Criteria. IEEE (1984)

[12] Wax, M., Kailath, T.: Detection of Signals by Information Theoretic Criteria. IEEE Trans. ASSP 33(2), 387–392 (1985)

An Investigation about the Threshold of AEV Stopping Criterion in BICM-ID System

Xiao Ying, Li Jianping, and Cai Chaoshi

Communication University of China

Abstract. Stopping criterion is proposed to overcome the disadvantages of decoding delay and complex iteration computation caused by iterative decoding in bit-interleaved coded modulation (BICM-ID) system with fixed iteration number. Average-Entropy-Value (AEV) scheme is one of the efficient stopping criterions with variable iteration numbers in BICM-ID system. When the average entropy value is smaller than the predetermined threshold, which is a sufficiently small positive number, the iterative process can be stopped. But how small the threshold should be? Is the threshold that previously used in the AEV scheme suitable? Whether there exists a more appropriate value? This paper is to find the answers according to the questions raised above and give a suitable range about the threshold of the AEV stopping criterion.

Keywords: BICM-ID, SISO decoder, Demapper, Stopping criterion, AEV, Threshold.

1 Introduction

Bit-interleaved coded modulation with iterative decoding (BICM-ID) is a smart scheme for its excellent performances both in AWGN channel and in Rayleigh fading channel [1]. The iterative decoding algorithm is proposed to overcome the disadvantages caused by bit-interleaving in AWGN channel through increasing the free Euclidean distance by the knowledge of other bit values [2] [3]. The iterative decoding algorithm is to achieve the global optimum through a step-by-step local search. Unlike the structure of turbo decoder, there is only one soft-input soft-output (SISO) decoder in the BICM-ID receiver. Thus, the iterative process is between the SISO decoder and the demapper. As the iteration number increases, the BER performance improves. However, it's not always so. Since the conventional iterative decoding algorithm in BICM-ID is a scheme with fixed iteration number, it gets very little performance improvement with any further iteration. Considering the unnecessary decoding delay and the complex iteration computation, to solve this problem caused by the fixed iteration, various stopping criterions are proposed [4-10]. However, almost all the schemes are proposed for turbo decoder, not for BICM-ID receiver. Reference [11] proposed a novel scheme for BICM-ID system named Average-Entropy-Value (AEV) stopping criterion, which can take a better trade off between the average iteration numbers and the performance degradation at high SNRs for BICM-ID system. Furthermore, it doesn't need any extra storage units. This stopping criterion is stopped when the AEV is smaller than a predetermined small

positive number, which is called the threshold. But how small the threshold should be? Whether there exists a more appropriate value to be the threshold of the AEV scheme? This paper has investigated many probable values with an extensive amount of simulations to find the answers and devises that the suitable range of the threshold of AEV stopping criterion is about $0.5*10^{-4} \sim 0.5*10^{-3}$.

The paper is organized as follows: Section 2 gives a general review about the model of the BICM-ID receiver. Section 3 introduces the AEV stopping criterion. Section 4 shows the investigation process about the threshold. Last, section 5 draws the conclusion of the paper.
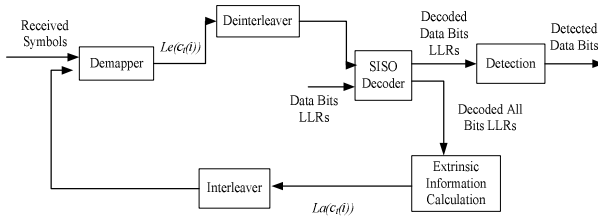
Fig. 1. The model of BICM-ID receiver

2   The BICM-ID Receiver

Fig.1 shows the receiver block of the BICM-ID system. The decoding process is between the demapper and the SISO decoder. The extrinsic information $Le(c_t(i))$ is deinterleaved before being applied to the SISO decoder

$$L_e(c_t(i)) = \log\left[\frac{P(c_t(i)=0 \,|\, r_t, L_a(c_t))}{P(c_t(i)=1 \,|\, r_t, L_a(c_t))}\right] - L_a(c_t(i)) \tag{1}$$

The a prior information $La(c_t(i))$, which is the output of the SISO decoder is then fed back to the demapper

$$L_a(c_t(i)) = \log\left[\frac{P(c_t(i)=0)}{P(c_t(i)=1)}\right] \tag{2}$$

We should note that $La(c_t(i))$ is set to be zero during the first iterative cycle. At the last iterative cycle, the hard decision, which is getting from the SISO decoder, is obtained based on the extrinsic bit information. The iterative process stops only when the AEV is smaller than the predetermined threshold.

3   The AEV Sopping Criterion

In the BICM-ID receiver that using stopping criterions, the iteration process is stopped when the uncertainty of the decoding is small enough. We all known that in information theory, entropy is a measure of uncertainty. Thus, reference [11]

proposed a novel stopping criterion named Average Entropy Value (AEV) criterion for BICM-ID system. The equation below is the entropy of the decoded bit u_i,

$$H(u_i) = -P_{ei} \log_2 P_{ei} - (1 - P_{ei}) \log_2 (1 - P_{ei}) \tag{3}$$

where P_{ei} is the a posteriori bit error probability of u_i,

$$P_{ei} = p(u_i \mid y) = \frac{1}{1 + e^{|L(u_i)|}}, \ P_{ei} \in (0, 0.5] \tag{4}$$

So, the AEV of all the decoded bits can be expressed as

$$\bar{H} = \frac{1}{N} \sum_{i=1}^{N} H(u_i) \tag{5}$$

We should note that when P_{ei} belongs to (0, 0.5], $H(u_i)$ is a monotone increasing function of P_{ei}. Therefore, in the AEV scheme we use $H'(u_i)$ instead of using $H(u_i)$, where

$$H'(u_i) = -P_{ei} \log_2 P_{ei} \tag{6}$$

Above all, the AEV stopping criterion can be stated as:

$$\bar{H} = \frac{1}{N} \sum_{i=1}^{N} H'(u_i) < Th \tag{7}$$

where Th is the predetermined threshold, which should be a sufficiently small positive number.

4   The Threshold Investigation Process
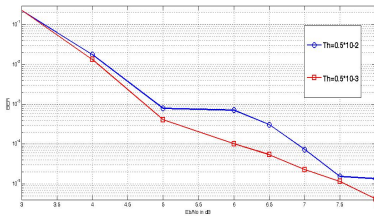
Fig. 2. BER comparison of $0.5*10^{-2}$ and $0.5*10^3$
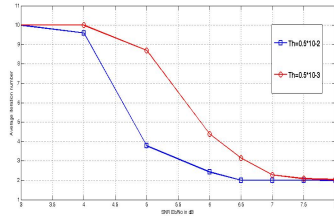
Fig. 3. Iteration number comparison of $0.5*10^{-2}$ and $0.5*10^{-3}$
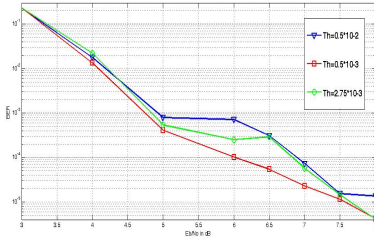
Fig. 4. BER comparison adding 2.75*10⁻³
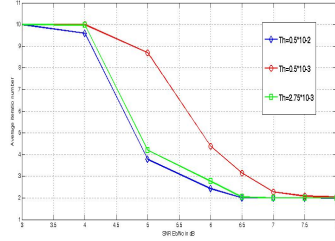
Fig. 5. Iteration number comparison adding 2.75*10⁻³
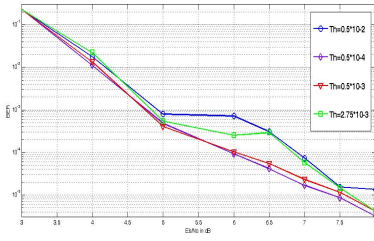
Fig. 6. BER comparison adding 0.5*10⁻⁴
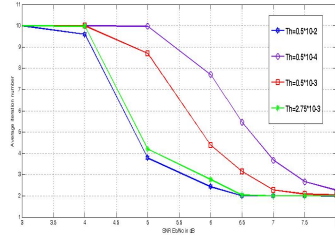
Fig. 7. Iteration number comparison adding 0.5*10⁻⁴
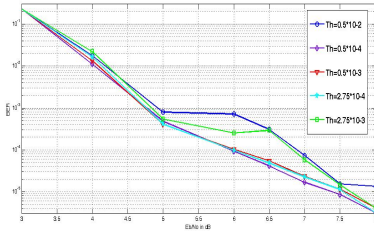
Fig. 8. BER comparison adding 2.75*10⁻⁴
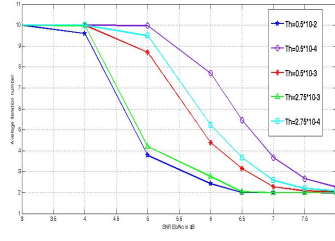
Fig. 9. Iteration number comparison adding 2.75*10⁻⁴
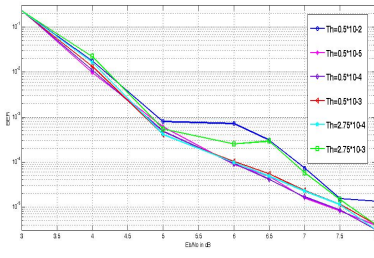
Fig. 10. BER comparison adding 0.5*10⁻⁵
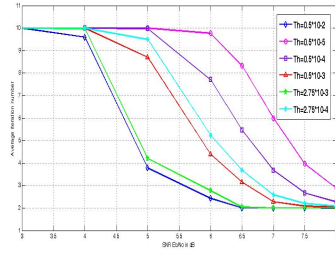
Fig. 11. Iteration number comparison adding 0.5*10⁻⁵

Reference [11] sets the threshold of AEV stopping criterion to be $0.5*10^{-3}$. Is it the most appropriate value? Is there exists a more optional value to be the threshold? Below, we will make a lot of investigations about the threshold via a large amount of simulations.

The parameters of all our simulations are as follows: rate 1/2 convolutional code, 8PSK modulation with SP mapping, Rayleigh fading channel, frame size 5114, maximum iteration number 10.

Compared with $0.5*10^{-3}$, we may want to know if there exists a loosened threshold which can largely reduce the iteration number without obvious performance degradation. So, our first try is $0.5*10^{-2}$.

From Fig.3 we can see that, the $0.5*10^{-2}$ scheme shows a great advantage over the $0.5*10^{-3}$ scheme considering about the iteration numbers. When SNR is 5dB, the iteration number can be reduced from 8.6 down to 3.8. However, the corresponding BER performance is extremely bad as shown in Fig.2.

Is the value between $0.5*10^{-3}$ and $0.5*10^{-2}$ optional? We then choose $2.75*10^{-3}$ as our second try. Fig.4 and Fig.5 makes a joint display that there isn't any distinct improvement using $2.75*10^{-3}$ as the threshold. The BER performance is still very bad. Thus, we turn to find a smaller threshold than the $0.5*10^{-3}$ scheme.

We choose $0.5*10^{-4}$ as the third try. Although the iteration number has a increase as Fig.7 shows, there exists a BER improvement. Fig.6 exhibits that when BER is below 10^{-5}, the $0.5*10^{-4}$ threshold scheme can get a more than 0.2dB coding gain compared with the $0.5*10^{-3}$ scheme. Therefore, the $0.5*10^{-4}$ scheme is optional when high-accuracy system is required.

How about the intermediate value that between $0.5*10^{-3}$ and $0.5*10^{-4}$ as the threshold? We choose $2.75*10^{-4}$ as the fourth try. Fig.8 shows that the $2.75*10^{-4}$ scheme performs almost the same as the $0.5*10^{-3}$ scheme in terms of BER properties. However, there is an obvious increase in iteration numbers as shown in Fig.9.

What's the performance that applying a threshold smaller than $0.5*10^{-4}$? Then, $0.5*10^{-5}$ is our next try. The $0.5*10^{-5}$ scheme conducts almost the same as $0.5*10^{-4}$ scheme in terms of BER performance as Fig.10 exhibits. Yet, the iterative number is greatly increased. Fig.11 shows that at the point of SNR 6.5dB, the iteration number has increased from 5.5 to 8.4. So, we can clearly know that a threshold that smaller than $0.5*10^{-4}$ isn't worth choosing at all.

5   Summaries

This paper has made a large amount of investigations about the threshold of AEV stopping criterion in BICM-ID system with extensive simulations as validation. It has verified that, in general, the value $0.5*10^{-3}$ is the most appropriate threshold. Yet, when high-accuracy system is required, the $0.5*10^{-4}$ scheme is more appropriate. The thresholds that either smaller than $0.5*10^{-4}$ or larger than $0.5*10^{-3}$ aren't suitable for the AEV stopping criterion. Therefore, the approximate range of the threshold for AEV stopping criterion in BICM-ID system is about $0.5*10^{-4}$~$0.5*10^{-3}$.

References

[1] Samahi, S.S., Goff, S., Sharif, B.S.: Comparative study for bit-interleaved coded modulation with iterative decoding. In: IEEE AICT, pp. 316–318 (May 2009)

[2] Li, X., Ritcey, J.: Bit-interleaved coded modulation with iterative decoding. IEEE Communications Letters 1(6), 169–171 (1997)

[3] Li, X., Ritcey, J.: Bit-interleaved coded modulation with iterative decoding using soft feedback. IEE Electronic Letters 34(10), 942–943 (1998)

[4] Hagenauer, J., Offer, E., Papke, L.: Iterative decoding of binary block and convolutional codes. IEEE Trans. Inform. Theory 42, 429–445 (1996)

[5] Shao, R.Y., Lin, S., Fossorier, M.P.C.: Two simple stopping criteria for turbo decoding. IEEE Trans. Commun. 47, 1117–1120 (1999)

[6] Scanavino, B., Maggio, G.M., Tasev, Z., Kocarev, L.: A novel stopping criterion for turbo codes based on the average a posterion entropy. In: GLOBECOM 2003, vol. 4, pp. 2051–2055. IEEE (2003)

[7] Yu, N.Y., Kim, M.G., Kim, Y.S., Chung, S.U.: Efficient stopping criterion for iterative decoding of turbo codes. Electron. Lett. 39, 73–75 (2003)

[8] Bokolamulla, D., Aulin, T.: A new stopping criterion for iterative decoding. In: Proc. IEEE Int. Conf. Commun., pp. 538–541 (June 2004)

[9] Wu, J., Wang, Z., Vojcic, B.: Partial Iterative Decoding for Binary Turbo Codes. IEEE Trans. Communications 57(11), 3298–3306 (2009)

[10] Li, S., Xie, L., Chen, H., Wang, K.: A new stopping criterion for Duo-binary Turbo codes. In: IEEE CMC, vol. 2, pp. 271–274 (2010)

[11] Zhang, S., Li, J., Cai, C.: An average-entropy-value-based stopping criterion for BICM-ID. In: IEEE International Conference on WCNIS 2010, pp. 98–101 (2010)

An Impoved Symbol Mappings on 16QAM Constellation for BICM-ID

Liu Na, Li Jianping, and Che Qing

Communication University of China

Abstract. Bit-interleaved coded modulation with iterative decoding (BICM-ID) is a spectral efficient coding technique. The technique is therefore very attractive for next generation of wireless communication. It has been shown that when interleaver and error-control code are fixed, symbol mapping which defined by the signal constellation and the bit labeling has a critical influence on the error performance of BICM-ID. This paper presents an improved symbol mapping which combines two M-PSK with different radius and phases called (4,12). Through comparison with the conventional symbol mappings in terms of BER performance over Rayleigh channel, numerical results show that the new symbol mapping can improve the performance of BICM-ID system more than 0.1dB at low-to-medium SNR regions.

Keywords: Bit-interleaved coded modulation with Iterative Decoding (BICM-ID), QAM, symbol mapping, Rayleigh fading channel.

1 Introduction

Trellis coded modulation (TCM), proposed by Ungerboeck in [1], is based on maximizing the Hamming distance between codewords, which can provide good performances for the AWGN channel but for fading channel. A different strategy that increases the time diversity of the coded modulation which is referred as BICM, was suggested by Zehavi in [2]. The improved time diversity at expense of reducing the free squared Euclidean distance, leading to a degradation over AWGN channels [3]. With a careful design of signal mapping, iterative decoded BICM (BICM-ID) overcomes the drawback of conventional BICM over AWGN by increasing the free Euclidean distance with the knowledge of other bit values [4].

According to Chi-Hsiao Yih's results [5], symbol mapping was recognized as the crucial design parameter to achieve an excellent performance for BICM-ID. In general, a symbol mapping is defined by the signal constellation and the bit labeling [6]. Several mappings on QAM constellation were introduced in [7] [8]. Modified set-partitioning labeling (MSP) proposed in [7], optimized under the harmonic mean criterion, showing the performance relatively close to that of turbo coded modulation with less complexity. The optimal Maximum squared Euclidean weight (MSEW) mapping proposed in [8] has the minimum number of symbol pairs which have the minimum squared Euclidean distances with Hamming distance-1. There is no one mapping having the best performance over all SNR regions. Many studies mentioned previously have concentrated on finding symbol mappings lowering bit error rates

(BER) at high SNR. In this paper, an improved symbol mapping called (4,12) is proposed, which can achieve better performance than other conventional mappings at low-to-medium SNR more than 0.1dB.

This paper is organized as follows. In section 2, we briefly review the model of BICM-ID. In section 3, the conventional symbol labelings and an improved symbol mapping are shown. Then, Section 4 shows the simulation results of BICM-ID performance with different mappings. Section 5concludes the paper.

2   System Model

The conventional BICM can be modeled as a serial concatenation of a convolutional encoder followed by a bit-by-bit interleaver and a memoryless modulator as shown in Fig.1. At the transmitter, the information sequence $\{u_t\}$ is encoded by a convolutional encoder. The coded binary sequence $\{c_t\}$ is then fed into a bit interleaver which not only break the sequential fading correlation but also increase diversity order to the minimum Hamming distance of a code.
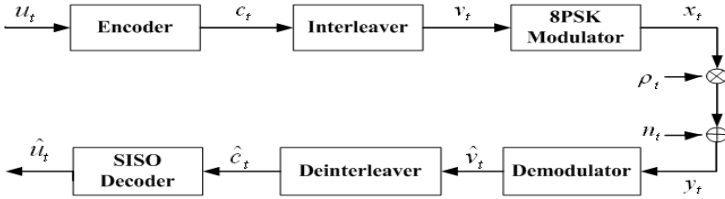
Fig. 1. BICM system model

After interleaving, the consecutive bits of the interleaved coded sequence are grouped to form $V_t = [v_t^1, v_t^2, v_t^3]$, which are mapped onto signal set χ of size $|\chi| = M = 2^m$ through the one to one mapping $\mu:\{0,1\} \to \chi$, $x = \mu(t)$ and the corresponding 8PSK signal at time t by x_t,

$$x_t = \mu(x_t), x_t \in \chi \tag{1}$$

Where the 8PSK signal set is $\chi = \{\sqrt{E_s} e^{j2n\pi/8}, n = 0,\dots,7\}$ and E_s is the symbol energy. For a frequency nonselective Rayleigh fading channel with coherent detection, the received discrete-time signal can be expressed as

$$y_t = \rho_t x_t + n_t \tag{2}$$

Where x_t is a transmitted symbol in M complex-valued. ρ_t is the Rayleigh random variable with $E(\rho_t) = 1$ presenting the fading amplitude of the received signal and nt is a complex AWGN with one-sided spectral density N_0. For the AWGN channel, ρ_t is set to 1.

The receiver of BICM-ID in Fig.2 uses demodulation and convolutional decoding as a su boptimal iterative method. The demapper processes the received complex symbols y_t and the corresponding priori loglikelihood ratios (LLR) of the coded bits.

$$L_a(c_k(i)) = \log[\frac{P(c_k(i)) = 0}{P(c_k(i)) = 1}] \tag{3}$$

And outputs the extrinsic LLRs

$$L_e(c_k(i)) = \log[\frac{P(c_k(i)) = 0 | r_k, L_a(c_k)}{P(c_k(i)) = 1 | r_k, L_a(c_k)}] - L_a(c_k(i)) \tag{4}$$
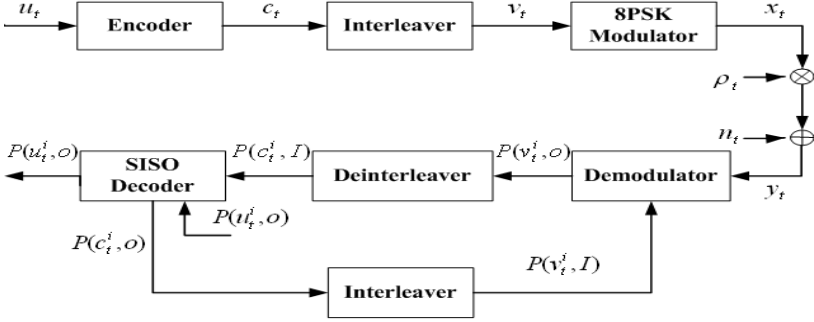
Fig. 2. BICM-ID system model

3   Signal Labeling

Different signal constellation labels are critical for optimazing different decoding methods. The comparison of five labels for 16QAM is shown in Fig.3. The influence of signal labeling to BICM-ID can be quantified by the two Euclidean distances [5] [9]. Unlike AWGN channels, there is no dominating term in the performance bound for the Rayleigh channel.

The asymptotic performance of BICM over Rayleigh fading [3] can be approximated by

$$\log_{10} P_b \approx \frac{-d_2(C)}{10}[(Rd_h^2(\mu))_{dB} + (\frac{E_b}{N_0})_{dB}] + const \tag{5}$$

Where P_b is the probability of bit error, $d_2(C)$ is the minimum Hamming distance of the code, R is the information rate and d_h^2 is the harmonic mean of the minimum squared Euclidean distance. For any M-ary constellation with a labeling map μ, d_h^2 can be calculated by

$$d_h^2(\mu) = (\frac{1}{m2^m} \sum_{i=1}^{m} \sum_{b=0}^{1} \sum_{x \in \chi_b^i} \frac{1}{\|x - z\|}) \tag{6}$$

Where m=log$_2$(M), $z = \hat{z}(x) \in \chi_b^i$ denote the nearest neighbor of x .Specifically,

$d_2(C)$ controls the slope of the probability of bit error P_b curve while d_h^2 provide the

horizontal offset. Given ideal feedback for each $x \in \chi_b^i, \chi_b^i$ contains only one term

$\tilde{z} = \tilde{z}(x)$ whose label has the same binary bit values as those of x except at the i'th
bit position.

From (6), the general rule is to design a labeling map μ that $\|x - \tilde{z}\|$ is larger than

$\|x - \hat{z}\|$ for all x (if possible) in order to achieve the iterative decoding gain. It is

preferable to have a labeling map that maximizes $\tilde{d}_h^2$ while having sufficiently large

original d_h^2 to make the first iteration work well.

For BICM-ID with convolutional code, numerical results from calculating the
harmonic mean of the minimum Euclidean distance are shown in Table 1. This gives
a quick comparison between various labeling schemes. It is shown that Gray yields
the best first round performance, however, the performance gain with feedback is very
small. MSP and MSEW yield great iterative decoding performance after feedback, but
they both have the poor d_h^2, which leading to a degradation at low SNR. The weaker
performance improvement of d_h^2 obtained by SP labeling is also shown for
comparison.

Table 1. Harmonic mean of the minimum squared Euclidean distance d_h^2 before and after
feedback and the d_h^2 asymptotic gain over six labelings

Labeling	d_h^2 (before)	d_h^2 (after)	Gain(dB)
Gray	0.492	0.514	0.19
SP	0.441	1.119	3.56
MSP	0.420	2.279	6.65
MSEW	0.400	2.364	1.77
Antigray	0.320	0.992	1.49
(4,12)	0.739	2.118	1.46

In this paper, we propose an improved type of symbol labeling scheme, which is
composed of two M-PSK with different radius and phases called (4,12). The ratio
between the radius of the outer and inner circles is approximate to 2.6, which is
larger than the ratio in [10]. The new signal constellation depicts in Fig.3. Via
computer simulations, it is vertified that the new mapping could provide the better
performance than Gray, SP, MSP and MSEW at low-to-medium SNR regions, due to
largest d_h^2.
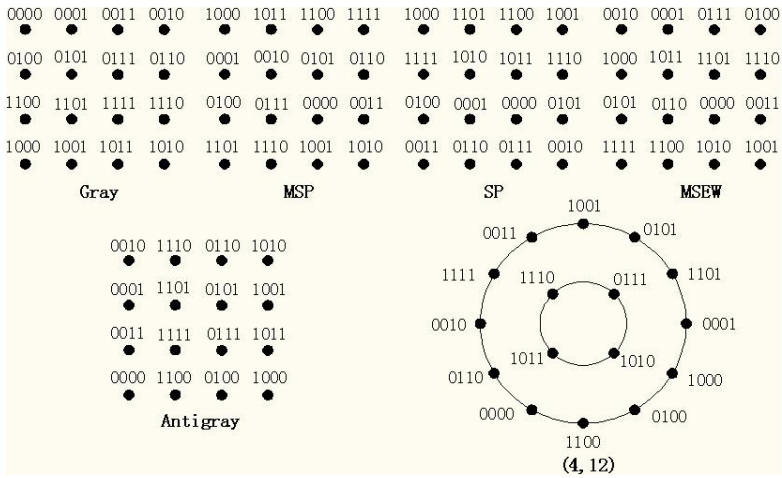
Fig. 3. Six labeling methods for 16QAM

4   Simulation Results

In our simulations, a rate-1/2, systematic convolutional (RSC) code with polynomials (131, 171) is assumed. The pseudorandom bit interleaver with a length of 5114 bits is used. Linear MAP algorithm is used for the decoding of each RSC code. Fig.4 and Fig.5 compare the BER performance with 10 iterations employing various mappings mentioned previously over Rayleigh channel.

The effect of signal labeling is shown in Fig.4. It can be observed that without iterative decoding, Gray yields the best performance. MSP and MSEW desperately achieve the worst performance at low SNR, although they have lower error floor. Furthermore, when the BER is close to 10^{-3}, Antigray mapping achieves at least more than 0.5dB than the conventional symbol mappings. And then, the (4,12) can yield 0.1dB code gains over Rayleigh fading channel compared with Antigray mapping in Fig.5.
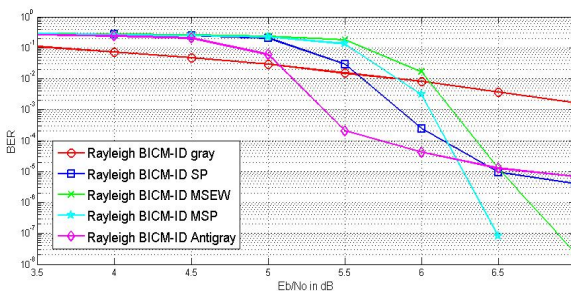
Fig. 4. Performance comparison of conventional mappings
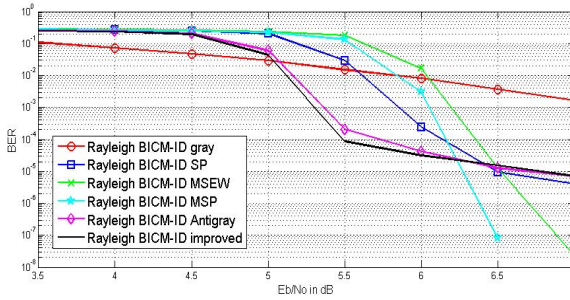
Fig. 5. Performance comparison of conventional and (4,12) mappings

5   Summaries

In this paper, an new symbol mapping (4,12) is proposed. We have analyzed it in terms of BER performance via theoretical analysis and computer simulation. To compare its performance with some previously mentioned signal labeling maps, such as Gray, MSEW, Antigray, SP, and MSP, some computer simulations were completed. Simulation results show that the newly proposed signal mapping can obtain more than 0.1dB code gains compared with the conventional mappings at low-to-medium SNR regions over Rayleigh channel.

References

[1]   Ungerboeck, G.: Channel coding with multilevel/phase signals. IEEE Trans. Inf. Thoery 28, 55–67 (1982)
[2]   Zehavi, E.: 8-PSK trellis codes for a Rayleigh channel. IEEE Trans. Commun. 40(5), 873–884 (1992)
[3]   Caire, G., Taricco, G., Biglieri, E.: Bit- interleaved coded modulation. IEEE Trans. Inform. Theory 44(3), 927–946 (1998)
[4]   Li, X., Chindapol, A., Ritcey, J.: Bit-interleaved coded modulation with iterative decoding and 8PSK signaling. IEEE Transactions on Communications 50(8), 1250–1257 (2002)
[5]   Yih, C.-H.: Performance optimization of bit-interleaved coded modulation with iterative decoding, pp. 737–742 (2002) IEEE 0-7803-7625-0/02
[6]   Muhammadand, N.S., Speidel, J.: Joint optimization of signal constellation and bit labeling for bit-interleaved coded modulation with iterative decoding. IEEE Communications Letters 9(9), 775–777 (2005)
[7]   Chindapol, A., Ritcey, J.: Design, analysis and performance evaluation for BICM-ID with square QAM constellations in Rayleigh fading channels. IEEE Journal on Selected Areas in Communications 19(5), 944–957 (2001)
[8]   Jun, T., Stuber, G.L.: Analysis and design of symbol mappers for iteratively decoded BICM. IEEE Transactions on Wireless Communications 2, 662–672 (2005)
[9]   Samahi, S.S., Goff, S., Sharif, B.S.: Comparative study for bit-interleaved coded modulation with iterative decoding. In: IEEE AICT 2009, pp. 316–318 (May 2009)
[10] Fang, W., Li, J.: Optimization of symbol mapping for Bit-Interleaved Coded Modulation with Iterative Decoding. In: IEEE International Conference on Communication Technology, pp. 346–350 (November 2010)

Coal Mine Ventilator Remote Monitoring System Based on the Fuzzy Control

Liang Wang[1], Yuanjun Wang[1], and Jitian Pei[2]

[1] Shandong University of Science and Technology, Qingdao, China 266510
{6057079,wangyuanjun723}@163.com
[2] Jining III Coal Mine, Jining, Shandong, China 272069
pjthyn@163.com

Abstract. For the reason of more complex ventilation system and more ventilation parameters needed in the coal mine, the standard mathematical model of ventilation system is hard to be set up. At the same time, the common practice cannot obtain the satisfactory effect. This paper put forward a remote monitoring system of coal mine ventilation based on a fuzzy control to realize the real-time monitoring of mine ventilation systems. The remote automatic control program of ventilator will be realized by Siemens modules and configuration softwares. The results show that the fuzzy automatic control system is useful, with the characteristic of safe, reliable and easy-operating. It provides a reliable guarantee for safe production of coal mines and the efficient operations of the fan.

Keywords: Ventilator, Fuzzy control, Remote monitoring.

1   Introduction

The ventilation plays a very important role in the process of coal mining. Mine ventilator undertakes the task of safeguarding the whole underground coal mine safe production. Because of the relatively poor working conditions, more level and depth, complex roadway distribution and high rate of fans accident, these are giving a great threat to the safety of underground workers. Therefore, the safety, reliability and economic operation of ventilator remote monitoring are of vital significance in mine production. To make the ventilator operate safely and efficiently, we must adopt remote automatic control system. The system structure goes through the transformation from the industrial control computer as the core of the centralized control model to the PLC as the core of a distributed control model. The conventional control method for automatic adjustment of the ventilator is difficult to achieve the expected effect. The remote monitoring system of coal mine ventilation system can overcome the various drawbacks in theory. The fuzzy control method has features of rapid response, simple control, without modeling, so the ventilation system based on fuzzy control has been got practical application.

2   System Working Principle

The basic idea of fuzzy control is to use computer to replace human experience control, and these experience with the fuzzy control rules is mostly expressed by language. Considering the nonlinear and multi-interference of this system, the fuzzy control method is able to avoid the difficult problem of modeling, and has better control performance, so adopting the program that fuzzy control technique adjusts remote automatically of Ventilator. This system collects all ventilator operation parameters to industrial PC, and then through the software for the fuzzy algorithm controls the start-stop of the ventilator, thus achieving unattended and intelligent target. The results show that the fuzzy control automatic adjustment system control has good effect. This proves that making full use the advantage of classical control method and fuzzy control method to control the air volume of fans is practical and feasible. And along with the manufacturers have introduced the fuzzy reasoning module， which makes the system more mature and stable[2].

3   Fuzzy Control Theory

3.1   Fuzzy Control Overview

Fuzzy Logic Control was short for Fuzzy Control. It is a computer numerical control technology based on fuzzy set theory, the fuzzy language variables and fuzzy logic reasoning. Fuzzy Control is essentially a kind of nonlinear control, belonging to the category of intelligent control. A major feature of fuzzy Control not only has systematic theories, but also has a large number of the application background[5].

3.2   Fuzzy Control System Element and Basic Principles

The fuzzy control system structure is similar with regular feedback control systems, including fuzzy controller, the input and output interfaces, executive body, controlled object and measuring devices. The difference is that fuzzy control system is based on fuzzy mathematics, fuzzy linguistic form of knowledge representation and fuzzy logic rules reasoning, using computer control technology to form a channel digital control system with feedback channel closed-loop structure. The structure of fuzzy control system, as show in Fig.1.
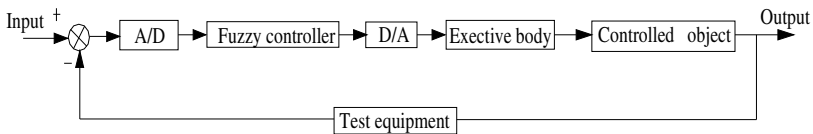
Fig. 1. The structure of fuzzy control system

Fuzzy control system can be divided into the following four parts[3]:

(1) Fuzzy controller: Actually, it is a micro-computer, the fuzzy controller is the core of the fuzzy control system, and it is mainly achieved by software on the computer in practice.
(2) Input /output interface device: Fuzzy controller through an input/output interface gets digital signal quantity from controlled objects, makes output digital signal of fuzzy controller decision through D/A transformation, then converts it into analog signals, and gives actuators to control the controlled object.
(3) Controlled objects and the implementing agencies: Controlled objects can be linear or nonlinear, constant or time-varying, also it can be single or more variables, delay or without delay and have a variety of strong interference. Controlled objects lack of accurate mathematical model suitable to select fuzzy control.
(4) Sensor: Sensor is a device which converts controlled objects or controlled variables of various processes to electrical signals. Controlled variables are often non-power, and sensors play a very important position in the control system, whose precision directly affects the precision of the whole control system.

Fuzzy control is a computer intelligent control based on fuzzy set theory, the fuzzy linguistic variables and fuzzy logic reasoning, and its core part is the fuzzy controller. Its basic principle, as show in Fig.2.
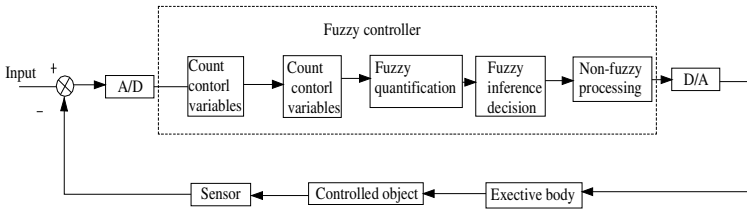
Fig. 2. Fuzzy control basic principle

4   Design Ventilator Fuzzy Controller

There are many kind of fuzzy controller structures, we use two-dimensional fuzzy controller according to the actual situation of the controlled object in this system[4].

4.1   Fuzzy

Ts is sample period of this system, obtaining its input response r and y in every sampling time, and getting error e and deviation rate train Δe, e and Δe as fuzzy controller input.

$$e(n) = r(n) - y(n)$$

$$\Delta e(n) = \frac{e(n) - e(n-1)}{T_s} \tag{1}$$

$$Y = \frac{12}{b-a}(x - \frac{a+b}{2}) \tag{2}$$

The actual deviation and deviation changes between a and b. Using the 2nd formula change them between -6 and +6. The exact amount will be converted into a blur. Interval discretization of the domain will be divided into n files, and each file corresponds to a fuzzy set, such as NL, NM, NS,O, PS, PM, PL.

4.2   Rule Base

Controlled objects are deeply understood, and the corresponding control rules are given:

if E= NL   and   ΔE= PL then U = PL
if E= NM   and   ΔE= PL then U = PM

if E= PM   and   ΔE= NL then U = NM
if E= PL   and   ΔE= NL then U = NL

4.3   The Fuzzy Relationship Matrix and Fuzzy Decisions

All the control rules will be tabulated, which constitute the rule base in Table 1.

Table 1. Control rule base

ΔE \\ E	NL	NM	NS	O	PS	PM	PL
PL	PL	PM	NL	NL	NL	NL	NL
PM	PL	PM	PL	NS	NS	NS	NS
PS	PL	PM	NS	NS	NS	NS	NL
ZE	PL	PM	PS	O	NS	NS	NL
NS	PL	PM	PS	PS	PS	PM	NL
NM	PL	PL	PL	PM	PM	NM	NL
NL	PL	PL	PL	PL	PM	PM	NL

4.4   Fuzzy Decision

This system uses the gravity center method for fuzzy decision-making. Fuzzy reasoning process is based on fuzzy control rules to get fuzzy relationship matrix R, then according to the synthesis of fuzzy reasoning rules to control the amount of output fuzzy sets Uij.

$$R_{ij} = (E_i \times EC_j) \times U_{ij} \tag{3}$$

In the formulas: Rij is the fuzzy relation between various rules. Ei is the fuzzy vector of error e. Ecj is the fuzzy vector of ec. Uij is output vector of fuzzy controller.

5   System Design

System software mainly consists of sensor intelligent identification, data collection and processing, algorithm, mechanical status, fault analysis and diagnosis, alarm and print

out to complete their corresponding function. Software system adopts configuration compilation. At the same time, signals are analyzed in time domain and frequency domain.It analyzes and judges the final data processing, outputting in the form of digital or graph etc.

The system uses high-performance sensors, selects the Siemens S7-300PLC, adopts Advantech industrial computer, and takes on-line UPS, to ensure that the system can operate regularly when the system power is switched. And through the switches, pass the system to office automation network and group company nets in the mine. System structure diagram, as show in Fig.3.
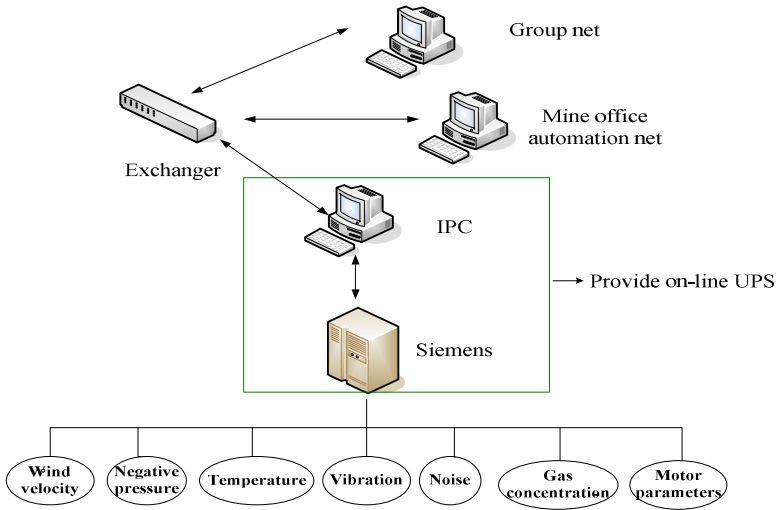
Fig. 3. System structure diagram

The system has mainly functions: 1. real-time monitoring ventilator system operation parameters, and touring display. 2. Analyzing the acceleration time domain and frequency domain. 3. Fuzzy control. Software controls real-time data parameters and the set reference. 4. The intelligent diagnosis. Use the intelligent fuzzy system inside of industrial PC, and then position accurately the common mechanical breakdowns in the ventilator. 5. Alarm print.

6   Summaries

Using fuzzy control method can well realize remote automatic monitoring of the whole mine ventilator, with unattended. Choosing appropriate pace can accelerate the operation speed. Real-time monitoring system organically combines ventilator information monitoring and office automation network together, which has provided protection for every relevant units in mine and grasping the information timely and accurately. It can guarantee the ventilator's data acquisition and fault alarm in the harsh environment, still can find position and type of the failure timely and accurately, which

has laid a good foundation for the next step to eliminate the failure, and has provided reliable guarantee for the safety production and high effective production. This situation has played an active role in improving the security situation in coal mining. But fuzzy control system theory has some important theoretical issues to be unresolved.

References

[1] Zhao, T., Wei, J.: The OPC Online Monitoring System. Coal Electrical Machinery (2010)

[2] Wu, X., Ren, Z., Ma, X., Li, S.: Coal Mine Ventilator On-line Monitoring System Research Status and Prospect. Coal Science and Technology 37(12)

[3] Li, J., Tang, W., Gu, R.: PLC and The Fuzzy PID Controller in The Ventilator Application of Air Control System. Jiangnan University Journals 5(4)

[4] Qin, X., Gu, S.: Control Ventilator Airflow Control Based on the Fuzzy PI. Electrical Application 27(16)

[5] Dai, L.: Local-ventilator Gas Emissions Research Intelligent Based on Fuzzy Control. Xian Science and Technology University Degree Thesis (2008)

[6] Zang, X., Wang, Y., Song, S., Luo, Y.: The Fuzzy Control Theory Used in Coal Mine Ventilation Safety Automation System. Chinese Security Science Journal 10(3)

[7] Fu, S., Li, H., Zhu, Q.: A Fault Early Warning and Software Development of the Main Ventilator in Mine. Journal of Beijing University of Technology (8) (2007)

[8] Guo, X., Ma, X.: Vibration Trend Prediction Based on GrayLSSVM Combination Model for Mine Main Ventilator. In: Chinese Control and Decision Conference. Inst. of Elec. and Elec. Eng. Computer Society, United States (2008)

The Research of Coal Mine Conveyor Belt Tearing Based on Digital Image Processing

Qing-Liang Zeng, Ji-Gang Wang, Liang Wang, and Cheng-Long Wang

College of Mechanical and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266510, China
{qlzeng,6057079}@163.com, wangjigang052@126.com,
wcllym@hotmail.com

Abstract. The application of digital image processing has been widely used in industrial detection. In coal mine production, the conveyor belt easily tearing. This will take advantage of computer on terms of quick and accurate. This uses the method of mixed programming with C++ & Matlab to process image signal, and designs the hardware and software of the detection system. This system will improve the reliability of detection of the belt tearing, the detection with intuitive, fasting and accurate.

Keywords: Belt conveyor, Tearing, Digital image processing, Matlab.

1   Introduction

The belt conveyor is the main equipment in the coal mine production process. In the five protections of the belt, the belt tearing is a very common failure. The reliable operation of the belt conveyor is very important in the economic benefits of the entire coal mine. Steel cord is widely used in the long-distance belt conveyor to increase the tensile strength. But because of the complex working conditions in mining areas,coupled with the belt itself is rubber, which is likely to result belt longitudinal tear.

With the development of computer application technology, which has been widely used in various fields. For the detection of underground belt conveyor, the application of traditional methods can no longer meet the modern mine safety management of high demand. Must be studied more   visual inspection methods and techniques based on machine vision.

In recent years, Matlab as a mathematical has been greater progress in the data processing, numerical computing, systems analysis, graphic display, graphics processing etc. In image processing, many algorithms have emerged. So in biomedicine, signal analysis, image and graphics processing has been widely used. By studying the various aspects of digital image processing in the application of technology, the reliability of the computer's high-speed and digital image processing technology. This paper proposes the coal mine conveyor belt tearing based on digital image processing and designs this system.

2   Research of the Coal Mine Conveyor Belt Tearing

Belt conveyor as a major coal mine production facilities is sub-stratified designed in the underground. With continuous delivery capability, high efficiency, easy control, etc. Using the steel cord belt inside the belt tension in order to strengthen capacity. But between the steel cord and steel cord ,the conveyor belt is composed entirely of rubber which low intensity. During operation, the belt often tear by metal, coal stone and other sharp objects. If it is failure to detect and stop conveyor, will cause belt greater damage and great economic losses.

There many reasons cause conveyor belt tear in long working distance. For example: foreign body through tear, materials compression tear, conveyor belt deviation tear, conveyor component failure tear, etc[1]. Conveyor belt tear occurred arbitrary and complicated underground conditions. In the event of tearing failure, requiring rapid and accurate detection more difficult.

Tear fault for the belt, the current detection methods can be divided into the main contact detection method and non-contact detection method. Contact detection methods are: linear detection, bar-type detection, oscillating roller detection, vibration detection, pressure detection, bandwidth detection[2]. Non-contact detection methods are: embedded detection, ultrasonic detection, X-ray detection, conductive coil detection, conductive rubber detection, image processing detection method.

3   System Theory and Composition

The underground belt conveyor detection method based on digital image processing is in the research stage. The main principle includes direct detection, diffuse reflectance method and blanking detection. The image signals are processed by DSP unit or industrial computer. Development of algorithms for image processing, mainly based on C #, C + +, Matlab, LabView, etc.

When the image signal acquisition, below the belt, through the bright light, will round exposure to the belt at the spot, adjust the lighting distance, to ensure that circular spot diameter slightly smaller than the belt width. In order to prevent the belt torn by blanking pollution source, especially let the light outer the belt at the place. Which will lead to spot not the standard round, but that does not affect the test results. When the belt tear, phenomenon in the spot will appear within a crack, crack at the spot light will be reduced. Collected image signal will clearly reflect this phenomenon, through image processing software to detect the degree tear.

The principle of the method is based on the use of high-brightness light source, collecting spot image signal. Not only alarms quickly and accurately in the belt tear, but can detect the fine cracks during working. It also can advanced protect the conveyor.

Through research of the belt conveyor, belt tear phenomenon more likely to occur in the nose and tail. So set two signal acquisition devices in the nose and tail. In order to ensure the accuracy of detection, it put other two devices uniform in the working distance. Fig.1 shows the structure of the belt detection system.
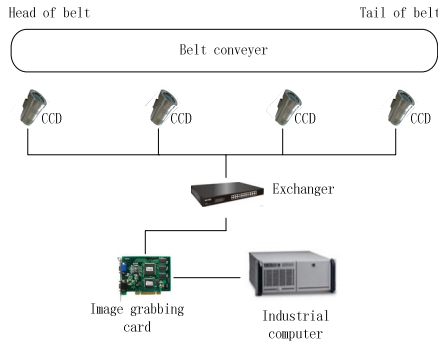
Fig. 1. Structure of the belt detection system

4 Image Preprocessing

In the whole belt detection system, image acquisition and preprocessing of signals play a crucial role. Because of at the mine working place, it is wet, high temperature, dusty and low visibility. With the impact of adverse environmental factors, making the image acquisition from the CCD low clarity, distortion and with lots of noise. Due to the impact of the external environment, in order to enhance the authenticity of the image, the image must be preprocessing. We can adjust brightness and contrast of the original image. Crack characteristics will be more obvious, and easy identification. Fig.2 shows the process of image preprocessing.
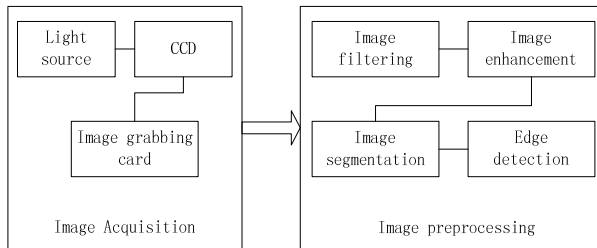
Fig. 2. Process of image preprocessing

Image preprocessing is divided into four parts: image filtering, image enhancement, image segmentation and edge detection and morphological processing.

Image Filtering. Image noise according to their causes, is divided into external noise and internal noise. Because of the complexity and low visibility, the noise in the system mainly for external noise[3]. System uses the median filter to filt out image noise. In the poor environment, median filter can improve the clarity of the processed image better than mean filter, and lay the foundation for subsequent processing.

Image Enhancement. In the belt tearing place, image will be more obvious difference. In order to display the crack more clearly in the image, the system sues the Laplacian operator to increase the cracks, expand the differences between light and dark and increase the brightness in the crack internal. Prepare for the image segmentation.

Image Segmentation and Edge Detection. System sues threshold method to separate tearing part and normal part in the image, detect the discontinuity in the local features of the image, then put them together into the boundary, the image is divided into different areas between these boundaries. Cracks in the image will be reacted more clearly. Edge detection is important part of the system. The common algorithm are as follows: Roberts algorithm, Sobel algorithm, Prewitt algorithm, Kirsch algorithm, Laplacian algorithm and LOG algorithm. In order to light the cracks in the belt, Roberts algorithm can accurately locate edge, the width of the boundary is unlike the other wide. So the system chooses the Roberts algorithm.

Morphological Processing. Although the image enhancement processing makes some noise suppressed, it must smooth binary images in order to remove the noise completely[4]. Mathematical morphology dilation and erosion methods remove the image interference point, when the belt further processing, the interference of other factors will not arise miscarriage of justice.

5   Belt Tearing Detection System Design

Hardware Design. To obtain a clear and accurate belt image signal, the system uses 30 million-pixel CCD area array camera, designs hardware driver circuit, get the belt image of 5Hz frequency. System chooses high-brightness and single color LED. It can get high quality and high contrast image signals. To be collected quickly and accurately convert analog signals into digital signals, it uses multi-channel high-speed image acquisition card, installed in the computer PCI expansion slot.

Software Design. System Software primarily process the collected image signal through the C++ and Matlab mixing programming method. Functions of Matlab image processing toolbox provide powerful tool for digital filtering and digital image processing. Making the digital image processing become easy and intuitive. In order to call Matlab library function in C++, first put m files into a C program, then generate C files to exe files by Matlab's mex batch file, finally through the C++ external functions called shellexec() implementation calls.

After image preprocessing, it needs further processing belt crack. Calculate wide of crack means calculate the share of the crack within the pixel. During calculating the width, process the image in two graphics by selecting gray threshold value[5]. Compare the calculated width with the pre-set width, when over the threshold, system will putout alarm and stop the conveyor. Fig.3 shows the process of system software design.
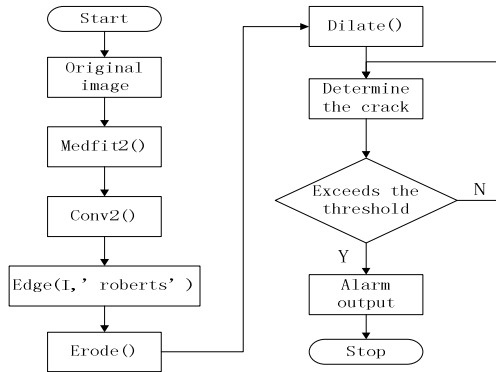
Fig. 3. Process of system software design

6   Summaries

Application of digital image processing technology in the biomedical, industrial production, military and office automation increase widespread and more mature. In the field of mine belt conveyor detection, the research is still initial stage. The method in this paper, based on computer processing technology, can greatly improve accuracy and avoid false positives. It can be detected in ground control center through the cable. Meanwhile, the system has good portability, can be applied to other industrial production testing of the belt. Of course, due to the complexity of the underground working environment, the system also need to be perfected according to the different working conditions.

References

[1] Xu, R., Lin, F., Miao, C.: Discussion on Monitoring and Measuring Method of Belt Longitudinal Tearing Based on Loading Chart. Coal Science and Technology (2009)
[2] Zhang, A., Sun, Y., Yin, Z.: Research status and tendency of longitude tearing protection for belt conveyor. Coal Science and Technology (2007)
[3] Qi, J., Tan, C.: Application of Digital Image Process in Belt Longitudinal Rip Vision Detection. Coal Mine Machinery (2009)
[4] Jia, Y., Cheng, L., Yu, G.: Digital image measurement of specimen deformation based on CCD cameras & Image J software: An application to human pelvic biomechanics. Information Optics and Photonics Technologies II (2007)
[5] Zhou, X., Chang, J.: Research Survey of Digital Image Processing Technology in Microstructure Analysis of Wood. Forest Engineering (2008)

A Method of Data Gathering and Processing for the Automatic Variable Spraying System

Shuhui Zhang[1], Lixia Wang[2], Jiangtao Qi[3], Hongjun Su[3], and Min He[3]

[1] Department of Mechanical Engineering,
Nanjing Institute of Technology,
Nanjing, Jiangsu Province, China
[2] School of Aeronautical Manufacturing Engineering,
Nanchang Hangkong University,
Nanchang, Jiangxi Province, China
[3] College of Bio-agriculture Engineering,
Jilin University Changchun,
Jilin Province, China

Abstract. This paper presents a method of gathering and processing the spatial information for the variable sprayer in precision agriculture. A S3C44B0X microprocessor of ARM7 series was took a signal gathering and processing core for the variable sprayer. The outline of the farm was drawn by using the GPS (Global Positioning System) and the GIS (Geographical Information System). After the states of weed, plant diseases or insect pests in the farm were investigated by the manual or the sensor, the prescription for automatic variable spraying was provided by the plant protect expert system. When the variable sprayer worked in the field, the real-time signals of the position and the ground speed of the variable sprayer, the pressure and the flux in the sprayer pipeline, were gathered and processed. At the same time, the prescription were referred to the position of variable sprayer , the flux was controlled by opening size of nozzle of the electric control valve in the sprayer pipeline according to the prescription and working speed of the variable sprayer. The results of the field test shown that the method of data gathering and processing could be competent for the automatic variable spraying in precision agriculture.

Keywords: Data gathering and processing, Variable spraying, precision agriculture, S3C44B0X microprocessor, Prescription.

1 Introduction

When a variable sprayer is working in the field, the GPS or DGPS is used to supply the real-time spatial positional information for the variable sprayer [1,2]. The states of weed, plant diseases and insect pests in the farm were investigated by the manual or the sensor with the DGPS. The management information database about the weed, plant diseases and insect pests in the farm was built by GIS, and a prescription of applying herbicide or pesticide was worked out by the plant protection expert system. According to the prescription and the sprayer's working speed, the automatic variable sprayer accomplishes automatic variable spraying by controlling flux which was controlled by

opening size of nozzle of the electric control valve in the sprayer pipeline [3-9]. The automatic variable spraying based on the prescription needs two steps to fulfill: The first step is to work out a prescription for automatic variable spraying. The electronic map of the farm is got by GPS. Then the farm is divided into many plots ,each of them is coded and named. The management information database of this farm is built and a prescription is presented by the plant protection expert system. The second step is to fulfill automatic variable spraying in the field. The first procedure has been solved by the re-development language MapBasic of the MapInfo Professional 7.5. This article researches the data gathering and processing of the variable sprayer when it is working in the field. The real-time gathering signals of the position, the speed, the pressure and the flux in the pipeline are the article's research keys.

2   Signal Gathering System

Implementation process of the automatic variable spraying is: after the signal gathering system collects the involved signals, the automatic variable spraying controller reads data in the prescription and outputs the corresponding voltage value to control opening size of nozzle of the electric control valve in the sprayer pipeline to realize automatic variable spraying[9].

The hardware of the signal gathering system of the automatic variable spraying consistes of a variable spraying controller, a speed signal collection unit, a pressure signal detector unit, a flux detector unit, and a set of DGPS receiver. The block diagram of the data gathering system was showed in Fig 1.
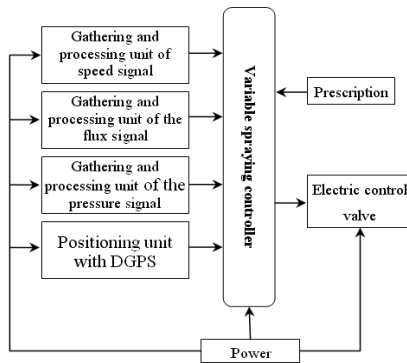
Fig. 1. Block diagram of the data gathering system

2.1   Variable Spraying Controller

The CPU of the variable spraying controller is S3C44B0X, a 32-bit ARM microprocessor which was produced by SAMSUNG Company. The S3C44B0X has six 16-bit timers, each timer works in interrupt mode or direct memory access model. The integrated UART module provides two independent asynchronous serial interfaces which are port PC and port PE. The S3C44B0X microprocessor undertook

two functions: one is to gather signals of the position and the ground speed of the variable sprayer and signals of the pressure and flux in the spraying pipeline; the other is to process these signals and output corresponding voltage value to control opening size of nozzle of the electric control valve in the sprayer pipeline to realize automatic variable spraying.

2.2   Gathering and Processing Unit of the Ground Speed Signal

The signals of the ground speed are gathered by means of a LM18-300NA-type magnetic induction sensor, the working voltage of which is 6 ~ 36VDC, output is a serial of 10V pulse signal (digital signal). The unit of gathering the ground speed signal is shown in Fig 2.
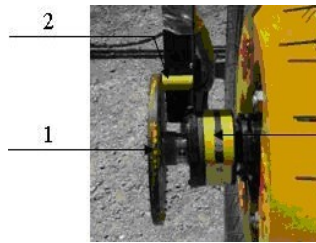
1 Pulse inductor   2 Magnetic induction sensor

Fig. 2. The unit of gathering the ground speed signal

In order to reduce the noise and interference, a TLP521 chip was used as filter to shape signals and transform signals from 10V level to 3.3V level to adapt to the S3C44B0X microprocessor. In addition, 74HC14 chip was used for obtaining a series of stable pulse signals which were directly input to the external interrupt port 2 (EINT2) on the S3C44B0X microprocessor after being shaped.

2.3   Gathering and Processing Unit of the Flux Signal

The gathering and processing unit was made of a LWGY-type flow sensor and a signal shaping circuit.

The key features of the LWGY-type flow sensor, shown as Fig 3, are as follows:
15mm nominal diameter,
$0.6 \sim 6.0 \ \text{m}^3/\text{h}$ in 0.5% accuracy,
Powered by 24VDC,
12V pulses output.

The flow sensor's outputs digital signals. In order to reduce external interference, the TLP521 chip is used to anti-interference processing and The 12v level signal output from the flow sensor was transformed into 3.3V level which is required by the S3C44B0X microprocessor I/O port. The transformed signals are inputted to the external interruption on the EINT3S3C44B0X microprocessor.

Fig. 3. The LWGY-type flow sensor   **Fig. 4.** The HX-L61-type pressure sensor

2.4   Gathering and Processing Unit of the Pressure Signal

The pressure signal gathering and processing unit in the spraying pipeline consistes of a HX-L61-type pressure sensor and an analog signal processing circuit. The HX-L61-type pressure sensor was used for detecting the pressure level in the spraying pipeline. The HX-L61 pressure sensor, shown in Fig.4, is selected considering fixing and liquid characteristic (pesticides or herbicide) in the pipeline. The key features of the HX-L61-type pressure sensor are:

　1.6Mpa maximum pressure,
　0.5% measurement accuracy,
　24VDC power voltage,
　4 ~ 20mA current output.

The analog signal processing circuit is used to amplify and convert analog signals into digital signals. The pressure signals are input the S3C44B0X microprocessor through the analog signal processing circuit and feedback to the spraying control system to build up a closed-loop control system.

As the pressure sensor outputtes a 4 ~ 20mA analog signal, and the S3C44B0X microprocessor only received maximum input voltage of 3.3V, an analog signal processing circuit have to use, which is made up of a operational amplifier OP07 chip and a TL431 chip after the conversion resistance. The signal of the pressure is input into the AIN [7] pin on the S3C44B0X microprocessor after the analog signal processing circuit.

2.5   Positioning Unit with DGPS

Two AgGPS132 receivers, produced by Trimble Ltd., were used to make of DGPS. The DGPS includes the following key features:

　1 meter accuracy after differential correction,
　NMEA output and RTCM input.

The positioning unit gets signals from the DGPS and recognizes   which plot the automatic variable sprayer work in. The automatic variable sprayer read NMEA - 0183 format of $GPRMC statements [10].

3   Gathering and Processing

3.1   Reading the Prescription for the Automatic Variable Spraying

After the prescription for the automatic variable spraying is given by the plant protection expert system, these prescription data are memorized in the NAND flash memory in a "*.txt format" file. In the control program of the automatic variable spraying, uses 'pfile = OpenOSFile (TextFilename, FILEMODE_READ)' sentence to open the prescription file in read-only way. After the prescription file is opened, the program uses 'while (LineReadOSFile (pfile, str) >2)' sentence to read the longitude and the latitude of the variable sprayer, prescription data and the shape parameters of the field in a row and then in a column respectively. After acquiring all the data above, the program used CloseOSFile (pfile) to close the prescription file.

3.2   Gathering and Processing Signals of Velocity and Flux

The velocity sensor and the flow sensor output the digit signals. These signals are input to the EINT2 pin and the EINT3 pin of I/O port on the S3C44B0X microprocessor's after the signal processing circuit. The signals of velocity and flux are got by programming the timer and the interrupt counter of the S3C44B0X microprocessor.

3.3   Gathering and Processing the Pressure Signal

The pressure sensor outputtes an analog signal within 4-20mA, so the A/D converter what is embedded in the S3C44B0X microprocessor is used to convert the analog signal to the digital signal. In order to improve accuracy and reliability of the signals obtained, the control program reads A/D converter 100 times and calculate an average, then inputtes the mean value to the variable spraying controller.

3.4   Gathering and Processing GPS Signals

Under the automatic variable spraying with the DGPS, the automatic variable spraying controller receives the GPRMC sentence to gather positional and velocity information of the automatic variable sprayer.

The GPRMC sentence's format is as follows:

$GPRMC, 013946.00, A, 4351.101496, N, 12519.821271, E, 0.05, 218.30, 111105, 4.5, W, A*20.

The $GPRMC sentence is a string which contains 77 characters. The position of the each data segment in the sentence is fixed. The time information is at the 8th to 16th byte of the string, and the latitude information was at the 20th to 30th byte, the longitude information is at the 34th to 45th byte, the speed information is located at 49th to 52nd byte. The time, the position and the speed could be read at the corresponding place through the $GPRMC sentence.

3.5   Driving the Electric Control Valve

The automatic variable spraying controller outputtes a voltage to drive the electric control valve to control flux after getting these input signals above. The relationship

among the flux, the pressure in the pipeline, the control voltage of the electric control valve was as following [9]:

$$q = 10000 \bullet Q \bullet v \bullet d \tag{1}$$

$$V = 0.0139q - 3.226 \tag{2}$$

$$P = 0.2766V - 1.4756 \tag{3}$$

Here,   q-- Flux, kg/s
           Q-- Prescription, kg/hm^2
           v-- Velocity, m/s
           d-- Working breadth, m
           V-- Drive voltage of the electric control valve, V
           P-- Pressure, MPa

The electric control valve is driven by a 0~10V voltage. The I/O port on the S3C44B0X microprocessor only outputtes digit signal, therefore a D/A conversion chip—DAC0832 is used to convert the digital signals into the analog signals. In order to keep the driving signals within the range of 0~10V, a signal amplifying circuit was designed by using a μA741 and a 50K resistance after the DAC0832 chip. When the variable sprayer worked at speed of 4.2km/h, the automatic variable spraying controller gathered the spatial information of position and velocity form the GPRMC sentence and recognized the plot which the automatic variable sprayer worked in, read prescription data which were saved in NAND flash, gathered signals of the flux and the pressure of the pipeline. And then output control voltage to control the opening size of the electric control valve according to formula (1), (2) and (3). The examination data of the flux and the pressure were saved in a *.txt file in the NAND flash. The data gathered were shown on the screen in time while the automatic variable sprayer was working in the field as Fig.5.

Fig. 5. The display window

4   Results and Conclusion

1) Field experiments of the automatic variable spraying herbicide were carried out, which applied the method of data gathering and processing for the automatic variable spraying in 2009 and 2010. The sum of working area was coved 10hm^2. The results of field experiment showed that the maximum of relative error of the automatic variable sprayer, equipped with the system and method of data gathering and processing, was less than 5% when the dosage was 275kg/hm^2~600kg/hm^2.

2) The advantages of the method of data gathering and processing are as follows: convenient operation, simple structure and precision, and good for protecting environment at present.

3) The system is composed of 5 independent signal gathering and processing units. So it has the high reliability and stability.

References

[1]   Ma, J.-Y., Pan, Y.-C., Zhao, C.-J., Li, X., Wang, J.-D.: A Control System for Field Variable Spraying Based on GPS and GIS. Microcomputer Information (embedded & SOC), 85–87 (2006) (in Chinese)

[2]   Zhang, S.-H., Qi, J.-T., Liao, Z.-J., Xu, Y.: Research and application of control system for variable rate fertilizer applicator based on CPLD. Transactions of the CSAE, 200–204 (August 2010) (in Chinese)

[3]   Wang, J.-H., Fu, Z.-T., Wang, X., Qi, L.-J.: Design of Variable rate sprayer controller based on AT89C52. Microcomputer Information (Control and Automation Publication Group), 8–10 (2006) (in Chinese)

[4]   Zhai, C.-Y., Zhu, R.-X., Sui, S.-T., Xue, S.-P., Shangguan, Z.-P.: Design and experiment of control system of variable pesticide application machine hauled by tractor. Transactions of the CSAE, 105–109 (August 2009) (in Chinese)

[5]   Geng, X.-Y., Li, Y.-M., Miao, Y.-B., Liu, C.-L.: Development of variable rate fertilizer applicator based on GPRS. Transactions of the CSAE, 164–167 (November 2007) (in Chinese)

[6]   Shi, W.-P., Wang, X., Wang, X.-Z., Zhuang, W.-D., Wang, C.: Variable Rate Spraying Technology on the Basis of GPS and GIS. Journal of Agricultural Research, 19–21 (February 2007) (in Chinese)

[7]   Yang, Q., Pang, S.-J., Yang, C.-H., Li, M.-T., Li, Y.-J., Yang, S.-M.: Variable rate irrigation control system integrated with GPS and GIS. Transactions of the CSAE, 134–138 (2006)

[8]   Meng, Z.-J., Zhao, C.-J., Liu, H., Huang, W.-Q., Fu, W.-Q., Wang, X.: Development and performance assessment of map-based variable rate granule application system. Journal of Jiangsu University (Natural Science Edition), 338–342 (July 2009) (in Chinese)

[9]   Wang, L.-X., Zhang, S.-H., Ma, C.-L., Xu, Y., Qi, J.-T., Wang, W.: Design of variable spraying system based on ARM. Transactions of the CSAE, 113–118 (April 2010) (in Chinese)

[10] AgGPS124/132 operation manual, pp. 153–163. Trimble Navigation Ltd., Sunnyvale (2000)

Research of Monitoring System
for Dangerous Goods Posture

Xue Han

School of Computer Science and Engineering, Hunan University of Science and Technology
Xiangtan, China, 411201
hanmei11020011@126.com

Abstract. This paper presents the design and implementation of wireless sensor network system based on three-dimensional acceleration which is used for real-time, accurate detection of the status of dangerous cargo in railway transport, and the feedback of test results to the train relevant personnel, so as to help the relevant personnel grasp of cargo status in transit real-time. When it comes to some danger, such as cargo dumping or hit, the relevant personnel can take appropriate measures to prevent the risk of further occurrence on time.

Keywords: Dangerous cargo, monitoring, three-dimensional acceleration, wireless sensor networks.

1   Introduction

Wireless sensor networks[1] system is the frontier hot research field in international, involving highly cross-multidisciplinary and integrated knowledge, which combines a sensor technology, embedded computing technology, modern networks and wireless communication technology, distributed information processing technology, etc[2-3].

Rail transport is the main channel of dangerous cargo. This approach produces great vulnerability about monitoring dangerous cargo, and inevitable security accidents caused by human error. The technical effect of this paper is as follows: The acceleration sensor we used can measure the acceleration of three-dimensional orientation, but also monitor the bumping and hitting of cargo effectively; Using the CTP protocol to form Ad hoc network, combined with GPRS networks and networks based on RS-485 protocol, the data collected are uploaded to the second-level monitoring centers, in order to achieve effective monitoring of the status of dangerous cargo on the train.

2   Overall System Design

Wireless sensor monitoring network transmits data within internal nodes using the ZigBee protocol, and uses CTP protocol Ad hoc to build sensor network. A transport carriage can form a separate monitoring network, and all the monitoring network measurement data connect the external data transmission network through gateway.

First-level monitoring center makes connection with the aggregation node using the approach based on RS-485 bus, and the secondary control center connects the wireless sensor monitoring network using external data transmission network through gateway. System figure is shown in Fig.1.
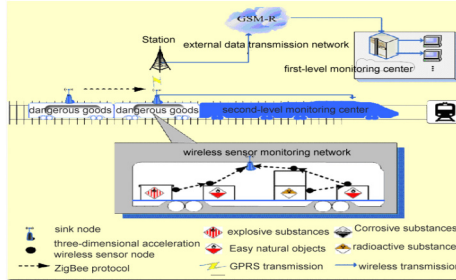
Fig. 1. The design of the whole monitoring system structure

Wireless sensor monitoring network, which is the core of the whole monitoring system, is responsible for cargo status data detection and status change detection, and report test results to the monitoring center real-time. Some sensor nodes are responsible for collecting three-dimensional speed data, and some are responsible for the sense of vibration and free fall. In the top and side of the box, at least three three-dimensional acceleration wireless sensor nodes need to be deployed and started, and in the compartment at the top, a sink node need to be deployed and started; The nodes which belong to the same box are divided into a group, and are assigned a specific group number, each group identified a node as the head node. All three-dimensional acceleration sensor nodes go into the initialization cycle after starting, each node obtains the initial sensing data of three-dimensional acceleration. If a single node is in stable status, the detection status will be initialized as the initial state. At the same time, the node will establish hierarchy tree wireless network through CTP protocol[3].

3   Introduction to Acceleration Sensor Network

3.1   Hardware Design

Wireless sensor nodes are used for collection and storage of cargo status data and inspection of cargo status, and send the status data and test results to sink node by wireless transmission. Its hardware structure is shown in Fig.2, the wireless sensor node includes a local storage module M25P80, three-dimensional acceleration sensor module MMA7455L, processor module MSP430F1611, wireless communication module CC2420 and the energy supply module [4].
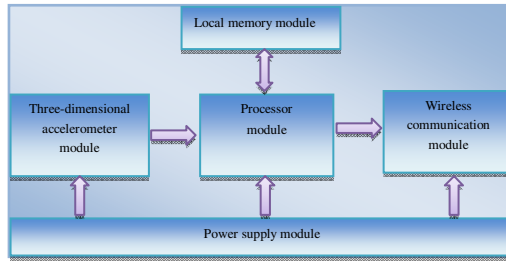
Fig. 2. The hardware structure of wireless acceleration sensor node

3.2   The Acceleration Sensor Chip MMA7455L

MMA7455L has four modes: the standard mode, measurement mode, the level monitoring mode and pulse monitoring mode, which can be selected by setting the mode control register's MODE0 and MODE1.

In order to better monitor the railway cargo status, we need to use the measurement mode of acceleration sensor MMA7455L. We can get rich acceleration data of X, Y, Z axis through this mode, which can be done data processing through the processor of sensor nodes or computer. As a result, we can calculate the object's current status, and even reasoning the cause of the accident. Some of the cargo transported on the train is fear of bumps, such as glass products, ceramic products, and explosive materials and so on. MMA7455L acceleration sensor's pulse monitoring mode can monitor the vibration of an object in a good way. As long as setting a reasonable time window and acceleration data thresholds, the vibration of the object can be sensed, and the processor can be informed by the form of break.

3.3   CTP Protocol Analysis

CTP（the Collection Tree Protocol）[5] is a tree-based sink protocol. Some nodes of the network are set to root, and the other nodes form the route of the root according to route gradient, through which the sink network to the root is formed. CTP has no address, and its nodes don't send data packets to fixed node, but by selecting parent node as the next hop, the root is selected implicitly, and the node form the root route according to route gradient. CTP provides the best multi-hop data transmission to the root. It has route choose mechanism for ensuring transmission reliability. But it does not make100% reliable, it do its best, and sometimes may not make it even try its best.

4   Design and Implementation of Gateway

The gateway module adopts the hardware architecture frameworks shown in Fig.3, mainly including processor module MSP430F1611, wireless communication module CC2420, power supply module, GPRS module SIM300 and RS-485 communication module. The wireless communication module CC2420 is mainly used for aggregating the sensing data of wireless sensor network. The GPRS module SIM300 is responsible for the data transmission of GPRS network, and is also responsible for the TCP/IP packets of the transferring data. The RS-485 module is used for achieving

wired communications with first-level monitoring center. The processor module MSP430F1611 is used for data protocol conversion and the control of other modules, so as to achieve transmission connection between two different networks.
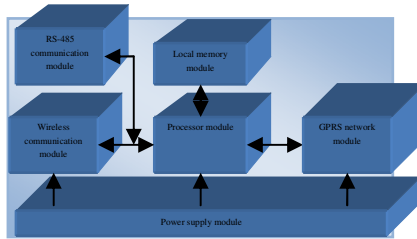
Fig. 3. The gateway module structure

5   Analysis of Three-Dimensional Acceleration Data

Firstly, we place 10 containers in the compartment, secondly, put three three-dimensional acceleration sensor node at the top and side of each container, then place a sink node at the top of the compartment , at last, deploy a wireless sensor networks consisting of 31 node equipment. Single sensor node communication range is less than 10 m, and need to go through multi-hop to transmit the data to the sink node. The node at the top of the container is the head node of the container.

5.1   Acceleration Analysis in Normal Status

We only analyze the collected data of nodes on the top of the container, because it is broadly representative, the collected data of nodes in other parts is only different in all directions. Acceleration sensors, has 4-G selected range, 8-bit data collection. As a result, when the acceleration is 1G, the number of collected data is 31. In container normal status, the collected data of a sample cycle in X, Y, Z axis are shown as Fig. 4. It can be seen from the figure, with train under uniform, the data of X-axis and Y-axis data in the table tend to zero at the top of the cargo, indicating that there is no acceleration in this direction; The data of the Z-axis in the table tend to 31, indicating that the acceleration of the Z-axis direction is 1G, so the container is in a stable and normal state and the bottom of the container is in the Z axis.
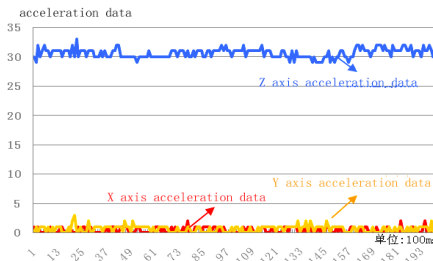
Fig. 4. The 3d acceleration distribution as cargo is stable

5.2   Acceleration Analysis of Container Overturned

We still only analyze the collected data of nodes on the top of the container. When the container overturns, before and after container overturned status the collected data of a sample cycle in X,Y,Z axis are shown as Fig. 5. As can be seen from the graph, the Z axis direction is 1G at first, and the two others are 0G, which means that cargo still in stable normal status. Then the acceleration of Y axis direction rising gradually, the acceleration of Z axis falling gradually, while the acceleration of X axis is still 0, it means cargo is overturning in Y axis. At about 121ms in the sample cycle, the cargo' acceleration in 3d begin to fluctuate acutely. At the moment into volatility, the acceleration of Y axis is approaching to 1G. Obviously it is a vibration, which results from the cargo turn to landing at 90 degree in Y axis. During the vibration, the acceleration of 3d direction has had a tremendous change, with the Y direction by the largest amount, maximum achieving "4G". After the vibration, the acceleration of X axis and Y axis turn to 0, while the acceleration of Y axis turns to 1G, it shows cargo have already completed its turnover. From stationary status to turn to touchdown moment, the acceleration of Y axis and Z axis change slowly and that means the turnover process isn't violent.
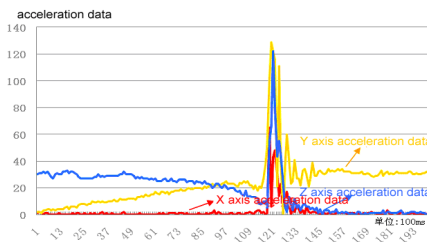
Fig. 5. The 3d acceleration distribution as goods is overturned

5.3   Acceleration Analysis from Shaking to Overturn

We still only analysis data collected on the top of the container. The status data before and after the container overturned, is sampled in a cycle in X, Y, Z axis shown as Fig.6. We can see from the figure, the status after overturned and the overturning moment of the ground, three-dimensional data curves distribution are basically the
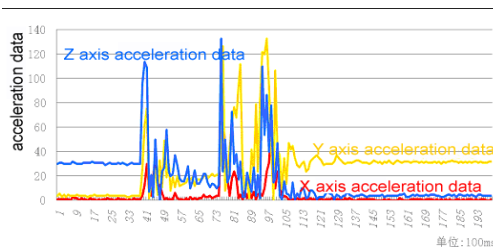
Fig. 6. The 3d acceleration distribution as cargo is hitted

same with the previous section. The overturned process of the previous section is relatively relaxed, but in this section the three-dimensional acceleration data changes drasticly, acceleration in Z axis and Y axis even more than 4G range, indicating that the container shakes violent in the Y axis direction and finally lead to the overturned container.

6   Summaries

It can be seen from this paper, cargo status monitoring system collects acceleration data in three directions of cargo accurately. By analyzing acceleration data sampled from a single node, the current status of cargo can be better detected, which play a role in sensing the cargo status. As acceleration is a vector, if three-dimensional acceleration data of the other two nodes is used, we can monitor the cargo status further through the vector synthesis.

References

[1] Sun, L.-M., Li, J.-Z., et al.: Wireless sensor network. Tsinghua University Press, Beijing (2005)
[2] Li, J.-Z.: Preface of special issue on wireless sensor network. Journal of Software 18(5), 1077–1079 (2007)
[3] Shen, B., Zhang, S.-Y., Zhong, Y.-P.: Cluster-Based Routing Protocols for Wireless Sensor Networks. Journal of Software 17(3), 1588–1600 (2006)
[4] Tang, Y., Zhou, M.-T., Zhang, X.: Evolution of Research of Wireless Sensor Network Routing Protocols. Journal of Software 17(3), 410–421 (2006)
[5] Zhao, Q.-L., Jiang, Y.-H., Xu, M.: Analysis and Comparison of Routing Protocols for Wireless Sensor Networks. Computer Science 2(36), 35–41 (2009)

The Application of Ontology Building Based on the Form Background Feedback Mechanism of Concept Lattice

Yan Hongcan and Wang Jian

Science College,
Hebei United University,
Tangshan, Hebei Province, China City, 063000
yanhongcan@heuu.edu.cn, qwertry163@163.com

Abstract. Ontology is a key technology of the realization of semantic web. This paper adding a feedback mechanism of form background of concept lattice to existing ontology building based on FCA, domain experts could according to the similarity of individuals to evaluate the a undetermined form background could be utilized or not. At the same time it has solve the problem of how to mix the logic components owl: same As and owl: different From of OWL and FCA. The purpose of paper is to solve the problem of increase the recall level and the recall precision of ontology which is built based on FCA.

Keywords: Concept lattice, ontology, similarity of concept, feedback mechanism of form background.

1 Introduction

Ontology is a key link of the realization of semantic web,and is the turning point from the description to reasoning of knowledgement[1]. So the building of ontology occupies a basic and important position. Although computer experts and domain experts has built many ontology also has many difficulties.

Althology the development level of the tools of ontology engineering has been impoved at certain degree, the building of ontology is also a trival and arduous task, it could cause the bottleneck os knowledgement acquisition, and it seriously influenced the development of semantic web. There has not been a mature and complete theory as a guidance for the methology of ontology-building, and the current ontology-building methods are always for concrete projects. It is difficult to stay the same standsrd of ontology-building for different ones, and it is not favour for the scale and standardized building of ontology, because of the lack of unified theory. The methods of ontology-building should be researched in-depth.

At present the international recognition of the methods of ontology building. Some of them are Skeletal Method, Enterprise Modeling Method, Methontology, KACTUS, Circlation Access Method, IDEF-5, Seven-feet Method[2].

This paper will use seven-feet method for ontology-building as the core connotation, and the purpose of it is the ontology-building by introducing FCA, and creating the feedback mechanism of form backgroud of concept lattice for solving the

problem of the mixing together os concept lattice and the logic components of ontology, the building of experimental ontology about medicinal plants.

2   From FCA to OWL

2.1   The Introduction of FCA

Firstly FCA is put forward by Wille, a German professor[3], and it is used for the discovery, displaying and squencing of concepts.In FCA the extendion of concept is considered the set of all objects of it and the connotation is the set of all attributes for such objects. So it could ralize the formalization of the concept, all the concepts and their eneralization and instantiated relationships constitute a concept lattice.

2.2   The Introduction of OWL

OWL (Web Ontology Language) is the standard of ontology description language which is recommended by W3C(World Wide Web Consortium). OWL is designed for all aspects of need, it is not only compatiblewith previous ontology has more powerful ability of semantic expression, and could ensure the decideable standards of DL (Description Logics), etc. The fundamental elements of OWL are Class, Property and individual:

(1) Class:*Class* defines a set of individuals which are sharing some attributes, *rdfs:subClass* could be uesd for making different classes as specific hierarchical structure.

(2) Property: *Property* could be used for stating the relationship of individuals or a individual and its property value. The Property could have its subProperty. The *rdfs:domain* of *Property* restricts the classes of individuals which has such properties. The *rdfs:range* restricts the value of the individuals.

(3) Individual:*Individual* is the example of class. It could use properties to connect a individual to another one.

2.3   The Extension of OWL Logic Components

The literature[13] defined the corresponding relation of concept lattice and the logic components of OWL (*rdfs:subClass,rdfs:domain* and *rdfs:type*), it preliminarily solves the problem of form concept lattice to the preliminary logic component of OWL.But there are many logic relations in practical application have been beyond the range of [13], such as the *owl:sameAs* and *owl:differentFrom*, which are very important for the later reasoning.So how to mix the two logic components and concept lattice together is the problem that this paper want to solve. The problrm of such two logic components is:

(1) owl:sameAs:Because of the vast regions in China, a object has different names in different regions in China,a object has different names indiffernt regions, for example the 徐长卿 (Radix Cynanchi Paniculati, Dicotyledoneae Gentlanales Asclepladaceae) also is called 一枝香 in China, if only distinguish such two names by string, they are irrenlent, but they are representative the same thing actually.

(2)owl:differentFrom:If the 徐长卿 and 马利筋 are differentiate only by properties 直立草本,叶对生 and 夏季开花, it shows that such two objects are representative the same thing, but they are not actually.

It could create a mechanism of form background to solve the problem. The purpose of this mechanism is to compute the correct number of properties to differentiate concepts (we could not infinitely increase the number of property, because of our finite compute resource).

It could use the known relations of concepts to reverse form backgrounds. We could use examples which is mentioned above to create a form background K_1 which is shown as Tab.1.

Table 1. Form background K_1

R	直立草本	叶对生	夏季开花
马利筋	×	×	×
徐长卿	×	×	×
一枝香	×	×	×

If we add two properties 祛风湿and 全株有毒, the new form background K_2 is shown as Tab.2.

Table 2. Form background K_2

R	直立草本	叶对生	夏季开花	祛风湿	全株有毒
马利筋	×	×	×		×
徐长卿	×	×	×	×	
一枝香	×	×	×	×	

Based on [13], we could add the mechanism of form background for the classification and clustering of objects. The mentioned plants in above example have a big disparity in utility layer. So it could be effectively differentiate by adding a couple of attributes. It should add many properties, if you want to differentiate plants in subspecies layer. And if the objects always have the same properties, so we could ensure that they are representative the same thing.

In addition the ontology which is for different domains, the object in that would be classified in different classes. For example, the research of phytology has many application aspects, such as phyyoecology and phytogeography. A kind of plants could be classified in different classes of different utility of ontologies, because of its utility of properties. So one otology should be discussed by computer experts and domain expert about its utility first of all.

2.4   The Calculation of Similarity of Concepts Based on the Mechanism of Form Background

In the process of actual search by computer, how to ensure the logic components *owl:sameAs* and *owl:differentFrom* is the key problem. So it is necessary import the

calculation of similarity of concepts in the process of building the otology. Actually the calculation of similarity of concepts is throughout the building and application of ontology. The search result is acquired by the similarity of concepts which is measured by a threshold. When the literature[5] computing the similarity of properties, it not only considers the similarity of the same property, but also some degree of similarity of different properties, and it is not correct assumption in a unified knowledgement background. It is improved by literature[6] and has gotten a formula of similarity of concept which is decided by individuals and properties corporately. But the formula which is given by literature[6] is only applicable for static formal background, this paper will create formulas of the similarity of individuals and the similarity of concepts with the dynamic formal background based on literature[6].

(1)The formula of similarity of individuals

This creation of formula is the key step in the feedback mechanism of form background. The paper refers literature[6] to create such a formula:

Definition 1. In such a form background (O, A, R), the similarity of two individuals e_1 and e_1 of that form background is defined as:

$$sim(e_1, e_2) = \frac{2 \times |I_1 \cap I_2|}{|I_1| + |I_2|} \tag{1}$$

The I_1 and I_2 is the set of relative properties for individual e_1 and e_2. If there are m individuals in this form background, we could create a $m \times m$ matrix of similarity of individuals by this formula. Domain experts could judge the superior and inferior of form background by the matrix, to wit domain experts will judge individuals is the same or not by their domain knowledge.

(2) The formula of similarity of concepts

If a form background is accepted by domain experts, it could construct a concept lattice by some algorithms[8]. And the similarity of concept could be computed by formula of similarity of individuals, it will be used for the description and reasoning of the concepts in domain ontology.

Definition 2. Two concept nodes (E_1, I_1) and (E_2, I_2) of concept lattice, and $E_1 = \{e_{11}, e_{12}, \cdots, e_{1m}\}$, $E_2 = \{e_{21}, e_{22}, \cdots, e_{2m}\}$, the formula of similarity of concept is defined as:

$$Sim((E_1, I_1), (E_2, I_2)) = \frac{1}{2} \frac{\sum_{i=1}^{n} \sum_{j=1}^{m} sim(e_{1i}, e_{2j})}{|E_1| + |E_2| - |E_1 \cap E_2|}, (e_{1i} \neq e_{2j}) \tag{2}$$

By this formula, we would have a quantization standard for the *owl:sameAs* and *owl:differentFrom.*

3   The Construction of Ontology Based on FCA

This paper provides a improved system of ontology building based on FCA.This system is semi-automatic.It needs the attend of domain experts in the first stage of this

system.The second stage could be automatic by some algorithms.And the system is as shown in Fig.1.

(1)The first stage of this system

The process of the first stage shows below:

1 Ensuring the scope of application of ontology;

2 Ensuring relative individuals and properties, creating the set of individuals and properties;

3 Choosing a part of properties in the set of properties and the set of individuals to construct a quarantine form background.

4 Acquiring the matrix of similaity of individuals from this form background by the formula(1).

5 The matrix will be evaluated by domain experts, if domain experts do not reply,return to 2 and adding some properties to construct a new form bavkground;in contrast it will go to the next step.

(2) The second stage of the system

1 From the form backgroun to concept lattice

The building of congcept lattice is the permise of the application of FCA,so the problem of the generation of concept lattice is a focal point in the application of FCA.After domestic and foreign scholars' research about that,there are some effective algorithms for generating the concept lattice.These algorithms fall into two categories: the Batch Algorithm and Incremental Algorithm. The Batch Algorithm includes Ganter Algorithm[9], Chein Algorithm[10], Bordat Algorithm[11]. The type of Incremental Algorithm is Godin Algorithm[12].It is usual algorithm for the application of generation of concept lattice. This paper'focal point is the building of ontology, so it will not be talk more.

2 From concept lattice to OWL

The literature[13] provides a algorithm of the transform from concept to OWLand it do not involve the similarity of classes(concepts),so it is necessary to add the value to evaluate the relationship of classes is *owl:sameAs* or *owl:differentFrom*.
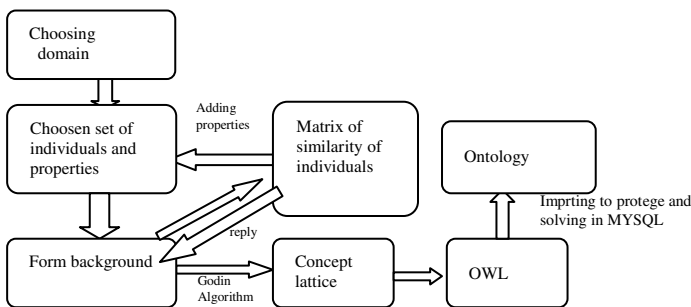
Fig. 1. The system of building of ontology which is adding the feedback mechanism of form lattice

4   The Experimentation and Simulation

Based on the simple example previously,we could construct a ontology of the medicinal of Asclepiadaceae, the final concept lattice shows in Fig.2. Because of the special requirement of nomination in Protégé software,so this paper creates a regulation to show the concept node:

(1) Using 1-12 for 12 individuals ,and 1-11 for 11properties in the talking domain.
(2)If the number is continuous,it will be representative by sign -,for example,the mumbers 2,3,4 could be representative with 2-4; If the numbers are not continuous, we use x to separate numbers,for example the numbers 4,7,9 could be representative 4x7x9.
(3)Using _ to separate the set of individuals and properties,for example,a concept node{E,I}={{1,2,3,6},{4,5}} could be representative as class_1-3x6_4x5.
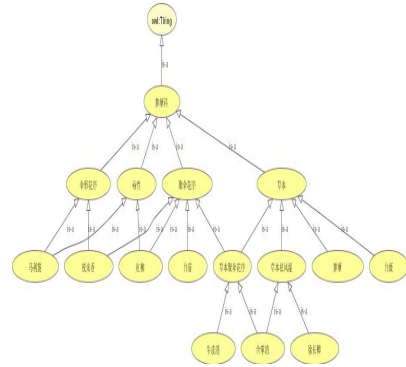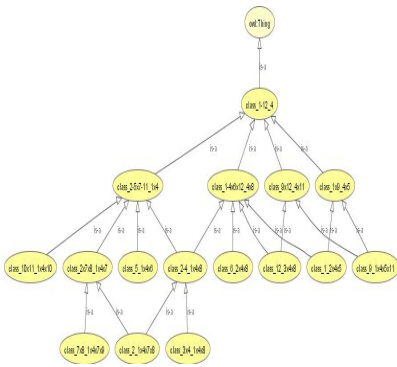
Fig. 2. The primitive Hasse figure of lattice   **Fig. 3.** The modified Hasse figure concept lattice

After the tagging of the concept nodes, the ontology has been constructed preliminarily, as shown in Fig.3. The result shows that the similarity of two concept nodes徐长卿and 合掌消 is 0.667, if the threshold is 0.5, those concept nodes could be considered as the same.

References

[1]   Song, W., Zhang, M.: A First Step Towards the Semantic Web. Higher Education Press (2004)
[2]   Noy, N.F., Mcguinness, D.L.: Ontology Development: a Guide to Creating Your First   Ontology   (EB/OL)   (2002),   http://protege/Stanford.edu/ publicatiolls//onto1.ogy-development/ontology101.pdf
[3]   Wille, R.: Restructuring lattice theory:an approach based on hierarchies of concepts. In: Rival, I. (ed.) Ordered Sets, pp. 445–470. Reidel, Dordrecht (1982)
[4]   Hu, H., Liu, D.-Y., Wang, S.-S.: Web Ontology Languange. Computer Engineering 6, 345–348 (2004)

[5]   Formica, A.: Ontology based concept similarity in Formal Concept Analysis. Information Science 176, 2624–2641 (2006)
[6]   Zhi, H.-L., Zhi, D.-J., Liu, Z.-T.: Concept Similarity Based on Concept Lattice. Computer Science 9, 156–157 (2008)
[7]   Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundation. Springer, New York (1999)
[8]   Li, T.: The Research of Ontology Model and Relative Algorthm Based on Concept Lattice
[9]   Ganter, B., Wille, R.: Formal Concept Analysis: Mathematical Foundations. Springer, Heidelberg (1999)
[10] Chein, M.: Algorithme de recherche des sous matrices premieres D'une matrice. Bull. Math. R.S. 13, 21–25
[11] Bordat, J.: Calcul pratique du treillis de galois d'une Correspondence. Mathematique, Informatique et Science Humaines, 31–47 (1986)
[12] Valtchev, Missaoui, R., Godin, R., Meridji, M.: Generating frequent itemsets incrementally: two novel approaches based on Galois 1attice theory. Journal of Experimental and Theoretical Artificial Intelligence 14, 115–142
[13] Cai, R.-T.: The Research of Automatica building of OWL Ontology from Text. Artificial Intelligence and Recognition Technology 2, 39–41 (2010)

Designation of Adult Education Administration System Architecture Design Based on C/S and B/S Mix Pattern

Yang Yuxin[1], Wang Jianliang[2], and Liu Azhu[3]

[1] Qinhuangdao Institute of Technology
Qinhuangdao, Hebei Province 066001, China
yangyuxin19630907@126.com
[2] Sandong Transport Vocational College
Weifang, China 261206
sdwfwjl@gmail.com
[3] Zhejiang University
Hangzhou, Zhejiang Province 310027, China
ReddyLao@gmail.com

Abstract. This architecture uses the construction which C/S and B/S unify. Using the connection pond visit technology, enhancing the database fully the visit efficiency, which improves the Web application, thus the reduced system expenses, raise the entire Web application system's operating efficiency. This architecture unifies the adult education administration management practical work and the systems engineering and the software engineering thought which based on this pattern under function module design. Optimizes and integrates the educational administration managerial data, maximum limit realizes data sharing. Make modern adult education administration work more scientific, networking and information-based construction.

Keywords: Connection pond, distributional buffer, business processing, data transmission, C/S pattern, B/S pattern.

1   Background

Nowadays technique of science is being developed and applied widely, the level of the talents people required by all kinds of vocation becomes higher and higher. As an important part of the higher education, adult education makes contribution to improving the culture level of the civil, setting up the life time education system and promoting the development of economy.

2   Structure

2.1   Structure Analysis

The present adult education system is merely divided into two kinds. One is C/S pattern, it consists of one or more clients, one or more servers, a under operating system and network. It is a distributive computing and compounding system which is

the main pattern of the traditional information administration. The other one is the B/S pattern. It bases on the web technique, thus makes it the main structure of current system. server turns the results into the instructions which can be identified by the client browser and returns it to the client.

2.1.1   B/S and C/S Mixed Pattern

Based on the case study in reality, the Adult Education Administration System is design in the B/S and C/S mix pattern. It gathers the advantages of both patterns. N-Tier Mode, open system, interconnection across different net and fast update of B/S. High interactivity, reliability and transaction handle ability of C/S. The mixed pattern includes all the advantage points of the two patterns to achieve a better system.

3   Main Technique

3.1   Database Connection Pond

The technique of database is becoming more and more important in the realm of information. Especially for the application of network and electronic commerce, the dynamic websites of them are all supported by the database technique. Database connection pond in JDBC is included to improve the system performance, big cost of the system since frequent database construction and destruction.

3.1.1   Theory of the Database Connection Pond

Set up a buffer pond for database. Store limited connections in the pong beforehand. When there need to make up a connection, take out one from the pond and return it after it is done. In order to meet the need of convenient application, the database connection is encapsulated which separate the handle logic and access logic.

The problem is what is the number of the connections in the pong will achieve the optimal performance. Thus the max and min connections are definited. The amount of min connections is that of connections when the pond is set up. If it is larger, the start of system will become slower but the respond time will decrease after startup. It is opposite if the amount is smaller. In reality, the speed lows down because there is not enough connections.

3.1.2   Techniques of the Database Connection Pond

3.1.2.1   Transaction Handle. Transaction Handle demands the operation of database satisfy the principle of "All-All-Nothing". Each transaction occupies a connection. The complexity of transaction handle falls down. A registry is generated when connection management announces that a transaction starts or ends. It separates the connections management and transaction handle. The registry is generated based on the actual condition. It can not be reused. Transaction users are distinguished by threads. Then the access of the database is through the query of the transaction registries. When the transaction is handled, the registry will be deleted from the list.

3.1.2.2   Encapsulation. Database handling is different from transaction handling. In order to gain a better application, operating interfaces are used to encapsulate the links. The ordinary links and transaction handling links use the polymorph of the oriented objects. They all implement interfaces according to their own characteristics.

3.1.2.3   Concurrency. Under the condition of multithreading, the connection management data must agree with connection inside data. Since Java provides functions to manage concurrency, key word synchronized can ensure the threads are synchronized and it is safe.

3.2   Technique of Cache

3.2.1   Distributive Cache Mechanism
Multithreading model is applied in distributive cache. Three kinds of threads service, main thread is in charge of the starting and maintenance of cache server, thread of service is in charge of receiving the requirements and accomplish them according to their types, cache management thread is in charge of the management of the local caches, are applied. The collaboration of the all the servers makes up the distributive mechanism.

3.2.2   Optimization
The core of the architecture, management of Java expends the function of the server to implement the data management. It consists of collection of the data, control of the alarm, data analysis, data access and interface configuration. These functions are encapsulated in the Mbean Module. It can be initialized through Agent inside the program or through web pages. Client applications run on the clients providing friendly interface. Data gained from JMX is presented on the web pages. Thus implement the corporation of the server and client.

In order to avoid the time-delay problems when clients collect the data, data is stored in the cache until it will be not used any more. Distributive cache is applied to manage the memory dynamically to promote the performance and stability of the architecture.

3.3   Technique of Data Transmission

The technique of the data transmission is applied to promote the running efficiency, reliability and safety. The data organization and data transmission are based on the Web Service.

3.3.1   Data Organization
Web Service is service oriented. The applications on the distributive servers on the Internet are integrated. Thus computers on different districts can work in collaboration to ensure the user all kinds of services.

First, WSDL interface is defined including the interface file and implementation file. Then register on UDDI Register Center. Data is organized according the format definite by the service interface file and the matching proposal. Users search for the services they require through the browser or client, position on the relevant WSDL

files, generate the interface on the client according to the description of the WSDL files and then accomplish the web data requirements through the interface. Web Service receives the requirements from the client, and then sends the data to client in XML through SOAP. At last clients complete the presentation.

3.3.2   Design of Data Transmission

Winsock helps to realize data transmission in C/S pattern. Client proposes requirements to server. After server receives the requirements, it provides the relevant services for the clients according to the TCP/IP protocol.

4   Design of Function Module

The departments of the administration are made up of educational administration, school rolls administration, teaching material administration and so on. Transaction Flow is divided into four parts.

Plan Establishment. Establish the enrollment and teaching plan according to possible enrollment student amount, the requirements from each district and education condition. Correspondence School Department can be set up in fit districts.

Data Collection. After the enrollment, the information of the new students, such as candidate number, name, sex, date of birth, ID, telephone, zip code and address, will be input into the system.

Data Handle. Set up students' files. Distribute students' number for the new students. Establish classes. Hand in the information of the new students. According to the teaching plan and schedule of each specialty, the courses are decided. Students select the fit courses. Teachers hand in the grades of the students in the limited time.

Data Summarization. Students get all the credits in the limited time. After certificated by all the departments, the students are allowed to graduate. Attach description should be added to record the paper and awards.

Most Education Administration Systems are based on C/S pattern which are beneath the firewall. But if the students want to query the grades, the remote access must across the firewall. Thus the B/S pattern is useful.

References

[1] Zou, Y.: Preliminary discussion on the information construction of colleges and universities. Literary and Pedagogic (07), 22–23 (2008)
[2] Liu, Z., Qiu, J.: Study and design of collectivity frame of figure campus. Computer Engineering and Design (29), 33–59 (2008)
[3] Zhai, X., Huang, K., Xu, X.: From the business architecture to the data architecture of Chinese universities. China Education Info. (07), 89–96 (2007)
[4] Shi, M.: On information construction of colleges and universities. China Higher Education Research (02), 23–29 (2007)
[5] Wu, H., Miao, C., Liu, Q., et al.: The research of web application security testing. Computer Security (16), 55–59 (2008)

The Design of the Control System of Navigation for UUV in Underwater Information Warfare

Lin Wei, Yuan Bingcheng, and Wu Pengfei

Dept. of Weaponry Eng.
Naval Univ. of Engineering
Wuhan, China
`Wish_monkey@163.com`

Abstract. This paper advances that the sailing control system based on data bus can be used in UUV, and the characteristics between 1553B bus and CAN bus are analyzed. Through the comparison of the two data bus we can get the conclusion that the CAN data bus is more appropriate in UUV. Their performance is simulated, and the result shows that CAN bus is a very well bus technology.

Keywords: 1553B data bus, CAN data bus, unmanned underwater vehicle (UUV), sailing control system.

1   Introduction

The advanced sailing control system is heading for the open, the use of COTS products takes the exchange of information as the centre, highly integrated and modular. It has gone through the mechanical hydraulic control modal, full-limit digital electrical control mode and based on data bus control mode. Based on data bus has been applied to other instances of large aircraft successfully, so the sailing system based on data bus control mode can be applied to UUV, the system structure shown in Fig.1.

The basic principles of the sailing control system that based on data bus UUV is the use of data bus to connect the intelligent measuring and control equipment of control sites at all UUV as a network integration, in order to achieve data transmission and sharing, and format various forms of automatic control subsystems to adapt to the actual needs, and improve the level of the intelligent, modularization and informatization of the UUV sailing management, in order to achieve reliable UUV sailing control.

The data bus is the link of field bus control system, the specific choice of what kind of data bus according to system's needs and the performance of the data bus. At present the more application of aviation are ARINC429 and MIL-STD-1553B.The former is widely used by civil aviation, it has two transmission rate, 12kbps and 100kbps, the bandwidth is limited, also a single source / multi-receiver traditional system, and its work is half-duplex, when the system is large-scale, its large data links will be very large and it is not conducive to saving space, are being replaced by ARINC629. While the MIL-STD-1553B is the military standard bus, it is a bus-type connection, additional equipment is relatively simple and very suitable for centralized

management of distributed processing systems. CAN bus is first applied to the automotive industry, its good performance is also applied by some aircraft. In this paper, the choice of data bus in UUV sailing control systems is mainly a choice between the MIL-STD-1553B and CAN bus.

2   Background

2.1   1553B Bus Communication Technology

MIL-STD-1553B is a bus that was developed in the late 70s the U.S. by American Army, and also the DAIS (Digital Avionics Information System) referred to the data bus. It is used widely in the aviation field. It also has been recognized by many countries and industries and formats the international standards, So it is widely used, not only in aircraft, but also tanks, ships, missiles, satellites and others have applied. It is a special bus, and has high price.

1553B bus communication systems are generally composed of several subsystems by the embedded bus interface and interconnected via the bus from the media, the various systems operate independently, resources and capabilities can be shared through the network [1-3].

1553B has 3 terminals, which's structure is shown in Fig.1:
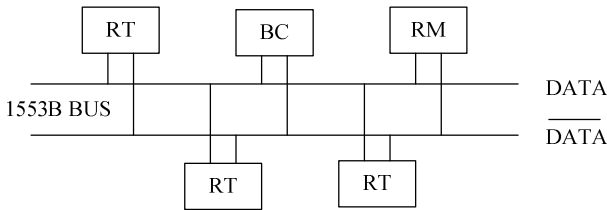
Fig. 1. The composition of the 1553B bus communication system

(1) Bus controller (BC): for scheduling management information transmission on the bus, there is only one bus controller on the bus at any time, however, it is the backup bus controller.

(2) Remote terminal (RT): the 1553B bus allows up 31 terminals to attached, each RT terminal is assigned a unique bus address, do not have the bus control function, but can be used as a backup bus controller.

(3) Bus Monitor (BM): the 1553B bus can have a bus monitor, it does not respond to any command of bus controller, but it can be used for monitoring and extracting the bus data for data analysis later.

The communication process is as follows: If a remote terminal send data to the bus controller , first, BC send a transfer command to RT, after a Short time interval, RT in response to it by the status word which receives the command, then send one or more (up to 32) data words to BC. In this case, it takes about 70 microseconds to send a data word from RT to BC. While the sequence of the data's transmission between RT is: first, BC sends a receive command to the receive RT, then sends a transfer

command to the send RT again. Send RT sending a status word first, that is ,transmit one or more (up to 32 data) data words to receive RT, when RT receives the data, BC will send a status word, thus ending the long-range data transfer between RT. Sending a data word from a RT to another RT takes about 120 microseconds, and therefore its cost higher. Of course, if the continuous transmission of 32 data words, the same as the system's overhead, so the bus will lower the percentage of time of information transmission [2].

2.2   CAN Bus Communication Technology

Late 70s, the use of data bus technology for distributed control field devices become a trend. CAN bus is one of the more extensive application of a field bus, which was designed originally by the German company Bosch for its automotive control system. The design is based on a simplified network communication to achieve high-speed signal transmission, and focus on reducing the wirings of the hardware circuit, improving the cost-effective and ensuring the reliability of the data's transmission in adverse electromagnetic conditions. Practice has proved that it has high reliability, strong anti-interference ability. Its flexible expansion capability, faster transmission speeds and more compact connection have been applied in many fields, and become the international standard [3].

In order to achieve openness, CAN bus system also refer to the ISO / OSI seven-layer communications model in the communication model, but only the data link layer, physical layer and application layer are used. Communication model shown in Fig. 2.
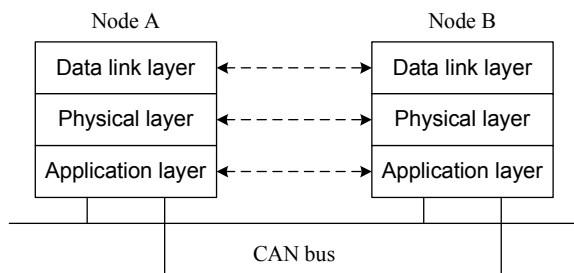
Fig. 2. CAN bus communication model

According to the CAN protocol ISO11898, it only regulates the data link layer protocols and physical layer. The hierarchical structure is shown in Fig.3 [4].

(1) Data link layer. Data link layer also subdivided into Logical Link Control sublayer (LLC) and media access control sublayer (MAC).

(2) Physical layer. Physical layer defines mode by which signals are transmitted. It's mainly about connect media, circuit electrical characteristics, data encoding / decoding, bit timing and synchronization of the description. CAN criterion does not define the drive / receiver characteristics of the physical layer, thus allowing to optimize the transmission medium and signal level's implementation depend on the specific applications.

(3) Application layer. The criteria in the CAN does not provide the relevant information about the application layer in real applications, such as the identifiers' allocation, network node status and other application layer protocols require the actual system customization. The appropriate application layer protocol for system reliability, real-time has great help. CAN bus in the application process, producing a lot of application layer protocols such as CANopen, iCAN, Devicenet, SDS and so on.
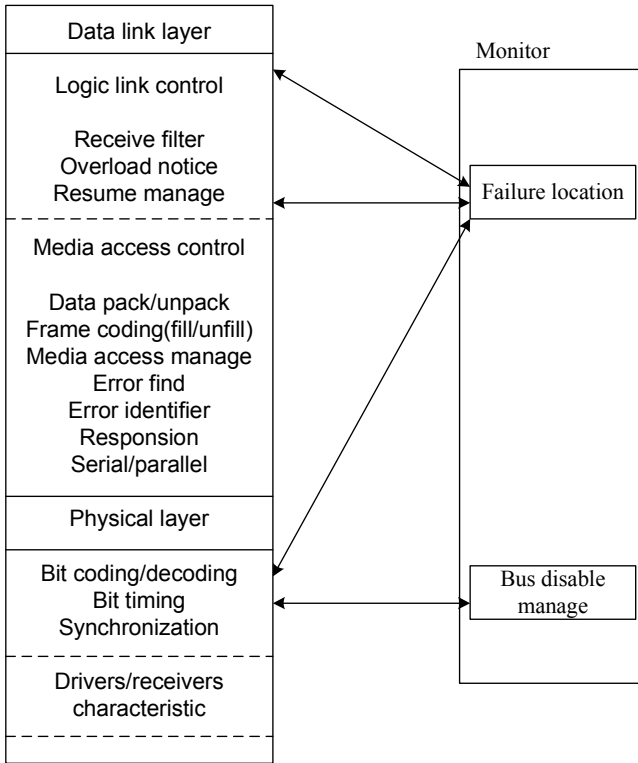
Fig. 3. The hierarchical structure of CAN-bus data link layer and physical layer

3 The Communication Capability Analysis between 1553B Bus and CAN Bus

By analysis between MIL-STD-1553B bus's and CAN bus's theories, 1553B bus and CAN bus system both can realize distributed processing, centralized control and real-time response; and have high reliability, maintainability, anti-jamming capability; The differences is mainly in the following points:

(1) Station address code was abolished on CAN bus, which adopt encoding blocks of communications data. So CAN bus can be linked to numerous end-node. In fact, due to electrical effects, it can be linked to 110 nodes. The 1553B network can attach 32 terminals;

(2) Communication mechanism is different. 1553B bus uses the master-slave asynchronous communication by command / response, all information's transmission on the bus are all controlled by commands issued from BC; and CAN bus's communication is more flexible and can be master-slave and multi-master type communications that does not require a node to schedule.

If the system's scheduling policy is not proper, the system's real time and reliability will be poor in Master-slave communication, and it is easy to impede the normal and effective communication. Multi-master communication type haven't scheduling, but need to avoid bus conflicts. CAN bus, which uses arbitration with a non-destructive technique (CSMA / CA), is similar to the Ethernet CSMA / CD' s solutions, but not all the same. If a conflict is detected in Ethernet nodes, all nodes are back to "monitoring" State. The CAN bus is based on the priority of the packet's identifiers for non-destructive arbitration. When faced with conflict, high priority packets can continue to send. Packets are not all need to stop transmitting.

CSMA / CD (Ethernet) and CSMA / CA (CAN bus) media access control protocol channel throughput expression on the MATLAB simulation results shown in Fig.4 and Fig.5.

$$S_{CSMA/CD} = \frac{aGe^{-aG}}{aGe^{-aG} + a(1 - e^{-aG} - aGe^{-aG}) + a(2 - e^{-aG} - aGe^{-aG})} \tag{1}$$

$$S_{CSMA/CA} = \frac{1 - e^{-aG}}{1 - e^{-aG} + a} \tag{2}$$

In the expression (1) and (2), a is the transmission delay, G is the channel traffic. According to the simulation, we can see, when the transmission delay and traffic gradually increases, CSMA / CD to control the effects of conflict visibly decreasing, while the CAN bus, CSMA / CA's non-destructive arbitration is able to maintain high channel utilization Rate. Visibly, CAN bus will play a very good effect in multi-master communications.

Under certain conditions, the 1553B bus and CAN bus system is equivalent to the M / D / I system (fixed line service time Poisson distribution system), the use of queuing theory simulation of the 1553B bus and CAN bus system response time changes, the conclusion is: the 1553B bus make a longer response time because of the Master-slave communication, and with the increase in the number of nodes ,the time increased significantly; while CAN uses CSMA/CA can solve the bus conflict better, its response time in the same Stations can be smaller [5].

(3) 1553B bus is special and expensive, but the CAN bus products are cheaper and easier to obtain. Although 1553B's rate has reached 50M/s or more, while the CAN bus is only 1M / s, but the CAN bus's communication is more efficient, Moreover, control information is generally small., the system's communication rate can be met as long as the number of nodes of the system is reasonable. What's more, the continuous development with the application, the rate of CAN bus will be the same as the 1553B from 1M / s to high rate.
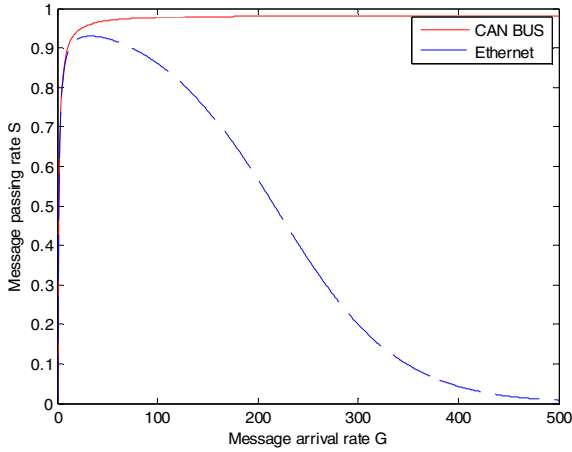
Fig. 4. The S-G simulation graph of CSMA/CD and CSMA when a=0.02
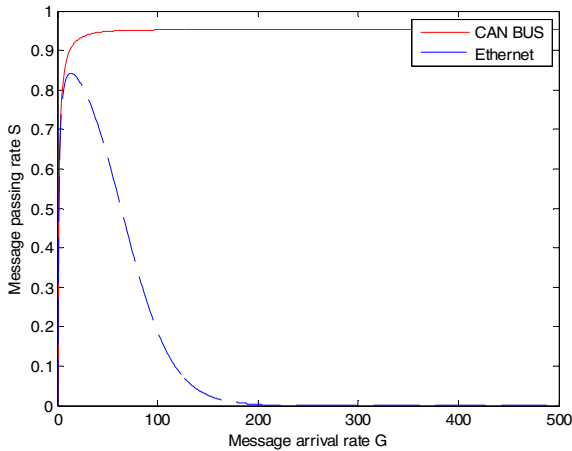
Fig. 5. The S-G simulation graph of CSMA/CD and CSMA when a=0.05

In addition, compared with the MIL-STD-1553B bus, CAN bus has a better performance in real-time characteristic and reliability. Sum up the analysis, CAN as the system's data bus has certain advantages. Therefore, in the study UUV sailing control system, we can consider the CAN bus as the data bus.

4   The Design of Sailing Control System for UUV Based on CAN Bus

Modular integrated electronic technology is used in the new UUV sailing control system, which will be open, standardized, comprehensive, electrification. It will

adopts a number of integrated modules and electronic control means to achieve the UUV sailing safety. These modules will be integrated with micro-processing chips, memory, power, signal processing, and some management sensors or some interface of other equipment, you can perform the appropriate tasks. By CAN data bus technology, you can connect them as a whole in order to achieve a integrative sailing control system which center is exchange of information.

The system does not stop working, each module can be directly attached to the bus which is named hot-swappable by means of CAN bus. System can work in multi-master or master-slave type communication. On the one hand each module sent the information to the UUV various parameters of the display and control units, on the other hand it can easily constitute the required virtual closed-loop controllable system. For example, Sailing subsystem, which uses CAN bus, first gather its own range of information, such as the Doppler sonar, Inertial navigation, GPS navigation information and so on, into a navigation computer, then the fusion data is upload to the integrative display and control units. The workload can be greatly reduced in the central controller.

CAN bus based control system for UUV navigation structure shown in Fig.6.
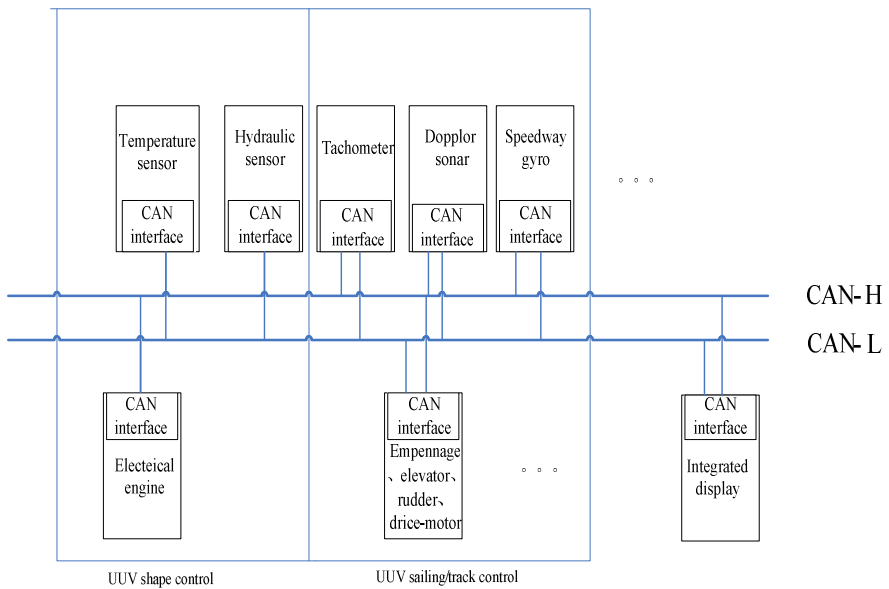
Fig. 6. Distributed UUV sailing control system structure based on CAN bus

CAN bus system is a multi-master system, host computers can copy each other, any node failure will not result in paralysis of the entire system. Because of a variety of fault monitoring and fault location methods which are adopted, reliability and timeliness are strong, and failure rate is reduced to 10^{-11}.

5   Conclusions

In this paper, according to the sailing control system's development trend, we have advanced the UUV sailing control system based on date bus, at the same time the paper mainly introduced MIL-STD-1553B bus and CAN bus and conducted a detailed analysis and comparison from various aspects. The result shows that CAN bus is a very well bus technology and can be used well in UUV sailing control system. In addition, according to some recent English literature [6], many light aircraft, or the larger aircraft, have made use of the CAN bus. It can be seen that the CAN bus data bus will become mainstream of the control bus mode probably.

References

[1] Xiong, H.-G., Wang, Z.-H.: Advanced Avionic Integration Technology. National Defence Industry Press, Peking (2009)
[2] Guo, Z., Xue, L., Lei, F., Wang, J.: An analysis of the CAN field bus MAC protocol. Journal of Huazhong University of Science and Technology (9), 64–66 (2002)
[3] Yan, M.-D., Chen, J.-P.: Design of Data Transmission Networks Based on CAN Bus for Automobile Electronic System. Journal of Chang an University (Natural Science Edition) 6(1), 86–89 (2006)
[4] Wu, K.-M.: The CAN Bus Principle and Application System Design. Peking Aerospace University Press, Peking (1996)
[5] Lian, B.-W., Li, Y., Zhang, Y., et al.: The Performance Analysis and Comparison on CAN Bus and 1553B Bus. Measurement & Control Technology 19(6), 47–49 (2000)
[6] Jalovecky, R., Bajer, J.: Development of the aircraft electronic system using CAN with CAN aerospace protocol (May 2009), `http://bbs.81tech.com`

Design and Industrial Practice of Computer Monitoring Control System for SRB Wastewater Treatment

Ying Ying

Zhejiang Water Conservancy and Hydropower College
Hangzhou, P.R. China 310018
`ylucy99@126.com`

Abstract. This paper investigates the importance of automated control in Sequencing Batch Reactor (SBR) wastewater treatment. A multi-stage computer monitoring control system for wastewater treatment, including sensing, control and management, was designed. The system construct and control-management functions are provided. A time-constant and sequence control of SBR tank (fundamental control) is demonstrated, and a technology parameters for stochastic injection during the automatic control process (high-level control). The proposed system has been used in a wastewater treatment plant at a selected city in Jiangsu Province and gained success in its operations.

Keywords: SBR wastewater treatment, Real-time control/management, Computer control system.

1   Overview of a Sequencing Batch Reactor—SBR Method

SBR method is the abbreviation of Sequencing Batch Reactor Activated Sludge Process, has developed rapidly in recent years a new wastewater treatment technology, new processes, but also internationally recognized as more advanced the production process.

In fact, SBR technology to be available in 1914, but the operation of the process more complicated, more variable parameters required for control, accuracy and reliability of the instrument has high requirements, limiting the development and the SBR process Promotion. With the computer control technology standards, automation equipment and on-line monitoring instrument of development, on the wastewater treatment process fully automated management and monitoring as possible, so the SBR method has spread rapidly in recent years [1].

SBR method is characterized by dividing wastewater treatment process into five main stages, such as fill, react, settle, draw and idle, and all of the stages in a single reaction tank, as shown in Figure 1. For continuous treatment of wastewater, the wastewater treatment plant may have several SBR tank, each SBR tank in space in descending order; the operation of each SBR tank is sequential in time, intermittent carried out. This in space and time are the characteristics of intermittent operation in sequence rather cumbersome to operate.

Treatment effect of SBR technology depends on its operating parameters, process parameters for better control; reaction process requires real-time monitoring, such as monitoring operation parameters DO (Dissolved Oxygen), PH (Potential Of

Hydrogen); requirements directly involved in the process logical and mathematical operations; according to the different nature to determine the optimal water flow and so on. The SBR method require automatic control system must be flexible application of the variable parameters, analog line testing, real-time display and call a variety of analog curve, SBR method for improved decision-making process of the database. Therefore, the need to adopt a process computer as the core of the automatic monitoring system, real-time control of the response phase of the time and blowers, pumps, electric valves and other equipment, transportation stops, to make the water meet the national emission standards, which became the SBR method can the key to any successful application.
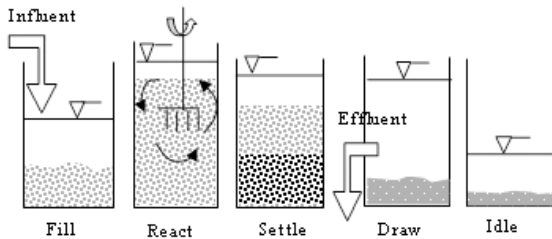
Fig. 1. Five stages in an SBR tank process period

2   Design of Computer Monitoring System

This design of SBR wastewater treatment test - control - management of computer monitoring system shown in Figure 2. From the bottom up, including: on-site control level - in the control room monitor level - director of Enquiry / management level. The system is successfully applied in a development zone, Jiangsu Province, the wastewater treatment plant.

From the network structure, is divided into two networks: the composition of the upper management of the computer, the director room of wastewater treatment plant and technology department; the lower layer for industrial monitoring, wastewater treatment plants by the central control room located in the industrial control Computer (IPC) and at the production site constitutes Programmable Logic Controller (PLC) components. These computers together into industrial Ethernet LAN 1 # ~ 3 # PLC access to the central monitoring computer Profile Bus local bus network.

Director of query / management-level system is mainly to achieve the following functions.

(1) Query the current value of the process parameters, settings, upper and lower limits, when the gauge is displayed when flashing.

(2) Query shows the current running status of process equipment, including: manual / automatic, work / stop, normal / failure.

(3) Query the current fault alarm and alarm information, including: the point of failure, fault type and nature of the fault start time, end time.

(4) Check the process parameters in real time curves.

(5) Query the historical data in the production process.

(6) Check the history of trend parameters.

In the main control room monitoring level to achieve the following functions: (1) Monitoring - the monitoring of process parameters: detection and display its value in the control room; on the monitoring of process equipment: In the control room to detect and display the operation status. (2) Remote control - the control room in the mouse and keyboard on the process equipment start / stop control, set the value of the process parameters set, changes. (3) Self-control - according to one or some of the conditions of process equipment for automatic start / stop control or closed-loop automatic control. In director query / management level and between the control room monitoring level, the use of the browser/server (B/S) structure, B/S structure in SBR Wastewater Treatment Management - monitoring and control system available as a direct Effect:

(1) B/S structure can easily access more than one browser, without having to add any additional browser application. Therefore, work can change the browser into operation at the wastewater treatment plant after the operation and management personnel from the factory independently. For example, if there is a need, you can also access the computer lab Industrial Ethernet LAN, and set into a browser function, so in the laboratory to the production process can also browse the data. This no doubt is quite easy to expand system functionality useful.

(2) Access through the Internet web browser on a remote (such as upper management) to observe the required data and information.

(3) The director room of the wastewater treatment plant, Technical Bureau and other departments of the wastewater treatment is required to navigate the production process, not the burden of information processing and computing tasks, the best to play this form of B / S's strengths, the most suitable for B / S architecture.

The SBR wastewater treatment system consists of multiple automatic control system. Due to space limitations, only the SBR tank timing control system following a brief introduction:
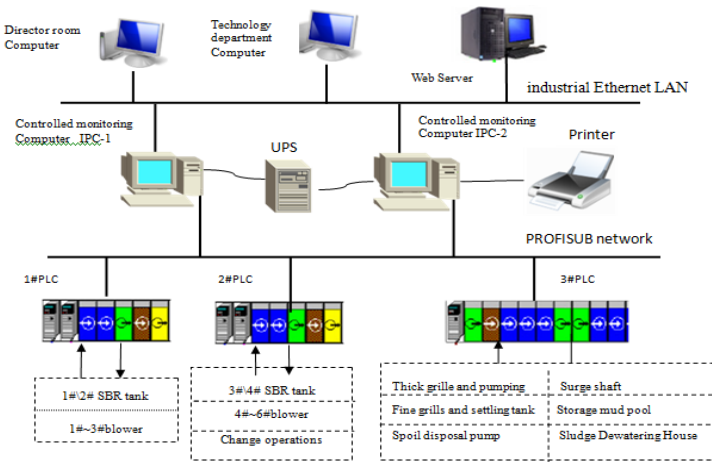
Fig. 2. Computer monitoring/management system framework for SBR wastewater treatment plant

3   SBR Tank Timing Control and Technological Parameters of the Random Cut

The wastewater treatment plant has 4 sequencing of the SBR reactor arrangement, each reactor in chronological order for fill, react, settle, draw and idle five stages (for a cycle) of operation, and the cycle for Shown in Figure 3.

In the time control, SBR reactor's basic cycle T is 4 hours, including fill T0 = 1 hour, react T1 = 2 hours, settle T2 = 1 hour, draw T3 = 0.5 hours, idle T4 = 0.5 Hours. The time can be set on the host computer or adjustment.

Taking into account the four SBR reactor relationship between the timing of work, 1 # SBR tank (1PLC control) and 3# SBR tank (2#PLC control) the timing difference ΔT = 1 hour, ΔT can be carried out in the host computer to re-set or adjusted.

1# SBR tank in a working cycle of each phase, the equipment and the relationship between the valve action described in Table 1 (2 # ~ 4 # SBR tank similar), PLC automatic control system should be interlocked to achieve timing Table 1 Control.
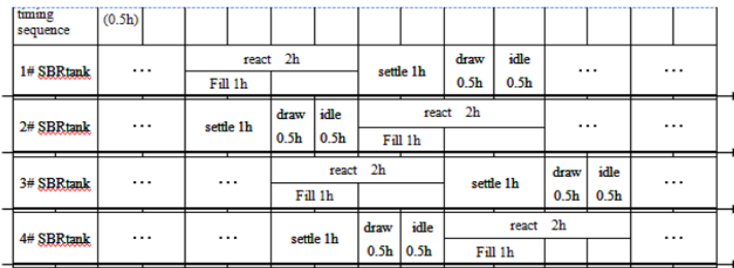
Fig. 3. A working period of an SBR tank

Table 1. Various equipment operations in 1#SBR tank working period

stages	1#、2# fill Electric valve	1# air inflow Electric valve	1# spoil disposal Electric valve	1# surplus sludge pump	1# backflow Electric valve	1# backflow sludge pump	1#、2# water decanter	1#~4# submersible mixer
				Various equipment Working state				
fill,	on	on	off	off	on	on	off	on
settle	off	off	off	off	off	off	off	off
draw	off	off	off	off	off	off	on	off
idle	off	off	on	on	off	off	off	off

The time to complete the SBR tank interlock controls the basic control requirements. In this control algorithm, the fill, react, settle, draw and idle time set by hand, SBR reactor process each segment and its implementation at all times be in strict accordance with the timing set, once the automatic state cut After entering the set timing, cannot change the set time parameters, to change these settings need to wait until the next operating cycle. This is clearly a fixed control reduces the operating efficiency and increase energy consumption.

The design and implementation of process control process parameters on the SBR random cut. Reactor under the fill or changes in concentration of organic matter, the flexibility to change the reaction time (especially in the react stage), in a timely

manner will cut the response time after the change control process without the need to wait until the next run cycle. Many of the concentration of organic matter in industrial wastewater vary greatly over time, sometimes several times or ten times the difference, while the energy consumption of SBR method mainly concentrated in the react stage. The system can be for different influent organic concentration, the right to change the reaction time, so that the processing of water quality while minimizing operating costs, reduces energy consumption, to prevent sludge bulking, SBR method to achieve a higher level of computer control.

4   Industrial Practice

This design has obtained the successful application in a wastewater treatment plant in Jiangsu Province a development zone. It is a project to the first phase of the project for the plant having 40,000 tons / d, two for the 80,000 tons / d, the forward for the 12 million tons / d, and the main process for the wastewater treatment plant based on SBR process.

5   Summaries

This design of SBR wastewater treatment test - control - management of computer monitoring system can satisfy the requirements of SBR wastewater treatment. In the wastewater treatment plant design proved successful application of the rationality and reliability. The main technical characteristics of the system are:

(1) Real-time control / management system using browser / server architecture. It will be better achieved at all levels of management and control of process parameters on the process of browsing queries.

(2) Basic control in the SBR tank (regular interlocking control), based on the complete control of the process parameters on the random intercept and the SBR method to achieve a higher level of computer control.

(3) The number of accused jointly coordinated control of equipment, a reasonable allocation of the equipment, loads, optimize utilization, and reduce energy consumption.

References

[1] Kim, J.-H., Chen, M., Kishida, N., Sudo, R.: Integrated real-time control strategy for nitrogen removal in swine wastewater treatment using sequencing batch reactors. Water Research 38(14-15), 3340–3348 (2004)
[2] Wang, D., Xu, Y., Du, Y., Meng, Z.: Analysis on technical characters of SBR. Journal of Harbin University of Commerce (Natural Sciences Edition) 19(6), 677–680 (2003)
[3] Zhou, D., Yuan, N., Ying, Y.: Design and application of intelligent monitoring and control system of the computer. Tsinghua University Press, Beijing (2002)
[4] Zhao, L.-X.: The Design of Urban Sewage Disposal Control System by PLC. Journal of Zhangzhou Normal University (Natural Science) 02, 60–63 (2008)
[5] Zhou, D., Yuan, N., Ying, Y.: Design and application of intelligent monitoring and control system of the computer. Tsinghua University Press, Beijing (2002)

Retraction: A New Method for the Visualization of Byzantine Fault Tolerance

Sun Ping

Zhejiang Water Conservancy and Hydropower College
Hangzhou, P.R. China 310018
`htbbs3@163.com`

Several conference proceedings have been infiltrated by fake submissions generated by the SCIgen computer program. Due to the fictional content the chapter "A New Method for the Visualization of Byzantine Fault Tolerance" by "Sun Ping" has been retracted by the publisher. Measures are being taken to avoid similar breaches in the future.

Early Fire Detection in Coalmine Based on Video Processing

Wanzhong Lei[1] and Jingbo Liu[2]

[1] Department of Electrical Information Engineering Henan Institute of Engineering,
Zhengzhou, 451191, China
`zzlwz@163.com`
[2] School of Electrical Engineering
Southwest Jiaotong University
Chengdu, 610031, China
`ljb79@126.com`

Abstract. Fire is one of the most serious catastrophic disasters in the coalmine. The early fire detection can help to avoid a disastrous fire. The existing temperature-sensed and smoke-sensed method may respond slowly to early fire, if it is far away from the sensors, or the setting values of sensor are unsuitable. Therefore, the detection method based on image and video processing is adopted to overcome the drawbacks of traditional fire detection method for coalmine. The paper analyses the status in the fire detection technology, and designs a structure to detect the early fire in coalmine. Firstly, image which comprises the potential fire region the potential fire region is detected by using frame differencing of monitor video, and denoised by median filter. Secondly, flame region is extracted by color information. Finally, Bayes classifier is employed to recognize fire combined with the dynamic features. The method can greatly improve accuracy of early fire prediction in coalmine, comparing with the traditional detection method.

Keywords: Video processing, Image processing, Fire detection, Coalmine, Flame recognition.

1   Introduction

Fire is one of the most serious catastrophic disasters in coalmine and it usually influences the safety production severely. It not only results in huge economy loss, but also threatens on miner security. If the fire can be warned by alarm early and be eliminated in time, the harm should be reduced greatly. Therefore, real-time fire detection is significant, especially in coalmine. There are some fire detection methods, such as the temperature induction, the smoke induction, the light induction and the complex pattern. In practice, these traditional methods may make leak forecast or wrong forecast in abominable condition for environmental disturbance, also cannot respond to the early fire timely. It is generally known that the environment of coalmine is very complex; thus, need to find new method to detect the early fire in coalmine.

The flame has abundant vision information, such as geometry parameters, luminance parameters, and flicker frequency. To overcome the previous problems

mentioned above, the image-based fire detection technology has been proposed based on the advantages of the image processing in recent years. Healey [1] use a model based on color information of fire while Phillips [2] further adds the time-varying property of the flame color in the detection process thus to reduce affection of background illumination. In [3], boundaries of flames are represented in wavelet domain and high frequency character of the fire region' boundaries is used as the fire detection feature. Other researchers use multiple features of flame to perform better. Liu et al. [4] detect the top point across the frame edge by using the FFT algorithm. Chen et al. [5] develop a set of rules over the RGB color space to detect the fire pixels. Marbach et al. [6] exploit a YUV-based color model to represent the fire in video data. Motion detection on the Y channel is used to detect the candidate fire pixels. The detected pixels are then verified based on their chrominance information U and V to determine whether the candidate pixels are the fire pixels or not. Horng et al. [7] develop an HSI-based color model to segment the fire-like regions for brighter and darker environments. Their fixed threshold method achieves 96.94% detection rate. It is expected that a large amount illumination variations is quite likely incurred during the incidence of fire. In [8], an alternative approach in color-based detection is to analyze the YCbCr color space instead of the RGB space, a fuzzy logic approach, which uses luminance and chrominance information, is used to replace the existing heuristic rules used to generate the potential fire region.

Although some problems and drawbacks still exist in the above systems, image-based technology has many advantages comparing with traditional technologies. It has a very good detection rate, can provide intuitionistic fire image information, and can be used in large space and abominable environment, so it is suit to early fire detection in coalmine. The method proposed in paper adopts [8] to detect the early fire in monitor videos of coalmine. It has many important military and commercial applications and has significant advantages over traditional fire detection methods. To avoid leading to a fire disaster, the fire alarm should be given as soon as detecting a burning fire early.

The rest of paper is organized as follows. In section 2, characteristics of fire used in paper are described. The proposed detection method is given in section 3. Finally, the conclusion is given in section 4.

2   Characteristics of Fire Used in Proposed Method

Unique visual signatures of fire are all essential for recognition. A region that corresponds to fire can be captured by spectral characteristics of the pixels and the spatial structure defined by their spectral variation within the region. The shape of a fire region usually keeps changing and exhibits a stochastic motion, which depends on surrounding environmental factors, such as the type of burning materials and airflow. To validate a real burning fire, color and dynamic features are used to distinguish other fire aliases [9].

2.1   Color

According to most previous fire detection methods and our experiments, we notice that color characteristics of fire are very distinct. The flames of general fire display reddish colors; the color of the flame will change with the different temperature. The

color shows range from red to yellow when the fire temperature is low, and it may become white when the temperature is higher. It may be the most powerful feature for detecting fire in video sequences. For a given general fire pixel, it is noticed that the value of the green channel is greater than the value of blue channel, and the value of red channel is greater than the green channel.

This color detection metric can be used to generate the potential fire region, further, which will be analyzed with the other non-color fire features described in follow.

2.2   Dynamic Features of Flame

Airflow usually makes flame oscillate or move suddenly, it demonstrates the changeable shapes of fire. This can reflect the corresponding effect especially on a variable flame area in an image. The size of a fire's area in an image cannot maintain to be constant for the shape of flames is changeable anytime owing to airflow. Besides, the flame always has a growth feature. The disorder of fires can be measured with the pixel quantity of flame difference between two consecutive images. These fire dynamic features include sudden movements of flames, changeable shapes, growing rate, and oscillation (or vibrations), etc. Based on the above analysis of fire, dynamic features will be used to identify a real fire for improving the reliability of detection. We utilize both the disorder characteristic (randomness of area size) of flames and the growth of fire pixels (boundary roughness) to check if it is a real fire [10]. They are given separated by follow formulas:

$$\Delta A_i = \frac{\left| A_i - A_{i-1} \right|}{A_i},$$

(1)

Where ΔA_i is the normalized area change for the ith frame image, A_i corresponds to the area of the fire blobs representing the potential fire regions. Fire is assumed if $\Delta A_i > \lambda_A$, where λ_A is a decision threshold.

$$B_R = P_S / P_{CH_S},$$

(2)

Where S is the convex hull of a set of pixels, P_S is the perimeter of S and P_{CH_S} is the perimeter of the convex hull of S. Fire is assumed if $B_R > \lambda_{B_R}$, where λ_{B_R} is a decision threshold.

3   The Proposed Method

The framework of proposed method is shown in Fig.1.

3.1   Potential Fire Region Detection

When fire occurs, the pixels of the two continuous images of video may change. After acquiring video data, we can use the frame differencing method to get the change region. Firstly, two continuous frame images are compared. If there is no evident change, then the two images are abandoned, and the system goes on detecting.

Conversely, if there is some pixels difference between them, then image should be collected, and the continuous frame image should be compared. If there is some change each time, then fire may occur, and the early fire detection may start. Described by formula as below:

$$\Delta f_i(x, y) = f_i(x, y) - f_p(x, y) = \begin{cases} 0, \Delta f_i(x, y) < T \\ 1, \Delta f_i(x, y) > T \end{cases} \tag{3}$$

Where, $\Delta f_i(x, y)$ is the difference between the two continuous frame images of the monitor video; $f_i(x, y)$ is the current image; $f_p(x, y)$ is the previous image, T is the threshold. When the gray difference between the current image and the previous image is less than the threshold T, we believe that there is no fire for the current image may simply change slightly because of the environmental effect.
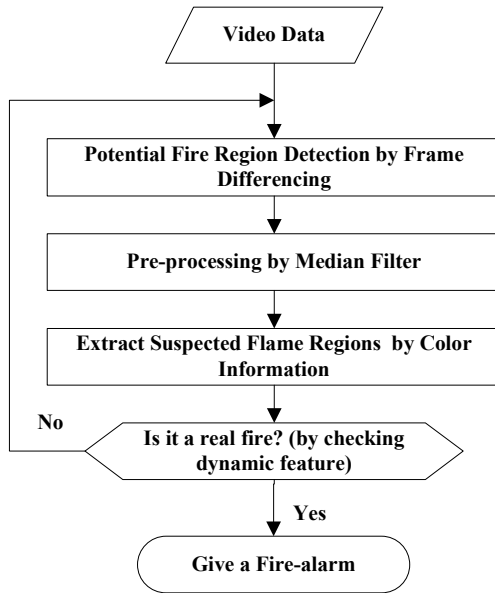
Fig. 1. The proposed early fire-detection algorithm

3.2   Median Filtering

As a typical non-linear filtering method, median filtering algorithm can remove noise at the same time retain a good image of the details of information. The basic principle of median filtering algorithm is to use the median value of all the pixel values in the filtering neighborhood of a certain point in part of an image, which is the value of the midpoint position item of all the pixel values sorted ascending or descending, to replace the value of the particular point. Medium value of the sequence of terms is defined as:

Assume a group of sequences $x_1, x_2, x_3, \cdots, x_n$, after sorted in ascending it becomes:

$$x_{i1} \leq x_{i2} \leq x_{i3} \leq \cdots \leq x_{in} \, (i = 1, 2, 3, \cdots, n), \tag{4}$$

Then its median value is:

$$y = Med\{x_1, x_2, x_3, \cdots, x_n\} = \begin{cases} x_{i(\frac{n+1}{2})}, n \text{ is odd number} \\ \frac{1}{2}\left(x_{i(\frac{n}{2})} + x_{i(\frac{n+1}{2})}\right), n \text{ is even number} \end{cases} \tag{5}$$

In this paper, median filter is used to smooth image. Median filtering not only can eliminate random noise in image mostly, but also can get a good distribution in the image and fuzzy image edge.

3.3   Suspected Flame Region Extracted by Color Information

We can easily extract the frame region by its color information. Suspected flame regions are selected from those potential fire regions according to the color information in the *RGB* and *YCbCr* color space. It is found that the power of R (red) component will increase and the power of Cr component is always greater than Cb when the flame appears. So, suspected flame region can be determined when it satisfies the following formulas [11]:

$$\begin{cases} Pf_{LR} > Pb_{LR} \, \& \, Pf_{LR} > Pb_{LB}, \\ Pf_{Cr} > Pb_{Cb} \, \& \, R > 100, \end{cases} \tag{6}$$

Where Pf_L and Pb_L denote the L band power of foreground (current frame image) and background (previous frame image) region respectively, and the subscript $R, Cr, and \, Cb$ denote the $R, Cr, and \, Cb$ color channel. To avoid the false flame detection due to the interference of background illumination or other environmental disturbance, the time varying property of flame geometry is taken into account. For the time varying geometry feature, angles of the four major corners (top, bottom, right, and left) of the flame region are calculated first. Calculating the angle by the formula:

$$\cos(Angle) = \frac{-(A^2 + B^2 - C^2)}{2AB}, \tag{7}$$

After that, the variation of the angle is estimated as follows:

$$\left|Angle_i - Angle_p\right| > Th, \tag{8}$$

Where the subscript i, and p denote the angle of the current and previous frame image respectively. If the variation of the angle is less than the threshold Th_{Angle}, the flame region is marked as the false result and is eliminated. Similarly, real flame region can be examined by analyzing the variation of flame center intensity in R channel. If its variance is less than a given threshold, it is removed from the flame set.

3.4   Fire Recognition

Bayes classifier is employed to combine the features discussed in the previous section. In order to classify the class fire from the class non-fire, and needs to estimate the mean and the variance of each class. Therefore For each frame image $f_i(x, y)$, naive set of potential fire regions is initially created based on the set of rules for color. For each potential fire region, a vector d of features is obtained as:

$$d = \begin{bmatrix} \Delta A \\ B_R \end{bmatrix}, \tag{9}$$

The Bayes classifier decision function in the fire class is described by:

$$f_1(d) = \ln P_r(b=1) - \frac{\ln |C_1|}{2} - \frac{(d-m_1)^T C_1^{-1}(d-m_1)}{2}, \tag{10}$$

Where ln represents the natural logarithm operation, b indicate a flag that represents one of the two possible classes, p is the Gaussian density of the vector in the fire class. Correspondingly, for the non-fire class, the decision function is:

$$f_0(d) = \ln P_r(b=0) - \frac{\ln |C_0|}{2} - \frac{(d-m_0)^T C_0^{-1}(d-m_0)}{2}, \tag{11}$$

Finally, the decision surface separating the two classes is:

$$f_{10} = f_1(d) - f_0(d) = 0. \tag{12}$$

4   Conclusions

In this paper, an eraly fire detection method based on image and video processing technology is presented in order to avoid or eliminate the fire disasters in coalmine timely. Both color and dynamic features are used to extract a real flame that is adopted for helping to validate the fire. Further, a fire alarm is given immediately when the fire alarm raising condition is met.

References

[1]   Healey, G., Slater, D., Lin, T., Drda, B., Goedeke, A.D.: A system for real-time fire detection. In: CVPR, pp. 15–17 (1993)
[2]   Phillips III, W., Shah, M., Lobo, N.V.: Flame recognition in video. In: Fifth IEEE Workshop on Applications of Computer Vision, pp. 224–229 (2000)
[3]   Toreyin, B.U., Dedeoglu, Y., Gudukbay, U., Cetin, E.: Computer vision-based method for realtime fire and flame detection. Pattern Recognition Letters 27(4), 49–58 (2006)
[4]   Liu, C.B., Ahuja, N.: Vision based fire detection. In: IEEE International Conference on Pattern Recognition, vol. 4, pp. 134–137 (2004)
[5]   Chen, T., Wu, P., Chiou, Y.: An early fire-detection method based on image processing. In: ICIP, pp. 1707–1710 (2004)

[6] Marbach, G., Loepfe, M., Brupbacher, T.: An image processing technique for fire detection in video images. Fire Safety Journal 41(4), 285–289 (2006)

[7] Horng, W.B., Peng, J.W., Chen, C.Y.: A new image-based realtime flame detection method using color analysis. In: Proceedings of IEEE Networking, Sensing and Control (ICNSC), pp. 100–105 (2005)

[8] Celik, T., Ozkaramanli, H., Demirel, H.: Fire pixel classification using fuzzy logic and statistical color model. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), vol. 1, pp. 1205–1208 (2007)

[9] Liu, C., Ahuja, N.: Vision-based fire detection. In: Proc. Int. Conf. Pattern Recognit., vol. 4, pp. 134–137 (2004)

[10] Paulo, V.K.B., Ebroul, I.: A Probabilistic Approach for Vision-Based Fire Detection in Videos. IEEE Transactions on Circuits and Systems for Video Technology 20(5), 721–731 (2010)

[11] Lai, C.L., Yang, J.C., Chen, Y.H.: A real time video processing based surveillance system for early fire and flood detection. In: Proceeding of 24th IEEE Instrumentation and Measurement Technology Conference (IMTC), Poland (2007)

Research and Application of the Heat Energy Recycling Technology in Mine Return Air

Li-Rong Wan, Jian-Liang Li, Liang Wang, and Guang-Yu Zhou

College of Mechanical and Electronic Engineering,
Shandong University of Science and Technology
Qingdao 266510, China
{Wanlr666,Lijianliang51,6057079}@163.com, Guangyu000@126.com

Abstract. The heat energy recycling technology in mine return air is a technology which is used to absorb and reuse the heat energy carried by coal mine return air. In this paper, according to the actual situation of some coal mine's return air well, the extracting method of heat energy in return air is put forward, the heat recycling system is designed and the heat energy is recycled through the use of the spray heat transfer technology. Finally through water source heat pump units, the heat energy is transformed to satisfy the needs of heating worker's bath water and winter heating for mine wellhead and workplaces, and thus the recycling of heat energy in mine return air can be realized.

Keywords: Mine, Return air, Heat energy, Recycling.

1 Introduction

Energy plays a crucial role in the development of our society. With the increasing demand for energy, the applications of renewable energy get more and more attention. Coal is the important foundation energy and raw materials in our country, and it has an important strategic position in the national economy. In the process of underground coal mining, there are a lot of harmful gases erupting out in coal seams, such as CH_4, CO, H_2S, CO_2, etc., and with the easily explosive coal dust they threaten the safety of underground workers. So in order to ensure the safety production of mine, mine must be ventilated. The fresh air in the ground is sent to the underground through the ventilator continuously to ensure the good working conditions. But mine is a huge regenerator and has rich geothermal resources. The air sent to the mine exchanges heat energy with the mine continuously, and eventually the temperature of the air reaches a balance with the geotemperature of the mine. Because mine's geotemperature is invariable basically, the temperature of mine return air remains stable basically annually. According to the incomplete investigations and statistics, the perennial temperature of mine return air is 18~28℃ or so, and its humidity is more than 90%. It is a kind of waste heat resources, which are stable and high-class. Under normal circumstances, mine's ventilation system only considers the safety of mine ventilation, so the mine return air all is discharged into the atmosphere. Along with it, the heat energy all is diffused into the atmosphere, causing a waste of energy. In view of this, this paper proposes a kind of technology about recycling the heat energy in mine return

air, which uses spray device to do it to replace the traditional coal-fired boiler for heating. Thereby the energy consumption and pollution of the coal mine enterprises are reduced, which further realizes energy conservation and emission reduction.

2   Determination of the Recycling Scheme of Heat Energy in Return Air

The main methods of extracting and converting heat energy from mine return air have the direct and indirect use of air source heat. The key to use air source heat directly is to choose air source heat pumps. The method of using air source heat indirectly is to use liquid mediums (such as water) to absorb the heat energy in gases, and then use the water source heat pump to extract heat energy from liquid again. The analysis of the scheme of recycling the heat energy in return air is as follows:

Considering the narrow space of the return air shaft, the big air quantity and a great deal of coal dust, if we choose the recycling scheme of air source heat pumps, as a result of using the surface heat exchanger, it is easy to be pasted by coal dust. At the same time, owing to the heat transfer of big air quantity, if we choose the surface heat exchanger, it needs a very large installation site and space, the system is complex, and the return air resistance is bigger, which affects the operation condition of mine main fan. And also R&D and application of large low-temperature air source heat pumps are difficult in our country. So the method of choosing air source heat pumps to recycle the heat energy from mine return air is not feasible.

If we choose the technology of using air source heat indirectly, which recycles heat energy from return air through the spray heat exchanger, it can control air resistance under the pressure of 200Pa, the project cycle is short, and the running resistance is small, which do not affect the normal work of the main fan. So in conclusion, this paper decides to use the spray heat exchanger to extract heat energy from mine return air.

3   Design of the Recycling System of Heat Energy in Return Air

Combined with the actual situation of some coal mine's return air shaft, the designed recycling system of heat energy in mine return air in this paper is shown in Fig.1.

The system mainly consists of water pump, wind wells diffusion tower, water spray device, recycling pool, water purifier and pool. Pool is parted in the middle, and the two parts are set respectively to low and high temperature areas. Water spray device is installed in wind wells diffusion tower. Recycling pool is connected to wind wells diffusion tower, and is connected to water purifier through pipes. Water purifier is connected to the high temperature area of pool, which is connected to the entrance of water source heat pump units. Water spray device is connected to water pump through pipes, which is connected to the low temperature area of pool through pipes. The low temperature area of pool is connected to the exit of water source heat pump units. Therefore the system structure is simple, it fully recycles the heat energy carried in return air, the utilization ratio is high, and it reduces the noise and pollution caused by return air. Its features are that water spray device has many groups of exhaust pipes and there are nozzles on exhaust pipes.
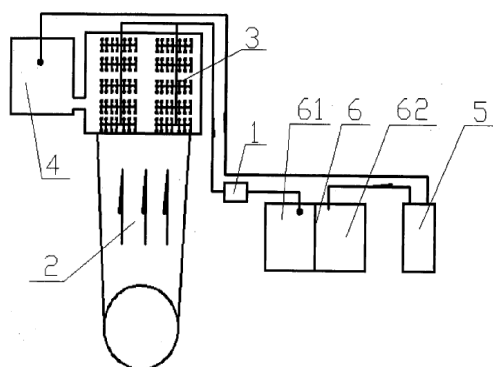
1. Water pump, 2. Wind wells diffusion tower, 3. Water spray device, 4. Recycling pool, 5. Water purifier, 6. Pool (61: the low temperature area, 62: the high temperature area)

Fig. 1. Recycling system of heat energy in mine return air

Its technological process is shown as follows: the low temperature water in the pool is pressurized to water spray device through water pump, and then is ejected to become vaporific through nozzles. After the low temperature water exchanges heat energy with the high temperature return air in wind wells diffusion tower to become the high temperature water, it is collected in the recycling pool. In order to prevent the high temperature water from being blown away and reduce water loss, there is a manger being installed in the upper portion of spray device. After the high temperature water is purified by water purifier, it is collected in the high temperature area of pool. After the water in the high temperature area exchanges heat energy with the outside world through water source heat pump units, it becomes the low temperature water, and then is collected into the low temperature area of pool for cyclic use.

4   Calculation about Recycling Quantity of Heat in Return Air

4.1   Recyclable Quantity of Heat

In the view of the actual situation of some coal mine, this mine's winter return air takes temperature: 22℃, relative humidity: 95%, air volume: 158m^3/s as basic parameters. After the heat energy conversion, these parameters of the return air become that temperature: 12℃, relative humidity: 100%, air volume: 158m^3/s. And then the available theoretical quantity of heat is:

$$Q = (i_1 - i_2)\rho\rho = (62.32 - 34.1) \times 1.2 \times 158 \approx 5350 \text{kw}$$

In the above equation: i_1 is the enthalpy in return air when its temperature is 22℃ and its relative humidity is 95%, kJ/kg; i_2 is enthalpy in return air when its temperature is 12℃ and its relative humidity is 100%, kJ/kg; ρ is the average air density of the return air, kg/m^3; L is the volume of return air, m^3/s.

It is observed that the theoretical quantity of heat extracted from the return air is greater than 3000kw, which is the heat needed by water source heat pump units, so it can solve the problem effectively that the heat source of water source heat pump units is insufficient.

4.2   Calculation about Exchanging Heat Energy through Spray Device

(1) Quantity of cyclic water

When the return air are under the conditions that the quality of converted heat is 3000kw, the volume of return air is 158m³/s, the temperature of return air is 22°C and the relative humidity is 95%, considering the actual situation of the mine, the inlet water temperature is $t_{w1} = 5°C$, which is the initial temperature of water, and we assume that the outlet water temperature is $t_{w2} = 12°C$, which is the terminal temperature of water. So the quantity of needed cyclic water is:

$$W = \frac{Q\rho_w}{1.163(t_{w2} - t_{w1})} = \frac{3000 \times 10^3}{1.163 \times (12 - 5)} \approx 3.69 \times 10^5 \text{ kg/h}$$

Therefore, when the temperature difference between inlet and outlet water is 7°C, the basic parameters of the designing spray heat exchanger is that the quantity of cyclic water is 4×10^5 kg/h.

(2) Calculation of parameters

From the structure of the chosen single stage spray chamber, it can be known that the type of centrifugal nozzle is selected as Y-1, d_0=4mm, $n = 13$pieces/（m²·row）, and spray chamber is two rows opposing spray patterns. When the mass velocity of the return air is $v\rho = 3.1$kg/(m²·s), the sectional wind velocity of the spray chamber is $v = \frac{3.1}{1.2} \approx 2.6$m/s .

Coefficient of the spray chamber is:

$$\mu = \frac{W}{G} = \frac{W}{3600\rho 6} = \frac{4 \times 10^5}{3600 \times 1.2 \times 158} \approx 0.59$$

In the above equation, G is the volume of return air, kg/h.

The heat exchange rate coefficient of the spray chamber is η_1, and according to its experimental formula, we can list the following equation:

$$\eta_1 = 1 - \frac{t_{s2} - t_{w2}}{t_{s1} - t_{w1}} = 0.745(v\rho)^{0.07} \mu^{0.265}$$

Taking the known conditions into the above equation:

$$\eta_1 = 1 - \frac{t_{s2} - t_{w2}}{21.4 - 5} = 0.745 \times 3.1^{0.07} \times 0.59^{0.265} \approx 0.70$$

It can be gotten that:

$$t_{s2} - t_{w2} = 4.92 \tag{1}$$

The thermal equilibrium equation is that $i_1 - i_2 = \mu c(t_{w2} - t_{w1})$. In the equation, C is the specific heat capacity of water and equals to $4.19 J/(g \cdot {}^0C)$. According to it, it can be gotten that:

$$62.32 - 34.1 = 0.59 \times 4.19 \times (t_{w2} - 5) \tag{2}$$

Combining (1) and (2) equations, it can be gotten that:

$$t_{w2} \approx 16.42^0 C, \ t_{s2} \approx 21.34^0 C$$

The total number of two rows of nozzles is:

$$N=2nA=2 \times 13 \times 61 = 1586 \text{ pieces}$$

In the equation, A is the cross-sectional area of spray chamber, $A = \dfrac{G}{3600 v\rho} \approx 61 m^2$.

And then the quantity of spray water through each nozzle is:

$$w = \frac{W}{N} = \frac{4 \times 10^5}{1586} \approx 253 kg/h$$

When the quantity of spray water through each nozzle is 253kg/h and the nozzle diameter is $d_0 = 4mm$, by checking the relationship chart about the quantity of spray water of Y-1 nozzle and the pressure of spray water and the nozzle diameter, it can be gotten that the water pressure needed before nozzles is 1.3atm, which is the working pressure.

(3) Calculation about the resistance of spray chamber

The resistance of front and behind mangers is ΔH_d:

$$\Delta H_d = \sum \zeta_d \frac{v_d^2}{2} \rho = 20 \times \frac{3.12^2}{2} \times 1.2 \approx 116.81 Pa$$

In the above equation: $\sum \zeta_d$ is the sum of the local resistance coefficients of front and behind mangers, and it depends on the structure of mangers. Generally we choose $\sum \zeta_d = 20$; v_d is the face velocity when air is blown to the fault surface of mangers. Because the frontal area of mangers is that the cross-sectional area of spray chamber deducts the borders of mangers. Generally we choose $v_d = (1.1 \sim 1.3)v$ m/s, and then $v_d = 1.2 \times 2.6 = 3.12 m/s$.

The resistance of nozzles exhaust pipes is ΔH_p:

$$\Delta H_p = 0.1z \frac{v^2}{2} \rho = 0.1 \times 2 \times \frac{2.6^2}{2} \times 1.2 \approx 0.81 Pa$$

In the above equation, z is the number of exhaust pipes and equals to 2 because of the two rows opposing spray patterns.

The resistance of water seedlings is ΔH_w :

$$\Delta H_w = 118 b \mu P = 118 \times 0.075 \times 0.59 \times 1.3 \approx 6.79 Pa$$

In the above equation: P is the water pressure before nozzles, atm; b is the coefficient depended on water-jet direction and air moving direction. When it is the downstream spray pattern of single row, it is thought that b= -0.22. When it is the upstream spray pattern of single row, it is thought that b= +0.13. When it is the two rows opposing spray patterns, it is thought that b= +0.075.

The resistances of the spray chamber is ΔH :

Because the resistances of the spray chamber consists of the resistance of front and behind mangers, the resistance of nozzles exhaust pipes and the resistance of water seedlings, therefore from the above calculations we can get that the resistances of the spray chamber is:

$$\Delta H = \Delta H_d + \Delta H_p + \Delta H_w = 116.81 + 0.81 + 6.79 \approx 125 Pa$$

4.3   Equipment Selection

According to the above design calculation and combining the specific characteristics of this project, we choose the Y-1 type of centrifugal nozzles with the number of 2000 and the YX-2(ABS) type of mangers. All pipelines use spiral steel pipes, and its system of laying uses direct burial. Heat preservation of pipelines use cyano-polyester-plastics heat insulation pipes, which is that pipe outside wall is brushed by a layer of rustproof and antiseptic cyanogen coagulation, the thickness of foaming polyurethane hydration layer is 50mm, and the strata externum uses FRP cap layer whose thickness is 1.5mm.

5   Applied Conclusion

The waste heat resources contained in mine return air are very considerable. The technology uses spray heat transfer device to realize recycling heat energy of mine return air, and solve the problem that the heat source of water source heat pump units is insufficient. Thus it can replace the traditional coal-fired boiler for mine wellhead and workplaces heating to reduce environmental pollution and personnel settings, so it gains good economic and social benefits. At the same time, it can remove coal dust and dirt in return air to reduce the return air pollution to the environment, so it has very good environmental protection benefits. This technology is also suitable to popularize and applied in the non-coal enterprises, and it can provide a very good reference value for them.

References

[1] Cheng, J.-S.: Mining Machinery. China Mining University Press (1997) (in Chinese)
[2] Xu, G.-C., Chen, J., Zhu, J., Chen, G.-F.: Talking about the use of coal mine waste heat resources. Coal Engineering (9) (2010) (in Chinese)
[3] Yang, R.-H., Zou, S.-H., Zhang, D.: Research of the using way of mine secondary heat resources. Mineral Engineering Research 25(4) (December 2010) (in Chinese)
[4] Liu, W., Sun, X.: Research of the heat recovery in mine return air. ShanDong Refrigeration & Air Conditioning (in Chinese)
[5] Di, J.-Y., Fan, J.-G., Mo, J., Zhu, D.-Y., Zhang, X.-L., Gao, Q.: The heat utilization device of mine retutn air. Chinese Patent (2009) (in Chinese)

2L-XMMS: An Efficient Method for Mining Infrequent Itemsets with 2-Level Multipul Minimum Supports

Han Xi-Qing[1], Dong Xiang-Jun[2,*], Jiang He[2], and Geng Ru-Nian[2]

[1] Administration Office, Shandong Institute of Commerce and Technology
xqhan@163.com
[2] School of Information, Shandong Polytechnic University
{d-xj,gengrnn}@163.com, Jianghe09@126.com

Abstract. Infrequent itemsets become very important when we study positive and negative association rules simultaneously because we can mine many valued negative association rules from them. In our previous work, we have proposed a 2LMS_inFS_FS model by assigning 2-level different minimum supports to every item to constrain frequent and infrequent itemsets respectively. But 2LMS_inFS_FS used basic Apriori algorithm to discover the defined itemsets, which is not efficient. This paper proposes an efficient model, 2L-XMMS (2-level XMMS) model, which is based on MMS (Multiple Minimum Supports) model, to improve the efficiency. The comparisons and the experimental results show that 2L-XMMS model are efficient to discover both infrequent and frequent itemsets simultaneously.

Keywords: Infrequent itemset, Frequent itemset, Negative association rule, Multiple minimum supports, Prune.

1 Introduction

Infrequent itemsets (inFIS) has been gradually attracted attention because they have many potential applications and can be used to mine negative association rules [1,2,3,4,5,6,7,8,9]. Theoretically, infrequent itemsets are those itemsets whose supports are less than a minimum support *ms*, and they are the complement of frequent itemsets. Because the number of infrequent itemsets is too large to be mined completely, in real application, some constraints are often given so as to control their number in a moderate degree. The constraints are different in different papers and the details see section 2.

In [8], XMMS (eXtended MMS) model was proposed to mine infrequent and frequent itemsets. XMMS model is got by giving a minimum support to constrain infrequent itemsets to MMS (multiple minimum supports) model in [10]. MMS model assigns every item a different minimum support to solve the problem caused by a single minimum support. The problem is a dilemmatic situation of how to set the single minimum support *ms*: if *ms* is too high, less frequent itemsets would be discovered and hence some valued association rules would be lost; if *ms* is too low, more frequent itemsets would be discovered and hence more rules would be generated, which would

---

* Corresponding author.

increase the difficulties for users to choose right rules. The MMS model, however, did not concern about infrequent itemsets.

Because of only using a single minimum support to constrain infrequent itemsets in XMMS model, XMMS model is faced with the same problem to discover infrequent itemsets as that using a single minimum support to constrain frequent itemsets mentioned above. In order to solve this problem, ref [9] proposed a model 2LMS_inFS_FS by assigning 2-level minimum supports to every item to constrain frequent and infrequent itemsets respectively. But 2LMS_inFS_FS uses basic Apriori algorithm to mine the defined itemset, which is not efficient. This paper proposes an efficient model based on MMS model, named 2L-XMMS (2-level XMMS) model for difference. The differences between this paper and [9] are in two aspects: 1) the algorithm in [9] is based on basic Apriori, while the algorithm proposed in this paper is based on MMS model, which is more efficient, and 2) we rewrite almost all sections in this paper. The experimental results show that 2L-XMMS is efficient and effective.

The rest of the paper is organized as follows: Section 2 is related work. Section 3 is 2L-XMMS model. Experimental results are in section 4 and section 5 is conclusions.

2   Related Work

The constraints to infrequent itemsets in [1] used single minimum support ms to divide frequent itemsets and infrequent itemsets, and the constraints   are as follows: for any itemsets A and B, $A,B \subset I$, $A \cap B = \Phi$, $s(A) \geq ms$, $s(B) \geq ms$, if $s(AB) \geq ms$, then AB is a frequent itemset; if $s(AB) < ms$, then AB is an infrequent itemset.

In [2], an algorithm for minimal infrequent itemset mining was presented. An itemset A is called a *minimal ms_infrequent* if $s(A) < ms$ and the supports of   all of A's proper subsets are not less than ms.

In [3], an algorithm for mining infrequent itemsets from weighted incremental updating database was proposed to deal with the incremental updating problem when anew database is inserted in the original database and the minimum support is not changed to mine frequent and infrequent itemsets.

Ref. [4] proposed a MLMS (Multiple Level Minimum Supports) model to discover infrequent and frequent itemsets. MLMS model assigns different minimum supports for itemsets with different length. Let $ms(k)$ be the minimum support of k-itemsets ($k=1,2,…,n$), $ms(1) \geq ms(2) \geq …, \geq ms(n) \geq ms > 0$, the constraints are as follows. For any k-itemset A, if $s(A) \geq ms(k)$, then A is a frequent itemset; and if $s(A) < ms(k)$ and $s(A) \geq ms$, then A is an infrequent itemset. Ref. [6] used the pruning strategy in [1] to prune the uninteresting itemsets in MLMS model. Ref. [5] mined positive and negative association rules from those infrequent and frequent itemsets discovered by MLMS model.

Ref [7] proposed a 2LS (2-Level Supports) model to discover infrequent itemsets and frequent itemsets. 2LS model uses two level supports ms_FIS and ms_inFIS ($ms\_FIS \geq ms\_inFIS > 0$) to constrain the frequent itemsets and infrequent itemsets respectively. For any itemset A, if $s(A) \geq ms\_FIS$, then A is a frequent itemset; and if $s(A) < ms\_FIS$ and $s(A) \geq ms\_inFIS$, then A is an infrequent itemset.

Ref [8] proposed a method called XMMS model to mine infrequent and frequent itemsets, which will discuss in details later. Ref [9] proposed a model 2LMS_inFS_FS based on basic Apriori. Ref [10] only discussed how to mine frequent itemsets based on multiple minimum supports.

3   2L-XMMS Model

3.1   Problem Statements

Let $I=\{i_1, i_2,..., i_n\}$ be a set of n distinct literals called items, and TD a transaction database of variable-length transactions over I, and the number of transactions in TD is denoted as $|TD|$. Each transaction contains a set of item $i_1, i_2,...,i_m \in I$ and each transaction is associated with a unique identifier TID. A set of distinct items from I is called an itemset. The number of items in an itemset is the length of the itemset. An itemset of length k are referred to as k-itemset. Each itemset has an associated statistical measure called support, denoted by s. For an itemset $A \subseteq I$, $s(A)=A.count/|TD|$, where $A.count$ is the number of transactions containing itemsets A in TD.

Let k-itemset $A \subseteq I$, $(A=\{a_1a_2,...,a_k\}) \subseteq I$. $MIS_{FIS}(a_i)$ and $MIS_{INFIS}(a_i)$ are two minimum support threshold to constrain frequent and infrequent itemsets of item i respectively, $MIS_{FIS}(a_1) \leq MIS_{FIS}(a_2) \leq ... \leq MIS_{FIS}(a_k)$, $MIS_{FIS}(a_1) \leq MIS_{FIS}(a_2) \leq ... \leq MIS_{FIS}(a_k)$, $MIS_{FIS}(a_i) \geq MIS_{INFIS}(a_i) \geq 0$ $(1 \leq i \leq k)$.

The definition of infrequent and frequent itemsets in XMMS model is described as follows.

Definition 1. Let ms be a minimum support for infrequent itemsets and $ms \leq MIS_{FIS}(a_1)$. If $s(A) \geq MIS_{FIS}(a_1)$, then A is a frequent itemset; and if $s(A) < MIS_{FIS}(a_1)$ and $s(A) \geq ms$, then A is an infrequent itemset.

In order to solve the proplem discussed in first settion, we replace the single ms with $MIS_{INFIS}(i)$. The definition of infrequent and frequent itemsets in 2L-XMMS model is as follows.

Definition 2. If $s(A) \geq MIS_{FIS}(a_1)$, then A is a frequent itemset; and if $s(A) < MIS_{FIS}(a_1)$ and $s(A) \geq MIS_{INFIS}(a_1)$, then A is an infrequent itemset.

Definition 2 allows users to control the number of frequent and infrequent itemsets easily by setting $MIS_{FIS}(a_i)$ and $MIS_{INFIS}(a_i)$ at a suitable value.

3.2   Algorithm

Now we discuss the algorithm to discover both infrequent itemsets and frequent based on definition 2. The following algorithm Apriori_2L-XMMS is designed by modifying MMS model (the algorithm *MSapriori* in [10]). The main difference between them exists in three aspects: 1) sorting I according to $MIS_{INFIS}(i)$ (not $MIS_{FIS}(i)$); 2) using $temp_k$ to store those itemsets that satisfy $MIS_{INFIS}(c_1)$ in each loop and then getting corresponding FIS_k and $inFIS_k$; 3) in function level2-candidate-gen() and candidate-gen(), using MIS_{INFIS} as the threshold. Algorithm *Apriori_2L-XMMS* is more efficient than algorithm *2LMS_inFS_FS* in [9] because algorithm *MSapriori* in [10] is more efficient than basic *Apriori*. We omit the explanations about *Apriori_2L-XMMS* here because of the limited space. Reader can refer to ref. [10] for more detailed explanations.

Algorithm Apriori_2L-XMMS

Input: *TD*: Transaction Database; $MIS_{FIS}(i)$, $MIS_{INFIS}(i)$: minimum support threshold;

Output: *FIS*: frequent itemsets; *inFIS*: infrequent itemsets;

1   M_{INFIS} = sort(*I*, $MS_{INFIS}(i)$); /* sort *I* according to $MIS_{INFIS}(i)$ */
2   *F* = init-pass(M_{INFIS}, *TD*); /* make the first pass over *TD* */
3   $temp_1$ = {<*f*> | *f* ∈ *F*, *f*.count/|*TD*| ≥ $MIS_{INFIS}(f)$};
4   FIS_1= {<*f*> | *f* ∈ *F*, *f*.count/|*TD*| ≥ $MIS_{FIS}(f)$};
5   $inFIS_1$ = {<*f*> | *f* ∈ *F*, $MIS_{INFIS}(f)$ ≤ *f*.count/|*TD*|<$MIS_{FIS}(f)$ };
6   **for** (*k* = 2; $temp_{k-1}$≠ ∅; *k*++) **do**
7   **if** *k* = 2 **then** C_2 = level2-candidate-gen(*F*)
8   **else** C_k = candidate-gen($temp_{k-1}$)
9   **end**
10   **for** each transaction *t* ∈ *T* **do**
11   C_t = subset(C_k, *t*);
12   **for** each candidate *c* ∈ C_t **do** *c*.count++;
13   **end**
14       $temp_k$ = {*c* ∈ C_k | *c*.count/|*TD*| ≥ $MIS_{INFIS}(c_1)$};
15   FIS_k={*c* ∈ C_k | *c*.count/|*TD*| ≥ $MIS_{FIS}(c_1)$ };
16       $inFIS_k$ = { *c* ∈ C_k | $MIS_{INFIS}(c_1)$ ≤*c*.count/|*TD*| < $MIS_{FIS}(c_1)$ };
17   end
18   *FIS* = ∪FIS_k; *inFIS* = ∪$inFIS_k$;
19   **return** *FIS* and *inFIS*;

Function level2-candidate-gen() is as follows:

    for each item *f* in *F* in the same order **do**
        if *f*.count ≥ $MIS_{INFIS}(f)$ **then**
            for each item *h* in *F* that is after *f* **do**
                if *h*.count ≥MIS_{INFIS} (*f*) **then**
                    insert <*f*, *h*> into C_2

Function candidate-gen() contains two steps: join step and prune step.

The join step is as follows:

insert into C_k
select *p*.item$_1$, *p*.item$_2$, …, *p*.item$_{k-1}$, *q*.item$_{k-1}$
from $temp_{k-1}$ *p*, $temp_{k-1}$ *q*
where *p*.item$_1$ = *q*.item$_1$, …, *p*.item$_{k-2}$ = *q*.item$_{k-2}$, *p*.item$_{k-1}$< *q*.item$_{k-1}$

The prune step is as follows:

  for each itemset *c* ∈ C_k **do**
    for each (*k*-1)-subset *s* of *c* **do**
      if (c_1 ∈ *s*) or ($MIS_{INFIS}(c_2)$ = $MIS_{INFIS}(c_1)$) **then**
        if (*s* ∉ $temp_{k-1}$) **then** delete *c* from C_k;

Table 1. Transaction database

TID	itemsets
T_1	A, B, C, E
T_2	B,D,E
T_3	B, C
T_4	A,B,D,F
T_5	A, C, F
T_6	A,C,D
T_7	A, D,E
T_8	A,B,E
T_9	A,C,D,E
T_{10}	B,C,D

4   Experimental Results and Comparisons

The transaction database is shown in Table 1. Let $MIS_{FIS}(A)=0.4$, $MIS_{FIS}(B)=0.4$, $MIS_{FIS}(C)=0.3$, $MIS_{FIS}(D)=0.3$, $MIS_{FIS}(F)=0.3$, $MIS_{FIS}(E)=0.2$ and $MIS_{INFIS}(A)=0.3$, $MIS_{INFIS}(B)=0.3$, $MIS_{INFIS}(C)=0.2$, $MIS_{INFIS}(D)=0.2$, $MIS_{INFIS}(F)=0.2$, $MIS_{INFIS}(E)=0.1$, the results generated by algorithm *Apriori_2L-XMMS* are shown in Table 2, where the itemsets with gray background are frequent itemsets and the itemsets with non background are infrequent itemsets.

Now we compare the results of *2L-XMMS* model with *XMMS* model. Given the same $MIS_{FIS}(*)$ and $ms=0.1$, the results generated by *XMMS* model are shown in Table 3. Because $MIS_{INFIS}(F)=0.2$, the itemsets whose supports are less than 0.2 are not infrequent itemsets in *Apriori_2L-XMMS*. We leave the corresponding position of these itemsets blank in table 2 to show the differences.

Table 2. *fis* and *infis* generated by 2L-XMMS model

1-itemsets		2-itemsets			
	s(*)		s(*)		s(*)
A	0.7	AB	0.3	BE	0.3
B	0.6	AC	0.4		
C	0.6	AD	0.4	CD	0.3
D	0.6	AE	0.4	CE	0.2
E	0.5	AF	0.2		
F	0.2	BC	0.3	DE	0.3
		BD	0.3		

3-itemsets			4-itemsets	
	ADE	0.2	ABCE	0.1
ABE	0.2		ACDE	0.1
	BCE	0.1		
ACD	0.2	BDE	0.1	
ACE	0.2			
	CDE	0.1		

Table 3. *fis* and *infis* generated by XMMS model

1-itemsets		2-itemsets			
	s(*)		s(*)		s(*)
A	0.7	AB	0.3	BE	0.3
B	0.6	AC	0.4	BF	0.1
C	0.6	AD	0.4	CD	0.3
D	0.6	AE	0.4	CE	0.2
E	0.5	AF	0.2	CF	0.1
F	0.2	BC	0.3	DE	0.3
		BD	0.3	DF	0.1

3-itemsets			4-itemsets	
ABC	0.1	ADE	0.2	ABCE	0.1
ABD	0.1	ADF	0.1	ABDF	0.1
ABE	0.2	BCD	0.1	ACDE	0.1
ABF	0.1	BCE	0.1	
ACD	0.2	BDE	0.1	
ACE	0.2	BDF	0.1	
ACF	0.1	CDE	0.1	

5   Conclusions

Infrequent itemsets are important because there are many negative association rules in them. Theoretically, infrequent itemsets are the complement of frequent itemsets and the number of infrequent itemsets is too large to be mined completely. So in real application, some constraints must be given so as to control the number of infrequent itemsets in a moderate degree. In this paper, we have proposed an efficient 2L-XMMS model by assigning 2-level minimum supports to every item to constrain frequent and infrequent itemsets respectively and designed an efficient algorithm Apriori_2L-XMMS to discover simultaneously both frequent itemsets and infrequent

itemsets based on MMS model. The comparisons and the experimental results show the validity of the algorithm.

References

[1]   Wu, X.-D., Zhang, C.-Q., Zhang, S.-C.: Efficient Mining of both Positive and Negative Association Rules. ACM Transactions on Information Systems, 381–405 (2004)

[2]   Haglin, D.J., Manning, A.M.: On Minimal Infrequent Itemset Mining. In: Proceedings of the International Conference DMIN, Las Vegas, Nevada, USA, pp. 141–147 (2007)

[3]   Dong, W.-J., Jiang, H., Chen, L., Liu, G.-L.: Incremental Updating Algorithm for Infrequent Itemsets on Weighted Condition. In: Proceedings in 2010 International Conference on Computer Design and Applications, ICCDA, vol. (1), pp. 36–39, 25–27 (May 2010)

[4]   Dong, X.-J., Zheng, Z.-Y., Niu, Z.-D., Jia, Q.-T.: Mining Infrequent Itemsets based on Multiple Level Minimum Supports. In: Proceedings of the Second International Conference on Innovative Computing, Information and Control, ICICIC 2007, pp. 528–532 (September 2007)

[5]   Dong, X., Niu, Z., Shi, X., Zhang, X.-d., Zhu, D.: Mining Both Positive and Negative Association Rules from Frequent and Infrequent Itemsets. In: Alhajj, R., Gao, H., Li, X., Li, J., Zaïane, O.R. (eds.) ADMA 2007. LNCS (LNAI), vol. 4632, pp. 122–133. Springer, Heidelberg (2007)

[6]   Dong, X., Niu, Z., Zhu, D., Zheng, Z., Jia, Q.: Mining Interesting Infrequent and Frequent Itemsets Based on MLMS Model. In: Tang, C., Ling, C.X., Zhou, X., Cercone, N.J., Li, X. (eds.) ADMA 2008. LNCS (LNAI), vol. 5139, pp. 444–451. Springer, Heidelberg (2008)

[7]   Dong, X.-J., Wang, S.-J., Song, H.-T.: 2-level Support based Approach for Mining Positive & Negative Association Rules. Computer Engineering 31(10), 16–18 (2005)

[8]   Dong, X.-J., Li, G., Wang, H.-G., Guo, Y.-B., Yang, Y.-Y.: Mining Infrequent Itemsets based on Extended MMS Model. In: ICIC 2007. CCIS, vol. 2, pp. 190–198. Springer, Heidelberg (2007)

[9]   Li, G., Wang, H.-G., Dong, X.-J., Yang, Y.-Y., Guo, Y.-B.: Infrequent Itemsets Mining Based on Two Level Multiple Supports. J. of Zheng Zhou Univ. (Nat. Sci. Ed.) 39(4), 94–97 (2007)

[10] Liu, B., Hsu, W., Ma, Y.-M.: Mining Association Rules with Multiple Minimum Supports. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 1999, San Diego, CA, USA, August 15-18, pp. 337–341 (1999)

Research and Design on Supervision System of Coal Mine Safety Based on Fieldbus

Zhu Li-Wang, Xu Jian-Bo, and Zhu Geng-Ming

School of Computer Science and Engineering, Hunan University of Science and Technology

Abstract. This paper researched the coal mine safety surveillance / monitoring technology systematically, and proposed an overall structure of the distributed safety supervision system which includes environment monitoring and person / vehicle location management. Then, it intensively analyzed the technical requirements of the system hardcore - supervision substation; and designed and developed the HW&SW of the substation, including circuits and relevant control programs of CAN bus interface and various function modules. Finally, this paper set up a simulation system in the laboratory, validated the feasibility of system framework and functions of the substation equipment.

Keywords: Mine safety, Fieldbus, CAN bus, Supervision substation, RFID.

1   Introduction

To protect the personal safety of the staff is the eternal theme of the coal industry. Those supervision systems of coal mine safety that have been put into operations played an important role to ensure the safe production of coal mining enterprises in china, but there are also some aspects to be improved which summarized as follows:

(1) The underground supervision substation (SST) is mainly used for methane monitoring, but a number of important environmental parameters and operation status of production equipments have not been included in the monitoring range [1].

(2) The backbone network (Host - SST) of some systems is based on RS-485 bus, so transmission distance and transmission speed are limited. The new industry standard for mine safety monitoring demands that transmission distance should not be less than 10km between the SST and transmission interface of the host computer without repeaters, therefore the traditional RS-485 technology have been unable to meet the relevant requirements [2].

(3) Most of the SSTs only monitor the environmental parameters and operating conditions, but person / vehicle locating and management system which could put forward instructions about the best routes for rescue and evacuation of persons after the incident needs to be re-built. Moreover, because of lacking information exchange, the effective coordination can not be ensured.

2   The General Framework of the Supervision System

2.1   Fieldbus Overview

For the entire supervision system of coal mine safety, data communication is the key to intelligentization , automation and networking, also an important tache for resources safety exploitation and informationization of mine safety management.

Accordance with the definition given by the IEC61158 Standard, fieldbus refers data bus installed among the field devices, as well as between field devices and automatic control devices in manufacturing or process area by means of digital, two-way, serial, and multi-point communications. Fieldbus has the following main technical characteristics:

(1) Openness system[3];
(2) Favorable mutual maneuverability and interoperability;
(3) A high degree of dispersion system architecture;
(4) Strong adaptability in field environment ;
(5) High ratio of performance to price [4].

At present, the more influential fieldbuses are FF, Profibus, HART, CAN, LonWorks and so on. Tab.1 shows the main technical characteristics of them.

Table 1. The main technical characteristics of several field buses

Features	FF	Profibus	HART	CAN	LonWorks
Typical scope of application	Instrument	PLC	Intelligent transmitter	Automobile	Automation
OSI hierarchy	1+2+3+8	1+2+7	1+2+7	1+2+7	1–7
Communication medium	Twisted pair wire, Cable, Optical fiber	Twisted pair wire, Optical fiber	Power signal line	Twisted pair wire, Optical fiber	Twisted pair wire, Power line, etc.
Medium access mode	Token Ring, Master-Slave	Token Ring, Master-Slave	Token Ring, Inquiry	Bit-arbitration	P-P CSMA
Error correction	CRC	CRC	CRC	CRC	CRC
Baud rate	2.5Mbps	1.2Mbps	1.2 Mbps	1 Mbps	1.25 Mbps
Maximum nodes	32	128	15	110	248
Priority	Embodied	Embodied	Embodied	Embodied	Embodied
Secrecy	-	-	-	-	Identification
Intrinsically safety	Yes	Yes	Yes	Yes	Yes
Transmitting range	1900m	1200m	1+2+7	10000m	2700m

2.2   The General Framework of the Supervision System

According to AQ6201-2006 (General-technical requirements of coal mine safety) issued by the State Administration of Work Safety and the relevant technical standards, in order to better solve the existing problems of previous supervision systems and provide a comprehensive monitoring platform that includes environment monitoring and person/vehicle location management, this article established an overall framework of supervision system for coal mine safety by integrating computer, digital communication and electronics technology, adopting three-layer distributed network architecture[5,6], as shown in Fig. 1.

[Notes] LAN: Local area network; CM : Coal machinery; EQ: Equipment; RFID: Radio frequency identification
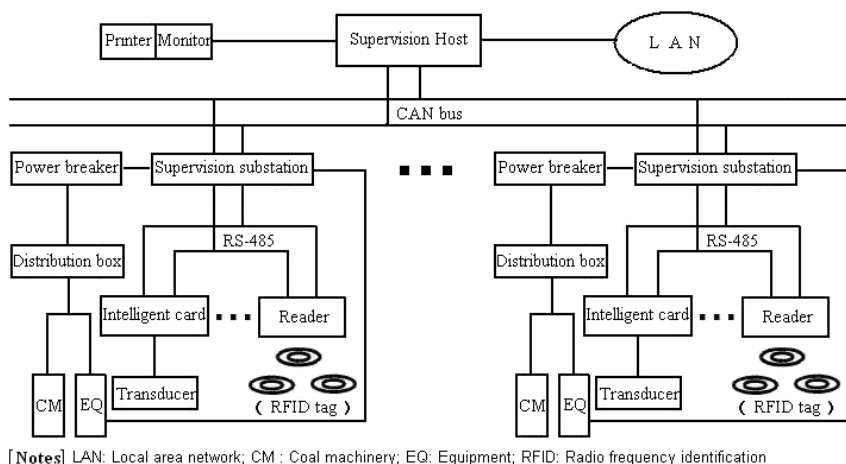
Fig. 1. The overall structure of the supervision system

The system consists of the ground supervision host (SHT), substation (SST), intelligent card, transducer, RFID tag &reader, and etc. SHT communicates with SSTs by field bus. Because CAN bus can reach 10km transmission distance and 1Mbps baud rate, therefore it is very suitable for communication between SHT and SST, so as to constitute the backbone network of the distributed control system. Most of the existing intelligent transducer products are equipped with RS-485 bus interface but CAN bus interface. RS-485 can reach 1200m transmission distance without repeaters, so it can covers a laneway or a mining face / heading face. Considering the technical feasibility and economic suitability, communication between substation and intelligent transducer may apply RS-485 bus. Types of transducer include methane concentration, CO concentration, water pressure, water level, and the tunnel wall surface displacement and stress, temperature, humidity, wind speed, air pressure, etc [7].

3   Hardware Design of the Supervision Substation

3.1   Structure of the Supervision Substation

The supervision substation is the hardcore of the supervision system[7]. According to the system framework and functional requirements of the substation, it can be divided

into the main control module, I / O module, human-computer dialogue and some other function parts, as shown in Fig.2.

3.2   CAN Bus Communication Interface

Supervision host communicates with substations by CAN bus interface. Adopting microprocessor AT89S55, CAN controller SJA1000, CAN transceiver PCA82C250, optocoupler 6N137, the CAN bus interface circuit structure can be shown in Fig.3. AT89S55 controls SJA1000 for initial setup and message transmitting and receiving, PCA82C250 is used to establish the physical layer of CAN protocol in order to achieve conversion between electrical levels and data.
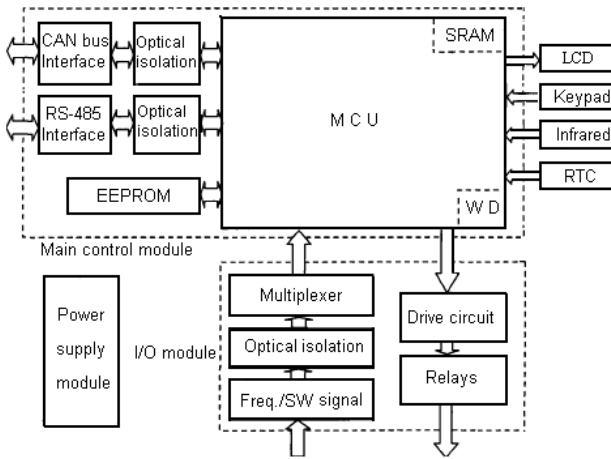
[Notes] RTC:Real-time clock; WD:Watchdog; Freq.: Frequency; SW: Switch

Fig. 2. The configuration of the supervision substation
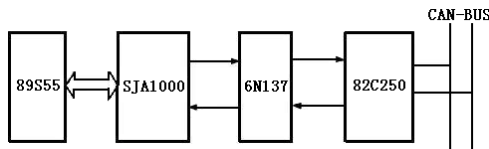
Fig. 3. CAN bus interface circuit structure

4   Software Design of the Supervision Substation

4.1   CAN Interface Communication Program

(1) Transmitting of message packets

MCU(AT89S55) first checks whether messages in the transmit buffer has been transmitted, if so, information collected by substation will be written to the transmit buffer of SJA1000, if success then set command register of SJA1000 and send messages, as shown in Fig.4.
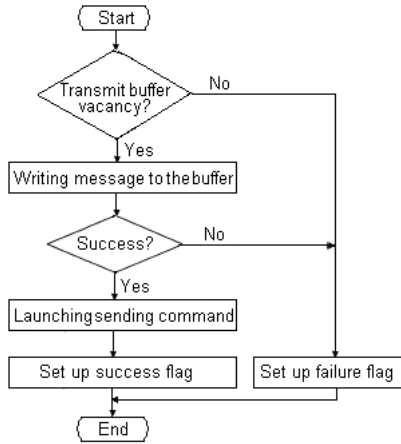
Fig. 4. Message-transmitting program flow chart

(2) Receiving of message packets

SJA1000 will move messages into the receive FIFO buffer via the acceptance filter, then generate a receive interrupt and set up a status flag for the receiver buffer. MCU processes interrupt, then read messages form the receive buffer of SJA1000, finally set the release flag in the command register of SJA1000 to release the buffer. The interrupt handler process is as shown in Fig.5.
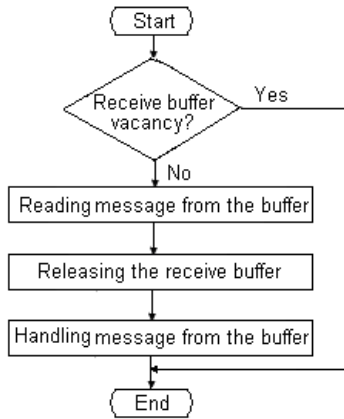
Fig. 5. CAN-Protocol message receiving process flow chart

4.2   The Main Program of the Supervision Substation

Tasks of the main program are as follows: system initialization, measurement data collection, person/vehicle location information uploading, implementing commands, and etc. Program functions to be performed include:

(1) Sending commands to intelligent card (or intelligent transducer);
(2) Collecting data form intelligent card (or intelligent transducer);
(3) Uploading data to the supervision host;
(4) Implementing commands sent by the supervision host.

4.3   Person / Vehicle Locating Function

Person / vehicle locating function is implemented by the radio frequency identification (RFID) tag & reader. RFID card reader fulfills the identification of person / vehicle and transfer the relevant messages to the supervision substation by RS-485 interface, then the substation upload the information to the supervision host through the CAN bus, finally personnel / vehicle location management function completed by the host software.

5   System Operation and Testing

After ten months of work, development for circuit board of the supervision substation was finished. In order to verify the feasibility of system framework and the main functions of the substation equipment, we set up a simulating mine safety supervision system in the laboratory, as shown in Fig.6, following equipments and parts are used: PC (as supervision host), USBCAN-II interface card (as USB-CAN converter), supervision substation, intelligent methane transducer, mine-specific RFID tag & reader, and so on.
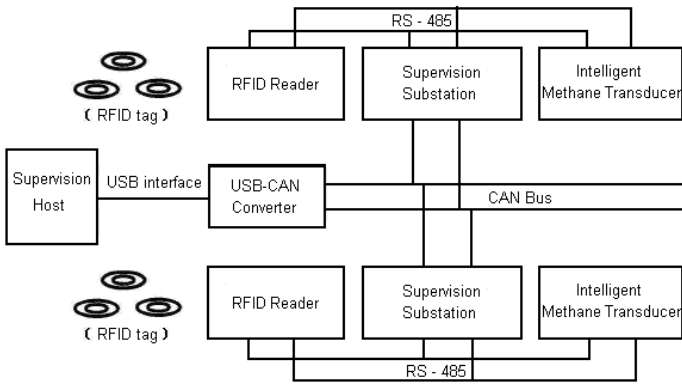
Fig. 6. Construction of the laboratory verification system

After testing and validation, the following functions have been implemented according to the relevant technical requirements.

(1) CAN bus communication between supervision host and substation;
(2) Data transmitting and receiving between supervision substation and intelligent methane transducer, Data transmitting and receiving between supervision substation and RFID reader;
(3) Necessary monitoring functions for coal machinery and equipments.

6   Summaries

This paper proposed a kind of the overall framework of coal mine safety supervision system by adopting reasonable structure and practical technology, then finished HW&SW design and development of the supervision substation. Finally, the feasibility of the system structure and functions of the substation equipment were validated. The developed substation equipment characterizes in fully functioning, reliable performance, convenient operation, and etc, so it can be suitable for mine safety supervision in coal industry.

References

[1] Lu, Z.-L., et al.: Summary of Mine Monitoring. Coal Technology, 85–86 (November 2008)
[2] Zhang, G.-S., Lin, A.-D.: Development Historical and Tendency of Mine Monitor System. Coal Technology, 8–9 (February 2009)
[3] Liu, W.-Q.: Study and Practice on Application of Fieldbus Technology in Coal Mine Supervision System. Journal of North China Institute of Science and Technology, 64–66 (June 2006)
[4] Liu, X.-R., Zhao, Y.-L., Feng, Z.-T.: Application of coal mine power monitoring system Based on Fieldbus. Shandong Coal Science and Technology, 57–58 (August 2008)
[5] Giron-Sierra, J.M., Insaurralde, C., Seminario, M.: CANbus-based Distributed Fuel System with Smart Components. IEEE Transactions on Aerospace and Electronic Systems, 897–912 (March 2008)
[6] Michael, J., Mike, B.: New CANbus Engine Control Panels Introduced. Diesel Progress, 110 (May 2007)
[7] Zhu, X.-S.: The Research and Implementation of the Coal Mine Safety Monitoring System Based on CAN-BUS. Xidian University, Xi'an (2007)

Methodology of Collaborative Design for PLM on Agricultural Machinery

Junming Hou and Dexu Yang

College of Engineering
Shenyang Agricultural University
`junming_hou@163.com`

Abstract. The advanced method of design and the increasing time-to-market of new products have place emphasis on the collaborative design to react to the global market. A product lifecycle management (PLM) methodology is proposed in the paper for the requirement. The methodology is to establish a structure for collaborative design of the project, design, process, proceed and product. Project represents the parts that relates to the equipment of product. Product is the key part of system, which all the other parts work for it. Firstly, the conceptions of collaborative design and the PLM are introduced. Secondly, the knowledge management and integration for system is applied for it. The methodology of collaborative design for PLM is introduced in detail. At last, an example of agricultural machinery is applied

Keywords: Collaborative design, PLM, agricultural machinery, knowledge management.

1 Introduction

With the development of the global market, enterprise meets the challenge of competition around world. The traditional mechanical manufacture style can not suit for global market, which needs a new method to meet the requirement. How to enhance the ability of enterprise's competition is a more important problem for enterprise. Computer support cooperative design (CSCD) is to apply computer technology to assist designer working in the engineering design field, which supports designer of different places work at the same time to complete the same task. King [1] considers that the fundamental issue of CSCD focuses on the computerization to establish a concept-sharing and seamless coordination among engineering design participants for concept formation. Collaborative design is a new method to meet the needs, which is a main method for networked manufacture. Product lifecycle management (PLM) links the different steps of the product design, which is computer aided engineering (CAD), computer aided engineering (CAE), and computer aided manufacture (CAM), product data management (PDM), and enterprise resources planning (ERP). Knowledge management (KM) system is present for collaborative design to heighten the speed of design.

Collaborative design works for PLM which integrate the enterprise and the associated knowledge efficiently. Design process is a knowledge-intensive and

collaborative task, which is the knowledge management process of the intensive process. The knowledge-intensive support is critical in the design process and is a key solution for future advantages in product development. Designers can share knowledge among the designers, manufacture process designers, quality management people and the assembly designer to achieve the best goal.

The product lifecycle is a methodology for product development. It is the goal that networks of participants exchanging services forms a concurrent model for integrated design. Contemporary design problems embody significant levels of complexity, which make it unlikely that a single designer can work alone on a design problem. The continuing growth of knowledge and supporting information and the ever-increasing complexity of design problems have led to the increasing specialization. There are many researchers on the PLM about the collaborative design.

2   Literature Review

Knowledge management is an advanced concept for collaborative design. A successful knowledge management system[2] enhances the way people work together, enables knowledge workers and partners to share information easily, so they can build on each other's ideas and work more effectively. The goal of knowledge management is to gather company proprietary knowledge in order to come up with the best decision-making, or to quickly seize the initiative with innovative ideas.

McGregor and Schiefer [3] introduced an approach to reclaiming and improving knowledge for an organization by establishing a framework that enables the definition of web services from a performance measurement perspective, together with the logging and analysis of web services. Kuo and Soflarsky [4] dealt with the development of an automated procedure for selecting motor carriers to minimize the transportation cost prepaid by a shipper. The aim of this paper is to present a methodology to design and integrate 'Collaborative Engineering Environments' supported by tools that enable cooperative work and intellectual capital sharing[5]. The project presented in this paper concerns the implementation in the Dutch construction industry of a methodology for sharing product information through a distributed object model [6].Hilary Cheng [7] presents an ontology-based approach for business intelligence (BI) applications, specifically in statistical analysis and data mining. The resulting knowledge from each experiment defined as a knowledge set consisting of strings of data, model, parameters, and reports are stored, shared, disseminated, and thus helpful to support decision making. Aeronautic industry and various auto manufactures have experienced substantial improvements and reduce design rework, reporting cost savings and reduction in development time as a result of the use of virtual environment technologies and digital models [8].Knowledge management (KM) [9] is a scientific discipline that stems from management theory and concentrates on the systematic creation, leverage, sharing and reuse of knowledge resources in a company. Knowledge can be generated by capturing existing (shared) knowledge through filtering, storing, retrieving and disseminating explicit knowledge and by creating and testing new knowledge [10].

3   Basic Concepts

3.1   Collaborative Design

Collaborative design is a complex process for modern design of enterprise, which assist designer to complete a task in a cooperative style. Collaboration is viewed as the next big wave faster e-Commerce, digital commerce and several other variants that emerged in the last few years [12-13]. There are seven steps to manage the collaborative design.

Collaborative design is defined as the following: Collaborative design is a process that the designers take part in the same task, which they are in different team. Designers complete the task in a contracting and parallel style, which they are in charge of their jobs. Finally, the task is complete in according to the requirement.

The collaborative design for the machine is a complex process. The characteristics of collaborative are the following:

1) The design is the cooperative. The design for mechanical production needs designers to work at the same time. Many designers take part in the same design. They are CAD designers, CAE analysing people, technological design people, and sale service people. The correlation of machine and the multi-subject make the process of design is collaborative. The information communication and feedback can be realized to achieve the final object.

2) The sharing and interaction of design information is realized. The serial and parallel relations all need the exchanging data and the sharing data. The amount of data is large. The accurate information is required in the collaborative design. So the data should be exchanged in a right way.

3) The complexity of design process is its characteristic. It is a complex process to develop a machine. The process can be divided into a few sub-processes. The parallel and serial relations are all exist. The layers and complexity of the activities make the design more complex.

Fig.1 shows the process of collaborative design.
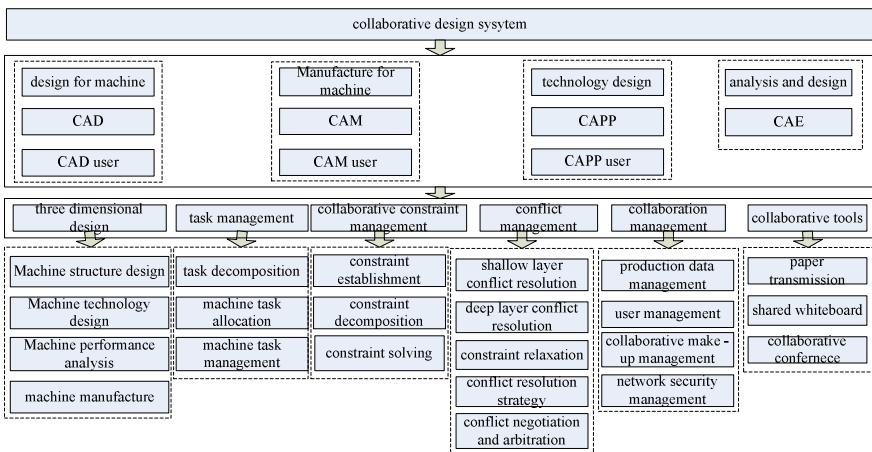
Fig. 1. Process of collaborative design

3.2   PLM Concepts

How to increase the ability of the collaborative design is a point problem. The collaborative design can support the product lifecycle management (PLM) to increase the efficiency of product development. PLM refers to the management of the product lifecycle from birth to obsolescence of a product [4]. During this lifecycle management process, the entities involved can be managed internal and external with internal and external collaboration facilitating the overall process. PLM is concerned with the product design, evaluation, market assessment, product strategy, manufacture process, product processing and the final lunch to mass production or customization depending on the industry. PLM is also related to CAD, CAM, CAPP, CAE, PDM and ERP.

(1) CAD is the main point for collaborative design, which includes the basic design tools (UG, CATIA, AutoCAD and Pro/E).

(2) CAM is the computer aided manufacture, which is the main object of PLM.

(3) CAPP is the computer aided process planning, which manage the manufacture process of product

(4) CAE is the computer aided analysis, which is to analysis the character of parts for machine.

(5) PDM is the product data management, which manage the data of design and manufacture process.

(6) ERP is the enterprise resource planning, which can manage the data of the enterprise of the whole data.

PLM is a concept that aims to integrate the various processes and phases involved during a product lifecycle, the basic process of design for PLM is the following: there is no single tool or package that can be termed as a PLM package. Organizations adopting PLM need to follow that is process driven. With such a framework, organizations that are embracing PLM will be able to successfully collaborate since the partnering enterprises will work as a seamless extended enterprise. All following a single unified process and not disjointed and fragmented processes are only integrated through data bridges. PLM is a concept that based on horizontal business processes as compared to vertical business units in organization. These business units are linked based on product life stages and processes. Teams and unit formation needs to be studied with an aim to provide an effective structure that can utilize the benefits of PLM and technology. PLM is a whole conception of the knowledge management system. All the distributed system is related to the PLM. In order to create a scalable PLM or collaborative strategy, studies on organization structure for PLM are operated.

3.3   Knowledge Management

Knowledge-based techniques can be applied effectively to achieve goals by leveraging multiple facets on a concurrent basis. These facets are the following [14]. Fig.3 is the knowledge technique for management.
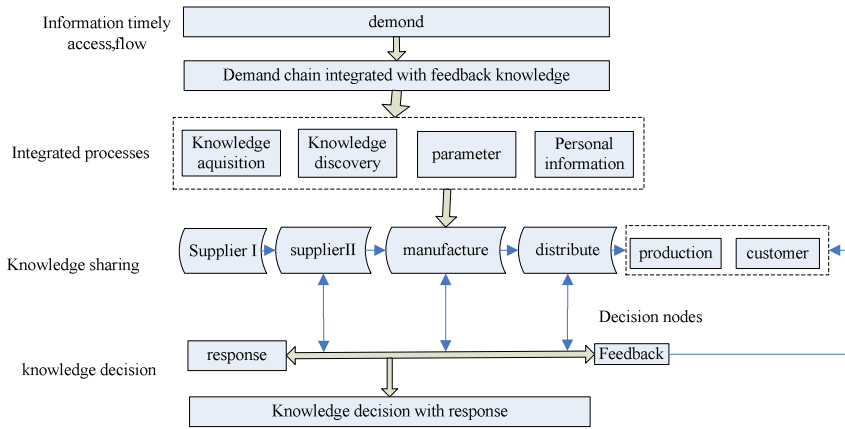
Fig. 2. PLM structure
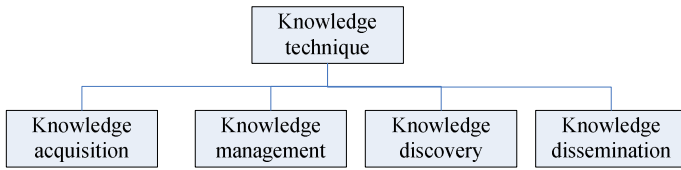
Fig. 3. Knowledge management structure

Knowledge acquisition, which characterizes the process of capturing information from various modalities, including people's minds and hand written documents into computer accessible media. Multi-agent system is applied in the collaborative design. Knowledge management, which deals with mitigating issues related to heterogeneities in underlying contexts of information coming from disparate sources such as multiple stakeholders, multiple projects and multiple stages of the process. Knowledge discovery, which involves using emerging techniques to analyze huge amounts of information and to automatically capture underlying knowledge that can provide better insights into the relevant information. Knowledge dissemination, which provides the automated extraction of the most relevant pieces of information from a huge computer based information infrastructure, with such extraction being tailored to the needs of different constituencies of users in each of the relevant set of organizations.

The integration of requirement and the reuse is the important part of the system. Fig.4 shows a diagram of the framework components. The arrows indicate general process flow: relationships can be created between all of the constituent elements. The remainder of this section discusses the constituent elements of the framework. The process for applying the framework begins with requirements capture. In the second stage, the requirements are analysed and selected and used to create a technical specification and product function structure. This specification is applied to a parameterised product specification and a product structure. The process-based knowledge reuse method is applied to the process and data transactions throughout the application of the framework.
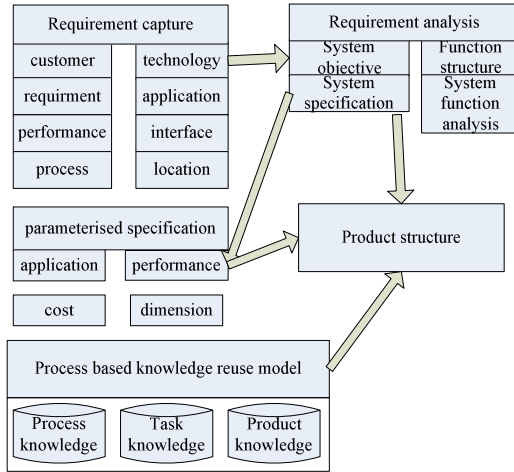
Fig. 4. Integration of knowledge requirement and reuse

4   Knowledge Integration

4.1   Methodology for Collaborative Design

The integration of PLM is to make projects, products, processes, proceeds and products into the same system, which assist designers and manager to enhance efficiency. The methodology is oriented to product, which is the critical part of PLM. The methodology is also to construct an approach to define the integration of system. The following is the definition of the projects, products, processes, proceeds and products. Project module represents all the entities relating to the organization, the resources (human and equipment). Project module is defined in the application environment.

Product represents all information which characterizes product contents in a systematic way. A product has different representations according to the predefined abstraction levels. For example, a connecting rode is a product in a physical way in a high abstraction level but it has a cylinder as the topological and geometrical representation in a low abstraction level.

Proceed represents abstracted definitions related to a sequence of physical steps which lead to the modifications of the product. For example, a manufacturing proceed can be represented in the Trade Environment by a set of manufacturing rules which can be refined into features in the Common Environment.

Fig.5 is the architecture of system. Process represents a succession of tasks whose implementation contributes to the modification of the product. Process representations are different according to the corresponding environments. In the application environment, a process is a workflow which defines the collaborative process between all the partners while a process in the Trade Environment is a set of logic rules for the trade.
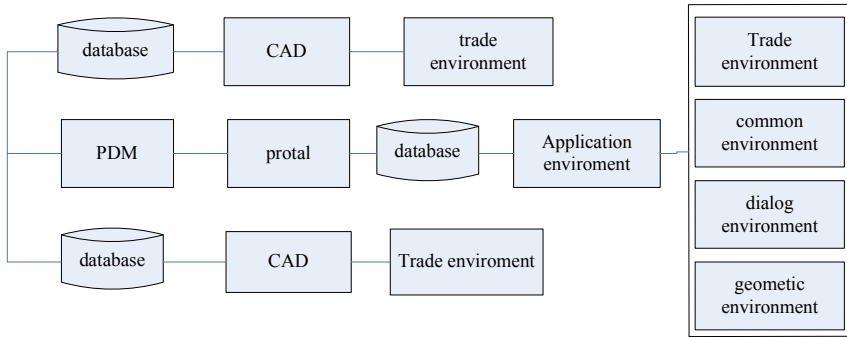
Fig. 5. System architecture

4.2   Methods of Knowledge Management in Collaborative Design

Fig.6 is a case diagram for collaborative design. The designer, designer manager and the computer leader are the main role of the system. They complete the design task and operate the database on different layer. The diagram expresses the relations of the roles for PLM.
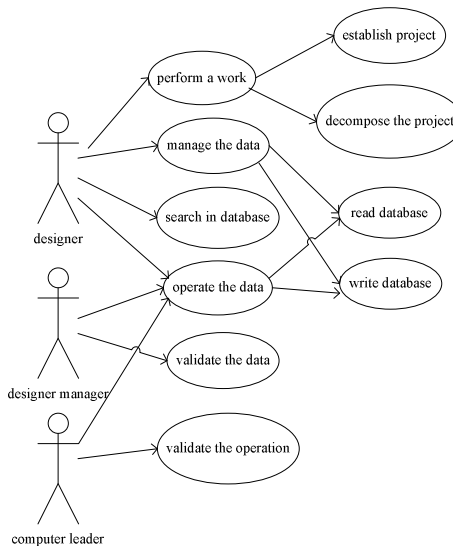
Fig. 6. Case diagram of collaborative design

There are seven steps for knowledge management in collaborative design. The following is the detail steps.

(1) The step takes records of all the equipment of the partners: CAD software and CAE software. The step can exchange the data and the communication among the partners.

(2) The step records the resources who are implicated in the development of the product. Different workspace is defined too. A workspace is a localisation on a portal where the actors exchange document between them. The designer can be used if the actor is allocated.

(3) The step records constraints, trade rules and scenarios. Each partner records their information in a private way. In this step, a vocabulary of each domain – ontology – is also defined if it is not already done or not complete if it already exists.

(4) In the step, system can record the information. These pages are launched when the partners want to resolve problems.

(5) The step records the global workflows. This step allows to record proceeds of co-design. At each stage, the beginning and end dates, resources allocated, as well as the responsible person who is supposed to validate the stage. The workflow is a suite of tasks, which calling to scenarios. The results in proceeds achieve modifications of the product.

(6) Collaborative step: the step is the last step, the designer go the work.

4.3   Example of Knowledge Management

(1) A shaft of agricultural machinery is designed with the methodology. As shown in the before, the methodology is present, the shaft is an important part of the machinery. The software is the Pro/E and UG. The partners work with the same task in the same time. The designer exchanges the data and communicates with others. The designer in different group takes the information and acknowledges the software of the partners. The designer collects the general information on human resources and material.

(2) In the beginning of the step, partners define their ontology in the trade environment. For example, one designer defines all his trade vocabulary in order to manipulate their parameters-clearance,gear parameter, etc. Then the designers apply their own information to during the production development. The main parameter and the relation information are all used for development. When one designer completes the task, the other designer receives the information. All the designers complete the work. Designers define the parameters in the form of agricultural machinery.
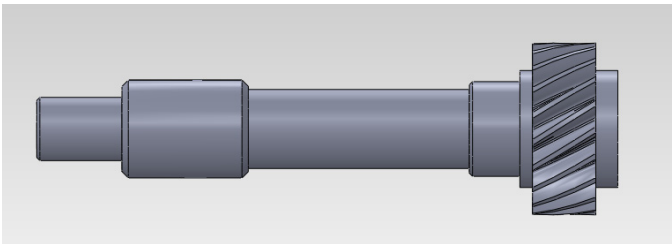
Fig. 7. Design for shaft of agricultural machinery

(3) The collaborative designers are launched when a problem occurs during the development of the product and requires the participation of other partners to resolve it. The collaborative design environment is established for the designers. When the task is modified, the other designer can see the modification. Then they complete the design task at the same time.

(4) An appointment is automatically inserted in sharing schedule in order to resolve the problem by videoconferencing if it is possible with the model available on the portal to annotate it. Collaborative information has been defined to optimise the product.

(5) All partners meet together in order to define the global workflow of the project. This workflow is then stored and shared on the portal. Each partner defines the task to bring its know-how in the early stage of the development. The actions that need a validation between different actors are defined as a collaborative task. At any moment of the development some tasks can be added to the workflow. Once these different steps are totally defined, the project starts.

(6) The prime manufacturer begins the design of the shaft. To assure a correct design of the part, the designer executes a design scenario by setting up graph numerical parameters and his constraints rules. Once the design task of the workflow is finished, the prime manufacturer, the analysis engineering and the subcontractor start a collaborative session like video-conferencing in order to set up the simplification parameters.

5   Conclusions

With the development of the increasing global market, collaborative design has been an important design style for manufacture. A PLM methodology of collaborative design for agricultural machinery is present in the paper. There are six steps for the PLM of collaborative design. The Case diagram is applied for the system. Knowledge management system is applied for the integration of PLM. The methodology for collaborative design is applied. At last, an example of agricultural machinery is showed to explain the methodology, which can assist to realize the collaborative design for PLM.

References

[1]   Trofino, J.: Voice activated nursing documentation: On the cutting edge. Nursing Management 24(7), 40–42 (1993)
[2]   Gupta, A.: A four-faceted knowledge-based approach to surmounting borders. Journal of Knowledge Management 5(4), 291–299 (2001)
[3]   McGregor, C., Schiefer, J.: A web-service based framework for analyzing and measuring business performance. Information Systems and eBusiness Management 2(1), 89–110 (2004)
[4]   Kuo, C.C., Soflarsky, F.: An automated system for motor carrier selection. Industrial Management and Data Systems 103(7), 533–539 (2003)
[5]   Mejiá, R., López, A., Molina, A.: Experiences in developing collaborative engineering environments:An action research approach. Expert Computers in Industry 58, 329–346 (2007)
[6]   van Leeuwen, J.P., van der Zee, A.: Distributed object models for collaboration in the construction industry. Automation in Construction 14, 491–499 (2005)

[7]   Cheng, H., Lu, Y.-C., Sheu, C.: An ontology-based business intelligence application in a financial knowledge management system. Expert Systems with Applications 23, 196–198 (2008)

[8]   Bochenek, G.M., Ragusa, J.M.: Virtual collaborative design environments: a review, issues, some research, and the future. In: International Conference on Management of Engineering and Technology, PICMET 2001 (2001)

[9]   Awad, E.M., Ghaziri, H.M.: Knowledge management. Prentice Hall (2003)

[10] Nemati, H.R., Steiger, D.M., Iyer, L.S., Herschel, R.T.: Knowledge warehouse: An architectural integration of knowledge management, decision support. Artificial Intelligence and Data Warehousing, Decision Support Systems (33), 143–161 (2002)

[11] Gray, J.J., McIntire, D., Doller, H.: Preferences for specific work schedules: foundation for an expert system scheduling program. Computers in Nursing 11(3), 115–121 (1993)

[12] Milan, J., Munt, C.: A modern, fully integrated hospital information system. In: Lun, K.C., Degoulet, P., Piemme, T.E., Reinhoff, O. (eds.) Proceedings of the Seventh World Congress on Medical Informatics, Medinfo 1992, Geneva, pp. 236–240. Elsevier Science Publishers, Amsterdam (1992)

[13] Tichkiewitch, S., Brissaud, D.: Integration of life-cycle constraints in design activity. Knowledge Intensive CAD, 119–130 (2000)

[14] Sharma, A.: Collaborative product innovation: integrating elements of CPI via PLM framework. Computer-Aided Design 37, 1425–1434 (2005)

[15] Lu, Y.-C., Cheng, H.: Automated optimal equity portfolios discovery in afinancial knowledge management system. In: Computational Intelligence in Economics and Finance, pp. 387–402. Springer (2003)

[16] Gzara, L., Rieu, D., Tollenaere, M.: Product information systems engineering: an approach for building product models by Reuse of patterns. Robot. Comput. Integr. Manufact. 19(3), 239–261 (2003)

[17] Eynard, B., Gallet, T., Nowak, P., Roucoules, L.: UML based specifications of PDM product structure and workflow. Computers in Industry 55, 301–316 (2004)

An Adaptive Threshold Method Based on the Local Energy of NSCT Coefficients for Image Denoising

Liu Xiyu[1], Yao Xiaolan[1], and Chen Xin[2]

[1] School of Automation,
Beijing Institute of Technology
liuxiyu1987@126.com, yaoxiaolan@bit.edu.cn
[2] Jinan Diesel Engine Co. Ltd.,
Jinan, China
chenxin106@126.com

Abstract. In this paper, we propose a new adaptive denoising method based on nonsubsampled Contourlet transform(NSCT). Traditional Wavelet provides only three directional components so that its geometrical property is not well, and Contourlet transform lacks translation invariance, therefore NSCT is developed. The proposed algorithm can adapt different thresholds on different scales and different directions. Further more, we use different thresholds in a directional subband according to local energy of NSCT coefficients to overcome the disadvantages of the unified threshold de-noising method and other fixed thresholds, which cause the image fuzzy distortion because of "over-killed". The experimental results prove that the algorithm outperforms existing schemes in both peak-signal-to-noise-ratio (PSNR) and visual quality.

Keywords: Nonsubsampled Contourlet transform, Adaptive threshold, Image de-noising, Translation invariance, Local energy.

1   Introduction

Images are inevitable to be corrupted by additive noise during acquisition and transmission.In order to get anticipant result, the image usually should be preprocessed before depth processed. That is to say, the additive noise should be removed while the important signal features being retained as much as possible.There are mainly two general classes of image de-noising methods, one is the space domain image de-noising ,and the other is the frequency domain image de-noising. As one of frequency domain de-noising methods, Wavelet Transform was once the main algorithm because of the characteristic of multiresolution, low entropy, multidirection, and the better approaching ability of singularity than Fourier transform[1].

Traditional Wavelet transform used in image processing   is generated by tensor product of two independent one-dimensional Wavelet functions, then it can only provide three directions: horizontal, vertical and oblique. Wavelet transform can not capture linear singularities well for its limited directions. For this reason E.J.Cades[2] prososed Redgelet transform and some scholars pioneered Curvelet transform. In 2002, a mutiscale, local, mutidirectional image representation method named Contourlet transform was put forward by M.N.Do and Martin Vetterli[3]. Contourlet

transform is a real 2-D image representation, which is constructed by Laplacian pyramid(LP) and directional filter banks(DFB), the former captures discontinous pointers, and the latter links them into linear strutures[4]. Image de-noising with the method of Contourlet transform introduces Gibbs-like phenomena around sigularities because of the lack of shift invariance. Due to the above reasons, Cunha, Zhou and Do proposed a more flexible and translation invariant method named nonsubsampled Contourlet transform(NSCT)[5], which is used in this paper.

Such frequency domain image de-noising metholds mainly have three developed aspects: 1) development of transform tools.2) more optimized choice of thresholds 3) improvement of the threshold function. The author introduces an adaptive threshold method, and divides the directional subband coefficients into three types according to the local energy: noise coefficients, image coefficients and neutral cofficients. Then they will be dealt with differently.

2   Nonsubsampled Contourlet Transform

NSCT is on the basis of Contourlet transform and it is divided into two shift invariant parts: nonsubsampled pyramid(NSP) and nonsubsampled directional filter bank (NSDFB). The former ensures the mutiscale property and the latter directionality. However, NSCT eliminate downsampling operation and upsamping operation during the decomposition and reconstruction of the image. Figure 1 shows the non subsampled filter bank(NSFB including NSP and NSDFB) structure and idealized frequency partitioning obtained with them.

NSP is constructed by interating two-channel nonsumsampled filter bank to achieve the multiscale decomposition. In each interation, we will get a bandpass subband and a lowpass subband which is the input of the next level.

NSDFB is also a two-channel nonsubsampled filter bank. All filter banks in the nonsubsampled directional filter bank tree structure are obtained from a single NSFB with fan filters[5]. An L-level tree structure decomposition leads to 2^L subbands with wedge-shaped frequency partition[6]. Fig.2 shows the process of NSCT: nonsubsampled pyramid decomposition (left) and nonsubsampled directional decomposition(right).
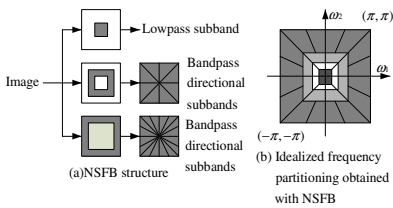
Fig. 1. NSCT structure
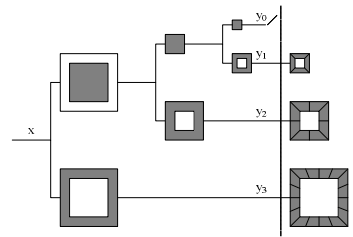
Fig. 2. Process of NSCT

3   Image Denoising Based on NSCT

3.1   Image Denoising Theory

Suppose that the size of the original image is $M \times N$,then the image with noise to be disposed can be expressed as:

$$\{g(i,j) = f(i,j) + n(i,j); i = 1,2 \cdots M, j = 1,2 \cdots N\}$$

Where, $\{f(i,j)\}$ is the original image signal and $\{n(i,j)\}$ is the noise signal. There are additive noise and multiplicative noise, and we choose the most common white Gaussian noise as the noise model in this paper. NSCT is linear transformation, so the coefficients aftter NSCT are divided into two independent parts: coefficients of original image and coefficients of noise. What's more, NSCT coefficients of original image is large and concentrated while NSCT coefficients of noise is small and dispersive. Therefore, a proper threshold methold can be set to deal with them. The image signal can be recovered after reconstuction with the new NSCT coefficients.

3.2   Adaptive Threshold

According to the analysis of image de-noising theory, we know that how to select the best threshold is the key. Denoho and some others give an universal threshold named VisuShrink[7] :

$$T = \sigma\sqrt{2\lg N} \tag{1}$$

Where, σ is noise standard deviation, N is the length of signal length, which is number of pixels in a image. Later, BayesShrink threshold, SureShrink threshold and many other methods are proposed.

Coefficients decrease along the decomposition levels after 2-D discrete Wavelet transform, and VisuShrink threshold becomes the following equation:

$$T = \sigma_j\sqrt{2\lg n_j} \tag{2}$$

Where, σ_j and n_j are   the noise standard deviation and coefficients amount in the jth level. The algorithm in this paper is an improved scheme from VisuShrink threshold. Considering that NSCT coefficients attenuate exponentially along the levels[8] and the primary function has its characteristic, We apply the method in NSCT domain. A new formula is given:

$$T_{sd} = \sigma_{sd}\sqrt{2\lg N / 4^s} * 2^{(1-s)} \tag{3}$$

Where, s represents the sth layer, d represents the dth direction, σ_{sd} represents noise standard deviation of the dth directional subband in the sth layer. σ_{sd} can be estimated by the robust extimator

$$\sigma_{sd} = Median(| C_{sd} |) \tag{4}$$

Where, C_{sd} denotes high-frenquency coefficients of the dth directional subband in the sth layer.

Equation (3) cosidered the levels and the directions, but it didn't considered the relativity inside the subband. Like traditional Wavelet, NSCT coefficients and energy concentrated. That is to say, probability of large coefficients emerge in the neighhood of large coefficient is bigger[9]. NSCT coefficients of original image is large and concentrated while NSCT coefficients of noise is small and dispersive. For these reasons, we amend Equation (3). Threshold at (i,j) in the dth directional subband of the sth layer is:

$$T_{sd}(i,j) = kT_{sd} \tag{5}$$

k changes with the local energy. Coefficients can be divided in to three types according to the relations between local energy and T_{sd}^2: noise coefficients, image coefficients, neutral coefficients. The first boundary is the coefficient whose average local energy equals T_{sd}^2, and the second boundary is the coefficient whose total local energy equals T_{sd}^2. Local energy with (i, j) as the center is calculated by

$$E(i, j) = \frac{1}{w \times w} \sum_{i, j \in B} Y_{(i,j)}^2 \qquad (6)$$

Where, $w \times w$ is the window size with (i, j) as its center, $Y_{(i,j)}$ is the coefficient at (i, j) of dth direcitonal subband in the sth layer. $Y_{(i,j)}$ is kept if $E(i, j) \geq T_{sd}^2$, while it is set to 0 when $w^2 E(i, j) < T_{sd}^2$. Otherwise, it is shrinked. Then we get a formula as follows:

$$k = \begin{cases} 1 & E(i, j) < \dfrac{T_{sd}^2}{w^2} \\ 1 - (E(i, j) - \dfrac{T_{sd}^2}{w^2}) / T_{sd}^2 & \dfrac{T_{sd}^2}{w^2} \leq E(i, j) < T_{sd}^2 \\ 0 & E(i, j) \geq T_{sd}^2 \end{cases} \qquad (7)$$

Window size is usually 3×3, from Equation (7) and (5), we know that threshold can be adjusted adaptively.

We introduce a new adaptive threshold method, and the concrete steps are shown below:

1. Perform multiscale and multidirectional decomposition of the image corrupted by additive white Gaussian noise with NSCT, and determine the number of levels and directions in each level.

2. Estimate the noise standard deviation σ_{sd}, and calculate T_{sd}.

3. Compute average local energy with Equation (6) at the position (i, j) in each subband, and get k using Equation (7) and T_{sd}.

4. Calculate adaptive threshold $T_{sd}(i, j)$ through Equation (5).

5. Use hard threshold function to deal with the NSCT coefficients of the image with noise.

6. Reconstruct the denoised image with new coefficients.

4   Experimental Results

In this section, we dispose images with different noise levels. The experiment can prove the superiority of the threshold method and the algorithm in this paper by comparing the results with other schemes such as Wavelet, Contourlet, and other threshold methods. We evaluate the denoising results subjectively and objectively. Subjective way means visual quality. And the objective evaluation criteria mainly has three method: peak-signal-to-noise-ratio (PSNR), mean square error (MSE) and signal-noise-ratio (SNR). In this paper, we choose PSNR. Suppose image is $M \times N$, then

$$PSNR = 10\log_{10}\frac{255^2}{MSE} = 10\log_{10}\frac{255^2}{\dfrac{1}{M \times N}\sum\limits_{i=1}^{M}\sum\limits_{j=1}^{N}(\tilde{f}(i,j) - f(i,j))^2}$$

where, $\{\tilde{f}(i,j)\}$ is the de-noised image.

a) original image   b) noisy image   c)VisuShrink method
                                          based on Wavelet

d) VisuShrink method   e) threshold in this paper   f) VisualShrink method
   based on Contourlet   based on Contourlet   based on NSCT

g) $3\sigma^{[5]}$ based on NSCT   h) threshold in this paper
                                        based on NSCT

Fig. 3. Results of different algorithms

Table 1. PSNR of different denoised images with different algorithm

image	σ	Noisy image (dB)	Wavelet VisuShrink (dB)	contourlet		NSCT	
				Visu-Shrink (dB)	Threshold in this paper (dB)	Visu-Shrink (dB)	3σ (dB)	Threshold in this paper (dB)
barbara	15	24.33	25.80	26.24	27.51	26.24	28.06	29.57
	20	21.86	24.57	25.65	26.37	25.12	26.81	27.76
	25	20.00	23.79	24.94	25.30	24.08	25.86	26.44
	30	18.49	23.19	24.34	24.35	23.40	25.09	25.36
	35	17.24	22.73	23.76	23.48	22.93	24.44	24.52
	40	16.17	22.31	23.40	22.65	22.60	23.89	23.93
lena.	15	24.17	30.39	28.69	30.23	29.53	31.06	32.50
	20	21.77	29.12	28.18	28.75	28.28	29.75	31.07
	25	19.90	27.99	27.43	27.45	27.47	28.80	29.91
	30	18.35	27.00	26.83	26.20	26.84	27.95	28.73
	35	17.08	26.17	26.33	25.18	26.34	27.35	27.96
	40	16.01	25.42	25.86	24.12	25.91	26.68	27.03

Fig3 shows the denoised images with different algorithms and the noise standard deviation is $\sigma = 25$. Fig 3 and Table 1 shows that the proposed algorithm achieves better performance. It can get both higher PSNR and Better visual quality. We can see block effects of Wavelet transform in Fig.3(a). Our threshold method based on Contourlet transform and $3\sigma^{[5]}$ threshold based on NSCT can also get good texture through Fig.3(e) and Fig.3(g). However, Contourlet transform has Gibbs-like phenomenon and we can see obvious oscillation points, while 3σ[5] method makes denoised image fuzzy distortion and worsen the weak edge because of "over-killed".

5   Summaries

This paper presents an adaptive threshold algorithm based on the local energy of NSCT coefficients. NSCT is shift-invariant and it has good geometrical property. The appropriate threshold can be chosen according to the local energy inside a subband while considering different decomposition scales and different directions. The experimental results show that our method is superior to other algorithms both in PSNR and visul quality.

References

[1] Liu, Z.: Image denoising via nonsubsampled wavelet based contourlet transform. Journal of Optoelectronics Laser 20(7), 954–958 (2009)
[2] Candes, E.J.: Ridgelet:theory and application. Department of Statistics, Stanford University, USA (1998)
[3] Do, M.N., Vetterli, M.: The Contourlet Transform: an efficient directional multiresolution image representation. IEEE Trans. on Imaging Processing 14(12), 2091–2106 (2005)

[4] Wang, X.-K., Gao, J.-H.: Image Denoising Method Based on Nonsubsampled Contourlet Transform and Bandelet Transform. In: The 1st International Conference on Information Science and Engineering, pp. 1278–1281. IEEE (2009)

[5] Da Cunha, A.L., Zhou, J., Do, M.N.: The nonsubsampled contourlet transform: theory, design, and applications. IEEE Transactions on Image Processing 15(10), 3089–3101 (2006)

[6] Li, M., Wang, T.: The Nonsubsampled Complex Contourlet Transform for image denoising. In: 2010 2nd International Conference on Computer Engineering and Technology, vol. 3, pp. 267–270. IEEE (2010)

[7] Donoho, D.L.: De-noising by Soft—thresholding. IEEE Trans. on Inform Theory 41(3), 613–627 (1995)

[8] Li, H.-J., Mei, X.: Adaptive threshold infrared image denoising algorithm based on NSCT. Computer Engineering and Design 29(12), 3230–3233 (2008)

[9] Zhou, D.-W., Shen, X.-L.: Adaptive Image Denoising Using Neighboring Wavelet Coefficients. Journal of Image and Graphics 13(11), 2112–2116 (2008)

A New Depth Extraction Method Based on Fusion of Motion Information and Geometry Information

Lijuan Wen, Zhiyi Qu, and Lei Shi

Lanzhou University

Abstract. In order to extract the accurate depth map which accords with human vision, a new depth extraction method based on fusion of motion information and geometry information was proposed in the paper. First, to extract the stationary background image from the video sequence with stationary background and moving foreground, the gaussian mixture model algorithm was adopted, then the paper adopted background difference method with adaptive threshold to extract complete moving objects area. Furthermore, the depth map of background image was computed based on geometry information, similarly, the depth map of moving foreground was computed based on motion information. Finally, the complete and precise depth map was obtained using the method based on fusion of the depth map of background image and the depth map of moving foreground.

Keywords: Background model, Moving objects extraction, Extracting the depth map by geometry information, Extracting the depth map by motion information, Linear perspective method.

1 Introduction

As improved quality of people life, the clarity of images and video is also higher and higher. At the same time, along with the everchanging development of hardware calculation, the transition (the transition is a new development direction in the images and video) from the planar display to the three-dimensional display is gaining attention from the academic circles. However, the key technology in the transition from the planar display to the three-dimensional display is to obtain the precise depth map.

At present extracting the depth map of image has two methods approximately: First, the depth map of image is computed based on geometry information, the method requires some geometry information such as baseline, evanishment line and so on. Second, the depth map of image is computed based on motion information, namely the depth map of image is obtained according to the motion information of adjacent images. The paper presents a new depth extraction method based on fusion of motion information and geometry information, the method improves the accuracy of depth map greatly and conforms to human vision.

2   Background Model Establishment and Moving Objects Extraction

2.1   Background Model Establishment

In the algorithm of using background difference method to obtain moving objects, the background structure is the key step. To reduce the influence to motion segmentation caused by the dynamic scene change, the majority of researchers devoted themselves to the research of different background model. Haritaoglu etc carried on statistical modelling for each pixel with the minimum strength value, maximum strength value and the greatest time difference value[2], and updated the background periodically. Karmann, Brand and so on adopted adaptive background model based on kalman filtering. It is quite complex to realize these methods, because background changes are not very complex in motion tracking, this paper adopted the method based on gaussian mixture background model mentioned in [1] to construct the background image.

2.2   Moving Objects Extraction

Although people's movement and light and other environmental factors change will have certain influence to the background, the change is small, it can be generally neglected, therefore this paper adopted background difference method. Background difference method is a technology which uses the difference of the current image and the background image to detect moving regions. Generally, It can provide the most complete characteristic data, and it is also far lower than optical flow method in computation[3].

The paper adopted the background difference method with adaptive threshold mentioned in [3] to extract moving objects image, because of the expansion operation in the method, the moving objects image obtained is not very precise. In order to overcome the defects, this paper also presented the algorithm proposed in [4] based on fusion of the mean-shift color image segmentation algorithm and moving objects image to extract the complete and precise moving objects.

3   Extracting the Depth Map of Background Image by Geometry Information

We had obtained the satisfactory background image, then the depth information of the background image was computed according to its geometry information.

First, extracting the evanishment line or baseline of background image(not all images have the evanishment line or baseline),then the geometry depth map was computed according to the evanishment line or baseline. This paper is aimed at the table tennis match video, therefore we need to obtain the evanishment line or baseline of background image. The baseline which is the horizontal line presented in [5] of the background image was extracted after the related operation. Generally speaking, due to the influence of gravity, the depth information of the object increases gradually from up to down, that is to say, the depth of the image below is greater than those above, or the characteristic of the nature makes the objects in the image below looks

closer to the human eyes than those above proposed in [8]. According to the theory above, we assumed the depth information is smallest in the position of the baseline(namely the baseline position is farthest to the human eyes), it is nearer to the human eyes from the baseline to down, and it is nearer to the human eyes from the baseline to up.

Suppose the line number that the baseline locates is m, then the following formula of computing depth information is obtained according to the deduction above.

$$d = \begin{cases} \dfrac{m}{h}*255 + \dfrac{m-i}{2*h}*255 & i \le m \\ \dfrac{i}{h}*255 & i \ge m \end{cases} \tag{1}$$

In the formula, m is the line number that the baseline locates, h is the number of the image lines (namely the image height), i is the current line number in the image, d is the calculated depth value of the current line number.

The depth map of the background image obtained by the formula above(see Fig. 1(a)), we can see that the depth map of the background image can not accord with human vision very well and also cannot response the depth information of the background image precisely. Therefore this paper adopted the linear perspective method proposed in [8] to optimize background depth map, namely, to obtain the optimized background depth map(see Fig. 1(c)),using the following formula to combine the figure 1(a) and the depth map obtained by the linear perspective method.

$$d_I = (d + d_L) * \frac{255}{512} \tag{2}$$

In the formula above, d_L is the depth value computed by the linear perspective method, d_I is the depth value optimized by the linear perspective method.

The linear perspective method is mainly to process the brightness values of the image. Using the brightness values of the image in the image processing can distinguish different objects, construct the outline of the objects and give the person the vertical deeply feels from far to near. The depth map obtained by the linear perspective method is shown in Fig 1(b).

Fig. 1. (a)   **Fig. 1.** (b)   **Fig. 1.** (c)

4   Extracting the Depth Map of Moving Foreground by Motion Information

Calculating the motion vector for the moving foreground objects can eliminate background mismatch points and reduce the computation [9].

The paper adopted adaptive diamond search algorithm proposed in [6] for motion estimation, and then used the method proposed in [7] to compute the depth value according to the depth information.

Adaptive diamond search algorithm for motion estimation, namely: proposes a new algorithm which according to space-time correlation of video sequence motion vector and center-biased character on the basis of DS search algorithm. First of all, the algorithm classifies the macro-block types and establishes qualificatory numerical value, then presents the prediction of initial searching point. The experiments proves that the algorithm improves greately searching speed on the condition of guaranteeing video quality. The formula of depth value calculated by depth information proposed in [7] is:

$$D(i,j) = C\sqrt{MV(i,j)_x^2 + MV(i,j)_y^2} \tag{3}$$

In the paper in [7], it treated motion as a sole depth cue, namely, it calculated the depth solely on the values of the X and Y motion vector values. The depth was estimated by the formula above. where D(i,j) is the depth value for pixel (i,j) and MV(i,j)x, MV(i,j)y are the X and Y motion vectors values for that pixel, respectively, and c is a custom defined scale parameter.

Because the difference value between the adjacent pixels in the extracted depth map is too big, it may cause distortion for the later computed stereo image and affect the effect of the stereo image. Therefore we proposed to carry on image segmentation for the moving foreground image, the paper presented the mean-shift image segmentation algorithm mentioned in [4] to segment the moving foreground image. After that we averaged the depth values in each division region to obtain the averaged depth map of moving foreground(see Fig. 2(a)).

Because the moving foreground depth map was computed only using the motion information, there may be some errors, and we could not make sure that the depth value in the image below somewhere was certainly bigger than that above. In order to overcome the defects, it needed to add the depth map of moving foreground and the depth map which depth value increases progressively from top to bottom and is irrelevant to the content together to optimize the depth map of moving foreground.

$$d_c(x) = \frac{y}{h} \times 255 \tag{4}$$

In the formula, y is the current line number in the image, h is the number of the image lines (namely the image height).The depth map(see Fig. 2(b)) was obtained by the formula above .

The adding processing results needed to be zoomed according to the formula below:

$$d_i = \left(d_f(x) + d_c(x)\right) \times \frac{255}{512} \tag{5}$$

In the formula, $d_f(x)$ is the depth value of moving foreground region in the depth map of moving foreground. Note: only the moving foreground region needs the optimization processiong, the background region does not need the optimization processiong.

From the Fig 2(c), we can see that the optimized depth map of moving foreground is quite satisfied through the optimization processiong.
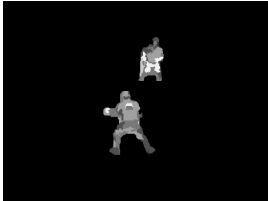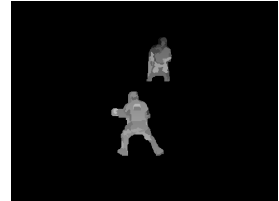
Fig. 2. (a)   **Fig. 2.** (b)   **Fig. 2.** (c)

5 Extracting the Ultimate Depth Map Based on Fusion the Depth Map of Background Image and the Depth Map of Moving Foreground

Fusing the depth map of background image $G(x,y)$ and the depth map of moving foreground $F(x,y)$ to obtain the fusion depth map $D(x,y)$ proposed in [9],the fusion procedure needs the mask image of moving foreground $A(x,y)$(the mask image of moving foreground is that the pixel value of the moving foreground region is 255 and the pixel value of the background region is 0),the fusion formula is defined as follows:

$$D(x, y) = \begin{cases} F(x, y) & A(x, y) = 255 \\ G(x, y) & A(x, y) = 0 \end{cases} \tag{6}$$

To obtain the ultimate depth map(see Fig. 3(a)), gaussian filter was applied to the fusion image. The depth map obtained in [9]((see Fig. 3(b))).By comparison, we can see that the depth map extracted by the method presented in the paper is more accurate and conforms to the human vision.
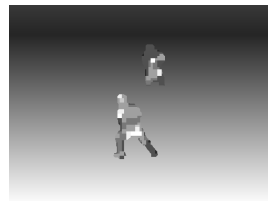
Fig. 3. (a)   **Fig. 3.** (b)

6   Experimental Results

The research used the video sequence of stationary background and moving foreground (130 frames) which is extracted from the video of World Pro-Tour Finals as the test video sequence. The paper made some improvement to the algorithm in [9],it presented the linear perspective algorithm to optimize the depth map, and to overcome the drawback which is the possible caused distortion for the later computed stereo image, the paper adopted the image segmentation algorithm to segment the moving foreground image, After that averaged the depth values in each division region to obtain the averaged depth map of moving foreground. The experimental results (see Fig. 3(a) and (b)) show the depth map extracted by the algorithm in the paper has better quality and improve that the algorithm in the paper is feasible.

7   Summaries

The algorithm in the paper made some improvement on the basis of the ideas in [9], it can improve the quality of depth map greatly and avoid the possible image distortion effectively for the later computed stereo image in the algorithm mentioned in [9], it can also conform to the human vision. The shortcoming is that the algorithm in the paper cannot realize the real-time extraction depth map. In the future, we may make some improvement to the existing algorithm, at the same time, enhance the real-time of the algorithm and quality of the depth map.

References

[1] Han, Y., Zhang, H., Li, L.: Shadow Suppression Algorithm Based on Gaussian Mixture Background Model. In: The 14th National Image Graphics Academic Conferences (2007)
[2] Long, W., Yang, Y.H.: Stationary Background Generation: an Alternative to the Difference of two Images. Pattern Recognition 23(12), 1351–1359 (1990)
[3] Tang, Y., Jiang, Y.: A moving human body segmentation method based on color image. Microelectronics & Computer 23(3) (2006)
[4] Song, S., Liu, D., Ceng, L.: A moving object detection algorithm based on the union of image segmentation and background model. Measurement and Control Technique 25(5) (2006)
[5] Harman, P., Flack, J., Fox, S., Dowley, M.: Rapid 2D to 3D Conversion. In: Proceedings of SPIE, vol. 4660, pp. 78–86 (2002)
[6] Wang, H., Kang, N.: A Adaptive Diamond Algorithm based on Motion Estimation. Computer Engineering and Application 43(26) (2007)
[7] Ideses, I., Yaroslavsky, L.P., Fishbain, B.: Real-time 2D to 3D video conversion. J Real-Time Image Porc. 2, 3–9 (2007)
[8] Sun, Y.: Key Technical Research of 2D to 3D Video Conversion. Shanghai Jiao Tong University 02 (2008)
[9] Huang, X., Huang, J., Wang, L., Li, D., Zhang, M.: A Depth Extraction Method Based on Fusion of Motion Information and Geometry Information. Zhejiang University 8(17) (2009)

Frame of a New Video Monitoring System for Home Safety

Shi Wen-Chong and Liu Mao-Hua

Hebei Normal University of Science & Technology

Abstract. The paper is intended to propose a practical solution for home video monitoring system, and focus on describing the structure of the system framework. It set out the present situation of domestic related products in the field of home safety, analyzed the lacks of some current video monitoring systems and products. The author thinks at present it is necessary to develop a video monitoring system for home safety in the type of short-range wireless communication +3G. The author built the framework of a home video monitoring system in the pattern of 3G mobile phone+3G network+Wi-Fi+NVS+IP Camera, proposed the developing idea of the system, described the development tasks of the system and made some important technical choices, selected Wi-Fi for short range wireless communication, h.264 for video compression. This innovative and unique feature is its simple and practical system structure, is ideal for residents use.

Keywords: Home Safety, Video Monitoring, Short-Range Wireless Communication, 3G, Wi-Fi.

1 Introduction

With the development of commodity economy in our country, people have expanded the scope of their activities. The time family members stay at home has been reduced, the residential risks of theft and fire have been increased. When leave home to engage in social activities, many residents want to know the home security circumstance urgently. At the same time, some traditional and unwieldy security means has gradually faded out of view of people; intelligent security has become the core of the intelligent home. The new home security system under the concept of Iot, i.e. wireless remote real-time monitoring ones, could become ideal home security means because of their ease of making, minimal impact to home environment, con-venience to implement monitor by mobile phone.

Domestic security monitoring passed simulation monitoring, digital monitoring and IPVS. Now with the 3G network gradually mature, mobile video monitoring technology gradually go to the front, with intelligence, wireless network, HD and other features. At present, video monitoring business has been recognized by the industry, more and more customers choose video monitoring system for promoting the enterprises' informationization cons-truction, video monitoring system is also widely used in the safe city, transit management, the typical application of 3G video monitoring for home security is mainly as following:

1.1   Mobile-Video-Telecom

This is the typical application case that 3G-mobile-phone-video-monitoring-system is used in the home security fields. Its operating is based on Internet and GPRS networks. It needs indoor wired connection, requires routers and WLAN, the mobile phone need install *Shou-shi-tong* software. Because it doesn't directly receive video information from the home Security facilities, and it works still by visiting a web site, this is indirect monitoring.

1.2   China Telecom's Global Eye

China Telecom's global eye is based on CDMA mobile communication network and mobile Internet. By CDMA, 3G mobile phone and computer terminal realizing real-time video monitoring and value-added services, it is a system in the model of MSP/CMS+VAU+ M_PU +M_CU+PSS+WAV-GW [1]. Usually there should be cable broadband installed in the family, and computer, router, door sensor, camera, and still an anti-theft & alarm machine. The mobile phone need install the client *global eye* software [2]. If there is no wired broadband at home, you need still connect the wireless platform of global eye services via the 3G of China Telecom EVDO network. For that the user need to buy Internet EVDO card. Actually the user gets video information from the camera via global eye platform on the fixed network, and the video information deliveries to the platform of the mobile global eye via wired broadband or wireless network. The User uses a mobile phone to browse the WAP pages and gets the video information. It is clear there is wired connection in the home facilities of the system. The user must bear the restrictions from the communication vendors and Internet, the initial investment is huge, and the charge will be higher.

1.3   Gateway of G3 Video Monitoring TDHOME

G3 is a service brand produced by China Mobile and based on the 3G technology TD-SCDMA standard. It is an LOGO and it is a branch of 3G too. G3 netbooks and G3 telephone provided a scope for G3 in video surveillance. The gateway of G3 video monitoring TDHOME can achieve high-speed wireless Internet, and with the function of video monitoring for home safety. TDHOME is consisted by G3 video monitoring gateway, ZigBee digital wireless door sensor, wireless network cameras, and other components. The mobile phone needs GPRS service. when a host for G3 video monitoring gateway is in the state of deployment, if the door is opened, ZigBee door sensor will be triggered. The trigger signal will be immediately passed on to the G3 video monitoring gateway host. It can quickly transfer the alert information by a piece of message to user's mobile phone. When a user's mobile phone receives the alert message, as long as the user click address link on the alarm message, he will be able to directly watch living video through the mobile phone in real time. The shortage of this gateway is that it cannot implement living video monitoring from multiple safetyty places all a day, and it is limited to mobile phone users. The lack of security means itself greatly limits its popularization and application.

   All the technologies mentioned above depend on 3G to realize remote video monitoring, and their lacks are they depend on Internet, WAP and GPRS too much, they don't realize the true system-wide wireless transmission, and have 'four high' shortages apparently – high investment, high starting point, high power, high fees.

And GPRS is not 3G! Clearly these are neither worthy of the wireless 3G remote video monitoring system, nor the ideal means for home security.

2   Framework of the System

This system will be civilian, be easy to setup, with few fees, and seek to make no use of Internet and GPRS. With a simple circuit, direct designing, to develop a system worthy of short-range wireless communication and video monitoring, with high efficiency and low power. Remote mobile wireless video transmissions must now rely on 3G technology, the core devices must be based on 3G; the main equipment in the system must connect wirelessly and form a WLAN. The system must have the characteristics and the corresponding functions. The framework of the system is as shown in Figure 1.

1. Camera   2. Sensor   3. Node for Video's transmitting

4. 3G video server   5. 3G Terminal (mobile phone)
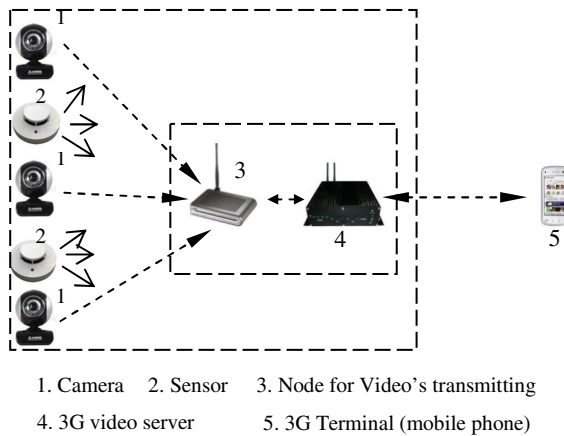
Fig. 1. The framework of the system

In the figure, 3G Terminal means 3G mobile phone generally, PAD can also be considered. It is used to receive and browse the real-time video information from home monitoring point and give control instructions to the system, set the parameters of the system, control the work status of the system if necessary. Need develop embedded software based on the original mobile system

Camera is the devices for capturing video information. It can be set at the top of perimeter doors, windows, or in study, bedroom, living room need to be monitored, or other appropriate locations. To capture high quality video information at night, the infrared camera can be used. Wireless camera is easier to be hidden obviously.

Sensor here means a sensor for smoke, or for temperature, for humidity, for light intensity, etc. It is mainly used to measure environmental parameters of home and alarm in time. The system will do video monitoring, and the compatible interfaces are also provided for varieties of sensors in order to upgrade and expand to IoT.

3G Video Server is installed indoors and it is the core component of the system, built-in 3G network communications control protocol and TCP/IP protocol. To save money, several households in the same unit above or below floor can share one server.

Obviously, it should have interface circuits, not only for 3G, but also for short-range wireless communication. To save video information as evidence for detection, it should have hard disks, to set up and modify user's permission or password easily at home, it should also have a keyboard and LCD monitor.

In the simplest case, when there is only one point to be monitored, and retaining video information is not necessary, the system will require only an IP camera supported Wi-Fi and a 3G mobile phone. But generally the points to be monitored are not unique. Even if from the perspective of video information storage only, the amount of data that an IP camera stores local video stream is only for a few hours, can't meet the needs of monitoring at any time and getting the evidence. So a 3G video server is essential.

When there are too many floors, too much walls blocking or in the case of more signal interference, the range of the wireless communication will be greatly reduced. Adding a video transporting node in the system can be considered. It is a wireless video transceiver, can avoid the interference from the crowded radio signal on the special band or various types of EDM or the household appliances. At the same time, it can also avoid the interference from transmitting frequency of it itself to other receivers. This is the key to keep the stability of video transmission and EMC, particularly it is also useful to the case of multiple points monitored, multiple 3G receiving terminal.

3   Key Technical Options of the System

Realizing short-range wireless communication for major facilities, and realizing the remote video trans-mission of 3G are the main aim of this system. Choosing the type of short-range wireless communication, the video compression standard, the embedded operating systems and developing the embedded software, controlling system's security are the core of this project.

3.1   Choosing the Type of Short-Range Wireless Communication

There are many kinds of the short-range wireless communication technology. For example, Bluetooth, Wi-Fi, IrDA, UWB, ZigBee, etc.. Because the effective transmission distance of Bluetooth, IrDA, ZigBee is less than 10 meters, they can't meet the need of wireless communication of a home with large area. The effective transmission distance of ZigBee may reach hundreds of meters (up to 134 meters), but its basic transmission rate is only 250kb/s, cannot meet the requirement of video transmission[3]. In a word, Wi-Fi can be selected here only.

At present, because of Wi-Fi with wide coverage (the radius can be up to hundreds of meters), high transfer rate (the maximum rate of generalized Wi-Fi is 54Mb/s) and less restriction to use, it is widely used in the industry. In the room-impaired case, the transmitting distance of Wi-Fi is up to 100 meters. For this system, this is the most important advantage. Generally while set the place where is centered on the video server and within a radius of 100 meters, a camera with Wi-Fi interface can wirelessly transmitting video signals easily. This is enough to a home often with area less than 200 square meters. Even to the thermocline-Villa with 2-3 floors, it can also meet the need. If the network of cameras formed of 'few cameras peer to peer'

could meet the demand, the cost is significantly reduced. But this system is designed to transmit video information, information's hysteresis and stability is very important, so the challenge is to provide sufficient guarantee of QoS.

The advantages of Wi-Fi was enhanced especially after broadband was used, as long as the user connected the broadband network in residential quarter and set AP, and the user place a PDA or others supported by WLAN in the area where there is the hot spots' radio wave, Internet could be connected with high speed. After a community was connected to a cable broadband network, it is no problem that a resident logins the broadband network with a computer at home, still login Internet wirelessly by AP and wireless network card. If necessary, a complete WLAN can be built, and one AP is set in the room. Currently an indoor AP's price with the coverage of radius from dozens to 100 meters is only hundreds of Yuan RMB, an AP was enough. So it is suitable for a home. The popularization and promotion of the products such as wireless Wi-Fi webcam laid the foundation for Wi-Fi's further application of the system.

3.2   Choosing the Video Compression Standard

In this system, the remote video monitoring by 3G mobile phone is real-time, the camera's collecting and transmitting video information, the 3G video server's receiving and processing the data should be controlled by the user's instructions at any time. Because the amount of information is very large, to keep the quality of the video, the data must be compressed during transmission. Reasonably choosing the video compression standard is essential.

There are two main series of compression standard currently, those are the ITU-T series of ones - h.261, and h.263, and h.264, and the ISO's MPEG series - MPEG-1, MPEG-2, MPEG-4. Every high version is the improving and upgrading of the low versions. And h.264 is the new digital video compressing standard co-sponsored by ISO and ITU after MPEG4. It retains the benefits of the past compression technology and has many new advantages, such as the lower bit rate, the superior fault tolerance and the good adaptability for network. Obviously, here h.264 should be   chosen, and the 'adaptability for network' is the most critical.

Currently, a high definition camera for h.264, or a wireless webcam for h.264, is a very common, The wireless Wi-Fi webcams above are also supported by h.264; DVR and NVR supporting 2 routes wireless digital camera and h.264 are produced by many companies in our county. These are conducive to successfully finish this system.

3.3   Choosing the Embedded Operating System

First, choose the OS for 3G video terminal (mobile phone). We select Google's Android. The reason is as following:

The OS for 3G video terminal must support 3G. Android 2.2 Froyo published in May, 2010 has the functions for 3G network sharing. And Android 2.4 Gingerbread published in Dec, 2010 runs faster.

The data on the market research [4] showed that since the second half of 2010, its market share has been rising in a high speed. It had become the market share leader in early 2011 (29%) of the OS. The similar mobile phone manufacturers are HTC, Motorola, Samsung, and LG, Sony Ericsson, Dell, Huawei, etc. These are the more

popular brands of 3G mobile phone. Although the market share of Iphone and BlackBerry is high (both 27%, Microsoft's Windows Mobile's share is only 7%), but they belong to Apple's and RIM's mobile phones respectively. For the wider application of this system, it must be made for more manufacturers of mobile phones and not tied to one brand.

The Open Handset Alliance constituted by Google and many enterprises, involves dozens of companies - the manufacturers of mobile phones and other terminals, mobile operators, the companies in semiconductor, software companies, etc. Its product line and industrial chain are consolidated and increasingly expanded. This benefits promoting the system in the future.

In the same light of the above reasons, the video server of 3G also uses Android system.

The cameras in the system need not 3G technology, supporting Wi-Fi only is enough. So the cameras can adopt Linux systems. Android and it are both open source systems. Since Android system is based on the kernel of Linux, the communication of the cameras and 3G video server, in another word, controlling the transmission of 3G video, is easy to achieve. A wired IP Camera may publish video information over the network while it is connected with Ethernet cable and the power. It will have Wi-Fi function as expected. An IP camera generally uses embedded Linux operating system. The system complies with the tradition so that more software resources can be referenced.

4   Summaries

The system will complete designing a wireless network, the algorithm of intelligent early warning, 3G video server, the front-end devices for video capturing, a group of software for mobile terminal, etc. Developing the core board with the functions of 3G communication and short-range wireless video transmission, developing a photographing small-plate with the capacity of wireless video information transmission, develop a set of embedded software… Building a video monitoring system of 3G mobile + 3G Network + NVS + Wi-Fi + IP camera for home security, is the ultimate goal of the system.

References

[1] Wang, P.: Structure and Working Principle of 3G Mobile Video Monitoring System (EB/OL),
      http://security.it168.com/2010/0729/836.shtml
[2] Zhang, G.: Mobile Global Eye and Its Applications. Modern Property Management (11), 101–103 (2010)
[3] Xing, C.: Review of Short Distance Wireless Communication Technology. Modern Electronic Technology (3), 65–76 (2004)
[4] Karlf, D.: Android's Market Share (EB/OL) (March 04, 2011)
      http://article.yeeyan.org/view/209293/177290

A New Real Time Data Acquisition System
for Low Hybrid Current Drive System

Zege Wu, Jia Fang Shan, and Handong Xu

Institute of Plasma Physics, CAS

Abstract. Low Hybrid Current Drive (LHCD) system is an important auxiliary heating system for Experimental Advanced Superconducting Tokamak (EAST). When EAST is discharging, LHCD outputs the microwave into the tokamak. In order to get and analyze the power of the microwave, a new real time data acquisition system was designed and materialized, and it can satisfy the requirement of the experiment after practical test.

Keywords: Real time acquisition, Double buffer, Reflect memory net, Qt.

1   Introduction

EAST is the first full-superconducting tokamak in the world, which started the assembly in 2003 and began to discharge in 2006. LHCD is an important auxiliary heating system for EAST, and when required in discharge it outputs microwave form 20 klystrons into the tokamak. EAST has successfully fulfilled the goal of H-MOD discharge in the end of 2010 with the assistance of LHCD. In order to get the status of the couple condition of the low hybrid wave with the plasma, the waveform of the microwave must be displayed during discharge. For later analysis, more accurate waveform has to be presented after discharge. So far, two systems, Real Time Display System and Interrupt Acquisition System, are working for this [1]. These two systems are acquiring the same signal; therefore there are great needless hardware and software redundancy in them. With the limitation of the bus technology and hardware, the current two systems are also obsolete and cannot satisfy some new requirements such as higher sampling rate for the experiment. All drive us to upgrade the two systems, and with the development of the hardware we bring up our new solution which is able to perform both functions.

2   New Solution

Fig.1 is the architecture of the new solution, which is a maser/slave structure. The slave computer equipped with the high speed data acquisition board and reflect memory board is in the lab field and used to acquire the microwave power. The maser computer with the reflect memory board is in the control room. Two computers are connected together through the reflect memory net and the Ethernet. The system works as follows. When the LHCD is outputting energy, the slave computer acquires the power data from the sensors and save the data form memory buffer into a file on

local disk. And at the beginning, the master computer will start a timer. Every time the timer expires, the master computer would ask slave computer for the power data of the 20 klystrons in LHCD. Receiving the request, the slave computer will send the current data in buffer to the master computer through the reflect memory net. If the cycle of the timer is short enough, for example 30 milliseconds, the master computer will paint waveforms which are moving forward during the discharge. Once the discharge is over, the slave computer sends the local power data file to the master computer through the reflect memory net and the master computer is able to present the waveform in more details.
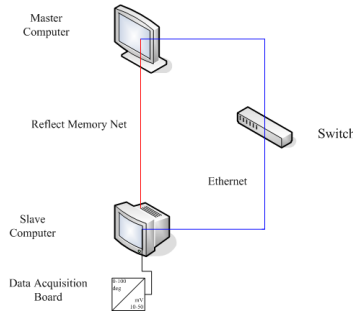
Fig. 1. Architecture of the new real time data acquisition system

3 Data Acquisition

The originally used data acquisition board, 16 channels 1MHz AC1820A with 128KB FIFO buffer on board, are scanning boards. With the development of the bus technology, the products based on PXI have much better performance. The ADLINK DAQ 2000 series are PXI-based high performance acquisition boards, which can satisfy the new requirement and provide some useful functions besides the high
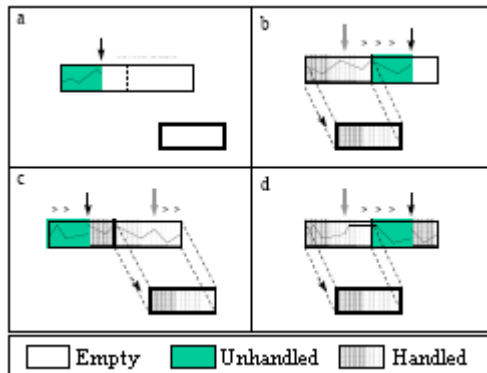
Fig. 2. Double buffer mode
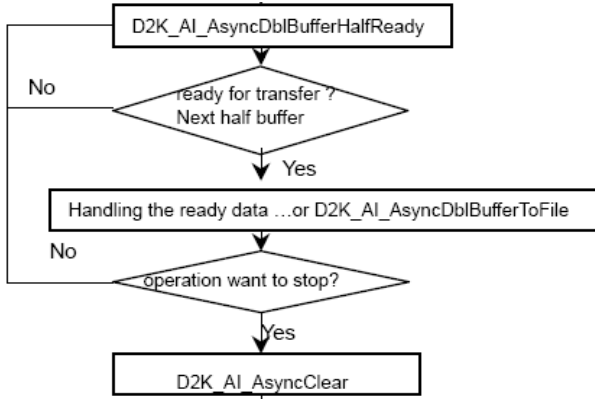
Fig. 3. Control flow of the double buffer

sampling rate. ADLINK not only provides programming libraries such as DLL for most Windows based systems, but also provide drivers for other software packages such as Lab VIEW. The ADLINK also provides the driver and the API which can be ported to many popular operating systems. After comparison, DAQ 2010 with four simultaneous channels and 1M FIFO on board is adopted to deal with the task of acquisition.

The DAQ 2010's software library, D2K-DASK, is able to use double-buffering technology for continuous data input, which is an especially efficient way solving the problem of long duration of discharge [2]. As Fig.2 shows, the data buffer for the double-buffering is a logical circular buffer, which is made up of two equal buffers. The board first writes data into the first buffer until it is full. When the board begins to write data into the second buffer, an event will be emitted to notice that the data in the first buffer can be processed according to the application need. After the board has filled the second buffer, the board returns to the first buffer and overwrite the old data. At the same time, an event will be emitted too to notice the data in the second buffer is available. This process in Fig.3 can be repeated endlessly and provides an efficient way for continuous data transmission.

4   Reflect Memory Net

The reflect memory net (RMN) is a communication net of high speed, real time and certainty, which is formed by Reflect Memory Board (RMB) connected with each other through optical fiber, and the computer or other device equipped with the RMB is one node in the RMN. The local memory on the RMB could be mapped into the user memory space in node, so operation such as read and write on the board memory is as easy as on the node memory. If one node's processor writes data into the memory on board, the new data will be broadcasted to all other nodes through the optical fiber in high speed at the same address and other nodes can read the new data from its local board in very short time. Thus, the memory on all boards can be

mapped into a global memory through the RMN, and the RMN actually is a distributed shared memory [3].

VMIC 5565 is selected as the reflect memory board, whose highest transmission rate and bandwidth can be up to 1.4Gbs/s and 2.12Gbaud [4]. And there are two kinds of topologies of the RMN: star topology and ring topology. Considering the amount of the RMB in use, the ring topology as shown in Fig.4 is chosen.
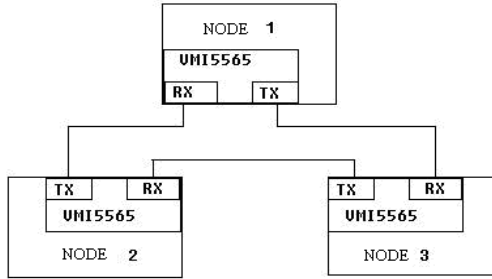
Fig. 4. Ring topology RMN

5   Software

Qt is a cross-platform application and UI framework [5]. It includes a cross-platform class library, integrated development tools and a cross-platform IDE. Using Qt, programmers can write applications once and deploy them across many desktop and embedded operating systems without rewriting the source code. The most important feature of the Qt is the signal/slot mechanism, which is different from the common callback mechanism and can provide great convenience for programming despite at cost of slight efficiency. The software on the master computer is developed with Qt on Linux. It is able to display the outline of the waveform in discharge and show the waveform in details after discharge. The software on the slave computer is developed on Linux without the UI. Besides the RMN, the master computer and the slave computer are also connected through the Ethernet. Using some easily obtained services based on TCP/IP such as telnet, some useful functions, for example, remote reboot and remote control can be easily realized and is helpful for the recover form the software failure.

6   Summaries

Considering the new requirement and the limitation of the current working system, we give our new solution. Choices about the hardware such as the RMN and software are made and both prove available. After practical test in lab filed, the new system is feasible and robust.

References

[1] Wang, D., Wang, M., Shan, J.: Distributed Acquisition Control System Based on QNX. Neuclear Fusion and Plasma Physics 27(4) (2007)

[2] ADLink Technology Inc. D2K-DASK Data Acquisition Software Development Kit For DAQ-2000 Devices User's Guide

[3] Sun, H., Xu, L., Yang, Q., Liu, X., Yuan, Z.: Computer Real Time Communication Network Based on Maped Memory. Guangxi Communication Technology 2 (2008)

[4] VME Microsystems International Corporation. VMICPCI-5565 Reflective Memory Board Product Manual

[5] Blanchette, J., Summerfield, M.: C++ GUI Programming with Qt 4. Prentice Hall, ISBN-10: 0-13-187249-4

Research on Application of Data Mining in Electronic Business

Duan Junwei[1,2] and Sun Hongzhi[2,3]

[1] Information Engineering Institute, Jilin Teachers' Institute of Engineering and Technology, Changchun China
[2] College of Management, Jilin University, Changchun, China
[3] China Petroleum Jilin Chemical Engineering and Construction Co., Ltd.

Abstract. In this paper, the feasibility of application of data mining in electronic business has been analyzed, and the research on the application that data mining is used in web design and client relation manage of electronic business in detail. Finally, a model of data mining system based on electronic business has been given.

Keywords: data mining, electronic business, decision.

1 Introduction

Concept of data mining was originally evolved from the "knowledge discovery from database" (KDD). But the word KDD first appeared in the Eleventh International Joint Conference on Artificial Intelligence which was held in 1989. After more than 10 years of development, data mining, whether in academia or industry, has received more and more extensive attention and obtained considerable development[1]. In academia, the data mining has integrated and promoted the development of the following computer science: machine learning, pattern recognition, statistical, soft computing, data visualization, intelligent databases, knowledge acquisition, expert systems, data visualization and high- performance computers and so on. In the field of retail, telecommunications, banking, securities, insurance, pharmaceutical and genetic research, it has obtained practical application[2].

With the rapid expansion of Internet in the world, and gradually penetrated into all areas of society, people can easily use her to e-mail, Web browsing, chat, download software, purchase of goods and so on. In view of the emergence of these opportunities on the Internet, many companies have put their operations to the Internet, establish their own websites. Through these websites, customers can interact anytime and anywhere with the enterprise conveniently, and implement traditional business activities, this is the electronic business. A complete commercial trade process contains the understanding of market condition before transaction, the inquiry, the quotation, sending orders, signing orders, sending and receiving delivery notifications, picking up vouchers, paying the exchange process and so on, in addition, it also involves in the behavior of certification of the administrative process , in the capital flow, the logistics, the information flow flowing, and only the process above has achieved paperless trade, that is, all non-human intervention, and uses the

electronic tools to complete, can be called a complete electronic commerce processes[2]. Electronic commerce has efficient, convenient, integrated, scalable, secure and coordinated merit. The enchanting charming of the electronic commerce has attracted all walks of life, the retail trade, the financial industry, the education industry, etc, all have started to follow this trend and develop their own electronic commerce system. Now, the government departments also have joined the ranks and began to advance government affairs' computerization, its fundamental mode and the concrete technology are the same with the electronic commerce.

2   Data Mining in the Electronic Commerce Application Feasibility Analysis

Electronic business enterprise to face the customer base is massive and complex, the market become bigger and more complex and changeable. This time, the enterprise is expected to more than ever to understand customer and market. Electronic business enterprise of market and customers are often very lack of knowledge. But, on the other hand, electronic commerce in the process of the functioning of the accumulated the massive related users' business behavior and the users themselves data. Electronic business is also facing a data rich but little knowledge embarrassment.

From the current development situation, the electronic business affairs will become a very important data mining application occasion. This was mainly due to the rapid development of one hand the Internet for data mining provides a new wider stage; On the other hand, the development of related technologies and gradually mature also make the application becomes possible.

A successful application of the data mining must have the following five conditions:

(1) Rich content of data.   Such as customers' buying records, because it has a lot of potential useful fields can be applied to mining algorithms, so it can be seen as "rich content" data;

(2) Large amount of data.   A reliable data model requires a lot of training data;

(3) Controllable and reliable data sets.   Manual input data and historical data integration for many problems, usually think completely automated data gathering that much better, will reduce a lot of data "noise";

(4) The results can be assessed.   Real and visible investment income is the most convincing;

(5) And the existing processing easy integration.   Have a computer application experience knows that a successful software projects should not only consider technological factors, also dealing with a lot of "non-technical" factors, such as a person's habits, laws and regulations and even some political considerations[3].

Electronic business affairs will become a very important data mining application field, mainly because it has the above five parts. Let us make a concrete analysis:

(1) The electronic business data is automatic recording, but not manual handling for example manual input and so on), therefore "the noise" are very few;

(2) The electronic business data is volume and rich content, electronic business contained the rich customer to trade the behavioral data and the website visit diary data greatly, these data have the rich content available for mining ;

(3) Electronic business has also recorded the data which in other business dealing is very difficult to record, for instance the customer has read the commodity record but not bought, such data provide the very good application place to use the electronic commerce to analyze customer's consumer behavior;

(4) Electronic business's data mining application effect is very easy to appraise. Such characteristics have decided the application data mining technology is not only feasible but also has the good prospects for development in the electronic business

Apply the data mining technology in the electronic business, Carry on the excavation to these data, may discover these "knowledge" valuably. The enterprise users may act according to these "knowledge", grasp the customers' tendency, trace the market shift, make the correct pointed decision, for instance improve the website, promote personalized pages to each kind of customer, or provide the preferential policy to the high outflow client base to carry on detains and so on. Establishment customer relations management system which take the data mining as the core for the electronic business, excavate the mass data which the electronic business system accumulates, will enable   the   enterprise to deal with the market shift promptly, seize the market commanding point [4].

3   The Application of Data Mining Technology in Electronic Business

3.1   The Application of Data Mining Technology in Electronic Business Website Design

The electronic business website is established to realize the electronic business goal on Internet, therefore electronic business website stand for enterprise's image on Internet. But for on-line consumer, if they are unsatisfied to enterprise's website, then they can't be possible to do shopping in this enterprise. The website design must make the website become consumer's core website, make the consumer visit the website repeatedly .To become consumer's core stand, you must be able to provide the personalized service, but to provide the personalized service, you need to depend upon the data mining technology. Through the data mining technology, provide the basis of design electronic business website to the website designer.

Using data mining in electronic business web design is mainly to mine the Web record, through mining the corresponding log file and data to find the visitors and customers of this site, and then adjust the Web site structure, in order to improve the service quality of the Web site. As for Web log, sequence mode analysis is the most important method, because Web log data record is according to the clients' visit time. Finding sequence mode can facilitate predicting customers' visit mode for electronic organizers, provide customers personalized service; Website administrator can classify visitors according to the sequence mode and only show the link of the sequence with the visitors often visit in on page , but direct the content that has not been displayed with a "more content". When visitor is browsing one page, it'll check

the sequential patterns which is fit for his browsing, and suggests "other pages that the customers often visit also" on the conspicuous position. Next, the correlation analysis method is also can be used to find out the relations of customers on the website access the various documents. For example: using correlation analysis may find, if the a customer visits page A, he will generally also visit page B，to build the correlation model of data mining on website, so that we can organize sites better, reduce the burden of filter information for customers.

Policymakers need to analyze and manage the visiting situation from various angles and levels, they need to know: what kind of customers visit the site, the methods that the clients like to obtain information through, how many times the site is visited, the frequency that the same customer visit the site, what buying behavior the customers often repeat, etc.[5] Data mining can find out the information above, to understand and increase the" appeal" of this website, to improve the visit times of it.

3.2   The Application of Data Mining in Electronic Business Client Relation Management

In the electronic commerce mode, the enterprise market management, sales management, customer management, etc., will change a lot, the competition between the enterprises brought by information network of advanced technology become more intense. But no matter how the business form develops, the enterprise should improve customer satisfaction in order to maintain and develop their own competitive advantages. Establish customer's trust to product and allegiance to the enterprise. Only by this way to get customers can realize enterprise profit. Moreover, under the electronic business environment, the stagger that enterprise on customer relation management appears more important.

Customer information is the foundation of commerce and customer relations management. If there are no detailed data of customers, we have no way to know about them, just don't have any superiority. Thus, according from the mass data, people need make use of the data mining to analyze customer's characteristic. They also need analyze enterprise and its operation law corresponding with market to raise enterprise's economic benefits constantly. If the enterprise takes the profits as its own goal, CRM is the most useful tool to achieve this goal, and DM is the best engine of this tool. You can use several following methods in the customer relations management:

3.2.1   Connection Analysis
Such as "80 percents of customers purchase commodity A and commodity B at the same time in his shopping" and so on For instance, Classic story of "diaper and beer' just uses this method. At first, we find they have high related coefficient which causes our attention, and then after deep analysis, we find out essential reason. If there is the relevance rule "A→B", when customers purchase commodity A on line, we can recommend commodity B to them immediately, namely   adjust the interlinking of B to a conspicuous position in order to give convenience to customers on their purchasing.

3.2.2   Sequence Pattern Analysis

In the e-commerce, the sequence pattern analysis may be used for discovering the following knowledge such as "after online purchasing   a commodity, and in period of time customer purchases commodity B on line, then purchases commodity C" thus form a behavior of customers "the pattern of A→B→C".

What may BE inferred is that after having bought a computer, he is likely to buy marking machines, scanner and so on. So when a customer links to an e-commerce website, we can look the customer's related record in order to analyze which pattern he or she supports. If it contains "A-B" sequence pattern, then we can adjust the e-commerce website dynamically, and put the related link in a striking position and give customers recommendation.

3.2.3   Classification Analysis

Classification analysis make an accurate description or establish   model or dig out the classification rules for each category   through in analysis electronic business sample customer data in the database , then use this classification rules to classify the records of other customers. For example, the credit card company based on customer's credit history, the cardholder is divided into different levels, and mark the level of the database assigned to each record. For each level, find their common ground, such as: "Above the yearly income 100,000 Yuan , foreign white-collar workers between the ages of 40-60, " the highest overall credit record. With this data mining results, the customer service departments have known that a potential value of new customers and have confident in the customer service investment.

3.2.4   Cluster Analysis

Cluster analysis Divides unclassified records into several categories, according to data of intrinsic different size and determine the category for each record, in doing not know that should divide into several categories of situations. The cluster analysis may discover the different client base from the customer basic storehouse, and describe the characteristics of different customer groups with the purchase pattern. In electronic business data mining, the customer is not willing to give own background information sometimes, therefore, can only come the cluster according to the customer browsing pattern, and extrapolates the customer hobbies and so on.

4   The Electronic Business of Data Mining System's Construction

As with other data mining systems, the facing data mining system for electronic business and electronic business platform are independent. Between the two parts interact through users. The system structure shown in Fig.1.

The whole system is divided into three parts: the electronic business part, the user, the design mining section. The data mining section is independent that is divided into three layers. The bottom is the data layer, namely, databases and other data sources. The data source is a relational database system of electronic business systems, expertise or other data sources; they are the most primitive data. The second layer is the mining level, including excavation of the database (usually a multi-dimensional database) and the mining engine. The data from the data layer can be carried out in their dug through data preprocessing. Database design and model by

loading the excavation database can be excavated. The top layer is the user interface and it can communicate with the user.[6] Guide the mining process; mining returned the result of mining to the user through the interface, the decision is usually visual.
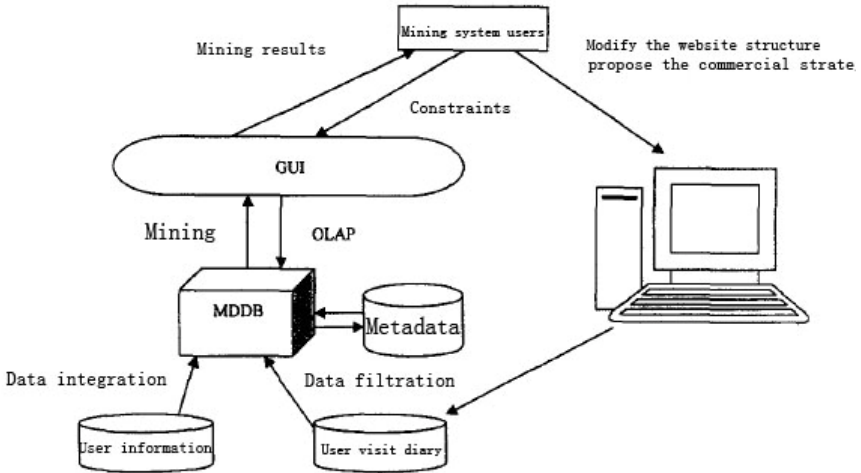
Fig. 1. Data mining system for electronic business model

In addition to mining section, also includes two parts of the user and electronic business systems. Electronic business system is a system that provides services for customers in the process of running it will produce a lot of information, which is the basis work for mining system. Users in the whole system are essential. Users in the mining process to interact with the data mining system, from the guide. In addition, after the outcome of the mining system, the user on in the data mining system as follows: put forward new or modified business strategy for electronic business site.

5   Summaries

In this paper, the actual situation of electronic business, data mining techniques on how electronic business website design and customer relationship management conducted a detailed study of the application, and gives a data mining system for electronic business model. But what kind of electronic business application specific data mining algorithms, to be more in-depth study further.

References

[1] Zhang, L.: A new efficient multi-level association rule mining algorithm. Northeast Normal University (Natural Science) 35(2), 15–20 (2002)
[2] Han, J., Kamber, M.: Data Mining Concepts and Techniques, pp. 13–18. Mechanical Industry Press, Beijing (2001); Fan, M., Meng, X., trans.

[3] Qi, M.: Electronic business Practical Guide, pp. 58–65. Higher Education Press, Beijing (2000)

[4] Berson, A., Smith, S., Kurthearting: Build data mining applications for CRM, pp. 109–116. People's Posts and Telecommunications Press, Beijing (2001); He, Q. trans.

[5] Liu, H.-Z., et al.: An ontology- based architecture for distributed digital museums (DB/OL), `http://ieeexplore.ieee.org/ie15/10231/32623/01526913.pdf`

[6] Corby, O., Dieng-Kuntz, R., Faron-Zucker, C.: Querying the Semantic Web with the CORESE search engine. In: LoPez de Mantaras, R., Saittaeds, L. (eds.) Proc. of the 16th European Conference on Artificial Intelligence, ECAI 2004 Subconference PAIS 2004, Valencia (August 2004)

Ancient-Coins Images Retrieve by Wavelet Transform and Relative Moments

Feng Xiao[1,3], Mingquan Zhou[2], and Guohua Geng[1]

[1] Institute of Visualization Technology
Northwest University
Xi'an, China
`xffriends@163.com`
[2] School of Information Science and Technology
Beijing Normal University
Beijing, China
[3] School of Computer Science and Engineering
Xi'an Technological University Xi'an, China

Abstract. A novel method of ancient coins feature extraction and image retrieve by wavelet transform and the relative moments is proposed, which firstly utilized the multi-scale wavelet feature template to achieve the polish output of color filter component and to realize gradient vector extended. Through multi-scale Wavelet transform, the boundary images at different scales of the color images of ancient coins were reaped; secondly, through the extracting of regional characteristics of each scale image based on the ten relative moments which were transformed from seven invariant moments, feature vectors of the image were formed; finally, feature vectors were normalized and its similarity was measured and then an image retrieval system was built. Experimental results show that the method has good versatility and robustness.

Keywords: Wavelet transform, Relative moments, Ancient-coin images, Multi-scale edge.

1 Introduction

With the increasing number of ancient coins unearthed, the study on how to achieve accurate and efficient image retrieval system of ancient coins has become a pressing research subject. Given that the unearthed ancient coins covered different historical periods and various kinds; for image retrieval, the image features to be retrieved are also very rich, it becomes essential to find an universal method of feature extraction and retrieval.

To Combine study status on the current computer-based image retrieval technology with the practical features images of ancient coins, on which this paper is based to propose a method of ancient coins feature extraction and image retrieve by wavelet transform and the relative moments. Having made full use of the effectiveness of wavelet transform for digital image edge detection and the consistency of relative

moments for the structure and the region of the image [1] [2], the method solved the issue of the inconsistency structure and region of complex digital images and improved the retrieval accuracy and versatility of image retrieval system.

2   Multi-scale Edge Detection

2.1   Multi-scale Wavelet Transform

A multi-scale version of this edge detector is implemented by smoothing the surface with a dilated convolution kernel $\theta(x)$. Suppose two-dimensional image is

$$f(x, y) \in L^2(\Re^2) \tag{1}$$

Where the kernel function (namely, the Gaussian function) is

$$\theta = G(x, y) \tag{2}$$

This is computed with two wavelets that are the partial derivatives of $\theta(x)$:

$$\begin{cases} \psi^1(x, y) = -\dfrac{\partial \theta}{\partial x} \\ \psi^2(x, y) = -\dfrac{\partial \theta}{\partial y} \end{cases} \tag{3}$$

The scale varies along the dyadic sequence $\left\{2^j\right\}_{j \in Z}$, We thus derive that the wavelet transform components are proportional to the coordinates of the gradient vector of f smoothed by $\bar{\theta}_{2^j}$:

$$\begin{pmatrix} W^1 f(x, y, 2^j) \\ W^2 f(x, y, 2^j) \end{pmatrix} = 2^j \begin{pmatrix} \dfrac{\partial}{\partial u_1}(f * \bar{\theta}_{2^j})(x, y) \\ \dfrac{\partial}{\partial u_2}(f * \bar{\theta}_{2^j})(x, y) \end{pmatrix} = 2^j \vec{\nabla}(f * \bar{\theta}_{2^j})(x, y) \tag{4}$$

Where * is Convolution, and ∇f is the gradient vector of f. We can see that the two modulus of this gradient vector is proportional to the wavelet transform modulus.

The algorithm in which the details were captured by smoothing the three color channels (R, G, B) respectively through multi-scale wavelet, utilizing different scales features to witness the detailed local changes in the true color images, provides more useful information than the method that processes the true color images directly, and it also provides more information for the signal local processing. The algorithm facilitates the accurate understanding of object perception.

In this algorithm, the Kernel function is the Gaussian function, that is

$$G(x, y, \sigma) = \frac{1}{2\pi\sigma^2} \exp\left\{-\frac{x^2 + y^2}{2\sigma^2}\right\} \tag{5}$$

the parameter of the Gaussian function $\sigma = 0.8$, template by 3×3. It is

$$\frac{1}{16}\begin{pmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{pmatrix} \tag{6}$$

When the scale was 2^j, the wavelet multi-scale transform inserted $2^j - 1$ zero(es) between the adjacent filter coefficients, which continued to decompose and formed a series of scales.

2.2   Extract the Edge Image

In the RGB system, each color point can be interpreted as a vector extending from the origin to that point in the RGB coordinate system. True color images have at least three components, pixels are vectors[3].

We take into account the fact that the color components are a function of coordinates (x, y) by using the notation:

$$C = \begin{bmatrix} c_R \\ c_G \\ c_B \end{bmatrix} = \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{7}$$

Let r, g and b be unit vectors along the R, G and B axis of RGB color space, and define the vextors:

$$\begin{cases} u = \dfrac{\partial R}{\partial x} r + \dfrac{\partial G}{\partial x} g + \dfrac{\partial B}{\partial x} b \\[2mm] v = \dfrac{\partial R}{\partial y} r + \dfrac{\partial G}{\partial y} g + \dfrac{\partial B}{\partial y} b \end{cases} \tag{8}$$

Let the quantities g_{xx}, g_{yy} and g_{xy} be defined in terms of the dot product of these vectors, as follows:

$$g_{xx} = u \cdot u = u^T u = \left|\frac{\partial R}{\partial x}\right|^2 + \left|\frac{\partial G}{\partial x}\right|^2 + \left|\frac{\partial B}{\partial x}\right|^2 \tag{9}$$

$$g_{yy} = v \cdot v = v^T v = \left|\frac{\partial R}{\partial y}\right|^2 + \left|\frac{\partial G}{\partial y}\right|^2 + \left|\frac{\partial B}{\partial y}\right|^2 \tag{10}$$

$$g_{xy} = u \cdot v = u^T v = \frac{\partial R}{\partial x}\frac{\partial R}{\partial y} + \frac{\partial G}{\partial x}\frac{\partial G}{\partial y} + \frac{\partial B}{\partial x}\frac{\partial B}{\partial y} \tag{11}$$

It can be shown[4] that the direction of maximum rate of change of $C(x, y)$ is given by the angle:

$$\theta = \frac{1}{2}\arctan\left[\frac{2 g_{xy}}{g_{xx} - g_{yy}}\right] \tag{12}$$

$$F(\theta) = \left[\frac{1}{2}[(g_{xx}+g_{yy})+(g_{xx}-g_{yy})\cos2\theta+2g_{xy}\sin2\theta] \right]^{\frac{1}{2}} \tag{13}$$

Along with the process, the gradients of the three components (R , G , B) were respectively calculated for each scale $C(x,y)$, and the image edge were formed through synthesizing the corresponding three components with each coordinates points (x,y) .

3   Feature Extraction Using the Relative Moment

3.1   The Relative Moment

In the process of image retrieval, the mathematical characteristics with the scale, translation and rotation invariance should be chosen to achieve the purpose of image retrieval identification. MK Hu had firstly proposed the continuous function and also justified the scale invariance, translation invariance and rotation invariance.

For the discrete digital image $f(x,y)$, its $p+q$ order moment is defined as

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{14}$$

its $p+q$ centre moment is defined as

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{15}$$

Where $\bar{x} = \dfrac{m_{10}}{m_{00}}$, $\bar{y} = \dfrac{m_{01}}{m_{00}}$ is the regional center of gravity.

In order to obtain the independent nature of image scaling, we can normalize the center of operations and the standardized the center distance could be expressed as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^{\gamma}} \qquad \gamma = \frac{p+q}{2}+1, \quad p+q=2,3,\cdots \tag{16}$$

seven invariant moments exported are as follows:

$$\varphi_1 = \eta_{20} + \eta_{02}$$

$$\varphi_2 = (\eta_{20}-\eta_{02})^2 + 4\eta_{11}^2$$

$$\varphi_3 = (\eta_{30}-3\eta_{12})^2 + (3\eta_{21}-\eta_{03})^2$$

$$\varphi_4 = (\eta_{30}+\eta_{12})^2 + (\eta_{21}+\eta_{03})^2$$

$$\varphi_5 = (\eta_{30}-3\eta_{12})(\eta_{21}+\eta_{03})\left[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21}+\eta_{03})^2\right]$$
$$+ (3\eta_{21}-\eta_{03})(\eta_{21}+\eta_{03})\left[3(\eta_{30}+\eta_{12})^2 - (\eta_{21}+\eta_{03})^2\right]$$

$$\varphi_6 = (\eta_{20}-\eta_{02})\left[(\eta_{30}+\eta_{12})^2 - (\eta_{21}+\eta_{03})^2\right] + 4\eta_{11}(\eta_{12}+\eta_{30})(\eta_{21}+\eta_{03})$$

$$\varphi_7 = (3\eta_{21}-\eta_{03})(\eta_{12}+\eta_{30})\left[(\eta_{30}+\eta_{12})^2 - 3(\eta_{21}+\eta_{03})^2\right]$$
$$+ (3\eta_{12}-\eta_{30})(\eta_{21}+\eta_{03})\left[3(\eta_{30}+\eta_{12})^2 - (\eta_{21}+\eta_{03})^2\right] \tag{17}$$

However, digital images which are digitized and discrete do not have scale invariance. Though Hu's invariant moments has the advantages of rotational invariance and uniqueness of describing the graphics, etc, the problem of large amount of calculation can be seen from the expression of φ_5, φ_6, φ_7.

   This paper took the relative moments to describe the shape features of objects, made them only relevant to the shape of the object and irrelevant to the structure of the scaling. According to Reference[5], the relative moments are defined as follows:

$$R_1 = \frac{\sqrt{\varphi_2}}{\varphi_1} = \frac{\sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{\mu_{20} + \mu_{02}} \qquad R_2 = \frac{\varphi_1 + \sqrt{\varphi_2}}{\varphi_1 - \sqrt{\varphi_2}} \quad R_3 = \frac{\sqrt{\varphi_3}}{\sqrt{\varphi_4}} \qquad R_4 = \frac{\sqrt{\varphi_3}}{\sqrt[4]{|\varphi_5|}} \qquad R_5 = \frac{\sqrt{\varphi_4}}{\sqrt[4]{|\varphi_5|}}$$

$$R_6 = \frac{|\varphi_6|}{\varphi_1 \bullet \varphi_3} \quad R_7 = \frac{|\varphi_6|}{\varphi_1 \bullet \sqrt{|\varphi_5|}} \qquad R_8 = \frac{|\varphi_6|}{\varphi_3 \bullet \sqrt{|\varphi_2|}} \qquad R_9 = \frac{|\varphi_6|}{\sqrt{\varphi_2 \bullet |\varphi_5|}} \quad R_{10} = \frac{|\varphi_5|}{\varphi_3 \bullet \varphi_4} \qquad (18)$$

The ten relative moments which are extended from the moment invariants unified the calculation of regional and structural features, having the proportion of translation and rotation invariance.

3.2   Normalization

The relative moment obtained by each scale of Images are totally 7×10, together to form a feature vector, that is:

$$F_{i \cdot j} = \left\{ f_{i,1}, f_{i,2}, \cdots f_{i,j} \cdots, f_{i,N} \right\} \qquad i = 1, \cdots 7; N = 1,2, \cdots 10 \qquad (19)$$

Since each component of feature vector has a different physical meaning and the variance, as well the range differ greatly, normalization must be performed rather than directly measuring the similarity lest great deviation will be caused.

   Here we took the Gaussian normalization method, suppose the feature column vector $F_{i \cdot j} = \left\{ f_{i,1}, f_{i,2}, \cdots f_{i,j} \cdots, f_{i,N} \right\}^T$ is a Gaussian series. Firstly we calculated the mean and standard deviation of the series to and then took the following formula for feature normalization.

$$F_{i \cdot j} = \frac{F_{i,j} - \mu_{i,j}}{\sigma_{i,j}} \qquad (20)$$

68% of the normalized result value fell on [-1,1]; while took $3\sigma_{i,j}$ for normalization, that is, to replace the denominator with $3\sigma_{i,j}$, 99% of its value fell on [-1,1]. In practice, we set the value outside the range -1 or 1 to ensure that all the eigenvalues $F_{i,j}$ fell within the range.

3.3   Similarity Measurement

Ancient-coin-image-retrieving mainly takes the method of similarity. Namely, to sort the image database based on discriminant rules and query requirements of the similarity and provide a number of images that are the most similar to the requirement. This article took the Euclidean distance to perform similarity measurement. Suppose the query image Q and the input image I, the distance between the two images is:

$$D(Q,I) = D(F(Q), F(I)) = \sqrt{\sum_{i=1}^{70} (F_{i,j}^Q - F_{i,j}^I)^2} \qquad (21)$$

The similarity of the two could be defined as:

$$Sim(Q,I) = (1 - D(Q,I)) \times 100\% \qquad (22)$$

The closer the distance between the two, the higher the similarity, that is, the more similar the two images.

4   Experimental Results and Analysis

To verify the feasibility and rationality of the method, this paper builds a database with the image of 256 ancient coins, which covers the main ancient coin types from the Spring and Autumn Period to the Qing dynasty.

In Fig. 1, the first image in the upper-left corner was a query image (sample image) and search results returned the top 8 images that are the most similar to the query image. According to Figure 1, even if the shapes are translating, rotating, scaling, they could still be good matches and the similarity was shown in Table 1. Though there are also knife-like ancient-coins and many other shapes of ancient coins in the image library, this time the retrieved images are all circular coins and the retrieval accuracy are of 100%.
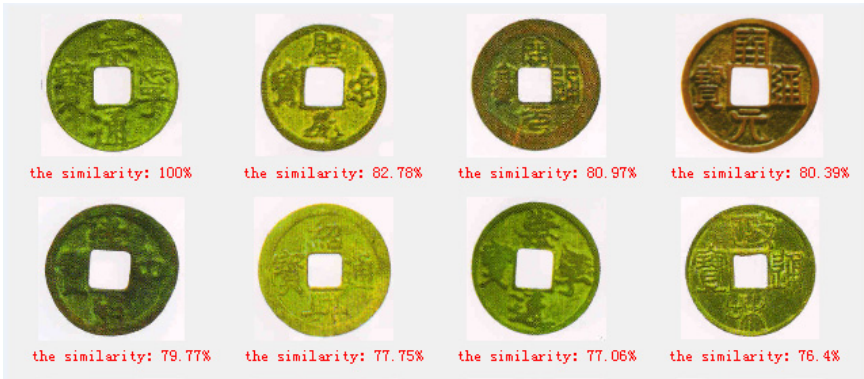
Fig. 1. Query Results

Table 1. Search results for Euclidean distance and similarity

image	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8
Euclidean distance	0.0	0.1722	0.1903	0.1961	0.2023	0.2225	0.2294	0.2359
similarity	100%	82.78%	80.97%	80.39%	79.77%	77.75%	77.06%	76.4%

5   Summaries

This paper synthesized the analysis methods of the image outer edge and the image area, proposed a method of ancient coins feature extraction and image retrieve by wavelet transform and the relative moments, constructed ancient-coin- image retrieval system that integrated the large-scale contour information and the small-scale detail information in the wavelet transform to recognize ancient-coin images. Experimental results show that the method has good versatility, as well as robustness.

References

[1] Hu, M.K.: Visual Pattern Recognition by Moment Invariants. IEEE Trans. on Information Theory IT-8, 179–187 (1962)
[2] Chen, C.C.: Improved Moment Invariants for Shape Discrimination. Pattern Recognition 26(5), 683–686 (1993)
[3] Gonzalez, R.C.: Digital Image Processing, pp. 247–249. Publishing House of Electronics Industry, Beijing (2003)
[4] Zenzo, S.D.: A note on the gradient of a multi-image. Computer Visual Graphic Image Processing (33), 116–125 (1986)
[5] Wang, B., Sun, J., Cai, A.: Relative Moments and Their Applications to Geometric Shape Recognition. Journal of Image and Graphics 6(3), 296–300 (2001)

The Key Data Mining Models for High Dimensional Data

Xiang Deng, Beizhan Wang, Haifang Wei, and Minkui Chen

Software School,
Xiamen University,
Xiamen, China
{dxl20dxl20,weihaifang1989,chenney447}@gmail.com,
Wangbz@xmu.edu.cn

Abstract. With the rapid development of computational biology and e-commerce applications, high-dimensional data becomes more and more powerful. Thus, it is an urgent problem of great importance when mining high-dimensional data. However, there are some challenges for mining data of high dimensions, the first one is the curse of dimensionality and the second one is the meaningfulness of the similarity measure in the high dimension space. In this paper, we present several state-of-art techniques for constructing three data mining models with analyzing high-dimensional data, these models include frequent pattern mining, clustering, and classification. And we discuss how these methods deal with the challenges of high dimensionality.

Keywords: High-dimensional data mining, Frequent pattern mining, Clustering, Classification.

1 Introduction

Data mining is focused by many fields, which including frequent pattern mining, clustering, classification and other models. These models can be used to represent, analyze and summarize large data set using potential models. And frequent pattern mining, clustering and classification are the most widely used data mining models. These existing frequent pattern mining algorithm, clustering algorithms and classification algorithms can have a good performance by analyzing not very high dimensional data.

The development of various new application domains, such as bioinformatics and e-commerce, underscores the need for analyzing high dimensional data [1], [2]. For example, in a gene expression micro array data set, there could be tens or hundreds of dimensions, each of which corresponds to an experimental condition and in a customer purchase behavior data set, there may be up to hundreds of thousands of merchandizes, each of which is mapped to a dimension. However, there are two challenges of data mining of high dimensions: (1) The curse of dimensionality and (2) The specificity of similarities between points in a high dimensional space diminish.

Many mining algorithms computational complexity will increase the dimension exponentially. As the dimension increases, generating more and more data, the traditional measure of similarity in high-dimensional space has become meaningless,

which makes the similarity measure based on the traditional mining algorithms, such as clustering, in high dimensional space can not be achieved.

This paper focuses on the state-of-art techniques for constructing these three data mining models. Section 2 describes frequent patter mining. Section 3 gives clustering methods. Section 4 gets classification approaches and Section 5 make a conclusion about these three data mining models.

2   Frequent Pattern

Frequent pattern mining is the most crucial step of association analysis. The association rules and frequent patterns first proposed by Agrawal in 1993, and put forward the concept of the Apriori algorithm for mining association rules [1], [2], [3]. Most algorithms for mining frequent patterns utilize the Apriori property stated as follows. First the frequent k-itemsets L_k, and two frequent k-itemsets have k +1- items that connect candidate sets C_k +1, the candidate set by computing the test of whether support for the frequent k +1- itemsets L_k + 1, until the frequent item sets can not be far [3]. The disadvantage is the Apriori algorithm will produce a large number of candidate sets and the need to repeatedly scan the database for the support of candidate itemsets calculation. To enhance the efficiency of the algorithm, the researchers made a lot of improved algorithms, such as Apriori-based algorithm AprioriTid, DIC, Partition, DHP [4] and based on frequent pattern tree FP-growth method [5].

The problem that the full set of frequent patterns contains significant redundant information and consequently the number of frequent patterns is often too large. To address this issue, Pasquier [6] proposed to mine a selective subset of frequent patterns, called closed frequent patterns. If the number of occurrences of a pattern is the same to all its immediate sub patterns, then the pattern is considered as a closed pattern. The CLOSET algorithm [7] is proposed to expedite the mining of closed frequent patterns. CLOSET uses a novel frequent pattern tree (FP structure) as a compact representation to organize the data set. It performs a depth-first search, that is, after discovering a frequent itemset A, it searches for super patterns of A before checking A's siblings [1], [2].

An algorithm for mining frequent closed pattern is CHARM [8]. The similar with CLOSET, CHARM algorithm also uses depth-first search strategy, The difference between CHARM and CLOSET is that CHARM stores the data set in a vertical format and using tidset for itemsets with the support set, and the introduction of diffset records correspond set of candidate nodes and its parent nodes with tidset. This can significantly reduce the huge memory. Through the IT tree (Itemset-Tidset Tree) space and services to achieve key double space search, the search space more efficient cutting. CLOSET and CLOSET+ [10] algorithm are based on the algorithm introduced a number of improvement strategies, such as bottom-up physical projection and top-down mix of the virtual projection strategy of combining a variety of pruning strategies such as project merged superset pruning, Item ignored (item skipping) and other space technology to improve the efficiency of the algorithm and the algorithm scalability [6].

These algorithms can find frequent closed patterns when the dimensionality is low to moderate. When the number of dimensions is very high, e.g., greater than 100, the efficiency of these algorithms could be significantly impacted [1], [2]. CARPENTER

[9] is therefore proposed to solve this problem. It first transposes the matrix representing the data set. Next, CARPENTER performs a depth-first row-wise enumeration on the transposed matrix. Rather than the conventional methods listed items, and much more effective pruning strategies to make the algorithm for high dimensional data sets of biological information showing a higher efficiency.

3   Clustering

Clustering is widely adopted data mining model and the purpose of cluster analysis is to divide the data set into different groups (clusters) based on the similarity of data objects, which can be used for e-commerce customer, the analysis of micro-biology, sequence analysis, geographic data analysis, pattern recognition and other fields. Researchers have proposed a large number of algorithms to suit different applications, according to algorithm ideas, traditional clustering methods can be divided into five categories: division methods, hierarchical methods, density-based methods, grid-based methods and model-based methods.

Traditional clustering methods to cluster high dimensional data set, the main two problems [1], [2]:

(1) In a high dimensional data set, for any point, its distance to its closest point and that to the farthest point tend to be similar. This phenomenon may render the clustering result sensitive to any small perturbation to the data due to noise and make the exercise of clustering useless.

(2) In a high dimensional data space the distribution of the lower dimensional data is not intensive, which make the distance between two data points is very similar. This is a common phenomenon, while the traditional clustering methods are based on distance, and therefore can not be build clusters based on distance in a high dimensional space.

In recent years, in order to solve these problems three methods proposed as follows.

3.1   Feature Transform [11]

This is a traditional method, including principal component analysis and singular value decomposition and other strategies. By linear combination of the method the dimension of the original data set to merge new dimension k such that k-means, such as the traditional algorithm for a class in this new dimension in k-effective clustering, which aims to reduce dimension. However, this method has three disadvantages: first, difficult to determine the appropriate k value; Second, there are a lot of high-dimensional space, independent of dimension and cover up the cluster, causing difficulties to the cluster; third, some clustering clusters are useless. Therefore, this method is only suitable for the majority of prior known dimensions are related to cluster high dimensional data space.

3.2   Subspace Clustering [12]

Subspace clustering algorithm to expand the feature selection task, try to set the same data found in different sub-space clustering.Subspace clustering must evaluate features

on only a subset of the data, representing a cluster. Parsons refer to this "measure of locality" measure to dense this context and further divide the two categories of subspace clustering algorithms based on how they determine a measure of locality with which to evaluate subspaces.

The bottom-up search method takes advantage of the downward closure property of density to reduce the search space, using an APRIORI style approach. Algorithms first create a histogram for each dimension and selecting those bins with densities above a given threshold. The downward closure property of density means that if there are dense units in k dimensions, there are dense units in all (k ¡ 1) dimensional projections. Candidate subspaces in two dimensions can then be formed using only those dimensions which contained dense units, dramatically reducing the search space. The algorithm proceeds until there are no more dense units found. Adjacent dense units are then combined to form clusters. This is not always easy, and one cluster may be mistakenly reported as two smaller clusters. The nature of the bottom-up approach leads to overlapping clusters, where one instance can be in zero or more clusters. Obtaining meaningful results is dependent on the proper tuning of the grid size and the density threshold parameters. These can be particularly difficult to set, especially since they are used across all of the dimensions in the dataset. A popular adaptation of this strategy provides data driven, adaptive grid generation to stabilize the results across a range of density thresholds.

The top-down subspace clustering approach starts by finding an initial approximation of the clusters in the full feature space with equally weighted dimensions. Next each dimension is assigned a weight for each cluster. The updated weights are then used in the next iteration to regenerate the clusters. This approach requires multiple iterations of expensive clustering algorithms in the full set of dimensions. Many of the implementations of this strategy use a sampling technique to improve performance. Top-down algorithms create clusters that are partitions of the dataset, meaning each instance is assigned to only one cluster. Many algorithms also allow for an additional group of outliers. Parameter tuning is necessary in order to get meaningful results. Often the most critical parameters for top-down algorithms is the number of clusters and the size of the subspaces, which are often very difficult to determine ahead of time. Also, since subspace size is a parameter, top-down algorithms tend to find clusters in the same or similarly sized subspaces.

3.3   Clustering Based on Pattern Similar [1], [2], [13]

In many practical applications, there is another similarity, although different data objects to a sub-space dimension is not very close to the values, but showing a consistent pattern, such as rate of change in the same or similar trend.

A pCluster consists of a subset of objects *U* and a subset of dimensions D where for each pair of objects ul and u2 in U and each pair of dimension dl and d2 in D, the change of ul from dl to d2 should be similar to that of u2 from dl to d2. A threshold is used to evaluate the dissimilarity between two objects on two dimensions. Given a subset of objects and a subset of dimensions, if the dissimilarity between every pair of objects on every pair of dimensions is less than the threshold, then these objects constitute a pCluster in the given dimensions. A novel deterministic algorithm is developed to find all maximal pClusters, which utilizes the Apriori property held on pClusters.

4   Classification

The classification is also a very powerful data analysis tool. In a classification problem, the dimensions of an object can be divided into two types [1], [2]. One dimension records the class type of the object and the rest dimensions are attributes. The classification model captures the intrinsic associations between the class type and the attributes so that an (unknown) class type can be accurately predicted from the attribute values. For this purpose of build the model, the data is usually divided into a training set and a test set, where the training set is used to build the classifier which is validated by the test set. This paper introduces two models developed for classifying high dimensional data: SVM and rule-based classifiers.

4.1   SVM

Supporting vector machine (SVM) [14] is one of the developed classification models, proposed in 1998. The success of SVM in practice is drawn by its solid mathematical foundation that conveys the following two salient properties [1]. First, the classification boundary functions of SVM maximize the margin, which equivalently optimize the general performance given a training data set. Second, SVM handle a nonlinear classification efficiently using the kernel trick that implicitly transforms the input space into another higher dimensional feature space. HowOver, SVM has two problems. On the one hand, the complexity of training an SVM is at least $O(N^2)$ where N is the number of objects in the training data set. So when the training data set is large, it becomes too costly. On the other hand, since an SVM essentially generate a hyper-plain in a transformed high dimensional data space, to identify the principal dimensions that are most responsible for the classification is very difficult.

4.2   Rule-Based Classifiers

Rule-based classifiers [15] offer some potential to try to address the above two problems in SVM. Two steps in a rule-based classifier: First, the training data set attributes and characteristics of mining frequent association rules between class attributes; then use the association rules for classification. Compared with SVM, it is easy to understand. Currently, rule-based classification is the classification of research areas. However, Rule-based classification is facing three major problems: First, what method that classification rules; the second is how to effectively prune the rules in the training data mining will certainly have a lot of rules, classification rules will be some other rules Implication, therefore the rules need to determine the relationship between the implication redundant rules and then cut through the pruning method to simplify the classification model; third is when a new data classification, and this may be a number of matching rules, How to choose the most effective rules for this classification[2].

   A rule-based classifier consists of a set of rules in the following form: A1 [l1, u1]∩A2[l2, u2]∩. . .∩Am[lm, um]→C, where Ai[li, ui] is the range of attribute Ai's value and C is the class type. The above rule can be interpreted as that, if an object whose attributes' values fall in the ranges in the left hand side, then its class type is likely to be C with some high probability. Each rule is also associated with a confidence level that depicts the probability that such a rule holds. When an object satisfies several

rules, either the rule with the highest confidence or a weighted voting of all valid rules may be used for class prediction. However, neither CBA [15] nor CPAR [16] are targeted for high dimensional data. An algorithm called CMAR [17] is proposed to generate rule-based classifiers for high dimensional data set [1]. CPAR produce less directly from the training data, high-quality prediction rule that the greedy algorithm while generating rules using pruning methods side, it is more efficient than the CBA and CMAR, if a data object to meet a number of classification rules, first for each category selecting the best k rules of section, and then calculating the average for each type of expected accuracy of k rules, choosing the highest class average expected accuracy of predictions as the data object. Experiments show that CPAR has higher classification accuracy and classification speed [2].

5   Summaries

The key techniques are listed in Tab 1. The purpose of data mining researchers is to seek an effective way of analyzing the high dimensional data, so this kind of research is still in progress. However, future research goal is still to resolve challenges for mining data of high dimensions, including the curse of dimensionality and the meaningfulness of the similarity measure. At the same time, it is very important to improve efficiency and adaptability of the relevant frequent pattern mining, clustering and classification in high dimensional data set.

Table 1. Techniques of Three Models

Models	Techniques
Frequent pattern	Apriori CLOSET CHARM CLOSET+ CARPENTER
Clustering	Feature transform Subspace clustering Clustering based on pattern similar
Classification	SVM Rule-based classifiers

References

[1] Maimon, O., Rokach, L.: Data Mining and Knowledge Discovery Handbook, pp. 803–808. Springer (2010)
[2] Chen, H., Wang, Y., Wang, J.: The Research and Advances on Mining High Dimensional Data. Computer Engineering and Applications 24 (2006)
[3] Agrawal, R., Imielinski, T., Swami, A.N.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216. ACM Press (1993)

[4] Brin, S., et al.: Dynamic Itemset Counting and Implication Rules for Market Basket Analysis. In: SIGMOD 1997, pp. 255–264 (1997)

[5] Han, J., Pei, J., Yin, Y.: Mining Frequent Patterns without Candidate Generation: A Frequent- Pattern Tree Approach Mining Frequent Patterns without Candidate Generation. Data Mining and Knowledge Discovery (8), 53–87 (2004)

[6] Pasquier, N., Bastide, Y., Taouil, R., Lakhal, L.: Discovering Frequent Closed Itemsets for Association Rules. In: Beeri, C., Bruneman, P. (eds.) ICDT 1999. LNCS, vol. 1540, pp. 398–416. Springer, Heidelberg (1998)

[7] Pei, H.M.: CLOSET: an efficient Algorithm for mining frequent closed itemsets. In: Proceedings of ACM Workshop on Research Issues in Data Mining and Knowledge Discovery (2000)

[8] Zaki, H.: CHARM: An Efficient Algorithm for Closed Itemset Mining. In: Proc. 2002 SIAM Int. Conf Data Mining, SDM 2002, Arlington, VA, pp. 457–473 (2002)

[9] Pan, C.T., Yang, Z.: CARPENTER: finding closed patterns in long biological data sets. In: Proceedings of ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (2003)

[10] Wang, J., Han, J., Pei, J.: CLOSET+: Searching for the Best Strategies for Mining Frequent Closed Itemsets. In: Proc 2003 ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining, KDD 2003, Washington, D C, vol. 08 (2003)

[11] Ding, C., He, X., Zha, H., et al.: Adaptive dimension reduction for clustering high dimensional data. In: 2002 Proceedings of Second IEEE International Conference on Data Mining, pp. 147–154 (2002)

[12] Parsons, L., Haque, E., Liu, H.: Subspace Clustering for High Dimensional Data: A Review. SIGKDD Explorations 6(1), 90–105 (2004)

[13] Wang, H., Wang, W., Yang, J., Yu, P.: Clustering by pattern similarity in large data sets. In: Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD, pp. 394–405 (2002)

[14] Vapnik, V.N.: Statistical Learning Theory. John Wiley and Sons (1998)

[15] Liu, B., Ma, Y., Wong, C.K.: Improving an Association Rule Based Classifier. In: Zighed, D.A., Komorowski, J., Żytkow, J.M. (eds.) PKDD 2000. LNCS (LNAI), vol. 1910, pp. 504–509. Springer, Heidelberg (2000)

[16] Yin, H.: CPAR: classification based on predictive association rules. In: Proceedings of SIAM International Conference on Data Mining (2003)

[17] Li, W., Han, J., Pei, J.: CMAR: Accurate and efficient classification based on multiple class- association rules. In: ICDM 2001, San Jose, CA, pp. 369–376 (2001)

Study on the Methods of Encoding for Raster Data Compression

Guo Zhao and Wang Zheng

College of Mining Engineering, Hebei United University
Tangshan 063009, China

Abstract. With the rapid development of GIS, the increasing raster data volume occupies a lot of storage space, which reduces the rate of conduct and analysis of GIS. So it is vital to study the methods of encoding for raster data compression. Generally speaking, the common methods are not of pertinence enough to compress the data as much as possible. So the aim of this paper is to conclude and analyze some kinds of raster data structures and its compress encoding, based on which, the paper puts forward a new method of encoding for raster data compression. It is spiral encoding, which can compress the buffer areas data in a large degree.

Keywords: Raster data structure, GIS, Compression encoding, Spiral encoding.

1   Introduction

Raster data structure is a kind of inlaid data model that stores the graphic data with a unified grid and image element to describe the special phenomenon with the continuous tiled regular grid [1], as show in Fig.1.

					2			
				2				
		1		2				
			2				3	3
		2					3	3
		2					3	3
	2					3	3	3
2								
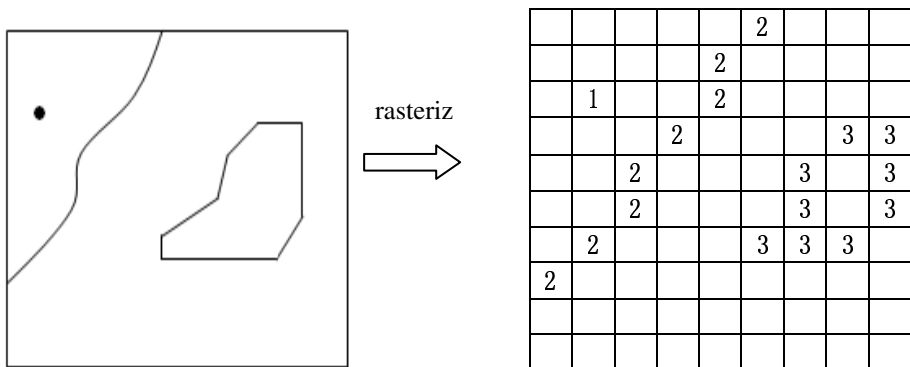
rasteriz

Fig. 1. Graphics of raster data

Raster data structure is used to describe the discretization of numerical value about two dimensional geographical elements. And every grid matches an attribute. The special of it is described with line and raw. The grid is square on average. But it can also be regular triangle, rhomb, rectangle, regular hexagon, and the regular hexagon is the most advantageous form to be the grid cell, which is analyzed and proved geometrically by Lin Hui. In the case of same data volume, the use of regular hexagon can improve the accuracy of the data. Or in the case of same accuracy it can decrease the data volume [2]. The accuracy of grid data depends on the side length of the grid. But the complicated entity can always be described with the unified grid, which results in the loss of information. So in order to decrease the loss of data, we usually ascertain the size of grid by making a use of the accuracy standard of the lease side length.

2   Run Length Encoding

Run length encoding is a method that combines the adjacent grids which own the same value (attribute) line-by-line and express it with the form (A_i, N_i). A is attribute. N is the number with the continuous same attribute. i is serial number, as show in Fig.2.

The idea of this method is simple and the method is easy to operate. So it can be used widely. But the pertinence of this method is not powerful enough. So the effect may be different with different picture, and the data volume may be larger than the data volume uncompressed.

4	7	6	6
6	6	5	5
3	5	5	5
5	7	7	7

binary mapping

serial number	binary sequence
1	（4, 1）
2	（7, 1）
3	（6, 4）
4	（5, 2）
5	（3, 1）
6	（5, 4）
7	（7, 3）
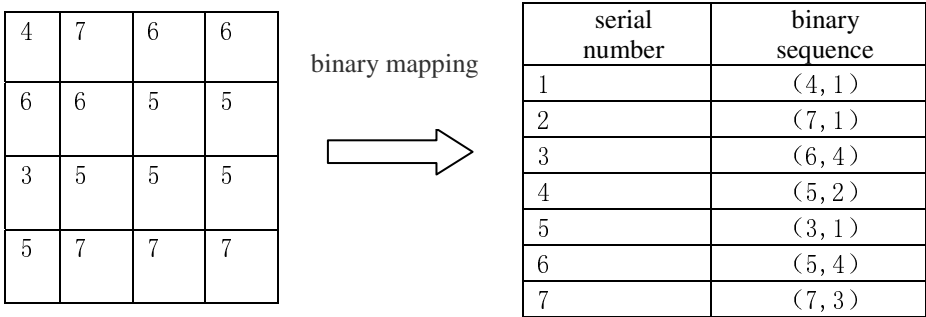
Fig. 2. Schematic diagram of run length encoding

3   Regular Quad-Tree Encoding

The method of quad-tree structure is to divide a grid map or picture into four parts equally, and then check the attribute (grayness). If all of the grid value is same in a subrange, stop the division. If not, divide the subrange into four subranges equally. As it is showed previously, we won't stop the division until all the subranges keep the same attribute or grayness, as show in Fig.3.

9	9	9	9	0	0	0	0
9	9	9	0	0	0	0	0
0	9	9	0	7	7	0	0
0	0	0	0	7	7	0	0
0	0	0	0	7	7	7	7
0	0	0	0	7	7	7	7
0	0	0	0	7	7	7	7
0	0	0	0	7	7	7	7
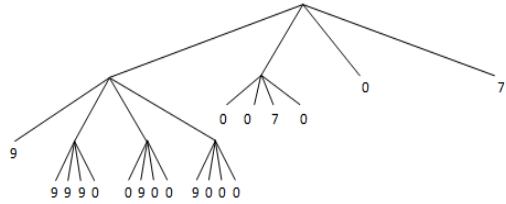
(1) quad-tree segmentation schemes   (2) quad-tree

Fig. 3. Structure of quad-tree

This method can calculate the sum of polygon efficiently, meanwhile decrease the internal storage and express the graphic structure accurately. And it is easy to realize. But the quad-tree can't be divided without stop, unless the graphic is the 2n*2n (n is integer) grid array, which limited the range of the images that can be managed.

4   Linear Coded Quad-Tree Encoding

To decrease the internal storage occupied by regular quad-tree up to the hilt, we put forward linear coded quad-tree. Compared with the regular quad-tree, we don't need to record midside nodes and use indicater. We only need to record the leaves nodes and express them with address codes. But this method also keeps the disadvantage of the limitation of managed images as the regular quad-tree does.

5   Wavelet Transform Encoding

Wavelet transform is a method that we can see a image as a wave, and divide it into easier wavelet [3]. To accomplish this transform, we should calculate the average value of the adjacent picture elements groups (such as 2,4,6,8 and so on). Meanwhile we should also record original elements groups and the difference (the wavelet coefficient) among the average picture elements groups. Most of the coefficients are 0 in an image. We usually round it into 0, even if it is not 0, for we can save lots of internal storage. Only by this way can we store the data in a lower resolution [4], as show in Fig.4.

C_0 is the discrete quantitative original image. C_3 is "vague image", which is lowpass filtered three times. d_1^1, d_1^2, d_1^3 are the series sub images which are highpass filtered three times horizontally. d_2^1, d_2^2, d_2^3 are the series sub images which are highpass filtered three times vertically. d_3^1, d_3^2, d_3^3 are the series sub images which are highpass filtered three times diagonally [5].
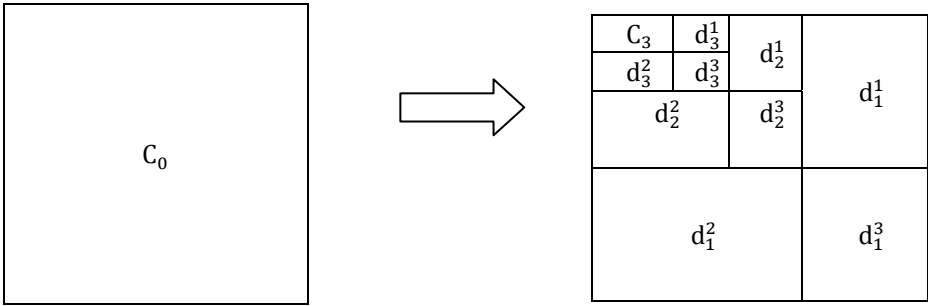
Fig. 4. Decomposed structure of C_0

The development of quantificat method wavelet transform produces a new arithmetic, which has more advantage than the existed image compression standard (such as JPEG arithmetic). To achieve the best property, a lot of expected speciality is required, such as orthogonality, symmetry and so on. In fact, it is hard to satisfy these expected specialities at the same time, which will limit the design of wavelet.

6   Spiral Encoding

Geography information system (GIS) has the powerful special analysis function, but the result of the analysis is always producing lots of grid images with butter areas occupying a large space of internal storage in a large degree, which slows down the rate of data transmission and the rate of analysis and conduct. So contraposing the character of buffer grid images and by studying the data compression encoding, I put forward a method, spiral encoding, that can compress the buffer grid images in a large degree.

There are mainly two types of the spiral encoding: the 2n*2n grid encoding and the (2n+1)*(2n+1) grid encoding (n is integer).

2n*2n grid encoding: We start to read the data from the image element that locates at n line and n row deasil. After this circle, we continue to read the data from the image element that locates at top left corner of the starting point ((n-1) line and (n-1) raw) deasil. Flowing the method above, we don't stop reading until all the data has been read. And the result should be showed as the run length encoding does, thus we express it with the form (A_i, N_i). A is attribute. N is the number with the continuous same attribute. i is serial number, as show in Fig.5.

(2n+1)*(2n+1) grid encoding: We start to read the data from the center image element. After it, we continue to read the data from the image element that locates at top left corner of the starting point deasil. Flowing the method above, we don't stop reading until all the data has been read. And the result should be showed as the run length encoding does, as show in Fig.6.

3	3	3	3
3	2	2	3
3	2	2	3
3	3	3	3

binary mapping

serial number	binary sequence
1	(2,4)
2	(3,12)
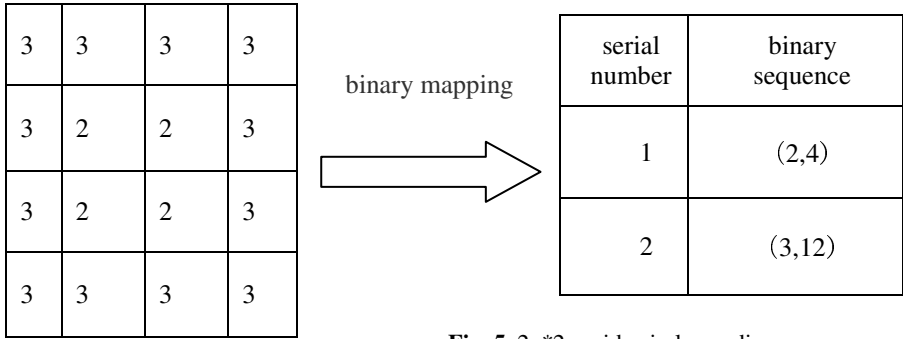
Fig. 5. 2n*2n grid spiral encoding

5	5	5	5	5	5	5	5	5	5	5
5	4	4	4	4	4	4	4	4	4	5
5	4	3	3	3	3	3	3	3	4	5
5	4	3	2	2	2	2	2	3	4	5
5	4	3	2	1	1	1	2	3	4	5
5	4	3	2	1	0	1	2	3	4	5
5	4	3	2	1	1	1	2	3	4	5
5	4	3	2	2	2	2	2	3	4	5
5	4	3	3	3	3	3	3	3	4	5
5	4	4	4	4	4	4	4	4	4	5
5	5	5	5	5	5	5	5	5	5	5

binary mapping

serial number	binary sequence
1	(0, 1)
2	(1, 8)
3	(2, 16)
4	(3, 24)
5	(4, 32)
6	(5, 40)
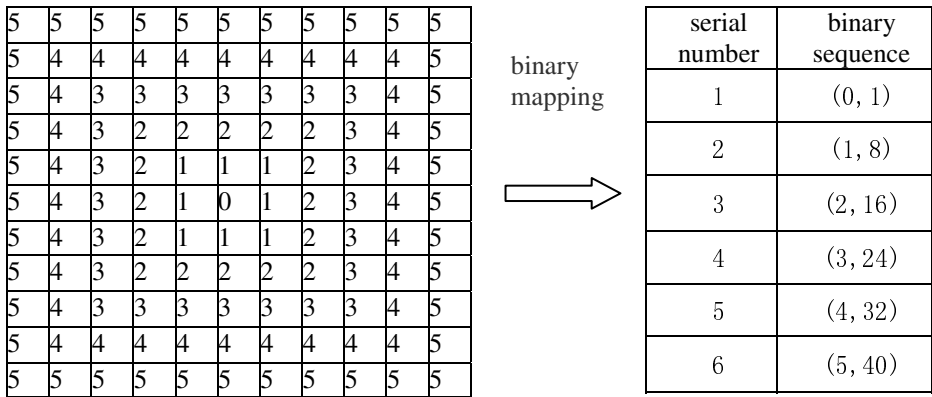
Fig. 6. (2n+1)* (2n+1) grid spiral encoding

5	5	5	5	5	5	5	5	5	5	5
5	4	4	4	4	4	4	4	4	4	5
5	4	3	3	3	3	3	3	3	4	5
5	4	3	2	2	2	2	2	3	4	5
5	4	3	2	1	1	1	2	3	4	5
5	4	3	2	1	0	1	2	3	4	5
5	4	3	2	1	1	1	2	3	4	5
5	4	3	2	2	2	2	2	3	4	5
5	4	3	3	3	3	3	3	3	4	5
5	4	4	4	4	4	4	4	4	4	5
5	5	5	5	5	5	5	5	5	5	5

binary

serial number	binary sequence
1	(5, 12)
2	(4, 9)
3	(5, 2)
4	(4, 1)
5	(3, 7)
6	(4, 1)
7	(5, 2)
8	(4, 1)
......	......
5 1	(5, 12)
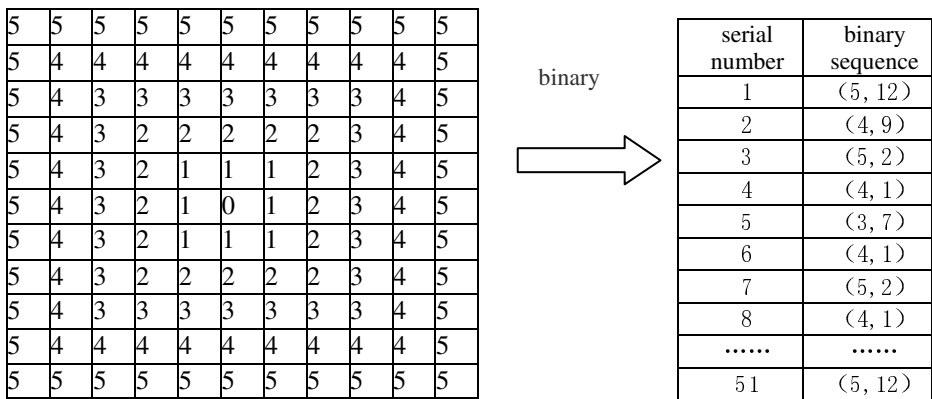
Fig. 7. Run length encoding

Different from the methods showed above, the spiral encoding possesses powerful pertinence and it can compress the buffer grid in a large degree, which will save lots of storage space and speed up the rata of analysis and conduct.

And then, let's make a comparison between spiral encoding and run length encoding.

The compression degree of run length encoding mainly depends on the character of grid data. So we can forecast it by estimating the unnecessary data Re:

Re=1-Q/m*n
Q is the times that the adjacent attribute values have changed.
m is the lines of the grid.
n is the raw of the grid.

If the Re value is greater than 1/5, the grid data can be compressed obviously. The effect of compression can be described by compression ratio, S=n/K (the maximal serial number). That is to say, the larger the compression ratio value is, the more obvious the effect of compression is [6].

The method above can also be used to spiral encoding.

So encoding the Fig.6 by run length encoding, we can conclude that $Re_2 = 0.58$ and $S_2 = 0.26$ comparing $Re_2 = 0.95$ and $S_2 = 1.83$ encoded by spiral encoding. So it is obvious that in the aspect of buffer grid image compression, the spiral encoding is more effective, as show in Fig.7.

7   Summaries

With the rapid development of GIS and the increasing data volume, there exist lots of methods of compression encoding. But the limitation that we know about spacial data and the way we study the spacial data leads to the average methods of compression encoding lack pertinence, which may not compress the data as much as possible and bring lots of inconvenience to GIS. So I suggest that we should create the method of compression encoding with more pertinence, which will save a great number of storage space and make our work more convenient.

References

[1] Huang, X.-Y., Ma, J.-S., Tang, Q.: An Introduction to Geographic Information System, p. 43. Higher Education Press, Beijing (2007)
[2] Lin, H., Peng, C.-H.: The Role of Raster Pixel Size and Shape in Geographic Information System. Remote Sensing Information (1) (2001)
[3] Addison, P.S.: The Illustrated Wavelet Transform Handbook. Institute of Physics Publishing, Bristol (2001)
[4] Chang, K.-T.: Introduction to Geographic Information Systems, p. 85. Tsinghua University Press, Beijing (2009)
[5] Chen, S.-S., Yang, L., Tan, M.-F.: A Dual-pyramid Data Structure for Image Coding Using Wavelet Transform. Journal of Image and Graphics (8, 9) (1997)
[6] Huang, X.-Y., Ma, J.-S., Tang, Q.: An Introduction to Geographic Information System, p. 46. Higher Education Press, Beijing (2007)

Research on Development of Agricultural Geographic Information Ontology[*]

Yongqi Huang[1] and Gaoyang Deng[2,**]

[1] Hubei Key Laboratory of Economic Forest Germplasm Improvement and Resources Comprehensive Utilization, Huanggang Normal University, Huanggang 438000, P.R. China
[2] Commercial School, Huanggang Normal University, Huanggang 438000, P.R. China
dgyhgnu@sina.com

Abstract. This paper first analyzes the reason why agricultural geographic information gives rise to semantic heterogeneity and solution thereof. Although OWL is the standard of ontology representation language in semantic web, it is insufficient in representing spatial characteristics, especially spatial relationship. Consequently it is pointed out to builds geo-ontology by virtue of three theories such as mereology, topology and location theory in this paper. This paper introduces mereology, topology and location theory, and then discusses how to adopt these three theories to builds geo-ontology. The outcome of experiment shows that solution put forward by this paper is feasible.

Keywords: Agricultural geographic information, Semantic heterogeneity, Mereology, Location, Topology, Geo-ontology.

1 Introduction

Agricultural geographic information is semantic interpretation of agricultural geographic data and attribute data, which consists of four interrelated characteristics, namely spatial position, spatial relationship, attribute property and temporal property. Agricultural geographic information is involved in each department or organization in society, which plays an important role in agricultural production and decision-making. Agricultural geographic information is distributed in heterogeneous systems, possessing different data formats and specifications and adopting different concepts and terms. This is to say, a semantic heterogeneity problem exists in agricultural geographic information[1]. Previously, owing to lack of efficient and reasonable technique means, it is difficult to completely share and integrate agricultural geographic information in digital agriculture application, which restricts application development of digital agriculture in a large part[2].

At present, sharing and integrating geographic spatial information have reached the stage of semantic interoperation and integration of spatial information based on ontology theory and technology. In other words, it is possible to solve semantic

---

heterogeneity problem of agricultural geographic information through ontology theory and technology[3]. Geo-ontology researches detailed connotations and hierarchy relationships of concepts of geographic spatial information in geographic information domain and offers semantic annotations of concepts so as to solve the problem derived from different geographic cognition and the intertranslation problem of logic description thereof, which can be applied to sharing and integrating geographic information.[4]

Geo-ontology is essentially independent of any concrete representing language, but it is necessary to choose a formalized ontology language to represent it. OWL, recommended by W3C, is the standard of ontology representation language in semantic web. As far as general ontology is concerned, semantic relationships that need to be represented are part-of, kind-of, instance-of and attribute-of. OWL provides abundant modeling primitives, so it is an ideal language representing ontology. As for geo-ontology, however, except for attributes characteristics of geographic concepts, it is more important to represent spatial characteristics, especially spatial relationship. Those spatial relationships are very important for geo-ontology, which are applied to spatial analysis, spatial query and spatial reasoning. OWL is incapable of dealing with spatial characteristics of geo-ontology. RCC (short for Region Connection Calculus) theory and n-intersection model have extensively been applied to topology description, but they still have a certain limitations[5]. In addition, geo-ontology should pay attention to positions, boundaries and whole-part relationships of spatial objects[6]. Consequently they are not suited to be basis of formalized representation of spatial characteristics in geo-ontology[7].

Because OWL is insufficient in representing geo-ontology, it is feasible to building geo-ontology by virtue of three theories such as mereology, location theory and topology. Application of those three theories may implement formalize representation of spatial relationships, spatial positions and spatial boundaries of geographic entities, and build a set of axiom system. Those axioms will be added to OWL as modeling primitives, which can provide formalize representation of spatial characteristics and spatial relationship in geo-ontology built by OWL.

2   Construct Spatial Relationship Based on Three Theories

Mereotopology is a formal theory of part-whole and connection relationships[8]. Space cognition and spatial reasoning of people depend on part-whole relationship in a large part, so part-whole relationship is very important in representing geographic space[9][10]. The primitive relation 'x is part of y' we symbolize by 'P(x, y)', which we take to be true when x is any sort of part of y, including y itself[11].

More generally, mereology cannot account for some very basic spatial relations, such as the relationship of continuity between two adjacent parts of an object, or the relation of one thing's being entirely inside or surrounded by another[12] [13]. To provide a systematic account of such relations will require topological machinery.

Topology, constructed on a mereological basis, is generally used to describe relative position relationship between geographic objects, which can be described by connection relationship[14]. Connection relationship is defined by boundary. Boundary problem is a very complex problem. Boundary can be divided into bond fide boundary and fiat boundaries. The central concept is the concept of boundary as

illustrated by the outer surface of a sphere. As a primitive, we assume 'x is a boundary for y', which we symbolize as 'B(x, y)'. We say boundary for, rather than boundary of, to allow for boundaries that are not maximal (corners, edge segments, parts of surfaces)[15].

Location, a theory based on Mereology, is used to research relationship between geographic object and spatial position occupied by it, which is very complex and an emphasis geo-ontology represent[16]. Geographic object is not equivalent to spatial position occupied by it. This is to say, geographic object is not representative of spatial position occupied by it, vice versa. The reason is that a space can be occupied by multiple geographic objects. We can use the notation 'L(x, y)' to indicate that 'x is located at y', that x stands to y in the primitive relational tie of *exact location*[17].

Through the above three theories, it is possible to implement formalized description of spatial positions, part-whole relationships, spatial topological relationships of geographic objects, and further realize spatial reasoning according to those formal axioms[18].

For example, we may construct eight basic spatial topological relationships between regions on the basis of the above three theories.

(1) Disjoint(x, y). The definition of Disjoint is exactly opposite to DP1. If there exists no z which sets up Part-of(z, x) and Part-of(z, y), x is disjoin from y, namely, Disjoint(x, y).

Disjoint(x, y) := $\neg \exists$z(Part-of(z, x) $\wedge$ Part-of(z, y));

(2) Touch(x, y). The definition of Touch is relatively complex, which needs to resort to Overlap and Interior Part definition DB3. Touch(x, y) indicates x only coincides with y on the boundary, not inside. Consequently, if x overlaps with y, and there exists no z which sets up that z is inside x and y at the same time, x touches with y.

Touch(x, y) := O(x, y) $\wedge \exists$z(IP(z, x) $\wedge$ IP(z, y));

(3) Overlap(x, y). Overlap can be defined by DP1, but DP1 is not strict Overlap relationship. In order to a strict Overlap relationship, we add the following restricts on the basis of DP1: x is not a part of y, y is also not a part of x; and there exists z which sets up that z is inside x and y at the same time.

Overlap(x, y) := O(x, y) $\wedge \neg$ Part-of(x, y) $\wedge \neg$ Part-of(y, x) $\wedge \exists$z(IP(z, x) $\wedge$ IP(z, y));

(4) Equal(x, y). If x is a part of y, y is also a part of x, namely part-of(x, y) and part-of(y, x), x is equal to y.

Equal(x, y) := Part-of(x, y) $\wedge$ Part-of(y, x)

(5) Contains(x, y) and Within(x, y). The definition of Contains(x, y) and Within(x, y) only resorts to Interior Part definition DB3. Within(x, y) means that x is inside y, which can be represented by IP(x, y). On the contrary, Contains(x, y) indicates that x contains y, which can be defined by IP(y, x).

Contains(x, y) := IP(y, x);
Within(x, y) := IP(x, y);

(6) Cover(x, y). In fact, Cover relationship is Inscribed-By relationship. Influenced by Egenhofer who represents this relationship with Cover, a number of researchers called it as Coverage relationship. Cover(x, y) means that x covers y, this is to say, x is inscribed by y. So y must be inside x, and the boundary of y overlaps with x.

Cover(x, y) := Part-of(y, x) $\wedge \forall z(B(z,x) \rightarrow O(y, z));$

(7) CoverBy(x, y). In fact, CoverBy relationship is Inscribing relationship. CoverBy (x, y) means that y covers x, this is to say, y is inscribed by x. So x must be inside y, and the boundary of x overlaps with y. Whose definition is exactly opposite to Cover(x, y)

CoverBy(x, y) := Part-of(x, y) $\wedge \forall z(B(z,y) \rightarrow O(x, z))$

According to the above eight new axioms, it is possible to not only represent spatial relationship between two known geographic concepts, but also judge spatial relationships between two arbitrary polygons by defined axioms and show them as ontology. We can edit polygon of arbitrary shape with open source software JUMP, then judge and output spatial relationship between them based on OWL format.

Here it should be pointed out that these three theories need to cooperate with each other to represent part-whole relationships, spatial postion relationships and topological relationships of agricultural geographic entities in agricultural geographic information domain. Any individual theory can not complete representation of agricultural geographic entities.

3   Development of Agricultural Geographic Information Ontology

The confine of agricultural geographic spatial information domain has considerable flexibility. As a result, when establishing domain ontology of spatial information we firstly need to determine domain confine, be clear about the aim of establishing ontology and specify application domain of ontology. In addition, we should apply pertinent classification methods to classifying domain knowledge too, for clear structure classification make organization of domain knowledge model orderly and has an important effect on reuse and integration of it.

A complete geo-ontology should consist of such primitives as geographic concept, geographic relationship, geographic axiom and geographic individual.

Geographic Concept Representation

In geo-ontology, concept or class is generally organized through hierarchy system such as lattice or tree. Concept lattice or tree and geographic concepts therein con constitute the chief problem, namely basic axis or main axis, and that concept attributes and relationships among concepts constitute auxiliary axis or secondary axis of geo-ontology. It is easily implemented on the basis of concept classification. The following is particular steps.

1) Determine primitive concepts and relationships, which are independent of other concepts and relationships, and applied to composed other concepts and relationships.

2) Adopt primitive concepts and relationships compose news new concepts and relationships, and classify them.

3) Find out attributes of individuals in each class, and establish relationships between individuals and attributes. Attributes include spatial attribute, temporal attribute, natural state, substance composition/covering property, purpose/use attribute and action/process attribute.

4) Determine semantic relationships among classes, and further label basic type of each relationship and mutual relationship among them like inverse relationship, transitive relationship and reflexive relationship.

Geographic Relationship Representation

Geographic relationship means a certain relation among geographic concepts in geographic domain. Relationships among geographic concepts include semantic relationship, temporal relationship and spatial relationship. Semantic relationship reflects correlative relationship of geographic concepts at semantic level, including equivalence relationship, general-special relationship, intersection relationship, non-intersection relationship and whole-part relationship. Spatial relationship is divided into distance relationship, direction relationship, topological relationship and similar relationship. In geo-ontology temporal relationship can be represented by interval algebra theory (Allen, 1983). Therein each time slot includes two endpoints. It may distinguish 13 basic temporal relationships through comparing relationships among endpoints of two time slots: Before(<), Meets(m), Overlaps(o), Starts(s), During(d), Ends(e), After(>), Met-by(m˜), Overlapped-by(o˜), Started-by(s˜), Includes(i˜), Ended-by(e˜) and Equals(≡).

In geo-ontology geographic relationship is generally described as ObjectProperty of geographic concept, composing secondary axis of geo-ontology. For each type of spatio-temporal relationship Moreover we can give it a definition and adopt rdfs:domain and rdfs:range of RDFS to stipulate its corresponding concept class, so as to constrain its application domain and confine.

Geographic relationship can be succeeded. At all events semantic relationships among higher level concepts are fixed, and can be succeeded by their all subclasses.

Geographic Axiom Representation

In geo-ontology axiom means commonly-recognized geographic laws, geoscience knowledge and some rules or constraints of geographic relationships in geoscience domain, so as to carry through reasoning and ensure consistency and integrity of geo-ontology. According to complexity of axiom in geo-ontology it can be divided into simple axiom and complex axiom. Simple axiom is used to ensure consistency and integrity of classes, attributes and individuals in geo-ontology, and can be represented by restriction rules (i) of ontology in itself. Those restriction rules consist of restriction rules of class, attribute and individual.

OWL-DL, a sublanguage of ontology language OWL, provides complete concept and classification definition, and also furnishes enough axioms operating classes. OWL-DL, however, is based on description logic and far from representing complex rules. For instance, it can not represent "if... then …"reasoning structure. On the other hand, "if... then …" is basic structure in reasoning. OWL-DL is incapable of representing spatio-temporal reasoning rules. For example, spatial entity A contains spatial entity B, and that spatial entity B contains spatial entity C, then A should contain C. But OWL-DL isn't able to define the reasoning rules like Contain(?A, ? B) ^ Contain(?B, ? C) => Contain (?A, ? C). Thus we should find out other solution methods to solve the representation of complex rules.

Geographic Individual Representation

Geographic individual means concrete instances concept or class in geo-ontology. Individual corresponds to concrete object in database system, composing concrete record of underlying database. The statements of individuals and their set constitute so-called knowledge base, which are important composition of geo-ontology. In geo-ontology we don't always give explicit statement, which depends on content and purpose of geographic concept model.
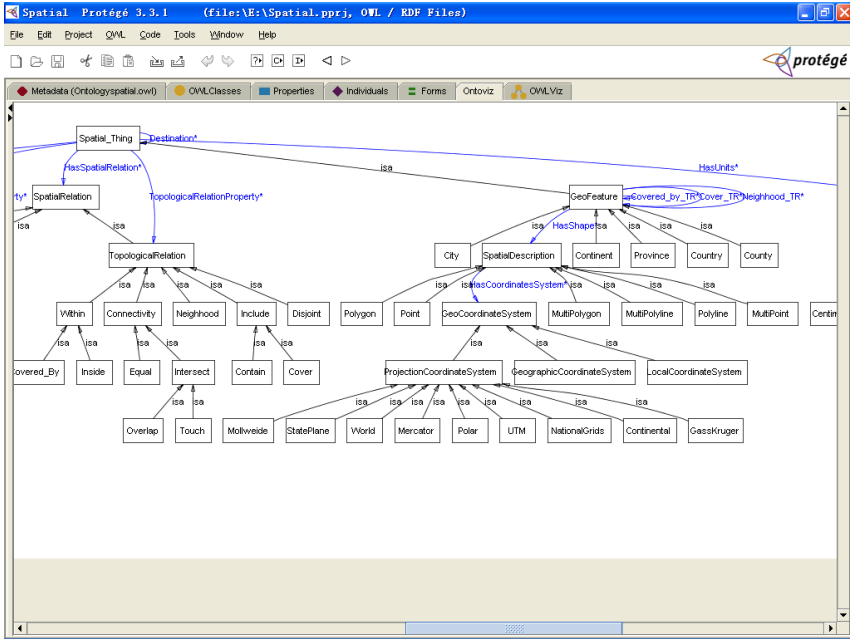
Fig. 1. A geo-ontology established in protégé environment

4   Summaries

Because agricultural geographic information is distributed in heterogeneous systems, a semantic heterogeneity problem exists in agricultural geographic information. Geo-ontology can solve the problem derived from different geographic cognition and the intertranslation problem of logic description thereof, which can be applied to sharing and integrating geographic information. Geo-ontology, however, except for attributes characteristics of geographic concepts, it is more important to represent spatial characteristics, especially spatial relationship. For OWL is insufficient in representing geo-ontology, it is feasible to building geo-ontology by virtue of three theories such as mereology, location theory and topology. This paper introduces mereology, location theory and topology, and then discusses how to adopt these three theories to builds geo-ontology. The outcome of experiment shows that solution put forward by this paper is feasible.

References

[1] Staford, J.V.: Implementing Precision Agriculture in the 21th Century. Journal of Agriculture Engineering Research 1(76), 267–275 (2000)

[2] Thompson, E.E., Duke, H.R., Westfall, G.W., et al.: Interdisciplinary Integrated Precision Farming Research. Precision Agriculture 3(1), 47–61 (2002)

[3] Fonseca, F., Egenhofer, M., Davis, C., Camara, G.: Semantic granularity in ontology-driven geographic information systems. Annals of Mathematics and Artificial Intelligenc 36, 121–151 (2002)

[4] Frank, A.U.: Tiers of Ontology and Constraints in Geographical Information System. Geographical Information Science 15(7), 667–678 (2001)

[5] Wolter, F., Zakharyaschev, F.: Spatio-Temporal representation and reasoning based on RCC-8. In: Cohn, A.G., Giunchiglia, F., Selman, B. (eds.) Proc. of the 7th Conf. on Principles of Knowledge Representation and Reasoning, pp. 3–14. Morgan Kaufmann, Breckenridge (2000)

[6] Muller: Topological spatio-temporal reasoning and representation. Computational Intelligence 18(3), 420–450 (2002)

[7] Erwig, M., Schneider, M.: Spatio-Temporal predicates. IEEE Trans. on Knowledge and Data Engineering 14(4), 881–901 (2002)

[8] Varzi, A.C.: Parts, Wholes, and Part-Whole Relations: The Prospects of Mereotopology. Data and Knowledge Engineering 20, 259–286 (1994)

[9] Asher, N., Vieu, L.: Toward a Geometry of Common Sense: A Semantics and a Complete Axiomatization of Mereotopology. In: Proceedings of the 14th International Joint Conference on Artificial Intelligence, pp. 846–852. Morgan Kaufmann, San Mateo (1995)

[10] Pietruszczak, A.: Pieces of mereology. Logic and Logical Philosophy 14, 211–234 (2005)

[11] Forrest, Peter: Nonclassical mereology and its application to sets. Notre Dame. Journal of Formal Logic 43, 79–94 (2002)

[12] Smith, B.: Fiat Objects. In: Guarino, N., Pribbenow, S., Vieu, L. (eds.) Proceedings of the ECAI 1994 Workshop Parts and Wholes: Conceptual Part-Whole Relations and Formal Mereology, pp. 15–23. ECCAI, Amsterdam (1994)

[13] Smith, B.: Boundaries: An Essay in Mereotopology. In: Hahn, L.H. (ed.) The Philosophy of Roderick Chisholm (Library of Living Philosophers), pp. 534–561. Open Court, Chicago (1997)

[14] Varzi, A.C.: On the Boundary Between Mereology and Topology. In: Casati, R., Smith, B., White, G. (eds.) Philosophy and the Cognitive Sciences, pp. 423–428

[15] Gotts, N.M., Gooday, J.M., Cohn, A.G.: A Connection Based Approach to Common-Sense Topological Description and Reasoning. The Monist 78, 51–75 (1996)

[16] Casati, R., Varzi, A.C.: The Structure of Spatial Location. Philosophical Studies 82, 205–239 (1996)

[17] Sider: Locations. Philosophical Topics 30, 53–76 (2002)

[18] Varzi, A.C.: Spatial Reasoning and Ontology: Parts, Wholes, and Locations. In: Aiello, M., et al. (eds.) Handbook of Spatial Logics, pp. 945–1038. Springer (2007); Hawthorne, J. and T.

Distributed Intrusion Detection Based on Outlier Mining

Wei Da and Han Shao Ting

Beijing University of Chemical Technology
Beijing 100029, P.R. China
`weidabuct@126.com, hsthst@mail.buct.edu.cn`

Abstract. With the rapid development of Internet and network technologies, intrusion detection system (IDS) is expected to be more intelligent. Generally, IDS in current use can rarely meet actual requirements in performance, accuracy and distributed characteristics. In this paper, we present a distributed network intrusion detection system, in which an improved outlier mining method on clustering is introduced. Experimental results prove that both traditional attacks like SYN flooding, and distributed attacks such as DDoS, can be detected effectively with visible accuracy rate and reliability.

Keywords: Intrusion Detection System, outlier mining.

1 Introduction

Intrusion Detection System (IDS) has been developed to be one kind of efficient security protection mechanism against network attacks, with various intelligent methods such as artificial immunity, revolutionary computing and data mining having been introduced in recent years [1]. However, as a rapid developing field, there still exists some problems with IDS from theory to application.

1) Performance. Since the diversity and rapid evolution of network attack methods, current detection methods have a unacceptable false-alarm and negative rate. In a large-scale network environment such as campus, most IDS suffer packets loss heavily, which will decrease the detection rate significantly.

2) Detection on distributed attacks. Compared with traditional attacks, distributed attacks are N to 1 attacks with a shorter performing procedure and a higher success probability. The behaviour pattern of a single node have little difference with normal communications, which can be ignored by traditional IDS. Along with the increasing cases of distributed attacks, current IDS solutions show a weak detection capability.

To solve these problems, an intelligent, real time detection algorithm needs to be proposed, which can be capable to recognize particular attack patterns in time. Data mining technologies have a great advantage in extracting new features and rules from massive data, in which clustering analysis is particular suitable for discovering inner

relations among data points. Based on the above considerations, in this paper a distributed network intrusion detection system with an outlier mining method on clustering is introduced to perform real time detection on local network.

2   Related Work

Introduce intelligence into intrusion detection will enhance its learning ability and adaptability. Current research has drawn two main lines, statistical methods and knowledge discovery. The former methods includes Bayes classification[2], hidden Markov model[3] and statistical learning[4], while the latter one involves some kinds of data mining methods.

Statistical methods assume that normal activities own some certain statistical patterns, and use some statistical variables to characterize system behaviour. Systems such as IDES, NIDES and EMERLD[5] have developed these statistical analysis technologies.

The main advantage of statistical methods lies in their learning capability of normal behaviour of the system, hence with a high detection rate and availability. However, this way can lead to a shortage of representation on time sequence relation, while most abnormal states that describe intrusions are events-sequence-sensitive. Besides, the intrusion patterns and normal patterns do not always comply with certain distributions, nor linearly separable, which will cause difficulties in applying some new statistical learning methods, such as support vector machine[6], to intrusion determination.

Considering intrusion detection as a procedure of data analysis, data mining methods can be used to discover implicit patterns or models of intrusion behaviours among massive audit data. No manual analysis or intrusion pattern coding is needed. The DARPA evaluation report shows that IDS based on data mining are better in performance than those based on manual analysis[7]. Many kinds of data mining methods, including association rules analysis, classification and clustering, has been applied to detecting intrusions in present.

As a unsupervised knowledge discovering method, intrusion detection based on clustering requires little priori knowledge of the data, but neither detection rate nor negative ratio meet the practical requirements, especially for intrusions of new and unknown types[8]. The main disadvantage of current intrusion detection methods based on clustering exists in the measuring method of data similarity, which can not describe the similarity of mixed-attribute data. This shortage will lead to a low detection accuracy.

The difficulty of applying data mining to intrusion detection remains in: 1) extracting attributes that effectively reflect properties characteristics of current system, and 2) developing an on-line data mining algorithm. Besides, those chosen attributes should be refined based on network security domain knowledge as well as specific application context.

3   Clustering

The definition of similarity between data objects and the difference between clusters is the most fundamental problem, which will directly affect the clustering result. Data used for intrusion detection contains mixed and classfied attributes, in which leaves a large margin for the refinement and improvement of existed measuring methods.

By dividing the data space as well as the distance of two data points into two sub-spaces respectively, namely numerical and classification spaces, the Minkowski distance of a linear space can be promoted as follows:

Given the cluster C, C_1 and C_2 of dataset D, and object $p=[p_1, p_2, …, p_m]$, $q=[q_1, q_2, …, q_m]$, where D_i denotes the set of   i-value of the attribute.

1)   Let $a \in D_i$, the frequency of a on D_i in C is defined as the including times of a by the C's projection on D_i, which is

$$\mathrm{Freq}_{C|D_i}(a)=|\{object|object \in C, object\ D_i=a\}|.$$

2)   Define the distance (or difference) between cluster C_1 and C_2 as

$$d(C_1, C_2)=\frac{\sum_{i=1}^{m} \mathrm{dif}(C_i^{(1)}, C_i^{(2)})}{m},$$

where $\mathrm{dif}(C_i^{(1)}, C_i^{(2)})$ is the difference of C_1 and C_2 on attribute D_i. Its value on a classified attribute D_i would be

$$\mathrm{dif}(C_i^{(1)}, C_i^{(2)})=\frac{1}{2}\sum_{p_i \in (C_1 \cup C_2)|D_i} \left| \frac{\mathrm{Freq}_{C_1|D_i}(p_i)}{|C_1|} - \frac{\mathrm{Freq}_{C_2|D_i}(p_i)}{|C_2|} \right|,$$

as the classified attribute is described by frequency set of different values. When comes to numeric attribute D_i, the value would be

$$\mathrm{dif}(C_i^{(1)}, C_i^{(2)})=|c_i^{(1)}-c_i^{(2)}|.$$

Especially, when a cluster owns only one object, we get the following definitions:

3)   The distance between an object p and a cluster C is

$$d(p, C)=\frac{\sum_{i=1}^{m} \mathrm{dif}(p_i, C_i)}{m},$$

where $\mathrm{dif}(p_i, C_i)$ stands for the difference of p and C on attribute D_i. Its value on a classified attribute D_i is

$$\mathrm{dif}(p_i, C_i)=1-\frac{\mathrm{Freq}_{C|D_i}(p_i)}{|C|}.$$

Its value on a numeric attribute D_i is

$$\mathrm{dif}(p_i, C_i) = |p_i - c_i|.$$

4)   The distance between two objects p and q is defined as

$$d(p, q) = \frac{\sum_{i=1}^{m} \mathrm{dif}(p_i, q_i)}{m}.$$

Its value on a classified attribute is

$$\mathrm{dif}(p_i, q_i) = \begin{cases} 1, p_i \neq q_i \\ 0, p_i = q_i \end{cases}.$$

On a numeric or sequence attribute, its value would be

$$\mathrm{dif}(p_i, q_i) = |p_i - q_i|.$$

The k-means algorithm, based on partition, is a classic method to solve clustering problems, which would be robust enough for massive data environments. However, this method can only process data of those without classified attributes, which are necessary for intrusion detection. In according with the above clustering measurement definitions, a improved algorithm is proposed in this section to process data of mixed attributes.

This modified k-means algorithm includes three steps:

1)   Initialization. The k objects will be choosed, and k clusters are built.
2)   Deliver each object to its nearest cluster.
3)   Re-compute the centroid of each cluster.
4)   Repeat step 2 and 3, till neither error nor change is significant, or no object is switched between clusters during iteration.

4   Outlier Mining

Outliers are those data that diverge from the average main part of the dataset, which can be suspected as produced by non-random factors. In intrusion detection, this implies that the possibility of an intrusion behaviour should be seriously considered. Outlier mining can be described as follows:

Given N data objects and expected outlier data amount, identify the primary n objects that are inconsistent with others. Whether those outlier can be correlated with an actual attack should be decided by domain knowledge.

Let dataset D be devided by k clusters $C_1, C_2, \ldots, C_k$, the outlier factor $OF(C_i)$ can be defined as the average distance of each two clusters, which is

$$OF(C_i) = \frac{\sum_{j \neq i} d(C_i, C_j)}{k-1}.$$

The OF(C_i) measures the deviation of the cluster C_i to the whole dataset whose value implies the degree of deviation.

After that, an unsupervised intrusion detection method based on clustering can be proposed based on the above outlier factor-by sorting the clusters in accordance with the outlier factor and marking a certain partition of clusters as normal or abnormal.

The steps can be described as follows:

1) Clustering on dataset T, get result $T_1=\{C_1, C_2, …, C_k\}$.
2) Compute OF(C_i) of each cluster C_i, and sort C_i in descending order of OF(C_i). Find the minimum value of b that satisfies

$$\frac{\sum_{i=1}^{b}|C_i|}{|T_1|} \geq \varepsilon.$$

Then, clusters of C_1, C_2, …, C_b can be marked as attack clusters, while the others as normal ones.

The ε can be regarded as an approximation of outlier data's scale. The less ε is set, the lower the detection ratio will be, and the same as negative ratio. Statistical experiences show that outlier data usually scales less than 5%. Generally the value of ε can be set as around 0.05. In a specific application environment, a more precise value can be obtained via performance demand and prori knowledge.

5   Simulation

By sending massive raw packets to a local host with specific attributes, a network intrusion can be simulated in a limited scale. Besides, benchmark can be set by using classic dataset, such as DARPA 98 and KDD'99. A network intrusion detection system NIDS based on the above design is implemented and deployed. The simulation's results are summarized as follows.

Build the model on DARPA 98 dataset week1 and week2, and test this model on other four weeks' datasets. The results on a value r=0.6-0.7 is shown in Tab.1.

Table 1. Detection performance on DARPA 98 dataset of NIDS

	Week1 (r=0.6)	Week1 (r=0.7)	Week2 (r=0.6)	Week2 (r=0.7)
Total Detection Rate	91.67%	87.56%	88.63%	89.11%
False Alarm Rate	4.38%	1.06%	4.25%	1.08%

The results of the simulating different attacks to the host with this NIDS of the above model is shown in Tab.2.

Table 2. Detection rate of different attacks

Attack types	r=0.5	r=0.6	r=0.7	r=0.8
guess_passwd(remote)	100.00%	100.00%	100.00%	10.55%
land	66.67%	100.00%	100.00%	100.00%
buffer_overflow(remote)	52.07%	71.86%	32.03%	7.10%
rootkit(remote)	38.00%	9.00%	0.00%	0.00%
teardrop	8.89%	5.44%	32.96%	14.32%
smurf	98.89%	97.66%	97.00%	95.78%
distributed SYN flooding	79.00%	100.00%	100.00%	100.00%
spy(remote)	0.00%	0.00%	0.00%	0.00%
DoS	98.10%	99.00%	99.15%	99.12%
TOTAL DETECTION RATE	93.58%	96.65%	96.53%	88.62%
FALSE ALARM RATE	3.47%	6.06%	5.76%	2.57%

6 Summaries

Traditional intrusion detection techniques are mainly based on domain knowledge and manual rules which can not work effectively in face of unpredictable network attacks. Therefore, data mining techniques are introduced and great achievements have been made in recent years. However, current approaches also suffer some shortcomings such as low time efficiency, low detection ratio, as well as high negative ratio and weak detecting ability for attacks of some new type. Based on previous research, here we proposed an improved outlier mining method on clustering, with a new cluster description and measurement as well as a modified k-means algorithm for online clustering. A prototype system is implemented, in which experimental results confirm us a satisfied performance.

References

[1] Stefan, A.: Intrusion Detection System: A Survey and Taxonomy (February 2006), http://www.mnlab.cs.depaul.edu/seminar/spr2003/IDSSurvey.pdf
[2] Heckerman, D.: Bayesian Networks for Data Mining 1, 79–119 (1997)
[3] Vaseghi, S.V.: State Duration Modeling in Hidden Markov Models 41, 32–41 (1995)
[4] Vapnik, V.N.: Statistical Learning Theory (1998)
[5] Dorothy Denning, E.: An intrusion-detection model. IEEE Transactions on Software Engineering 13, 222–232 (1987)
[6] Berges, C.J.C.: A tutorial on Support Vector Machines for Pattern recognition 2, 1–47 (1998)
[7] Lippmann, R., Fried, D., Graf, I., Haines, J., Kendall, K., McClung, D.: Evaluating intrusion detection systems: the 1998 darpa off-line intrusion detection evaluation, 12–26 (2000)
[8] Eskin, E., Arnold, A., Prerau, M.J.: A Geometric framework for unsupervised anomaly detection: Detecting intrusion in unlabeled data (2002)

Research and Implementation of Full-Text Retrieval System Using Compass Based on Lucene

Conghui Zhang and Shubo Zhan

State Key Laboratory of Networking and Switching Technology of
Beijing University of Posts and Telecommunications
Beijing, China
conghuizhang@gmail.com, zhanshubo@cintel.com.cn

Abstract. Lucene is a high efficient, open source Java full-text retrieval libarary, which has been widely recognized for its utiliy in the implementation of Internet search engines and local, single-site searching.Compass is a Search Engine Framework built on top of Lucene, it simplifys the work with Lucene, without suffering any performance or loss of capabilities.Firstly, the paper analyzes the architecture of Lucene;secondly, it studies how the Compass works;finally, the article descibes a simple full-text retrieval system using Compass based on Lucene in detail.

Keywords: Lucene, Compass, full-text retrieval.

1   Introduction

Lucene, the free, mature and open-source project of Apache software fund, is an excellent search engine for full-text retrieval. Lucene is not a complete full-text retrieval engine, it is just a java tool package. In this situation, a great number of other full-featured search products use Lucene as the underlying search engine implementation. Compass, the open source Java Search Engine Framework, is one of the outstanding full-featured search applications. It leverages Lucene to expose a broad range of search features. Through Compass we can develope full-text retrieval system more efficiently and easily.

2   Architectur of Lucene System

Lucene enables to add indexing and searching capabilities to applications and hides the complicated implementation of indexing and searching behind a simple-to-use API. In this way, Lucene acted as an independent layer and applications sit on top of it, which is depictured in Figure1. By layer upon layer object-oriented progressing, Lucene achieved to be a high performance, loose coupling, easy to secondary exploitation search engine system.
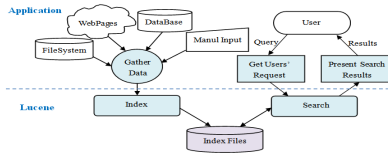
Fig. 1. Lucene Application Structure

Illustrated in Figure 2, the whole architecture of Lucene system is made up of External Interface, Index Core and Basic Structure Encapsulation the three major components. Of which the Index Core part is apparently the most important.
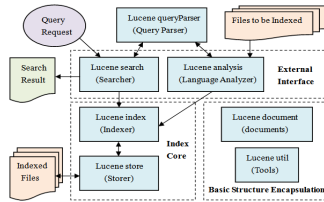
Fig. 2. Lucene System Architecture

3   Operation Mechanism of Lucene

(1) Lucene Indexing

To index data with Lucene, first of all, we must convert it to a stream of plain-text tokens, the format that Lucene can digest. Therefore, other patterns of files must be converted into text before being indexed.

Secondly, create a document object with its relevant field objects. Then call the method addDocument to store the data for indexing.

Next, use the specified analyzer to tokenize every word of indexes so that data stream can be more suitable to be indexed.

Last but not least, write the result back into index files.

The process of building full-text indexes is presented in Figure 3.
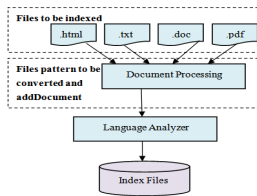
Fig. 3. Lucene Indexing Mechanism

(2) Lucene Searching

As shown in Figure 4, the query string requested by the user will be parsed by the QueryParser class in the beginning. Not only just parse the string into the corresponding query object for searching, but also tokenize every word of the indexes using the particular analyzer that is the same as the one during the indexing process.

Finally, search the built index files for query results, which will be ecapsulated in TopDocs class and returned back to users.
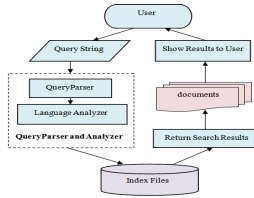
Fig. 4. Lucene Searching Mechanism

4   Compass Framework

After a brief introducing of Lucene, it is obviously revealed that Lucene is merely a low level usage and API. For designing a strong, mature and required full-text retrieval system, a secondary development on Lucene is in need.

(1)   Benefits of Compass Framework

Compass is a complete full-text retrieval search system built on top of Lucene designed to simplify the integration of search engine into any Java application.

1)   Easier for Developers: Compared with many Lucene APIs, Compass exposes a very simple API. It has a unique interface, with all operations available through it.
2)   Transaction Support: Lucene is not transactional. It causes troubles when attmptting to integrate Lucene with other transactional resources. Compass provides support for two phase commits transactions, such are read/committed and serializable.
3)   Rich Mapping Technologies: Compass supplies rich mapping techanologies to map a Java object model into the search engine and step further to provide seamless integration with ORM libraries, data grids, and others.
4)   Faster Updates: It is nontrivial for Lucene to perform an update, as it must delete the old Document first, then create a new Document. Due to transactional support and the fact that the saved data in Compass must be idendifiable, Compass execute a simpler and faster update.
5)   Index Fragmentation: While creating a Lucene enabled application, somtimes the index actually comprised of several indexes. Compass will automatically fragment the index into a couple of sub indexes or it can be mapped several aliases to the same sub index.

6)   All Fields Search: Compass creates an "ALL" field by default which serves as the default search field to search on all the fields stored in a Document when just working with Lucene is impossible.

(2)   Modules of Compass Framework

The overall Compass modules is composed of three parts, which are Compass Gps, Compass Spring and Compass Core, outlined in Figure 5.
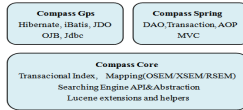
Fig.5 Compass Modules

Compass Core aims to be usable in different environments, and simplify the key implementations done against a search engine. Transaction management, kinds of mapping technologies, search engine API, search engine abstraction and other Lucene extensions are kept in Compass Core.

Compass Gps aims to integrate with kinds of content sources. The main feature is to integrate with different Object/Relation mapping(ORM) frameworks (iBatis, JDO, OJB, Hibernate)and allow for nearly transparent integration between a search engine and an ORM view of content that stores in a database.

Compass Spring aims to integrate seamlessly with the Spring Framework. Integrating with Spring in Compass is similar to the ORM frameworks integration with Spring. Moreover, Compass integrates with Spring transaction abstraction layer, AOP support, and MVC library.

(3)   Compass and Lucene
1)   Object Searching Engine Mapping(OSEM)
     OSEM is one of the prime features for using Compass. It provides support for mapping Java Objects to the Search Engine. Compass OSEM technology is the process of saving an object into the search engine and retrieving an object from the search engine.
2)   Relationship between Compass and Lucene
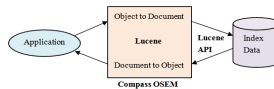
Fig. 6. Compass OSEM Based on Lucene

Figure 6 explicitly pictured the intimate relationship between Compass and Lucene. Compass allows to use most of Lucene classes without any changes. For example, if the application has a specialized Query, Analyzer, or Filter, we can use them directly within Compass.
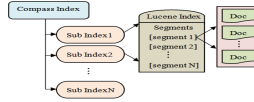
Fig. 7. Compass Index Structure

Seen from Figure 7, a Compass Index can be divided into several sub indexes. Every sub index has its own entire functional index structure, which mapps to a single Lucene index. The Lucene index part maintains a meta-data file about the index (called segments). The segments can be a single file or multiple files, which holds several Documents.

5   Full-Text Retrieval System

After a concise understanding of Lucene and Compass, we begin to design a typical full-text retrieval system using Compass based on Lucene.

(1)   Architecture of Full-Text Retrieval System
Referring to Figure 3 and Figure 4, the complete structure of a full-text retrieval system consists of two basic parts: searching and indexing, depicted in Figure 8.
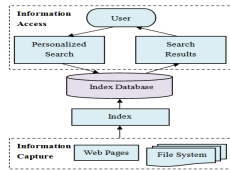
Fig. 8. Full-Text Retrieval System Architecture

In the searching process, users request the personalized search, and the index database will then return back the specified results. In the indexing process, information will be extracted to be indexed and saved in the index database according to what the developers have customized.

(2)   Design of A Full-Text Retrieval System
   1)   The Index Processing: Take the Class Book for example, we use Compass OSEM to implement the mappping, then create the index. The main code is as follows:

```
@Searchable
public class Book {
private String id;
    private String title;
    private String author;
@SearchableId
```

```java
    public String getId() {
        return id;
}
    public void setId(String id) {
        this.id = id;
    }
  @SearchableProperty(index = Index.TOKENIZED, store = Store.YES)
    public String getTitle() {
        return title;
    }
    public void setTitle(String title) {
        this.title = title;
    }
……
}
```

The annotation "@Searchable" above is to state that the class below will be indexed. "@SearchableId" is declared to uniquely identify the object within its mapping definition. "@SearchableProperty" is to state that the Class property needs to be mapped. The annotation "index" is to decide whether to be indexed or tokenized and the annotation "store" is to decide whether to store the propery in the database.

Later, configuring and building Compass for the mapped Class Book. The crucial code is as follows:

```java
CompassConfiguration cfg = new CompassConfiguration();
cfg.setConnection(path);
cfg.addClass(Book.class);//add the mapping information
compass = cfg.buildCompass();//build Compass
```

After that, starting to create, delete and update the index. The example code of creating the index is as follows, delete operation is very similar to it and update operation is to delete the index first then create a new index.

```java
public void createIndex(Book book){
    CompassSession session = null;
    CompassTransaction tx   = null;
    try{
        session = compass.openSession();
        tx = session.beginTransaction();
        session.create(book);
        tx.commit();
    }catch(RuntimeException e){
        tx.rollback();
        throw e;
    }finally{
        if(session != null){
            session.close();
        }
    }
}
```

2) The Search Processing: The key code for searching operation is as follows:

```
session = compass.openSession();
tx = session.beginTransaction();
CompassHits hits = session.find(queryString); // searching results
int n = hits.lenth();
if(n==0){
return Collections.emptyList();
}
for(int i = 0;i<n;i++){
   CompassHighlighter highlighter = hits.highlighter(i);
       //use Compasshighlighter to emphasize the hit content
   Book book = (Book)hits.data(i);
       book.setTitle(highlighter.fragment("title",book.getTitle()));
   books.add(book);
  }
  hits.close();
  tx.commit();
  return books;
```

3) Running the Full-Text System
We enter the string "Harry Potter" in the text, displayed in Figure 9.

Fig. 9. Input Query String

The results present the information of the book that contains the title and the author, exhibited in Figure 10.
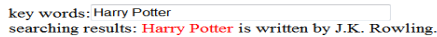
Fig. 10. Search Results

6   Summaries

Although Lucene is not a full-featured search application, it is a high reliable and extensible full-text information retrieval toolkit. Based on Lucene,Compass, the integrated search product is used to design the full-text retrieval system. As advantages mentioned above, Compass makes the process of developing convenient and most of Lucene classes and applications can be directly migrated to Compass. The full-text retrieval system we designed is very simple, for further develpleing, Compass provides various ways to adapt to all kinds of requirements. It needs conscientious future research.

References

[1] Hatcher, E., Gospodnetic, O.: Lucene in Action. Manning Publications Co., Greenwich (2005)
[2] Lucene2.4.1API (March 2009), `http://lucene.apache.org/java/2_4_1/api`
[3] Lucene China, `http://www.lucene.com.cn/about.html`
[4] Compass2.2.0 API (April 2009),
    `http://www.compass-project.org/docs/2.2.0/api/`
[5] Compass2.2.0 Reference (April 2009),
    `http://sourceforge.net/projects/compass/files/compass/2.2.0`
[6] InfoQ, `http://www.infoq.com/news/compass-11`
[7] SpringSideWiki,
    `http://wiki.springside.org.cn/display/springside/Compass`
[8] IBM developerWorks, `http://www.ibm.com/cn/zh/`

Research and Design on the Database Used for Aircraft Conceptual Design

Wei Chenglong and Liu Hu

Beijing University of Aeronautics & Astronautics
Beijing 100191, P.R. China

Abstract. Based on the need for data in aircraft conceptual design, this paper made the research on data storage model and data analysis model. Then the structure of database that included the database server, the interface and the connector was described. The database server was developed by the software of SQL Server, the application program by VB.NET and the connector by ADO.NET. The practicality of the database system was verified by a case study in the end.

Keywords: Aircraft, Conceptual design, Database, Expert system, SQL Server, VB.NET.

1 Introduction

As a complex project, aircraft design calls for the experience embeded in the previous aircraft since the series design is so common. For designers, accumulating data is vital[1]. The aircraft's configuration is frozen when the conceptual design ends, so that conceptual design plays a very important role in aircraft design and has great influence on aircraft's final performance. Conceptual design needs a large amount of data to determine and optimize many parameters. So knowing the prevsious design konwledge has great significance to conceptual design.

In the aviation industry, conventional data management is conducted by handbooks, which is quite work-consuming and incovenient to use and share [2] .Since 1960's when database technology was invented, it is used extensively in many fields now. It makes data management easy and improves the sharing of data. For instance, Americans developed air dynamic database named DATCOM which efficiently guided the configuration optimization of F-22 [3]. Therefore, to manage data efficiently and use it coveniently, database technology is a good choice.

2 Data Classify and Model Design

2.1 Data Classification

From defining the design requirments to determining the detail parameters, conceptual design involves a lot of data. So classifying these data is necessary, as the beginning of database design.

At the beginning of conceptual design, general data like dynamic coefficients, atmosphere paraters and airfoil paraters is needed, which is essential to design[4].

Then, according to the process of conceptual design, previous aircraft data that is available is classifed as follows.

Performance parameters that are used in the phase of deciding design requirments. For military aircrafts, mainly political, economical and cultural factors are considered. While civil aircrafts emphasize the market competitiveness on the condition of meeting the security requirments.

Basic design parameters: include takeoff gross weight, wing loading, thrust-weight ratio and so on. These parameters are vital and determine what the aircraft looks like.

Geometry parameters of different parts: include geometry parameters of fuselage, wing, tail wing etc. Finally the aircraft design work is given in project files with demensions and other requirments, so that this is one of the most important data.

Data of avionics. In morden aircraft, avionics play an important role to performance and arranging them is an important job in conceptual design.

2.2   Data Storage Model

"Hierarchy model" for data storage[5] is used as shown in Fig.1. Using "hierarchy model" instead of a large spreadsheet makes the management easy and convenient for building expert system which will be dicussed later.

According to hierarchy model, data from up to down is classified as   Information sources, Attribute Groups, Attribute Segments, Attributes, which are performance requirments, basic parameters and deminsion of parts correspondingly in the process of conceptual design, see details in Fig.1.
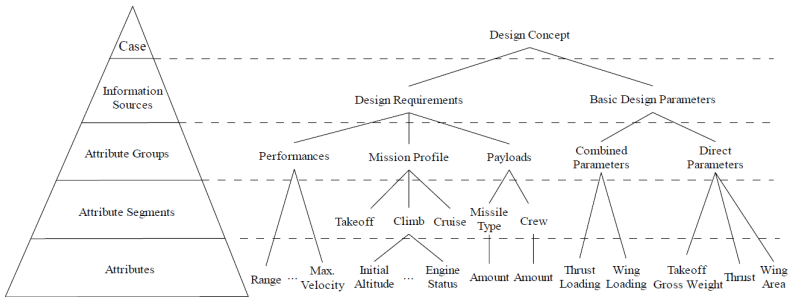
Fig. 1. Hierarchy model and examples of the information

2.3   Data Analysis Model

In the application programme, expert system is designed to aid the concept-selecting. Based on case-based reasoning(CBR) model, expert system is builded.

The process of CBR model includes four steps: retrieve, reuse, revise and retain, as shown in Fig.2 [6]. The data needed in CBR model is just mentioned in 2.1, i.e.,
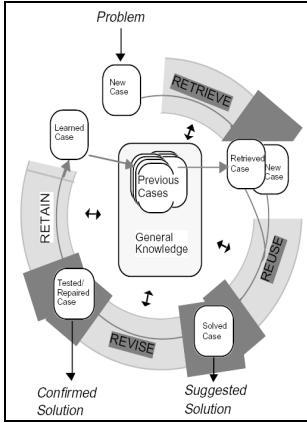
Fig. 2. The CBR cycle
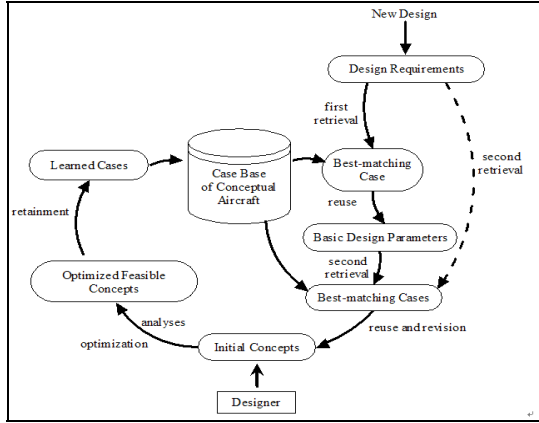
Fig. 3. Flow of CBR for aircraft conceptual design

general data and case information. The flow of CBR for aircraft conceptual design is discussed in Ref.7, and the flow is shown in Fig.3.

First, a new case is defined in parameters according to the design requirments. Then retrieve "best-matching" cases in the database.

In Ref.8, a method named nearest-neighbor (NN) method is talked about, which is used to retrieve "best-matching" cases. NN method is:

$$S(T,P) = \sum_{i=1}^{m} w_i \bullet s(a_i^T, a_i^P)$$

where $S(T,P)$ is an overall similarity of the target problem T with an old case P, a_i^T and a_i^P are the ith attributes of the target problem and the old case, respectively; $s(,)$ is pair-wise similarity measure along an attribute and it can be a function, rule, or heuristic; m is the number of cases; w_i is the weighting factor representing degree of importance of the ith attribute towards the determination of overall similarity, and usually $\sum_{i=1}^{m} w_i = 1$.

The "best-matching" cases are selected by the value of S. Ref.7 improved NN model to be hierarchy NN model. To match the hierarchy model, hierarchy NN model is chosen in this paper.

Second, design the basic parameters according to the "best-matching" cases which are selected in last step. Then, retrieve the "best-matching" cases again by basic design paremeters and verify the design requirments.

Finally, revise and retain the design according to the retrieving results.

Notes: considering aircraft is a comprehensive and multidisciplinary project, designers play an important role in expert system because men are more innovative than any system.

3   Database System Design

3.1   Database Structure

Database is a stage to realize data storage, maintenance and analysis. This need the corresponding developing-software and data access technology. The structure of database is shown in Fig.4.

Fig. 4. The Struncture of database syetem

The structure of the database selects Client/Server(C/S) model, which is used extensively in industry databases. The advantages of C/S model are quick-responding and strong performance.

Software of SQL Server to save data is used, which serve as server in C/S model. With this software, not only normal data but also image-data can be saved quickly. According to hierarchy model mentioned above, save the data by programming language. Considering that the data of airfoil is complex, Ref.9 gives a saving method by decomposing the data of airfoil.

Application programme works in the client-side, which is developed by VB.NET in this paper. VB.NET is a object-oriented programming language. The application programme can be developed conveniently by the visualized tools provided by VB.NET.

The technology of ADO.NET is chosen to realize data access, which is the newset data transfer model provided by Microsoft corporation. ADO.NET is the connector between data and application programme, by which users can retrieve, analysis the data[10].

3.2   Design Requirments

Database designing has to meet the design requirements:

(1) Normalization requirement: the data model has to meet third normal form.
(2) Integrity constrains: include entity integrity, referential integrity and integrity difined by users.
(3) Security demand: this is important to keep the system working.
(4) Data recover mechanism.

4   Database-Function Design

To simulate the reality as much as possible, object-oriented technology is utilized. Functionally, database system can be decomposed into several modules: definition of
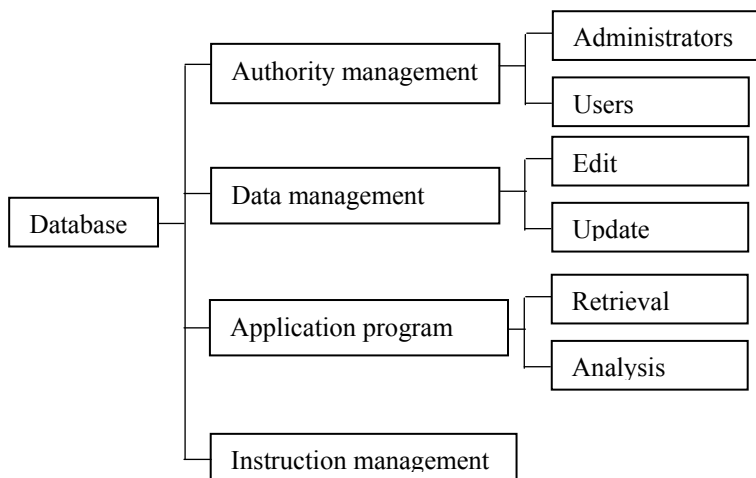
Fig. 5. Functions of database system

authority, data management, application program management and system instruction management, as shown in Fig.5.

1) Data editing function. Administrators can browse, modify and update data. And data-updating has two ways to conduct according to the mount of the data, i.e., updating by hands and by the special tools of SQL Server.
2) Data analysis function. Users can retrieve data by setting the constraints and analysis it. The result can show in several forms, such as curve, table etc.
3) System maintaining function. This includes account protection and data recovery mechanism.
4) Subsidiary function and management of instruction. Subsidiary functions include printing, data-outputting etc. Help files is to help users learn system by themselves.

5   Application Introduction

The database for aircraft conceptual design can assist the optimization of aircraft concept efficiently. In addition, database can track the developing tendency through updating data of the latest aircrafts, which is explained by an example.

Business aircraft is a new field compared with civil transportation. With the development of the global economy, the mount of business aircrafts increases quickly. However, conventional data can't meet the need of designing business aircraft. For example, there's error between traditional data and the reality while deciding the value of payload weight for single passenger.

To civil aircraft, we can get the value of payload weight for one passenger from traditional handbook is 90kg/man. Collect the data of business aircrafts appeared after

1990s', as shown in Fig.6. Horizontal axis stands for the value of passengers, the vertical axis payload weight. It could be concluded that the value of payload for single passenger exceeds 100kg/man in most of business aircrafts.

Example like this is easy to obtain. Nowadays, technology is developing quickly and many parameters of new aircrafts diverge very much from the traditional ones. In this situation, it is necessary to construct a database for aircraft conceptual design.
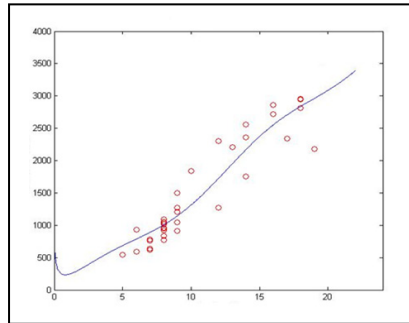
Fig. 6. Payload and passenger

6   Summaries

Aircraft design is a comprehensive project, and the development of a new aircraft concept needs both creativities and previous design knowledge. This paper provides an approach of developing database that is used in aircraft conceptual design. Naturally, there are still many detailed problems in the process of developing the database, which need to be investigated in future study.

References

[1]   Gu, S.: Aircraft conceptual/preliminary design. Beijing University of Aeronautics and Astronautics Press (2001)
[2]   Wan, M., Li, X., Liu, Y., et al.: Development on database of aircraft sheet metal forming. Journal of Beijing University of Aeronautics and Astronautics 30(2), 118–121 (2004)
[3]   Gillard, L.J.: AFRL F-22 dynamic e wind tunnel test result. AIAA99-4015 (1999)
[4]   Wang, G., Wu, Z.: Data Management in Conceptual Aircraft Design Process. Journal of Beijing University of Aeronautics and Astronautics 28(4), 455–458 (2002)
[5]   Tian, Y., Liu, H., Luo, M., et al.: Research on Case Base of Parameterized Large Civil Aircraft. Acta Aeronautica et Astronautica Sinica 31(11), 2202–2208 (2010)
[6]   Aamodt, A., Plaza, E.: Case-Based Reasoning: Foundational Issues, Methodological Variations, and System Approaches. Artificial Intelligence Communications 7, 39–59 (1994)
[7]   Liu, H., Lu, X., Luo, M.: Case-Based Reasoning for Developing Initial Concepts. In: 44th AIAA Aerospace Sciences Meeting, vol. 23, pp. 17825–17834 (2006a)

[8] Cheng, C.B.: A Fuzzy Inference System for Similarity Assessment in Case-Based Reasoning Systems: An Application to Product Design. Mathematical and Computer Modelling 38(3-4), 385–394 (2003)

[9] Bai, J., Wang, H., Liu, H., et al.: The Development and Application of Database Used for Conceptual Design. Aircraft Design 29(5), 1–4 (2009)

[10] Gong, P., Yuan, K., Yang, Z.: Database Technology and Application (principles + SQL Server+VB.NET). Higher Education Press (2008)

The Influence of Spatial Relationship between Nodes and Information on Information Spreading in Social Network

Shengbing Zhang, Wandong Cai, Yongjun Li, Bo Wu, and Zhilin Luo

School of Computer Science
Northwest Polytechnic University
Xi'an, Shaanxi 710129, China
zhangshengbing@nwpu.edu.cn

Abstract. Social network has been developed as a new platform of communication for information. The research of process of information spreading on social network is beneficial to understanding the information spreading features of social network. Based on the analysis of Micro-Blog data, it has been discovered that the nodes which have different spatial features have different effects on the information spreading process, and the information has a high probability to spread from one node to another when the nodes are in the same geospace or nearby.

Keywords: Social Network, Information Spreading, Spatial Influence.

1 Introduction

In recent years, social network has been highly developed at home and abroad, and became one of the most influential network applications. Social network plays an important role in the research of information spreading and interpersonal relationship, and become increasingly valuable. With the developing of social network, there arise some new network applications such as blog and Micro-Blog, which further strengthen the status of online social network on the research of information spreading.

Online social network is based on the feature of people clustering, which is different form information clustering which is the feature of some other online application such as Bulletin Board System (BBS). In the online social network, the people with some same or similar features build and keep their circle of friends.

Since the information is permitted and meaningful in the spread way of "word of mouth" in social network and has higher reliability, the receivers may consider the information carefully and share the information with their friends, which will accelerate the information spreading. For example, in the succor of Haiti earthquake, local people use social network to spread the earthquake information, which provide the up to the minute information for the succor. However, as the platform of communication for information, social network also help some bad information being spread such as rumor and panic. Such information has bad influence on the society

and impedes the management of the government. Therefore, the research of process of information spreading on social network is beneficial to understand the spreading features of social network information. It provides the foundation for the research of monitor and control of public feelings and information spreading of online social network.

This paper researches the social network based on the data of Micro-Blog, which is a new type of social network. Micro-Blog is a platform of communication based on interpersonal relationship. People could build their own community by the client API and public and share about 140 words information with others. In China, Sina Micro-Blog is the most famous Micro-Blog with biggest number of clients. The data in this paper all come from Sina Micro-Blog. We research the influence of the space relationship between information and nodes on the coverage of information by the analysis of the data.

2   Relate Works

In recent years, the researches of online social network mainly focus on the aspects as follows: (1) The measurement of social network topological features[1]; (2) User privacy protection[2]; (3) The measurement and analysis of user behavior[3]; (4) Personal search and recommendation[4]; (5) Information spreading and so on.

The existed research has proved that online social network has the features of small world and scale-free, which present that the average distance between nodes is small and there are some nodes with many neighbors. All these features play an important role in the social network information spreading and have brought the attention of scholars.

In online social network, there exist the influences between each other. The information published by one person with a big influence on others may have a high probability to be received by others. The reference [5] refers to an unsupervised model, which define the relationship strength by behaviors between nodes and the features of nodes. This model supposes that the relationships between nodes are mutual independent, which does not conform to the reality. Reference [6] predicts the influence between nodes by the linear regression model according to the behaviors between nodes and social network topological information, and the results show that the behaviors between nodes play an important role while the topological information is in a secondary place. However, this model only used data in Facebook to do the experiments and the applicability on other social network need to be proved in the future. Reference [7] researches the influence of accordance of space and time of nodes on the influence of nodes, and discovered that with the increasing times of appearance at the same time, the influence is increasing sharply, and with the increasing time interval of appearance at the same time, the influence is decreasing. This result is similar with reality, and since this reference has considered the influence of spatial location and time on the influence of nodes, it is beneficial for our research.

The existed works usually suppose the probabilities that the information transfer form one node to another are same. However, it is far different from the real-world situation and we must understand how to calculate the information spreading probability. In this paper, we analyze the Micro-Blog data, research the influence of spatial relationship of information and nodes on the information spreading, and discuss the influence of spatial features of nodes on information spreading probability between nodes.

3   Experiments and Analysis

Social network information spreading process can be summarized as follows: Firstly users publish information and notify their neighbors; then some neighbors comment, forward and share that information based on personal interests; finally their neighbors see the message, repeat the process and thus information in social networks has been spread. We believe that information spreading process of Micro-Blog is totally based on the above process. In the Micro-Blog environment, users are the Micro-Blog content publishers, and neighbors are the followers of the users. Information is spread by the comments, forwarding, sharing of followers in social networks. Therefore when we research the information spreading process of social networks, the impact of user comments, forwarding and sharing should be considered.

Information spreading between the nodes is influenced by many factors, such as the number of mutual friends, age, whether there are common interests and hobbies, and so on, and all these factors have different level influences. This paper focuses on the influence of spatial location of the users and their followers. What effect could be made if two spatial locations are consistent in the information spreading? This problem is to be discussed as follows.

Micro-Blog recorded the user information of comments, forwarding and sharing of the followers, which makes our experiments possible. In addition, the data of followers' features and the topology information of Micro-Blog could be crawled, which provide the database for the analysis of research of information spreading. From February 28, 2011 to March 6, 2011, a total of 16,982 piece of data have been crawled online, and in order to avoid the influence that in weekend, there are more online users than in working days, we have collected the dataset with a time span of one week. The dataset includes the users form the common Micro-Blog users to star users which have more than two thousand followers, which is basically covered the common range of users in Micro-Blog. Considering the unique influence of star users on the information spreading which requires specialized study, the star users which have tens of thousands of followers are not discussed in this paper.

The collected dataset also includes the user's registration information, followers' registration information, and followers' behaviors to the users' published information. When Micro-Blog users post a message, their followers will receive the information pushed by Micro-Blog system. Then followers can forward and comment this information.
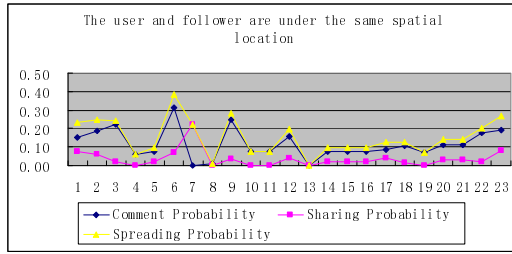
Fig. 1. Comment and sharing probability when the user and follower are under the same spatial location

If carried out these operations is considered the "followers" were An information dissemination behavior. Compare the Micro-Blog user registration information of the location code and followers' location code information to determine whether they are in the same geographical space. Fig.1 shows the information spreading probability when their spatial locations are same.

In fig.1, X-axis is arranged by the number of followers in a less to more order. Y axis is comment and sharing probability when the user and follower are under the same spatial location. Spreading probability is the sum of the forwarding probability and comment probability. It could be seen from the figure that when the number of followers is small, fluctuation in the graphics is relatively large, and with the number of followers is increasing, the graphics tends to be gently. This is because when the sample size is small, the influence of a random operation of a communication on the spreading probability will be significantly affected, and when sample size is increasing, the influence of random operations on the probability is lowed, which could more accurately reflect the true situation.
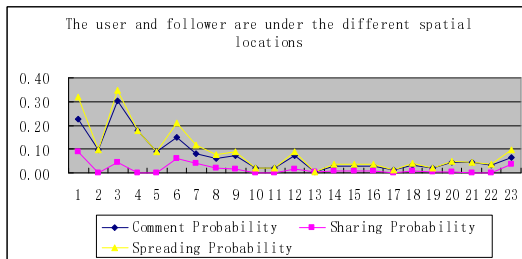
Fig. 2. Comment and sharing probability when the user and follower are under the different spatial locations

In fig.2, X-axis is also is arranged by the number of followers in a less to more order. Y axis is comment and sharing probability when the user and follower are under the different spatial locations.

Fig.3 (a) and Fig.3 (b) shows the comparison of sharing and comment probability under the same spatial location and different spatial locations. It is similar with fig.1 and fig.2 that the influence of spatial features on information spreading is inconspicuous, which is because when the sample size is small, the influence of a random operation of a communication on the spreading probability will be significantly affected, and when sample size is increasing, the influence of random operations on the probability is lowed, which could more accurately reflect the true situation. It should be noted that, in the Micro-Blog information spreading process, a significant comment is complex, and forwarding operation is relatively simpler, which could explain why, in Fig.3 (b) in the two lines are more closer: Forwarding is simple no matter whether the information published in the same place or not, and whether the followers are very interested in the information. Therefore, the probabilities of the two lines are close. Fig.3 (c) presents the information spreading probability under same spatial and different spatial locations. Since the information spreading probability is the sum of forwarding and comment probability, the influence of followers who is not so interested in the information and just forward it at random could be decreased.

(a)   (b)

(c)
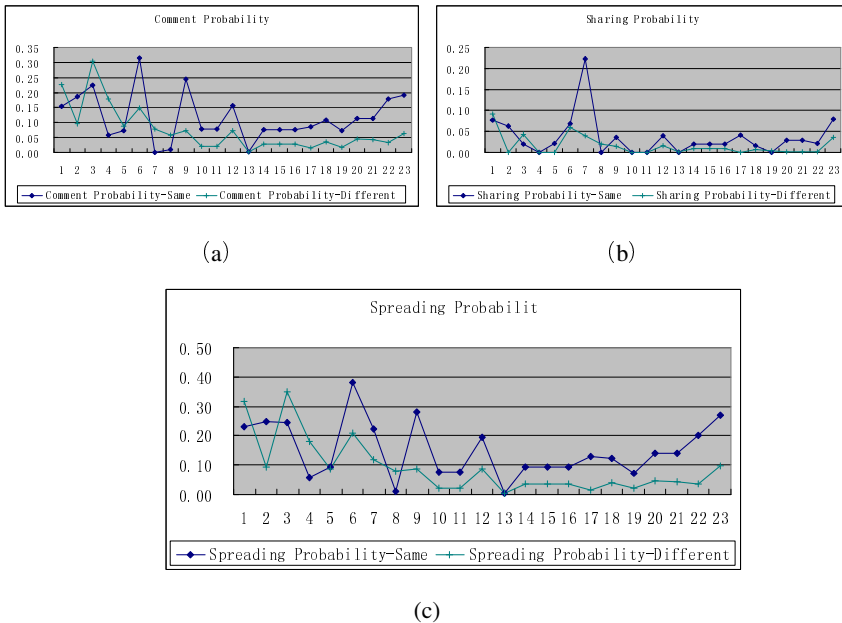
Fig. 3. Sharing and comment probability under the same spatial location and different spatial locations

Fig.4 is a summary of all data. From this figure we can easily see that, when information and communicators is in the same spatial location, the probability has been far greater than the probability in different spatial locations, even more than doubled, no matter whether it is a forwarding or comment. This result in a social
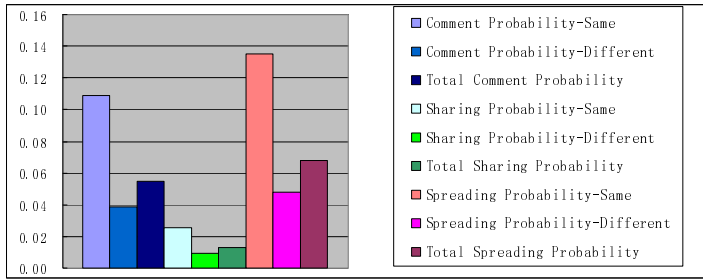
Fig. 4. A summary of all data

network is quite consistent with real-world situation: people are always more concern about the things around them.

4   Summaries

From the discuss above, we could conclude that when the information and nodes are in the same spatial location, the information spreading probability is big, and different spatial features have different effects in the information spreading process. In the future, we will build our evaluation methodology of key node, information spreading control mechanism and so on according to the result.

References

[1] Nazir, A., Raza, S., Chuah, C.-N.: Unveiling Facebook: A Measurement Study of Social Network Based Applications. In: Proceedings of Internet Measurement Conference, IMC 2008, pp. 43–56 (2008)

[2] Krishnamurthy, B., Wills, C.E.: On the Leakage of Personally Identifiable Information Via Online Social Networks. In: Proceedings of the 2nd ACM Workshop on Online Social Networks, SIGCOMM 2009 Workshop, pp. 7–12 (2009)

[3] Schneider, F., Feldmann, A., Krishnamurthy, B., Willinger, W.: Understanding Online Social Network Usage from a Network Perspective. In: Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement Conference, IMC 2009, pp. 35–48 (2009)

[4] Sarkas, N., Das, G., Koudas, N.: Improved Search for Socially Annotated Data. In: Proceedings of VLDB 2009, pp. 778–789 (2009)

[5] Xiang, R., Neville, J., Rogati, M.: Modeling Relationship Strength in Online Social Networks. In: Proceedings of the 19th International Conference on World Wide Web, WWW 2010 (2010)

[6] Gilbert, E., Karahalios, K.: Predicting Tie Strength with Social Media. In: Proceedings of CHI 2009 (2009)

[7] Crandalla, D.J., Backstromb, L., Cosleyc, D., Surib, S., Huttenlocherb, D., Kleinberg, J.: Inferring Social Ties from Geographic Coincidences. PNAS (Early Edition) (2010), http://www.pnas.org/cgi/doi/10.1073/pnas.1006155107

Safe and Easy Access to a Computer File-Sharing Program Laboratory Practice and Research

Lishuang Wei

Department of Mathematics and Computer Science Liuzhou Teachers College
Liuzhou, Guangxi 545004, China
wls_6812@163.com

Abstract. In order to meet the teaching needs of network security, public laboratories for computer file sharing methods were studied. Student lab machine's operating system is Windows XP, server operating system is Windows 2003. Associated set by a batch to achieve the student machine mapped network drive, and reboot the system after the student computers automatically connect with the server without having to re-enter your password and automatically map a network drive. Modify the group policy prohibit students from Network Neighborhood shared between machines to improve security. Students practice files stored on a network drive, the teacher simply copy and paste on the server operating folder, you can easily and quickly recovered the students practice files. After years of practice has proved: that the method is very practical and effective.

Keywords: Network security, file sharing, mapped network drive, automatically connected.

1 Introduction

School computer labs are common, use of different time periods of classes and teachers may not be the same, take the experimental task is different. Inevitably, the process of teaching students in jobs related to access and file sharing. How do teachers within the shortest possible time to collect student work and clean up is completed, the laboratory to consider the problem solved. In addition, teaching students inevitably want to test, to prevent students through the student computers to share resources between the Network Neighborhood or instant messaging software to cheat, but also laboratory to solve problems. Similar problems of information technology in primary and secondary education also exist [1], and in enterprises and institutions also have a file transfer and sharing each other's needs [2].

To solve the above problems, and some third-party software, some by programming [1] [2]. The process of programming complexity, the need to achieve those who have some programming basics. Teachers with a multimedia teaching software to send files and to submit student work, students receive job after sending teachers for students to receive, because time is short recess, the shortest is only 5 minutes, in just a few minutes is difficult to finish!

Rapid sharing with batch processing, in which there is no user account password, in the end the students were using a batch mapping a network drive, the need to machine the corresponding batch of students drag the "Start"menu, start the computer to automatically map [3]. However, there is a drawback to this method: batch stored in the student computers hard drive, start menu will put the batch, the students easily found from the Start menu, the location of the batch file, and open it to view the content of the batch, A security risk. If the batch hidden, restart the computer if the batch does not work, can not automatically connect with the server. By conventional methods in the windows set up a shared form [4] [5] [6], restart the computer and asked for a password. Through practice and research to improve the program, making sharing more secure; in the student-side methods of mapping a network drive some adjustments were made to restart the system can automatically map a network drive without prompting for a password.

2   Solutions

The same software environment, the server operating system is Windows Server2003, file format for NTFS or FAT32, students operating system is Windows XP Professional SP3.

(1) different laboratories using different network segment, to ensure that the teaching of non-interfering different laboratories.
(2) server, student machines are equipped with dual system, two systems installed in different partitions. Usually use a system of teaching, a system test students use.
(3) the two systems basically the same software installed, the examination system is not installed with the instant messaging software such as QQ and so on.
(4) teaching system usually shared with a batch, the user account on the server without a password. Test students using the system user account on the server need to reset your password, to improve security.

Described below, test students only used the system implementation process, focusing on some changes, other parts of the See reference [3].

3   Implementation Process

3.1   The Operation of the Server

Mentioned below, the server name: server1, address: 192.168.4.1. 56 users in the construction of the server, account name were user1, user2 ...user56, password unity is 123456. Then create two folders, one called "questions" and share a folder called the original name, Everyone group has Read permission; another non-shared folder, named "Test Folder. " In the "test folder " to create 56 folders, folder names are: 1,2,3 ... 56. Set each folder to share, share name of the original folder name, a folder corresponding to only assign unique user, the user has full control permissions to it. 1

corresponds to user1, 2 corresponds to user2 ... 56 correspond to user56. Student's host name followed by C1, C2 ... C56, mapped network drive, the 1 on the application of C1, 2 C2 ... 56 corresponds to the corresponding C56.

(1) batch add users with password

The same number of student computers to establish the number of users, use the command net user to complete:

net user username password /expires:never /passwordchg:no /add

username is a custom user name, password is a custom password. Parameter expires: never is set up the account never expires. Parameters passwordchg: no specified user can not change their passwords. Parameters add the new user is added to the user account database. Such as:

net user user1 123456 /expires:never /passwordchg:no /add

This statement creates a file named user1, password is 123456 users, this account never expires and the user can not change the password. Consistent with the statement that the number of users, just change the username on it. Note the use of space, it relates to creating a user's success or failure.

(2) Set "Password never expires"

(3) create a shared folder and set its properties

(4) Group Policy settings

Group Policy to ensure that the "Network access: Sharing and security of local account model" to the default security settings for "Classic - local users authenticate as themselves."

3.2   Client (Student Machine) Operating

The super user to log on locally as Administrator students machines, machines for students to map a network drive. Set of two maps in the student machine actuators, one for the student store files, one for students to get teachers to share files. User account on the server set a password, we tried two methods for students to drive mapped network drive, were given the following description:

3.2.1   Batch Achieved by Mapping a Network Drive

(1) mapping a network drive for students to store files, write the command net use map network drive command line:

net use [devicename|*] [Server IP] [password] [username]

Example: Students host machine named C1 batch file [1.bat] in the command:

net use I: \\192.168.4.1\1 "123456" /user:"server1\user1

net use H: \\192.168.4.1\ Questions "123456" /user:"server1\user1"

I: and H: is displayed on the student machine name of the 56 computer-driven unity of the folder with read-write map I:, read-only permissions to the folder with maps H:, other students or less the same batch on machine , to the corresponding folder name and user name.

(2) Double-click the batch file to run writing good, you can create a map, rename the network name of the drive was I: and H:.

(3) the batch file on the hard disk and drag the "Start "button in the startup items.

With this method mapped network drive, restart student computers will automatically connect, do not re-enter the password. Batch file must exist on the hard disk, the students can be found by starting a batch of content items, there are security risks.

3.2.2   Mapped Network Drive through Windows

(1) mapping a network drive

① Name of the title bar to open the window for the server name

Open the [Explorer], in the address bar [\ \ server name]. Such as input: \ \ server1, you can open the title bar name [server1] window, we call the server name window. The right of the child window of the window shows the server all the shared folders.

② Set the connection status and password

Right-click the server name in the window corresponding to the shared folder, click on the shortcut menu [mapped network drive], pop [Map Network Drive] dialog. Select the appropriate drive letter I:, the folder is named default. In [the Reconnect at Logon] A selection box ticking, Click the [other user name], pop [connection status ...] dialog box, enter the corresponding user name, password, first, do not fill, it is very important. Then click [OK] to return to [a mapped network drive] dialog box, then click [Finish] button, pop-up [to connect to server1] dialog box, then the [Remember my password], to check the , Fill in the appropriate password, then click [OK] button.

(2) Disconnect the network drive

Will be on a mapped network drive further off. In the Server Name window open [Tools] | [disconnect a network drive], selected on the steps to create a network drive, disconnect.

(3) and then map a network drive

In the Server Name window, right-click the corresponding folder from the shortcut menu, click [Map Network Drive], simply select the appropriate drive name, click [OK] button to complete the mapping. The need for two maps, a name for I:, corresponding to the shared folder can read and write, another name for H:, the corresponding read-only share files.

(4) modify the network drive name

In Explorer, rename the network drives were named I: and H:.

By the above method mapped network drive, restart the machine after the students no longer prompted for a password, automatic connection, students can not see the password machine. Compare the two implementations, we used the method described in this section mapped network drive.

3.2.3   The Collection of Student Files and the Contents of Shared Folders Clear

(1) the collection of student files

Simply the collection of student files "test folder" copy and paste into the designated areas can be.

(2) removal of [test folder] in the student data

① Establish [bulk delete the folder. bat] file

rd [/s][/q][drive:]path

/ s used to delete the directory tree. In addition to the directory itself, it will also delete the specified directory of all sub-folders and files. / q for quiet mode, with / s to delete the directory tree does not require confirmation.

② Remove the shared folder

Students disconnected from the server machine and run "batch delete the folder. bat " file, you can delete the student folder.

③ to re-create the shared folder

Run a batch file to re-create the corresponding folder, you can delete the folder will be re-created, and share the attribute will be automatically restored.

3.2.4   Prohibition of Neighborhood Students to Access through the Internet between Computers

Running on each student in "gpedit.msc", open the [Group Policy] window. Expand [Windows Settings] | [Security Settings] | [local policy] | [User Rights Assignment], right [to access this computer from the network], pop [Access this computer from the network properties] | [Local Security Settings] | Select [Everyone] | [delete] | [OK] | [is], the security settings in the default user group [everyone] to delete, to prohibit students from machine through Network Neighborhood between the purpose of the visit. Group policy and registry and then locked to prevent unauthorized modification of students.

4   Summaries

After years of practice, the program described in this paper file sharing, safe and easy. In the process of teaching students to store files in a specified network drive, the students need teachers to receive the file copy on the server side folder on the line. Public schools and university computer lab, teachers and students mobility. This method does not need extra cost, simple, economical, practical, enterprises and institutions also have some practical value.

References

[1] Chen, Y.: My room teaching document management solutions. Information Technology Education (11), 112–115 (2005)
[2] Chen, L.: Peer to peer file sharing system, local area network design and implementation. Computer Knowledge and Technology (4), 88–90 (2011)
[3] Wei, L. (cream), Li, X., Wei, T., Xie, Y.: Quick and easy with room to achieve a batch file sharing access scheme. Computer Knowledge and Technology (3), 120–122 (2011)
[4] Dragon Studios, New Introduction to Chinese version of Windows XP and enhance, pp. 283–284. Posts and Telecom Press, Beijing (August 2007)
[5] Kyushu book source, Chen, Y., Liang, W.: A school will magic Break Windows XP, pp. 172–173. Tsinghua University Press, Beijing (2007)
[6] Zhong, Z.: Windows XP 666 cases of classic techniques, p. 187. Publishing House of Electronics Industry, Beijing (2007)

Moving Objects Mining in Videos
with Distinctive Local Feature Configurations

Ma Chao and Shen Wei

College of Engineering and Technology, Northeast Forestry University

Abstract. A novel approach to mine distinctive spatial configurations of local features which occur frequently on moving objects from a given video clip is presented. In order to cope with high dimensional spaces of local feature descriptors and clutter of videos, a dimensionality reduction method and a motion segmentation aided mining method are proposed. Furthermore, a new transaction construction mechanism is introduced in mining procedure. Comparative experiments shown on data from TRECVID 2010 demonstrate that method presented in this work can exceed the performance of current state of the art techniques.

Keywords: Video mining, Local features, Dimensionality reduction, Moving objects.

1 Introduction

Mining frequently occurring objects and scenes in videos have been demonstrated to be significant to many applications such as video retrieval and browsing, semantic video annotation, video surveillance etc. Compared with global features, local features are more robust for video mining considering different scales, clutter and part missing.

However, there still exist two aspects of problems at least in this kind of method. Firstly, descriptors of local features are generally parameterized in very high dimensional spaces. This makes it difficult to measure distance of feature vectors, and limits the performance of video data mining in terms of speed and scalability. Therefore some researchers proposed enhanced algorithm in order to guarantee the accuracy of feature matching while reducing the time cost.[1] proposed PCA-SIFT descriptor to reduce the dimensionality of SIFT descriptor by conducting PCA on normalized grads block. Whereas [2] pointed out that descriptor obtained from PCA-SIFT shrinks the dividing ability of SIFT descriptor especially in rapid image matching procedure. Furthermore, although Gradient Location and Orientation Histogram (GLOH), which conducts PCA after obtains SIFT descriptor in polar coordinates, is shown to outperform SIFT and the other descriptors, but the complexity of computing is increased[3].

Secondly, local features are extracted in advance without prior knowledge of objects, and that results in large amount of features most of which are irrelevant to interesting objects especially when the objects are tiny. So the signal-to-noise ratio of the total set of features is very low, which severely reduces the efficiency of some

higher-level processes. [4] add transactions from the negative training set to the database, in order to discard itemsets that appear frequently on both objects and background. But this method is carefully hand crafted by manually annotation of all these background data.

In this paper, we attempt to tackle these problems by treating descriptor design as non-parameter dimensionality reduction problem, and these descriptors are only extracted from moving objects. Comparative experiments demonstrate that our approach has better performance than current techniques such as SIFT, SSD and PCA-SIFT considering the speed and the scalability.

2   Depiction of Moving Objects with Condensed Local Feature Configurations

The overall depiction of moving objects consists of four processing stages (Fig.1). Firstly, preprocessing is conducted, such as shot boundary detection, moving objects segmentation. Next the condensed descriptors of moving objects is extracted. Finally, we construct the depiction of moving objects with configurations of these descriptors. Now describe each stage in detail.
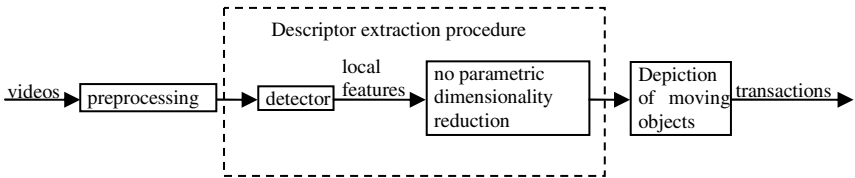
Fig. 1. Processing stages in depiction of moving objects

2.1   Prior Processing and Local Features

Foundation of the following processing stages is the prior processing of the videos. Videos are partitioned into shots and four "keyframes" are picked per second within each shot. This ensures a dense and uniform sampling. Then use a Difference of Gaussian (DoG) detector to extract regions in each keyframe. Combined with moving information of macro block, only the regions belong to moving objects are described with a SIFT-descriptor. Next, some research provide quantized codebook method that is constructed by clustering the SIFT descriptors. But this clustering process sacrifices the dividing ability of these descriptors which is the inherent uncertainty of clustering process.

In order to solve this problem, these features are soft-matched by assigning them to all codebooks when the distance between the feature and the center c of codebook is below a threshold $d_{\min}$. Then each descriptor can be described by a set of codebook labels as

$$x_i = \{c_j \mid d(R_i, c_j) < d_{\min}, j \in 1...N\} \tag{1}$$

where N is the total number of appearance clusters.

2.2   No Parametric Dimensionality Reduction

In order to reduce the dimensionality of descriptors, we tend to find discriminative projections of these descriptors in the original space. a criterion function constituted as

$$J(w) = \frac{\sum_{"i"non-match"j"}(w^t x_i - w^t x_j)^2}{\sum_{"i"match"j"}(w^t x_i - w^t x_j)^2} \tag{2}$$

In which the numerator represents square error of descriptors that doesn't match, and the denominator represents square error of descriptors that does match. So the whole criterion function represents the ratio of variance between the non-match and match descriptors along direction w. The direction w^* that maximize the ratio of $J(w)$ is what we seeking. In terms of covariance matrices form, equation 1 can be rewritten as

$$J(w) = \frac{w^t S_{non-match} w}{w^t S_{match} w} \tag{3}$$

Where

$$S_{non-match} = \sum_{"i"non-match"j"}(x_i - x_j)(x_i - x_j)^t \tag{4}$$

$$S_{match} = \sum_{"i"match"j"}(x_i - x_j)(x_i - x_j)^t \tag{5}$$

It is easy to show that a direction w that maximizes $J(w)$ must satisfy

$$S_{non-match} w = \lambda S_{match} w \tag{6}$$

Furthermore, the k eigenvectors associated with the largest k eigenvalues λ are extracted which is equivalent to [5] but without the local weighting functions. Then the linear projection matrix U can be constructed with these k eigenvectors, and the descriptors can be projected to subspace $x' = U^t x$.

There are two more problems should be concerned. Firstly, if projections w that are essentially in the noise components of the signals, but appear to be discriminative in the absence of sufficient data, the projection matrix described above might be overfitting. This problems can be tackled with a modified cost function, where $J_m = \frac{w^t S_{non-match} w}{w^t S'_{match} w}$ $S'_{match} = U \Lambda' U^t$ is a regularized version of

$S = U \Lambda U^t$ with its eigenvalues clipped against a minimum value.

Avoiding linear dependence may eliminate redundancy in representing the subspace. So another issue of interest is maintaining pursued projections orthogonal to one another. This can be easily achieved by adding linear constraints to criterion function (3). If w_k is the kth orthogonal projection, this constrained optimization problem can be expressed as $w^t w_1 = 0; ...; w^t w_{k-1} = 0$. Where $w_1, ... w_{k-1}$ are the set of orthogonal projections already obtained.

2.3   Depiction of Moving Objects

In order to incorporate spatial information of local descriptors associated with moving objects, a informative description is necessary. We treat these descriptors as items and encode spatial arrangement directly in them, and this yields a much stronger descriptor. We create transactions with selected items along with the neighborhood around a limited subset of these items. We set presenting frequency of items between f_{min} and f_{max} frames, and there must exist matching items in adjacent frames. These restrictions can eliminate some unimportant regions in keyframes and reduce the number of transactions, so the runtime of mining algorithm mentioned in next section will be reduced.

(a)

$\{ \text{tl } x_3 , \text{tl } x_5 , \text{tm } x_5 , \text{tr } x_7 , \text{ml } x_{15} , \text{mm } i_{central} , \text{bl } x_{21} , \text{bm } x_{21} \}$ (b)
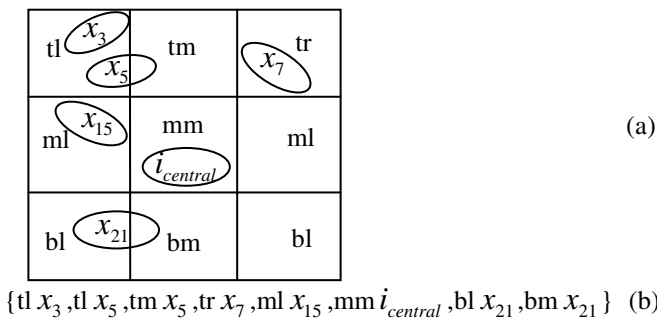
Fig. 2. (a) Creating transaction from neighborhood with 9 tiles. Circles represent local features, and letters indicate codebook they are assigned to. (b) Transaction.

Transaction of each item is created together with its surrounding nearest items and their rough spatial configurations. Instead of using a k-neighborhood like other research present we use the scale of central region which is restricted by item $i_{central}$ to define the size of neighborhood. More precisely, the square region which is proportional to the scale of central region belongs to the neighborhood of item $i_{central}$. Subsequently, each neighborhood is split into 9 tiles as shown in Fig.2(a). We label the tiles {tl, tm, tr, ml, mm, mr, bl, bm, br}(for "top-left", "top-middle", etc.), and append to each item the label of the tile it belongs to. Fig.2(b) shows the example of transaction created for $i_{central}$ from its fixed neighborhood. Since the neighborhood of each central item is split into subsection, this result in a very large number of items, but no changes is needed to facilitate the frequent itemset mining algorithm.

3   Mining Distinctive Local Feature Configurations

Based on the depiction of moving objects mentioned in the previous sections, the frequent and distinctive configurations of local features can be found. We chose the

APriori algorithm to implement frequent itemsets mining procedure. Transactions generated in section 2.3 are input to mining algorithm, and only the so called "closed repetitive gapped itemsets" are mined[6].

The output of the APrioti algorithm is a set of interesting local features with strong spatial constraints. Since these spatial configurations are less likely to appear by coincidence, the frequent itemsets may correspond to distinctive objects. Furthermore, a motion segmentation algorithm is implemented along with local feature detection procedure, so only the moving distinctive objects remained here.

The frequent feature configurations mined above are then clustered with method proposed in [7] to describe the object class. So given a novel video clip, we can now match the mined configurations to a specific object class.

4 Performance Studies

4.1 Design of Experiments

Results were presented with three main experiments. First we explore dimensionality reduction of local feature descriptors on the moving objects, To depict the results, ROC curves were plotted and the false positive rate at 95% true positives.. Second, we evaluate the performance of configuration mining method. Finally, the CPU-time of our mining procedure is measured.

The experiments are conducted on the data sets provided by TRECVID 2010, and 5,000 example video clips were drawn randomly from it.

4.2 Evaluation of Condensed Local Feature Descriptors

Results from our proposed algorithms were presented. For comparison purposes results for SSD and PCA were also included. The ROC curves presented in Figure 3 show the performance of these different methods.
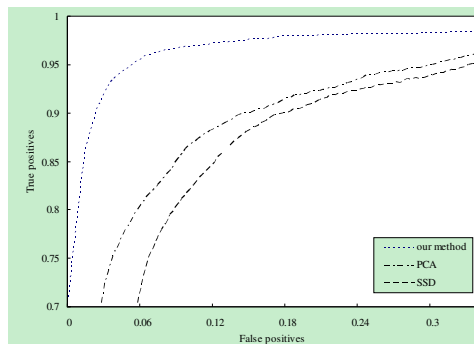
Fig. 3. Selected ROC curves for three descriptors

It can be seen that our dimension reduction method outperform raw SSD (33.7% error rate) and PCA (30.1% error rate). PCA reduces the dimension of descriptors dramatically but only get a small improvement over SSD in precision. However, our method gives a high precision with error rates of < 7% false positive at 95% true positive. Meanwhile, our method has far fewer dimensions (17 dimensions) than SIFT (128 dimensions), SSD (1024 dimensions) or PCA (28 dimensions).

4.3   Performance of Configuration Mining

Motion segmentation aided configuration mining is evaluated compared with a fixed 40-neighborhood method described in [8]. It is noted that there are only 10734 transactions in our method, while there are more than half a million transactions in 40-NN method. This difference can be explained that many detected regions related to static objects are omitted in our method, so the clutter can be eliminated in a significant degree. Additionally, our method can get more frequent itemsets with a high support threshold, and more condensed and reasonable clusters which can be confirmed by manual verification.

Table 1. Comparison of mining methods. (Regions: total number of regions detected interesting in video data set. #T: total number of transactions. s: threshold of support. #FI: total number of frequent itemsets. Clusters: number of clusters).

Method	Regions	#T	s	#FI	Clusters
Motion Seg.	$6.05*10^7$	10734	0.083	39460	24
40-NN	$7.82*10^9$	544364	0.0002	375	67

4.4   Running Time

The running time measurements are given in Figure 4. The time is measured for the whole frequent and distinctive configuration mining stage including depiction of moving objects, but before clustering stage. Results of the measurements demonstrate the scalability of our mining method, that is said that most frequent and distinctive configurations of moving objects can extracted from large amount of candidates in a matter of seconds.
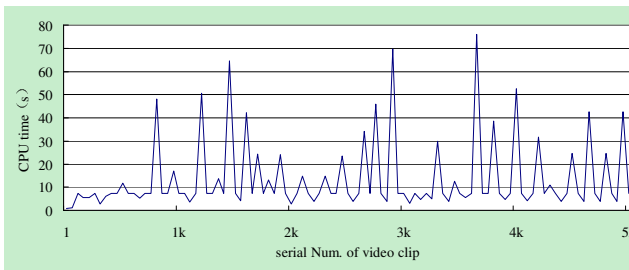
Fig. 4. CPU time of mining procedure

5   Conclusions

We have proposed a new discriminative method for mining moving objects frequently appearing in videos. Experiments show that our mining approach based on dimensionality reduction of local feature descriptors is a suitable and efficient tool for video mining. Restricting the interesting central region by moving macro blocks has proven to be useful for omitting clutter.

References

[1] Ke, Y., Suk, R.: PCA-SIFT: a more distinctive representation for local image descriptors. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 506–513 (2004)

[2] Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(10), 1615–1630 (2005)

[3] Bay, H., Tuytelaars, T., Van Gool, L.: SURF: Speeded Up Robust Features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006)

[4] Quack, T., Ferrari, V., Leibe, B.: Efficient Mining of Frequent and Distinctive Feature Configurations. In: International Conference on Computer Vision, ICCV (2007)

[5] Chen, H.-T., Chang, H.-W., Liu, T.-L.: Local discriminant embedding and its variants. In: Proc. of IEEE Conf. on Computer Vision and Patter Recognition, vol. 2, pp. 846–853 (2005)

[6] Ding, B., David, L., Han, J.: Efficient Mining of Closed Repetitive Gapped Subsequence from a Sequence Database. In: ICDE 2009 (2009)

[7] Ma, C., Shen, W.: Clustering navigation patterns using closed repetitive gapped subsequence. In: International Conference on Logistics Systems and Intelligent Management, ICLSIM 2010 (2010)

[8] Minogue, K., Gondry, M.: Come Into My World. EMI (2002)

Research on Information Memory of Location Server with SIP in Distance Education System

Shuxian Liu and Qing Cui

Information Science & Engineering Institution,
Xinjiang University
Urumqi Xinjiang, China
`lsx623@163.com, cuiqing@xju.edu.cn`

Abstract. This paper introduces how to realize fast redirection by location server in the distance education system. And explains in detail about the method how to store the information of user register in the location server of the distance education system.

Keywords: Session Initiation Protocol (SIP), location server, information memory, distance education system.

1 Introduction

1.1 Definition of SIP

SIP is Session Initiation Protocol, which is a signaling protocol used to implement instant message based on IP network[1].

SIP use the textual encoded mode which is the best feature of SIP when comparing with other VoIP and the existing standard in video telecommunication domain[2][4].

1.2 Four Components of SIP System

SIP system has four main components:

SIP UA (User Agent): It is the end-user's equipment. Users employ UAC (User Agent Client) to send messages. UAS (User Agent Server) responds to those messages [3].

SIP RS (Registrar): It is a data base with all the clients' proxy location in the domain.In the process of SIP communication the server will search the IP addresses and other related information of participates and send them to the SIP proxy server[3].

SIP PS (Proxy Server): It is used to accept the conversation request from SIP UA and to inquire SIP RS to get the address information of receiver UA[3].

SIP RDS (Redirect Server): It allows SIP proxy server redirect the SIP invitation information to the external domain[3].

2   Fast Redirection with SIP by Location Server

This is the most efficient way,but it need a equipment to add,which is location server.Location server is not the protocol entity of SIP, but it plays a very important role in systems of using SIP protocol to realize function for user location. Location server accepts, stores and returns the possible information of location about requested user to requesting user. Location server accepts registered information of users from registrar in local domain and other domains of system, analyses, collects and stores all of users' address information. When users registing, information of users is writed in registrar.Location server can store the information of users which be sent by registrar of local domain at regular intervals. The system can have more than one location server, it based on the number of users. If the requested user not belong to the same domain, it will be redirected by redirect server, redirect server can inquire about the address of requested user to location server directly and get the information.

In the Fig.1, user li registers and sends it's information of "I am li, my address is sip:li@202.160.123.112" to SIP proxy server of the domain named company.com or registrar,the proxy server or registrar stores the information, at regular intervals, the proxy server or registrar sends the information of registering to location server to store.
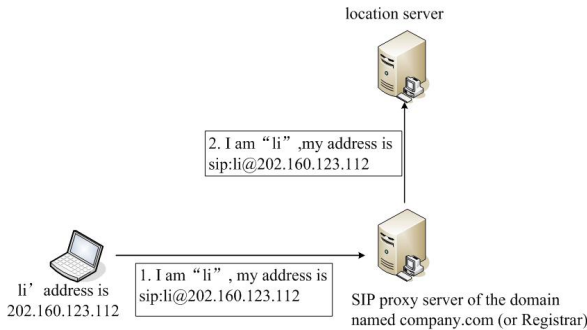
Fig. 1. Proxy server sends registered information of user to location server
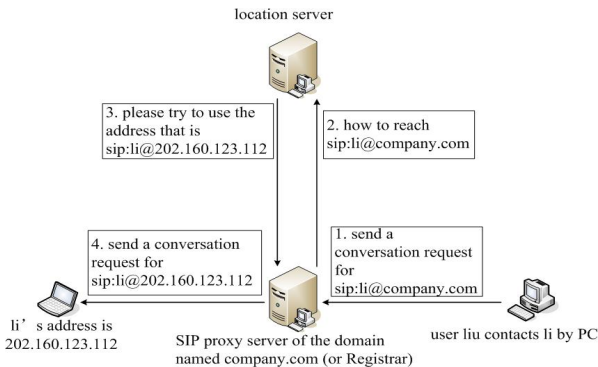
Fig. 2. Proxy server ask location server about user's address information

In the Fig.2, user liu contacts li by personal computer,the user proxy client in the personal computer accepts order and sends conversation request for li to SIP proxy server of the domain named company.com or registrar, and the proxy server inquires about "how to reach sip:li@company.com" to location server directly, the location server returns the information of "please try to use the address that is sip:li@202.160.123.112" to the proxy server of the domain named company.com or registrar,the proxy server or registrar sends the conversation request for sip:li@202.160.123.112 and user liu can converse with user li.

3   The Method of Information Memory in Location Server

3.1   The System Divided into Areas

When there are a lot of users, in the Fig.3, the system can be divided into areas, each area has some domains, and each domain has four components: SIP UA, SIP RS, SIP PS, and SIP RDS. But each area has and only has one location server.Because the scale of the distance education system is not too big, location servers can be connected by full mesh, and the way of unicast, that is to say the way of point to point can be used to send the information of synch.
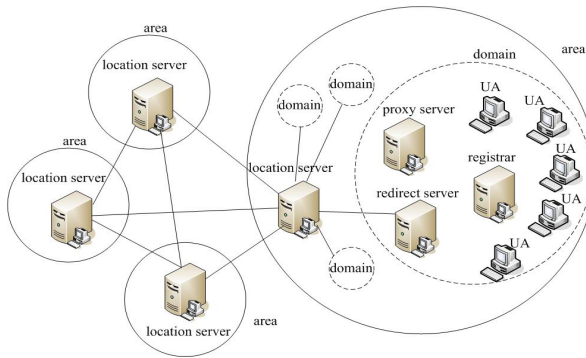
Fig. 3. The system divided into areas and has several location servers

3.2   The Method of Information Memory in Location Server

All users' register informations are be stored in location server, so how to store the register informations will affect searching efficiency directly that the information which stored in location server, that is to say, will affect efficiency of fast redirection with SIP directly.There are many methods to store information, for example, to use sequential mapping and sequential search, to use ordered list and binary search, or index by area and domain to store etc.

This paper introduce the method of tree structure based on time efficiency first that is called M2Tree. The M2Tree is a m-nary tree which stores all users' register information of the system, and m is the number that areas or domains in the area to be divided, that is to say, m=max{the number of areas to be divided, the number of

domains in the area to be divided}, and each node of user register information only has two children.

In the Fig.4, the root is the node of system, the root most has m children, to wit, these nodes that are the area information nodes about areas divided in the system are called first layer node. Each first layer node(area information node) most has m children too, and these children are domain information nodes about domains devided in the area, called second layer node,and each second layer node (domain information node) most has m children. Third layer, fourth layer…are nodes of user register information node in the system, and each node most has two children. The subtree that root is third layer node is an balanced binary tree, and consist of nodes of user register information. So, the node described by C language is:

```
struct M2Tree_AD_Node
{
    char *name;
    struct M2Tree_AD_Node *children[m];
};
    struct M2Tree_Node
{
    char *name;
    char *address[4];
    struct M2Tree_Node *left,*right;
};
```
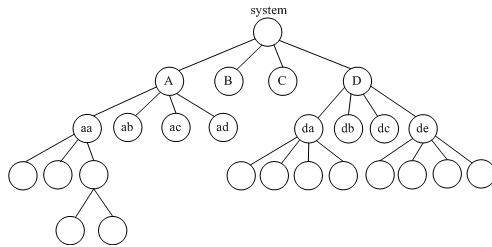

Fig. 4. The Example of M2Tree

4   Summaries

SIP protocol is selected as the next generation network (NGN) framework agreement, multicast based on SIP protocol will be widely used in multi-party video conferencing, instant messaging and distance learning system.

This article explains in detail about how to realize fast redirection by location server in the distance education system.As the number of users increases, the system can be divided into several areas, and set one location server in each area. And introduces in detail how to store the information of user register in the location server of the distance education system.

References

[1] Zhang, Z., Zhang, Y., Liu, Y.: SIP, Protocol and Applications. Publishing House of Electronics Industry, Beijing (2005) (in chinese)

[2] Camarillo, G.: SIP Demystified. Post & Telecom Press, Beijing (2003)

[3] CGIC, http://www.boutell.com/cgic/

[4] Siduan, F., Han, X., Long, Q.: Technology core and developing in SIP. Journal of Software, Seventeenth Periodical 2, 239–250

[5] Rosenberg, J., Schulzrinne, H., Camarillo, G.: Request for Comments: 3261. SIP: Session Initiation Protocol (2002)

Framework and Construction Contents of Digital Mine

Seng Dewen[1] and Shu Yueqing[2]

[1] Hangzhou Dianzi University,
Hangzhou, P.R. China 310018
`sengdw@163.com`
[2] Zhejiang Post & Telecommunication Construction Company, Hangzhou 310018
`shuyueqing@163.com`

Abstract. The basic goal of digital mine construction is to achieve efficient, safe, green development of mining resources. Research of digital mine must focus on the green development of mining resources exploitation. The construction of digital mine is based on the information integration technologies. Mining spatial data collection should be timely and reliable. Standardized system for constructing a digital mine, that is shared rules, standards, norms and policies are the most important in the construction of digital mine. The construction of digital mine must start from the reality of modern business operations of digital mine. Concepts update, training, increasing input, science and technology research organization, the formation of dominant force, and the promotion of the steady and healthy mining are the strategic implementation and specific countermeasures of the construction of digital mine.

Keywords: digital mine, mining engineering, system framework, digital construction.

1 Introduction

Digital mine construction is a large complex system works, its long-term goal is to [1-4]: for the exploitation of mining resources and the environment, and the whole process of production process control, advanced digital information technology, combined with large-scale intelligent machinery and equipment to replace traditional manual or mechanical operations, mine production and management of control of resources and the exploitation of the digital environment, technical equipment, intelligence, visualization of production process control, information transmission network, production management and decision-making more scientific.

With the level of technology development in today's world and mine at the present stage of technical equipment and management level, the construction of digital mine also undergo a long process. Therefore, the overall plan, step by step is the only way for China digital mine construction. At present, the specific objectives of digital mine construction is [5-8]:

(1) Using sophisticated computer software system, and mining resources, mining optimization, design, production planning and exploitation of the digital environment, modeling and visualization.

(2) The establishment of optical fiber, cable or wireless communication leak multimedia communication network as the main form of voice, video and data transmission network system differently, to achieve mine distributed data sharing.

(3) The use of advanced sensor network technology, to achieve mine production processes, equipment, safety and environmental monitoring and other data mining the automatic collection, intelligent analysis and visualization.

(4) Using Ethernet, PLC intelligent control and video surveillance system, and upgrading the mine, transport, ventilation, drainage systems and equipment such as intelligent centralized monitoring.

(5) The use of advanced production management and control system, production personnel mines and to locate and track mobile devices, as well as intelligent production scheduling and control, mine production to enhance the management and scientific decision-making.

In recent years, with coal, nonferrous ore, non-metallic minerals, represented by the traditional Chinese mining industry has been moving in the direction of digital mine made a solid pace, digital mine construction tasks, principles, procedures become increasingly clear, comprehensive shop digital mine construction open. The initial stage of digital mine construction, and its main tasks include: the establishment of MGIS business platform, dynamic management of mineral resources, and the mine visual elements to support the dynamic input-output analysis [6-10].

2   The Structure of Digital Mine

2.1   Data Layer

This is the data acquisition and storage layer. Data acquisition techniques including the use of various forms of data access and pretreatment; data storage, including various types of databases, data files, graphic files libraries. This is the layer for subsequent layers to provide some or all of the input data.

2.2   Model Layer

It is the presentation layer to show the mineral properties, such as three-dimensional space and two-dimensional block models, geological models, mining field model, geographic information system model, virtual reality animation models. The data processing layer is not only intuitive presentation of the image, but also for optimization, simulation and design input.

2.3   Simulation and Optimization Layer

Such as process simulation, optimization, design optimization and planning solutions.

2.4   Design Layer

That is aided design layer. This layer is the optimal solution into executable programs or directly provides the means for design.

2.5   Execution and Control Layer

Such as automatic scheduling, automatic monitoring and control process parameters, such as remote operation.

2.6   Management Layer

This layer includes management information systems (MIS) and office automation.

2.7   Decision Support Layer

Or more layers according to various information and data provided by processing results of correlation analysis and forecasting, to provide decision makers at all levels of decision support.

3   Composition of Digital Mine System

3.1   Data Collection and Management System

This system is responsible for data acquisition, processing and management. The system uses a variety of technical means (such as RS, GPS, conventional survey, sampling, testing, etc.), access to various forms of spatial and attribute information, the necessary pretreatment of information (such as graphics and digital data, or vector, Image analysis, coordinate transformation, data integration, etc.), to establish a database for data storage and management. The system software includes a number of pre-package and database of information and its management software system.

3.2   Deposit Model System

The function of the system is built on the spatial properties of ore digital model. To form, can be a massive models, wire frame model or solid model. Of mine production is most important is quality model, impurity model and value model, which is delineated ore, ore grade calculation and design, plan. International use of three-dimensional block model is the most, because most of the optimization algorithm based on this. This model at this stage of the main data source is drilling, trenching and sampling hole. Mining three-dimensional modeling system to the object-oriented methods and concepts to provide a "digital mine" expression of the necessary tools to produce three-dimensional scene, enabling the surface and down hole three-dimensional display of various entities, including: the ground, buildings, water, communication facilities, Trees, roads, fences; underground drilling, formation, Rock Lane (half arch), Coal (trapezoid) and the mechanical and electrical equipment and other objects [11-14]. Users can view any angle seam, roadway, drilling and ground spatial relationships between objects, to achieve visualization of multi-source data mining integration.

3.3   Geological Model System

Its function is based on drilling or remote sensing, telemetry information to establish deposit structure model, such as faults and fracture zones, litho logy, structure, etc., at this stage mostly to wire-frame model.

3.4   Optimization and Simulation System

The system is optimized algorithm and computer simulation, and programs on the important parameters to optimize and evaluate the system generally includes a number of independent single-function software package, such as optimization of the final state; production capacity optimization; and economic benefits based on ore The stripping (dig) plan optimization; the best industrial grade dynamic optimization; optimal estimation of ore grade; scheduling optimization; waste rock optimization; mining method parameter optimization; beneficiation process parameter optimization; equipment renewal life optimization; system simulation .

3.5   Aided Design System

Its function is the interaction with the user, the program will optimize the system to produce optimal results can be processed for the implementation of the program, or in the case alone without the formation of optimal solutions in terms of production of mines, the system is to measure, plan and mining engineering Pen, paper, plan meter and a calculator. To test collection, set blasting zone, exploitation and promotion plan, exploration and production, storage and all levels of mine blast calculation of the amount of work, and the amount of work involved in mining, grade calculation and drawing, all completed in the system. The system should have a high degree of live, interactive and convenient interface that will combine a high degree of systematic work, and the highest possible degree of automation, so that the original work to be completed in a few days, the program can be completed in minutes.

3.6   Scheduling System

To GIS-based mine, as mine of information in public office and decision-making platform and application software integrates various types of mines and model integration of public carriers, integrated scheduling and visual control of mining data streams, functional flow and business flow. Its function is based on the computer, GPS, wireless communication technology to achieve the automatic exploitation of transport to optimize scheduling. Function by reducing total transportation and mining, transport equipment, waiting time, improve mining equipment, transport system efficiency and reduce harvesting costs. The nerve center is the optimal scheduling system scheduling software systems, including vehicles equipped with route optimization, maintenance optimization, scheduling rules for becoming excellent.

3.7   Management System

The system is all walks of life in the application of information technology, the fastest growing area, now has a considerable degree of popularity. Mining enterprise management system includes MIS systems (financial management, personnel management, equipment management, spare parts and supplies management, energy management, comprehensive statistics, cost accounting, etc.) and office automation systems. This system is the last client - server form, in recent years mostly by technology-based Web site in the form instead.

3.8   Decision Support System

The variety of internal information systems (including floors above and the processing results of the data provided), and external information (mainly the market and policy information), the correlation analysis, inference and prediction for decision makers to provide a variety of production and management strategies Decision support to improve the timeliness and scientific decision-making.

4   Construction Contents of Digital Mine

4.1   Basic Network Platform

Based network platform not only to meet internal digital content transmission and the need for wired and wireless communications, we need to meet the needs of the Internet, through a broadband fiber optic ring network, field bus and commercial databases, multimedia databases, to achieve real-time database of information collection, transmission, Storage, analysis, decision making, control, distribution and queries, not only to ensure the interconnection of information, but also to ensure the reliability of the information, security, and timeliness.

4.2   Mine Data Warehouse

Mine data warehouse is not the simple accumulation of various databases and assembly, must be achieved from the scattered data to the standardized data integration, and its technology is the key for geological, mineral deposits and production of information complexity, mass, heterogeneity, dynamic and multi-Source, multi-precision, multi-temporal and multi-scale characteristics and a variety of professional applications, the establishment of sub-standard theme layer, establish a standard database, database structure, the definition of standardized metadata, index data and the data cube, to facilitate import and export data And automatic mining, eliminating mine "information island" and the data redundancy and achieve a variety of reference for the rapid retrieval and forwarding, real-time database must take the initiative to achieve a variety of events triggered and to ensure timeliness.

4.3   Mine Industrial Automation Systems

Mine industrial automation systems, including underground conveyor belt monitoring system, monitoring system to enhance the main auxiliary shaft, mine power monitoring system, mine drainage control system, gas drainage monitoring system, the main engine room ventilation and monitoring systems, room air pressure monitoring system, water supply dust Monitoring system, grouting monitoring systems, mine monitoring system cooling, air volume control and control systems, production measurement systems, surface production control systems, which correspond to the number of mines in the leg and hand, although able to complete a variety of actions to be controlled, but In the brain, where the brain is the system integration platform software, which can effectively through collaborative data warehouse, safety supervision system, safety management systems, production systems and scheduling systems, and automatic

control and remote control, to achieve real integration of management and control, rather than Simple connection of each subsystem and accumulation.

4.4   Safety Monitoring System

Safety mine safety monitoring system is the eyes, ears and nose, including wired communications, wireless communications and mobile communications systems, mine personnel positioning systems, video surveillance systems, security monitoring systems production environment, hydrological monitoring system, underground pressure monitoring system Fire control system, rainfall monitoring system, surface movement monitoring system, through which the data warehouse, safety management systems, scheduling systems to provide real-time information for decision support system.

4.5   Production Technology Management System

Digital production technology management system is the core of mine, including geographic information systems to measure, mining, collaborative design, intelligent ventilation system, transmission and distribution of geographic information systems, industrial pipe network geographic information systems. But we need to note that, this production technology management is not the manual method to move the computer, and to achieve scientific and standardized management and high-intelligence support. For example, to measure and survey of geographic information system is not drawing, but through the refinement of the mine surveying and geological exploration, to deposit, geological, hazards, and the precise delineation of roadway engineering and visual computing, you can automate a variety of Measurement correction, the error is expected, Survey Adjustment, surface subsidence is expected to computing, and to achieve WYSIWYG all calculations, and can automatically generate any of a variety of mineral scale map for the mine production, safety and management of geological protection; Three-dimensional geographic information system based on mining, including collaborative design system to complete roadway layout, rock movement is expected to protect the pillar (pillar), is supporting selection and mining, excavation, machinery, transportation, communication, security and other design And generate the relevant documents; intelligent ventilation system's main purpose is to ensure normal production and disaster during the time period of demand for the wind; Transmission and Distribution Geographic Information System's main function is arbitrarily complex network to achieve optimal device selection, fault Current calculations, voltage and energy loss calculation, the capacitor current calculation, calculation of dynamic stability thermal stability and protection setting calculation, with the power supply SCADA system to achieve remote communication, remote sensing, remote harmony remote control; other systems and so on. These are imitated by artificial means cannot be completed.

4.6   ERP System

Mining ERP system is the sophisticated management information system, including human resource management systems, equipment management systems, materials management systems, distribution management systems, production planning management system, property management, budget and cost management, financial

management systems, Office automation systems. Its characteristics are to be integrated seamlessly with other systems, such as human resource management, production planning, material requirements, budget and cost management must be based on the design results, production technology, equipment management, materials management, distribution management to the help of Geographic Information systems to achieve optimal distribution of site management and other subsystems are similar, either based on data from other systems, or for other system services that can be seamlessly integrated through the data warehouse, or to become an information Island.

4.7   Integrated Command and Control System

Digital integrated command and control system mine are not to monitor and control, industrial automation, personnel location and communication system integration simple, plus some, such as production scheduling logs and reports. But on the three-dimensional visualization platform, monitoring and control, communications and personnel location systems, and production technology, management, security management, and strong mining ERP integration.

5   Conclusions

Mining enterprise is different from other enterprises, their production both underground workplaces, but also the ground; and underground and ground works with each other, the dynamic changes. Therefore, the number of mine construction is a complex, large and long-term system engineering, to be phased construction. In the initial stage of construction, the first is to develop MGIS, and technology needs of major mines around the key issues and professional development of modules and applications. This paper researched the framework and the construction contents of digital mine. The components of a digital mine were analyzed and discussed in detail.

References

[1] Seng, D.W., Li, Z.X.: 3D visual modeling system for mineral deposits. Journal of University of Science and Technology Beijing, 453–456 (May 2004)

[2] Kreuseler, M.: Visualization of geographically related multidimensional data in virtual 3D scenes. Computers & Geosciences, 101–108 (July 2000)

[3] Marschallinger, R.: Three-dimensional reconstruction and visualization of geographical materials with IDL- examples and source code. Computers & Geosciences, 419–426 (June 2001)

[4] Seng, D., Song, Z.: 3D interactive visualization system for cojmplex geologically related data. In: First International Conference on Intelligent Networks and Intelligent Systems, 519–522 (November 2008)

[5] Seng, D.W., Wang, H.X., Yue, G.Y.: 3D modeling and visualization of complex geological structures using OpenGL. Boundaries of Rock Mechnics. In: Proceedings of the International Young Scholars Symposium on Rock Mechanics, pp. 935–938 (May 2008)

[6] Smith, G.L., Caris, C., Soole, P.: Virtual mine prototype: interactive visualization methods using VRML, Java and an internet browser. Australia CSIRO Exploration and Mining Report No. 462F (1998)

[7] Rizk, A., Mather, K.: Simplifying simulation modeling through integration with 3D CAD. Journal of Construction Engineering and Management, ASCE, 475–483 (July 2000)

[8] Seng, D.W., Liang, X.: Visualization of large scale geographically related data in virtual 3D scenes with OpenGL. In: Wang, C., Zhong, S., Wei, J. (eds.) Second International Conference on Space Information Technology. Proceedings of SPIE Bellingham, pp. 1027–1033. SPIE, WA (2007)

[9] Lucilla, C.F.: Visualization of 3D information with digital holography using laser printers. Computers & Graphics, 309–321 (June 2001)

[10] Seng, D., Song, Z., Wang, H.: Construction and Visualization of Complicated Objects for 3D GIS. In: Proc. of SPIE Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Geo-Simulation and Virtual GIS Environments, vol. 7143 (June 2008)

[11] Seng, D.W., Li, Z.X., Li, C.M.: Application of marching cubes algorithm in visualization of mineral deposits. Journal of University of Science and Technology Beijing, 203–207 (March 2005)

[12] Seng, D., Li, Z.: Design and implementation of a 3D simulation system for geological and mining engineering. Journal of Liaoning Technical University 27(1), 9–12 (2008)

[13] Dollner, J., Hinrichs, K.: An object-oriented approach for integrating 3D visualization systems and GIS. Computers & Geosciences, 67–76 (June 2000)

[14] Seng, D.W., Li, Z.X., Li, C.M.: Volume rendering techniques for visualization of mineral deposits. Journal of Liaoning Technical University, 473–476 (April 2005)

Research on the Application and Status Quo of Digital Mine and Sensing Mine

Seng Dewen and Shu Yueqing

Hangzhou Dianzi University
Hangzhou, P.R. China 310018
sengdw@163.com
Zhejiang Post & Telecommunication Construction Company, Hangzhou 310018
shuyueqing@163.com

Abstract. Digital Mine and Sensing Mine are the result of integrating various disciplines including mine science, computer science, 3S techniques, artificial intellection and internet of things techniques (IOT), which will radically change the traditional mine production and our lifestyles. The functions, contents, main characters and the development situation and problems of digital mine are analyzed. The hierarchy and system framework of digital mine operation system and the application of IOT in digital mine, that is, sensing mine, are discussed. The key techniques of the application and construction of digital mine are studied.

Keywords: digital mine, mining engineering, sensing mine, IOT, 3D visualization.

1   Introduction

Digital mine takes computers and network as the core means to achieve the mine information acquisition, storage, transmission, presentation, processing and applications in various production processes and management and decision making. The basic framework of digital mine should be digital data acquisition and management systems, data applications, scheduling systems, management systems and decision support systems and other components. The construction of a digital mine will start from basic theories and models, mine spatial data mining and mining data warehouse construction, mining network transmission platform, a variety of related business processes, professional integrated software construction [1-3]. The key techniques of digital mine include 3D spatial data acquisition and management, spatial data model and data structure, data warehouse technique and data mining technique, 3D visualization and virtual reality and 3S, etc.

The construction of digital mine is a complicate system engineer. It involves in various technical fields of mines and organizational departments and cannot achieve immediately. We should combine with the information situation and needs of the mine enterprises in our country, macro-grasp, plan, and work out the basic framework and objectives of the overall building goals, and implement and promote step by step, to achieve a sustainable development of our mines.

For the mining industry, to go high technology content, good economic returns, low resources consumption, little environmental pollution and human resource advantages into full play to the new road to industrialization, digital, information is the inevitable trend of development. Implementation of digital management, considering the production, operation, management, environment, resources, security and efficiency and other factors, so that mine planning and management more efficient, richer performance practices, more information, higher analytical ability and accuracy, thereby enhancing the mine production and management of the timeliness, effectiveness, level of optimal allocation of resources, comprehensive strength, and promote the sustainable development of mining, to improve the overall efficiency, market competitiveness and adaptability of the goal. In recent years, through continuous scientific exploration, mining enterprises in China have been part of a digital management [4-7].

In the process of implementing digital mine, mine of information for the complexity, mass, heterogeneity, uncertainty, and dynamic, multi-source, multi-precision, multi-temporal and multi-scale features, need to build a unified comprehensive information System platform to achieve integration of mining data visualization, real mine of information sharing and interoperability, as mining exploration and production of decision support.

2   Contents and the Basic Features of Digital Mine

Since mine is a resource for the development of object discrete production systems, the main job is surface deep and complex geological conditions, poor environment, large-scale accidents have occurred, to the mining production in China caused heavy losses, but also seriously endangers the physical miners Security. China's mines in mine surveying, planning, design, production, management, monitoring the whole process of information mining areas and developed a growing gap between countries, China has neither the mine as a mine of information resources, one of the important co-ordination of strategic resources Development and use, but no system performance and stability, information resources sufficient information infrastructure, mining, nor the formation of "digital mine" building specifications, many scholars and business and technical personnel mines on the number of the definition, meaning and function is still relatively vague, This section defined by a complete digital mine, briefly describes the technical features of digital mine, building content, technical specifications, system functionality [8].

There are no accurate figures on the mine, the accepted definition. The ultimate goal of Digital mine is to achieve mine truly safe, efficient and economic exploitation. And mining development, the ultimate goal should be not only to meet human mineral Demand for resources, but also to adapt to ecological and environmental carrying capacity of the system to achieve the goal of sustainable development. This is the scientific concept of development in the concrete embodiment of mining engineering. Consolidated Mining and mine information technology developed different strategies and different academic ideas The concept of digital mine the expression of the concept of digital mine can be expressed on the grounds of three beginner to advanced levels: mining digital information systems, reflect the true overall mining and related phenomena virtual mine, no mine's remote control and automation of mining operations.

Mine digital information system includes the following subsystems: surface mining and mineral deposit model visualization information systems, mining engineering geology, hydrogeology and rock mechanics data acquisition, processing, transmission, storage, display and distribution of integrated systems engineering of exploration, mine planning Decision optimization system and mining, mining equipment operating status of major information systems, production processes monitoring and scheduling system, mining environmental change and disaster early warning information systems, mine operations management and information systems analysis of economic activities. At present, China has built a number of key mines or under construction includes the contents of digital information in different systems of mine [9].

Remote operation and automation of mining: in the above figures on the basis of, additional equipment automation, intelligent, underground communications and automatic real-time positioning and navigation technology, the production process of remote control mining operations and automation of mining, to the whole process from a single mine, Mine production of the digital office technology.

The core of digital mine is the time reference in a unified framework and space, scientific and orderly organization, management, maintenance and real three-dimensional visualization by various means to obtain mass, heterogeneous, heterogeneous, multi-dimensional and dynamic mining information And to establish a mine of information distributed sharing, coordination and use of mechanisms, the formation of a variety of flexible and convenient digital methods and simulation tools to maximize data mining and mining potential and play a role, and throughout the mine planning, production, operation and Management of the entire process to ensure scientific decision-making mines and modern management. Mine the true number of mine is a unified whole and the understanding of related phenomena and digital reproduction, is the number of mines and numbers is an important part of China. It is the height of the final performance information for the mining, automation and efficiency, as well as no mining and Remote mining [4].

3   The Construction of Digital Mine

Mine is a digital time and space in a unified framework, through digital and three-dimensional modeling, to achieve all the objects above and below ground mining visualization of transparent management; can be simulated through the mining process, to achieve early warning of disasters ahead to avoid disaster , reach the essence of security; available through meticulous management, to achieve optimal use of enterprise resources, reduce production costs and achieve high yield and efficiency; the ultimate goal is to achieve the visualization of the mining process, automation, intelligent and even unmanned. Therefore, the "digital mine" of the building in disaster warning, save energy, optimal use of resources, environmental protection and recycling of mine has a great significance.

Digital mine is a national strategic resource security system is an important component. Digital mine construction will enable a comprehensive and detailed grasp of the distribution of mineral resources utilization and the level of protection for our industry, combined with the international market, that can achieve reasonable use of two resources at home and abroad, two markets, scientific, quantitative prediction of

the future Supply and demand situation, the establishment of an effective supply of strategic resources allocation strategies and safeguards. This is China's market economy development, but also effective participation, the use of increasingly global resources market. Mine is mining the development of digital high ground, who occupied the high ground, would control the development trend of the industry and initiative, so the number of mine construction and development developed rapidly, and in the number of mines, based on real-time process control, Real-time management of resources, mine information network construction, new machinery automation applications and intelligent automatic control of.

Digital mine construction is to enhance the international competitiveness of mining an important measure. China's mines the overall low level of mining technology, compared with the international advanced level, there is a big gap. Its outstanding performance is lagging behind some of mine equipment, mining small scale, low labor productivity, inefficient mines and mineral resources utilization rate. Digital mine construction will quickly improve mine design, decision-making and management of the scientific level, through the use of high-tech new equipment to increase ore production to increase productivity, improve product quality, reduce production costs, enhance the international competitiveness of China's metallurgical and mining purposes. Digital mine construction is an important cornerstone of sustainable development. Digital mine construction can be of mineral resources (including mining and reserves) out of the digital performance, this information can be judged according to priority development needs of those resources can ensure sustainable development of the national economy, can do a steady supply of mineral resources, rational use and Low environmental damage, can solve the situation of mineral resources make ends meet, living beyond the critical situation, the supply of resources and increasing consumption of resources to meet the basic balance [10].

Mine was to evaluate the number of mine important data resources based on the ecological environment. Digital mine construction will encroach on the land caused by mining, surface form damage, destruction of vegetation, land degradation, soil erosion, desertification, and dust pollution, water pollution, mining landslide, debris flow dump, tailings dam, Regional ecological landscape, etc. for the destruction of digital information through the scientific analysis, development of standardized mine ecological reconstruction, the establishment of a balanced dump reclamation evaluation system, hazard assessment and prediction system and modern management system, so that mine waste To rational use, significantly improving the ecological environment in mining, the maximum possible preventive measures, development of mineral resources and environmental development.

4   Research of Digital Mine Abroad

United States, Canada, Australia and other mining countries in terms of digital mine started earlier, the United States first proposed the concept of Digital Earth, and then quoted by many experts and scholars. After many countries in the world with their actual number of mines were further put forward the development of planning and development goals.

20th century, early 90s, the Canadian nickel mining company began to study the remote control technology, the goal is to achieve remote control operation of the entire mining process. The United States has successfully developed a wide range of mining scheduling system, using computer, wireless data communication, scheduling optimization, and global positioning system technology for open pit production of computer real-time control and management, and successfully used in industry, has led to open pit Almost no one realized mining. Canada has set out a proposed in the vision 2050, that is, in remote areas of northern Canada to build a unmanned mine, all mine via satellite control equipment, to achieve mechanical crushing and automatic mining [5].

Since the 20th century, some of the world mining industry in many developed countries have developed a number of mine-building software, and has been successful in many mining applications. One representative of the software are: the United Kingdom developed Datamine mining software, developed by the Australian Maptek Vulcan software, developed by the Australian Surpac software developed by Intergraph U.S. interactive graphics system developed by the Canadian Lynx MINCAD system, Rockwell developed WhittleFour2D open pit optimization design software. Abroad has now reached a level of: Continuously and automatically complete the heading face of a loop hole or face a row of deep mining hole in the drilling sector. Drilling accuracy is improved, a significant increase in working hours by the operator substantially reduced. The maintenance of mobile equipment in place still needs to go down the direct intervention of workers; underground mine development of radio communication systems has been completed, there are several automatic positioning and navigation systems used in the test mine.

Many countries are conducting trials, the establishment of demonstration mining area, but were undoubtedly true for the production of the mine caving method adopted. Ordered mine the miners of such a long period, and the operation is relatively simple and easy to automate. If the remote control to achieve the overall mining and automated mining target, although the individual operations in the past decade has been achieved impressive results, but the overall number of mines still in the construction of infancy. Construction of the largest digital technical difficulties mine the uncertainty in its object, it is difficult to accurately measure and control. Deposit mining is a complex, changing, information hiding, is difficult to predict the huge system, so from the information collection, transmission, processing, integration, display to automatically control the production process used, involving a wide range of areas, need more interdisciplinary, innovative and Accumulated experience. It should be recognized, digitized mining is mining development objectives and direction, rather than a specific project.

But the number of mining can greatly increase productivity, reduce production costs, so that enterprises in the metals market downturn is still a competitive edge; able to adapt to the growing number of deep mining conditions in which miners away from the high temperature, the threat of rock burst And other harsh operating environment, a radical improvement in the safety and health of miners, to achieve the office of mining. Therefore, digital mine of great significance and a strong attraction, but also has broad room for innovation. In this respect China is only in the gestation stage.

5   Problems in the Development of Digital Mine

For a long time, China's mines has been in the primary level of mechanization of labor-intensive, extensive management, technology and equipment behind, decision-making, design, production, etc. are dependent on many aspects of the experience, science is not high. Mainly as follows: mining geology, surveying, mining, technical methods are backward, human factors design of a larger establishment, construction, production automation and control low level of job security is low, the low level of production and visualization, communication methods are backward, backward production facilities Workers lack training in safety and reasonable operation, management philosophy obsolete, the system is not perfect, and slow speed of information transmission. These problems, mainly due to lower levels of the mine management inadequate management innovation, management ideas old, weak management infrastructure, management tools and methods behind, leading advanced technology, systems and means of introduction, building efforts is not enough. The resulting production is not high science, safety is low, not enough rational exploitation of low productivity, upload issued lag, lack of effective security measures and early warning mechanisms. To protect the safety of mine production is reasonable, efficient, enhance digital construction of the mine is particularly important [7].

China mining enterprises decision-makers, managers and engineers, information technology in the mines there are still conservative, short-term benefits, heavy soft and hard light "has nothing to do with an armchair, " and various other unhealthy obstacles The process of building mine of information, hindering the healthy development of digital mine. Research in science and technology workers cannot be part of the scene depth, behind closed doors, cannot combine well with the scene, but also affect the number of mine-building process is an important reason. At the same time that many of the non-scientific and technological workers overburdened, and I feel impulsive, difficult to concentrate on real scientific research. Over time, not only unsustainable Digital mine construction, mining, science and technology for sustainable development will be seriously affected.

National Digital Mine is still in early stage of development, mature and integrated to the unified management of spatial information, real-time dynamic information and management information base platform has not been reported, analysis of the main reasons for the developer involved with the profession. Currently, the direct promotion of the development of related digital mine developers have 3 categories: one for the geological survey system developers, the vector from the early into their mapping system, and gradually developed into a professional geological feature measuring system, and some integration Some management functions, such developers claiming to be the leader in digital mine; followed by the mining automation system integrators, who from the early mine safety monitoring system, developed to the full integration of integrated automation system of mine, all mine production will be Real-time information link control in hand, claiming to be a number of these developers, practitioners of mine [4]; third is mine information management, developers, they are from the office automation, marketing, equipment and labor management module and other start, will Mining sections of the management processes of the business information. Since these three developers involved in coal mining in different business units, each taken by the technical line, application platforms vary, resulting in the

current system difficult to integrate various information resources cannot be shared, spatial information is difficult to form a unified, real-time information and management information Platform.

From the above analysis shows that the development of digital mine is not an overnight thing, the need for collaborative professional development, technology and equipment to solve intelligent, 3DGIS support technology, automatic collection of different sources of information technologies, the integration of multi-source fusion of heterogeneous information , Three-dimensional modeling and visualization techniques, spatial and attribute data organization and management of centralized or distributed and shared publishing technology, basic analysis of information processing, engineering applications of basic information and other key technologies. The development of these technologies is uneven, there is a gradual process of development, so the number of mine construction is step by step manner. According to the objective requirements and the current mining technology, I believe that building in order to describe the main framework for spatial information mining, mining safety integrated real-time information and management information based on the number of mines is a digital information platform for the development of the mine a milestone in the road. Development of standards and digital mine description of the standard interface for third-party developers has become the consensus of all kinds.

6   Conclusions

This paper studied the theory and applications of the technical problems in the construction of digital mine and the sensing mine. The key technologies were emphasized. Through the research of digital mine home and abroad, the problems of digital mine were analyzed, and the solutions were proposed. Through the research of digital mine contents, features, and the basic framework, the construction of digital content mining, construction principles, the goal of building and construction specifications were presented. The paper also studied the key technologies in the construction of a digital mine, including three-dimensional data acquisition, data model and data structure, data warehouse and data mining, visualization and virtual reality technology, 3S and integration technology.

References

[1]   Seng, D.W., Liang, X.: Visualization of large scale geographically related data in virtual 3D scenes with OpenGL. In: Wang, C., Zhong, S., Wei, J. (eds.) Second International Conference on Space Information Technology. Proceedings of SPIE Bellingham, pp. 1027–1033. SPIE, WA (2007)

[2]   Kreuseler, M.: Visualization of geographically related multidimensional data in virtual 3D scenes. Computers & Geosciences, 101–108 (July 2000)

[3] Seng, D., Li, Z.: Design and implementation of a 3D simulation system for geological and mining engineering. Journal of Liaoning Technical University 27(1), 9–12 (2008)

[4] Seng, D.W., Li, Z.X., Li, C.M.: Application of marching cubes algorithm in visualization of mineral deposits. Journal of University of Science and Technology Beijing, 203–207 (March 2005)

[5] Seng, D.W., Wang, H.X., Yue, G.Y.: 3D modeling and visualization of complex geological structures using OpenGL. Boundaries of Rock Mechnics. In: Proceedings of the International Young Scholars Symposium on Rock Mechanics, pp. 935–938 (May 2008)

[6] Seng, D., Song, Z., Wang, H.: Construction and Visualization of Complicated Objects for 3D GIS. In: Proc. of SPIE Geoinformatics 2008 and Joint Conference on GIS and Built Environment: Geo-Simulation and Virtual GIS Environments, vol. 7143 (June 2008)

[7] Seng, D., Song, Z.: 3D interactive visualization system for cojmplex geologically related data. In: First International Conference on Intelligent Networks and Intelligent Systems, pp. 519–522 (November 2008)

[8] Marschallinger, R.: Three-dimensional reconstruction and visualization of geographical materials with IDL- examples and source code. Computers & Geosciences, 419–426 (June 2001)

[9] Smith, G.L., Caris, C., Soole, P.: Virtual mine prototype: interactive visualization methods using VRML, Java and an internet browser. Australia CSIRO Exploration and Mining Report No. 462F (1998)

[10] de Kemp, E.A.: Visualization of complex geographical structures using 3-D Bezier construction tools. Computers & Geosciences, 581–597 (1999)

Retraction: Application of Amphibious Technology in the ReutoMail

Sun Ping

Zhejiang Water Conservancy and Hydropower College
Hangzhou, P.R. China 310018
`htbbs3@163.com`

Several conference proceedings have been infiltrated by fake submissions generated by the SCIgen computer program. Due to the fictional content the chapter "Application of Amphibious Technology in the ReutoMail" by "Sun Ping" has been retracted by the publisher. Measures are being taken to avoid similar breaches in the future.

Research on Web Application Software Load Test Using Technology of TTCN-3

Ying Li and Qinghua Liu

Jiao Zuo University Jiaozuo, China
`liying19771224@163.com,`
`jzliuqh@sohu.com`

Abstract. Nowadays, TTCN-3 has been widely used in many domains, including telecommunication, automobile, medical equipment, etc. However, in software domain, it doesn't begin. To resolve this problem, study on the applicability of using TTCN-3 on testing web application software was made in this paper. The main part of TTCN-3 is core language for describing test behaviors and control. Test script programmed with core language and necessary assistant entities including Test Adapter and Codec form a whole Testing System. Experiment projects of testing web application software using TTCN-3 was developed in research, which includes load test. Besides that, data was also collected for final evaluation. The study and attempt of using TTCN-3 on web application software test not only provided a good reference for the using of TTCN-3 in software domain, but also had a directing effect on the improving and optimizing of TTCN-3.

Keywords: TTCN-3, Web application software test, Load test.

1 Introduction

With the rapid development of IT industry, application is becoming more and more complicated, causing more development failure. Therefore, people have more understanding of the test: test is no longer an idea after the matter; it is becoming more and more important and indispensable in the development process. People desperately need a kind of effective and flexible method of testing; TTCN standard is developing in this background.

TTCN - 3 can be applied to various black box testing of interactive system and distributed system .Due to TTCN - 3 appears very recently, so is not widely used. From worldwide, mainly concentrated in Europe And today its main test field is in telecommunications, for instance, in telecommunications aspects it includes ISDN, Atman in mobile communications aspects it includes GSM, UMTS ,and in the Internet it includes IPv6, SIP, and systems based on Curtain reality, test applications of TTCN - 3 in the field of communications is relatively successful, including MOTOROLA, ALCATEL, Vodafone, O2 and many other international large telecommunications enterprises are using it. In comparison, the application of TTCN - 3 in software testing is still a blank. Since we have just realized this point, we hope to introduce TTCN – 3

into IT software testing, but first need to analyze and evaluate its abilities to determine whether it is suitable for IT software testing.

Web application software has an important characteristic that it allows several users to access and use through the network at the same time, it involves the withstand ability of network, namely software can maintain maximum numbers of its performance, namely "load". Load testing means through gradually increase the system load, the change of performance of test system, and finally make sure the system can withstand the maximum load test, provided that it satisfies the index. For Web application software, a key point of load test is how to simulate access of concurrent users.

2   Overall Structure

2.1   Analysis of Overall Structure

The main four entities that need to be realized by using TTCN – 3 are SUT, TE, TA and CD.

The realization of SUT is completely a process of writing JAVA program, and basically it has no relationship with TTCN - 3. In fact, in real test SUT is provided by user and need not to be developed by testers, here is mainly due to the research needs.

TE is test code or test scripts written by using TTCN - 3 core language. General test code includes elements of language described in section 2.2. Here, I divide these main language elements into three categories: types, including type, template, and signature; configuration, including port and component; behavior, including test case and function's describes data transported in the process of test, including test of properties of operation environment and test of behavior.

For TA, because it's not standard content published by the ETSI, so needs fully customization to realize. However, TTworkbench tools includes a Test Adapter developed and named by TestingTech company; this type has realized interfaces such as TriCommunicatioSA, TriPlatformPA, TciEncoding, etc, which will be used for SUT and testing system communication and test execution methods such as resend, triMap, etc. So while realizing TA, only need to inherit the Test Adapter type, and rewrite some methods according to different mechanisms of SUT.

The realization of CD is similar to TA. TTworkbench contains a Abstract Base Code type, and it realizes two very important methods encode and decode used for sending and receiving data of codec, respectively, and make it become comprehensible format for SUT and testing system. Therefore, in the realization of CD can inherit Abstract Base Codec, and rewrite encode and decode methods according to the sent and received data.

Simulating access of concurrent users is the key to load testing. Before the test, I made an investigation of existing popular automatic load testing software (including commercial and open source), and installed Load Runner, No-Load and Meter three kinds of tools. In use process, I found that testing mechanism of these three kinds of software is the same: they will record client operation, and store them in script, then establish several virtual users and simulating scene of hundreds of virtual users

operating at the same time, and record response time of solving each matter, resource use of middleware server, database loaded; according to these test results can analyze where bottlenecks are. The above process is "recorded playback". Generally, using this simulation test tools will have a background processing process, and through this agent process test tools can monitor and get communication information of system users and server in various communication protocols; test tools might use a type C or other scripting language to produce a test script, this script records request process of customer to server, and then test tools can replay the visit process and receive server response.

2.2   Overall Structural Design

The ultimate purpose of this study is to evaluate TTCN - 3, and test is just a necessary measure. This means that the test can reflect characteristics of Web using software test. So, I chose common "submit - confirmation (XML format)" function in Web application software as test object of load testing.

In terms of simulating concurrent user's access, needs to make some improvements to record and playback modes: first, interaction between client and server in test is relatively simple and has been settled before; so test cases (test case) defined in TE will describe user's access to the server. Concurrent user needs to be realized in TTCN-3 configuration by using Parallel Test Component (PTC) .As adaptive middleware, TA will realize interaction between different virtual users and server, and call methods in CD to finish conversion of data format. The working mechanism of TTCN – 3 is as shown in fig.1:
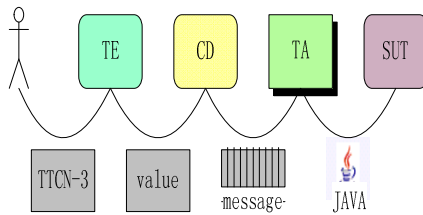
Fig. 1. Diagram of working mechanism of TTCN-3

Because of the test by using TTCN – 3 in this topic is realized through TTworkbench developed by TestingTech company, so first introduce an important entity -- Test Adapter (TA). Test adapter, which is similar to middleware characteristics, is written to fit different measuring system and testing system. TA realized some interfaces in TRI to fit different communications between measuring system and testing system. Actually, standard TTCN - 3 does not include the Taman TA is an entity defined by TestingTech company for supporting TTworkbench.But at present, TA has no universal realization; so the realization of TA is very important to testing.

3   Design and Realization of System

The working mechanism of 3 - TTCN   can simply described as test data written by testers in TE convert into a word throttling through decode method in CD ,and this word throttling being sent to receive port of connecting/mapping through send port defined in Texan receiving port give word throttling to corresponding TA of SUT; TA transform this word throttling into comprehensible format for SUT by calling decode method in CD; then SUT will compare received data with expecting data, if the same the test is passed, otherwise the test is a failure.

3.1   SUT

The subject of the system being tested is XMLServlet class which inherits Served and written by JAVA language and deployed in tomcat server of Apache; after client submitting page input user ID, click "Submit" button, and the XMLServlet class will establish HTTP connection to the server and send the user ID to server. XML Servelt class will get return information by using GET method and display it on client browser page in format of XML file.

3.2   TE

Define a constant as amount of scheduled concurrent users to facilitate changes.

Send data: because system requires users to enter user ID of a string type, so define the type for sending data as char string. Expectation to receive data: message returned by server displayed as XML file format, but message that get from testing system is check message of user ID, namely to return user ID to user for confirmation. Therefore, the expected type of definition is chartering .Template: defines a template, because the load testing focuses on load not on data.

Configuration: porting the system being tested interaction information between the user and the server are string types, which makes the port of the test system must   based on information(message-based); and the information have only two kinds of string and receiving string, so need only to define a port, and has the ability of two-way interaction (input).Component: defines the main test component and system test component, respectively, in which the main test component contains a timer as time limit of handling system.

Testing: test cases (testacies): firstly, declare a parallel test component (PTC) array as platform for interaction between each virtual users and server; then mapping this port in PTC and system test port to establish association. After that each PTC calls test behavior function for test and get the result.

Function: due to interaction between each virtual user and server is same, so defines a behavior function test to reduce write content of code, and also reflects the reusability of TTCN - 3 codes.

3.3   TA

For the visit of real concurrent users on simulation network, need to use thread mechanism the on TA, in which interaction   between each virtual user and the server is a separate thread, thus can realize concurrent of services .Concrete realization method is to define in TA a receiver Thread class which inherit Thread class, rewrite run method to make thread complete interaction activities between testing system and SUT, and directly organize receiver Thread class through rewriting of resend method to make it start. Actually, the originally written TA has no introduction of concept of thread, but depending on functions of SUT to realize resend method of Test Adapter class, and make test system can communicate with SUT and complete execution of test case. But graphical test log display testing process under this kind of circumstance is not true concurrent operation, because all return information return after all sending complete; obviously, this is a process of sequence execution. Therefore, to put forward the method of using thread to realize visit of concurrent user .After the modification of the TA, obtaining graphical test log display, and information transmission   between test system and SUT happening in cross , which accord with the reality, and is a real simulation of the scene of visit from concurrent users.

3.4   CD

Due to send data is char string type, which belong to the basic types, so need not to rewrite encode method. Although return data is also chartering type, but due to return data is content labeled by XML document, so it must be drawn from the XML document, and compare it with the expected data to determine the success or failure of the test; so must rewrite decode method. In view of that the origin of returned data is XML format file, adding jdom_b9.jar in class library, judo is a special tool for analyze XML format data, and use it can easily realize data extraction.

   To sum up, structure of load testing is shown as below:
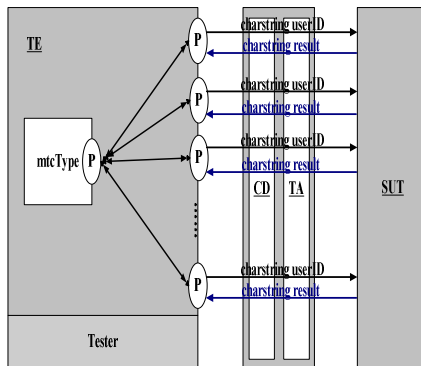
Fig. 2. Overall structure of load test

4   Test

Test starting, the system will dynamically establish 15 parallel testing components, of each test component will run a test case, and judge according to results of data matching to justify if the whole test is success or not. From the graph, we can see that the matching results executing on every test component is correct, and so the whole testing operation is a success.

5   Conclusions

Overall, the load testing is a relatively simple test, its SUT, test system and the scale of test scripts are relatively small, but due to the system being measured is universal system, which can reflect the feasibility of TTCN - 3 in load test. This paper is a research on testing method and technology of TTCN - 3, with analyzing characteristics and main points of using TTCN-3 to realize test; and obtain data from developing specific test items, experiencing different studies of TTCN-3 method and technology. Specific work includes the following content: extensive study of TTCN – 3 core language, and know well content and structure of TTCN - 3 test code written by TTCN – 3 core language, using TTCN - 3 test method and technology to develop experimental test program of load test, and understand the system of TTCN - 3test method. Based on test data and developing obtained experiences, and eventually made a conclusion that TTCN-3 is, in overall, suitable for Web application software testing, but also exists defects.

References

[1] ETSI ES 201 873-1 v3.1.1, Methods for Testing and Specification (MTS), The Testing and Test Control Notation version 3, Part 6: TTCN-3 Control Interface (TCI), France: ETSI (2005)
[2] Grabowski, J.: TTCN-3 – Testing and Test Control Notation
[3] Schieferdecker, I.: A TTCN-3 Test Platform for Reactive Systems
[4] Quality and TTCN-3, http://www.testingtech.de/ttcn3/quality.php
[5] Grabowski, J.: TTCN-3 – A new Test Specification Language for Black-Box Testing of Distributed Systems
[6] Schieferdecker, I., Grabowski, J.: Conformance Testing with TTCN

Design and Implementing of Database in Product Configuration Management System on Mass Customization

Yanling Zheng and Chunying Wang

Department of Electronic and Electrical Engineering, Baoji University of Arts and Sciences, Baoji, Shanxi 721016, China

Abstract. The main task of product configuration is that according to the rules of the configuration and limits of configuration conditions, checking parts searched products information database from for assembling, mutual constraint, etc, eventually outputting an effective Bill of Materials (BOM). In the paper, it is introduced how to create single-layer BOM in relational database. Based on object-oriented technology of relational database, model of data structure is constructed corresponds to the product configuration system and database design of the product configuration system is implementing. it is an important role for Product Configuration Management System on Mass Customization that the effective management of the product data, making full use of the present resources and rapidly realizing design of customized product, and provides reference for database design of similar system.

Keywords: Database design, Mass Customization, Relational database, Object-oriented Database, Product configuration.

1 Introduction

Modern manufacturing technology develops gradually the enterprise production mode into mode of diversification and individual demand. With the growth of individual demand, the enterprise's production and operation mode must be transformed from original independent production and sales into production according to customer's demand or production order, Mass Customization emphasizes using low cost instantly or fleetly in order to satisfy client's individual demand, and this product by Mass Customization must be provided for customers at the cost of mass production. In 1993, Pine discusses the theory of Mass Customization; he think Mass Customization will become the most important and the most competitive advantage of production mode in the 21st century[1].

Product configuration is an important method of achieving products diversification by designing and manufacturing new products, restructuring old products based on the product information management, it improves resource utilization of existing products, enhance capability of enterprise developing new product, and meet customer's diversification and individual demand. For the deficiencies of product configuration database module in the present commercialized software's system,

building a more intelligent and performance higher database management system, and strengthening product data and information management support, it is easy to realize data integration and exchange in many systems.

2   Storage Technology of Single-Layer BOM

Definition of product structure is that product structure trees are constructed by analyzing enterprise management mode and the hierarchical relationships between product components and structure. In product parts structure tree, product node is a root node, components compose product is the father node, and parts node located in the parent with the next layer node include all nodes leaf node of tree, which can also be derived down their child nodes to become parts node. How to use two data sheet describe the product structure based on single-layer BOM database storage, obviously, It is highly important that splitting the relation table from product tree structure for realizing BOM storage technology, On the contrary, reconstructing the product structure tree from the single-layer BOM is key for browsing and inquiring data.

1. Single-layer BOM

Single-layer BOM is different from multi-layer BOM[2]. Single-layer corresponding relation of adjacent parts is recorded in single-layer BOM, namely corresponding relation of the father and the son, and as multilayer BOM repeatedly increase a record of father-son relation. This structure of Single-layer BOM can reduce data redundancy, and improve data maintenance work. Single-layer BOM usually adopts two tables to record relation of product structure, database table structure are shown in Tab1,Tab2.

Table 1. Parts table

Parts coding	Version number	Parts name	Drawing number	Related properties	……
CX901	1	B	2	…	

Table 2. Product structure

Parent coding	Parent Version No.	Child coding	Child Version No.	Quantity	……
PX901	1	CX901	1	1	

Parts table in which attribute information of all the parts object are record is the basis of the whole product BOM table. Parts coding and its version number compose the only identification code of parts [3], which is the primary key of parts table and only identify parts in the whole product structure. Subsequent fields, such as parts

name, drawing number, ect, it record the auxiliary attributes of parts. In the BOM table structure, the parent coding, parent version, Childs coding and its Version number compose of primary code, which uniquely identify parts of different levels in the whole product structure. The same parts in different assembly relationship can be distinguished by the parent items and version number,   to avoid the data redundancy, parts with the same parent-child relationship in assembly are added up and not recorded repeatly in database. products are only child coding and child version number, its parent coding and version number are null. Therefore, single-layer BOM can describe the product tree structure. Connects operations is completed for Tab1and Tab2 in order to obtain all the parts information of product structure, then parts attribute information fill in the corresponding data table, so a BOM can be gotten.

The product database table obtained from disassembling graph of the product structure tree is that adverse use of structure tree reconfiguration method. The process of Reconstruction product structure tree is a process of tree traversal. There are two traversal methods: depth-first traversal and breadth first traversal, the former include the preorder traversal, midorder traversal and postorder traversal [4]. Different traversal method access the tree nodes according to different sequences, so the results the results are also different. Depth-first traversal access nodes according to the branch of a tree structure, returned when each branch visit to leaf nodes; Breadth-first traversal is in accordance with the tree depth (layer) in order to access node, it is that nodes are accessed from the zero depth node to gradually each layer node, if some nodes have the same depth, nodes are accessed from left to right. Considering the multi-level product structure and single-layer BOM storage properties. In this paper, method of breadth-first traversal is used to reconstruct product tree.

3   The Object-Oriented Technology Based on Relational Database

Mechanical product design process is a complicated, involving many disciplines, theoretical applications to realize many functions, such as design of Cycloidal Needle Wheel Decelerator function optimization, design of key parts structure optimization, and the fatigue life of important components in virtual experiment etc. Large amounts of data is made in be using series of CAD/CAM/CAE tools for pursuing these functions, such as 2D or 3D models of parts, engineering drawing, various of important parts data, series data of product configuration design etc. In order to realize the functions of product configuration design, it should is first that the data structure of design created, which correspond to configuration system in PDM system.

Class and subclass: Product configuration design system is a application software of product-oriented projects, in which all data is organized around products, projects, and display as a tree links. Tree links is internal links of all information based on PCD system database. it could be very easy to browsing data.

Hierarchy Structure of classes: Each class in this system has its own attributes, and display in the property card[5]. Classes of lower layers inherit all the properties of parent classes, but also they can have their own attributes. Therefore, the more are

layer from the highest layer class the data structure of the to the lower level class, the more are detailed class information.

4   The Object-Oriented Technology Based on Relational Database

In product configuration management system, all the data can be regarded as objects, such as a project, a product, a parts, etc. then, according to the requirements of enterprises or customer, defined levels of those data.
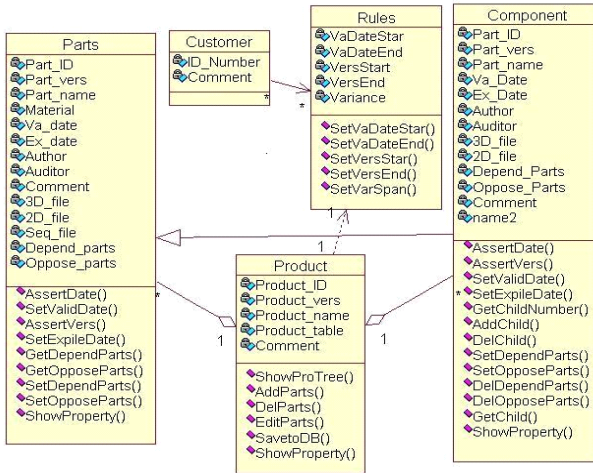
Fig. 1. Objects relationship in product configuration system

Based on object-oriented technology, there are five objects in configuration system by analysis, such as part objects, component objects, product objects, rule objects and customer object. Relationships between objects is establuished by using UML (The Unified Modeling Language), it is shown in figure1. UML is object-oriented modeling method widely used, it can quickly establish object models of various system, and distinguish the relation and interaction between them, finally, generating software code framework. Symbol " ⟶ "represents the composition relationships of the overall and individual in aggregation, for example, part objects and product objects, as shown in Fig 1. Arrow point from individual to overall and indicate quantity relationship, "1" denote exactly corresponding relation, "*" denote multiple quantity corresponding relations. Symbol " ----⟶ " represents dependencies relations, such as prduct objects and rule objects in Fig1. The arrow " ⟶ " represents association relationship, like Customer and Rules objects in figure1. Hollow arrow " ⟶ " represents generalization inherit relationships, Component object will inherit most of the attributes.part objects.

5   The Structure of Database Tables and Relationships

There are five tables in all products information characteristics table, namely as: Part_table, File_table, Relation_table, Product_table and Product_list.

In order to smoothly complete the product configuration design system based on the knowledge base and rule base, we must also create the following database tables: Products Matching Rules Form (ConfigRule), the variable value table-definedV variable conditions of table, definedVC-user and user rights table, which reflect the object association between the relational tables. The relationship between database tables is shown in Figure 2. PK represents Primary Key of data tables, FK represents Foreign Key of data tables, The arrow " $\longrightarrow$ " indicate dependencies relationship between tables, and n and 1 denote   corresponding quantity relationships   between the various tables tables in database.
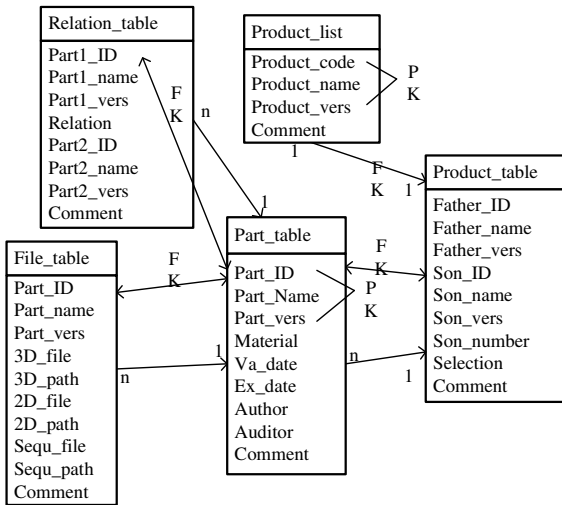
Fig. 2. The structure and relationships of database tables

6   Summaries

In this paper, single-layer BOM table creation is Implemented, relationships between objects is estabulished by using UML, and create database tables of product information in database of product configuration management system on Mass Customization. Based on object-oriented technology of relational database, database design of the product configuration system is implementing and database constructed by meet user requirements. it is an important role for Product Configuration Management System on Mass Customization that the effective management of the product data, making full use of the present resources and rapidly realizing design of customized product, and provides reference for database design of similar system.

References

[1] Pine, J.B.: Mass Customization: The New Frontier in Business Competition, pp. 86–88. Harvard Business School Press, Boston (1992)
[2] Deng, L., Zan, X., Wu, L., Wei, H., Mao, L.: A Product BOM Management Method Based on Single Layer BOM. Journal of Chongqing University (Natural Science Edition) 25(6), 5–7 (2002)
[3] Ping, Y.C., Tsung, T.C.: Analysis of assembly through product configuration. Computer in Industry 44, 189–203 (2001)
[4] Preiss, B.R.: Data Structures and Algorithms with object-oriented design patterns in C++, pp. 113–215. Electronic Industry Press, Beijing (2003)
[5] Huang, L.S., Chen, H.P., Zheng, Q.L., et al.: A New Dynamic Data Model for Object-Oriented Database Systems. Journal of Software 12(5), 735–741 (2001)

Study on Edge Extraction of Objects in Microscopic Stereo Matching

Junming Chang, Jun Tao, and Bin Yan

School of Mathematics and Computer Science, Jianghan University, Wuhan,
Hubei, P.R. China 430056

Abstract. Based on traditional approaches, there is much trouble in edge extraction of objects in microscopic stereo matching, such as not thinning to one pixel, breaking the connectivity of ridge, generating spikes and so on. In order to overcome these disadvantages, a simple and effective method is presented. By analyzing character of node, some templates are proposed. This method can effectively solve the less thorough thinning problem. Also the proposed edge linking approach is good at extracting the complete edge. The algorithm provides the basis for later microscopic stereo matching.

Keywords: Thinning algorithm, templates, microscopic stereo matching, edge extraction.

1 Introduction

A large number of feature extractions are specially prepared for matching in the stereovision. Image edge is one of the characteristic primitives, which is the most widely used in stereo matching. Image edge has a large number of useful information of the original image, such as direction, step character, shape, etc. These can reflect the characteristics of objects [1]. After the microcosmic object is imaged by the microscope, the microcosmic images from CCD are different from the normal images. The microcosmic image is relatively fuzzy with the low contrast. In microscope imaging process, the image is easily with noise and shadow because of light uneven. So the feature is difficult to be extracted out and fault pixel is easy to be generated up, with much edge spikes in the extracted objects.

The edge of the detection is often used the airspace differential operator. At present, the edge inspection operator has: Sobel operator, Roberts operator, Laplacian operator, Prewitt operator, Canny operator, etc [2]. These operators are primarily used in so images with obvious edges or simple pictures. Most of the extraction algorithms can achieve better results [3]. However, for the edge of the complex, lighting the microscopic image, the effect is not very ideal. For example, the edge is fuzzy, is not one pixel width, is too poor to be lost, and is whole inconsecutive [4]. These defects in three-dimensional microscopic match will appear in the characteristics of the right picture can't find the phenomenon of miss match, which is unable to meet the target on the location accuracy, brevity, the outline with the requirements of the real-time.

With the above analysis, the paper proposes a thinning method based on template which is used to resolve one pixel edge, and a smart tactic for the connection of the inconsecutive edges [5].

2   Edge Thinning

2.1   Edge Thinning Algorithm

The further refinement of all detected edges will not only help reduce the amount of image storage, but also adapt as the endpoint, crossing points and the relationship between the sub-critical parts of the connection feature extraction. To ensure the extraction of the edge of the extreme points of its neighborhood to remove the edge of the isolation, this paper used the method of morphology in the bone thinning edges. Expression as follows:

$$S(A) = \overset{k}{\underset{k=0}{U}} S_k(A)$$
$$S_k(A) = (A\Theta kB) - (A\Theta kB) \bullet B$$

A test that has been one of the edges, B said that the structure of 3×3 elements, said continuous on A k-corrosion, k is the corrosion of the empty set A is the number of iterative, said the results of the first k-refinement, that the final Obtained detailed edge.

2.2   Edge Thinning Optimization

On the one hand, edge thinning leads to the emergence of new glitches, new cross-point and end point; on the other hand, will exist at the intersection of the redundant pixels, which is the non-single-pixel. This phenomenon of cross-edges like the "T" or "+" character often encountered, and it is difficult to define nodes and points and it takes considerable trouble for the feature point extraction. Microscopic three-dimensional image of each match is small but not negligible differences. Intersection point of non-single-pixel processing, and glitches edge connector is removed the key. The lines in Figure 1 marked the crossing point after the refining of non-single pixel. Used for edge connectivity with the premise of removing redundant pixel method, used in the microscopic three-dimensional template matching method is more appropriate. To remove these non-single-pixel points, the paper uses a template approach. The non-single pixel in Figure 1 can be used to remove the template in Figure 2, that is to remove the image in the template with the location of the structure in Figure 2, the same point, * indicates the current position of points, 1 point on the edge of the target image, 0 points on the background image.

Fig. 1. The refinement of non-single-pixel in the crossing point

0	0	0
1	*1	1
0	1	0

0	1	0
0	*1	1
0	1	0

0	1	0
1	*1	1
0	0	0

0	1	0
1	*1	0
0	1	0

Fig. 2. The template of non-single-pixel without refinement

3   Edge Connection

3.1   Connection Algorithm

After the above refinement, the endpoint can be defined as the edge itself and there is only one neighborhood of 8 pixels, is a starting point and end point edge. Node is defined as the one in the itself edge, and 8 in the neighborhood of two or more pixels. There is a connection point between the two edges. The key point in the edge connector is determined primarily on the node and endpoint determination.

Edge connection is divided into the following steps:

①   Marking the starting point: for each edge pixel, the first to determine the key points, and to mark them.

②   Searching the edge: from each point on the field in its 8 non-zero search, and for the same mark, until you reach the end node; recorded in the course of each endpoint to traverse the length of the node.

③   Connecting edges: all belong to the same edge of the line, that is, connect the dots marking the same to form a continuous edge.

④   Removing burr: Comparison of 4 for each node adjacent to the edge direction, and then take a traverse point threshold, the branch will be removed less than the threshold, the connection of other branches to form the edge.

3.2   Edge Refinement

After treatment, the refinement of the previous image is a single-pixel image, but after removing the burrs, the node appeared in a number of non-single pixels. Figure 3 shows that the burr is present after removal of non-single pixel. Of such non-removal of a single pixel in the same way as the previous refinement, only the use of different templates. Figure 4 is to remove such non-single-pixel template in the Figure 3.

Fig. 3. The non-single-pixel of deburring

0	0	0		0	0	0		0	0	0		1	0	0
0	*1	0		0	*1	0		1	*1	0		1	*1	0
0	1	1		1	1	0		1	0	0		0	0	0

0	1	1		1	1	0		0	0	1		0	0	0
0	*1	0		0	*1	0		0	*1	1		0	*1	1
0	0	0		0	0	0		0	0	0		0	0	1

Fig. 4. The template of non-single-pixel by deburring

4   Experimental Example

The edge extracted from the microscopic image in the lab is shown in the Fig. 5. The line segment is identified as the intersection with amplified effect. The non-single-pixel based on the proposed template in the Fig. 4 is shown in the Fig. 6. Line is marked the removal of the pixel. Figure 7 and Figure 8 are processed through the algorithm a continuous, single-pixel image edges.
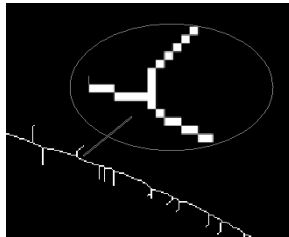
Fig. 5. The node with non-single-pixel

Fig. 6. The removement of non-single-pixel

Fig. 7. The edge with nodes and spikes
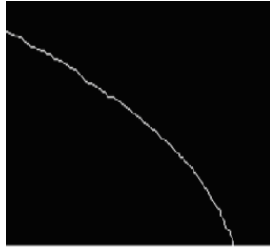
Fig. 8. The edge with inconsecutive pixel points

5   Conclusions

Image edge is one of the characteristic primitives, which is the most widely used in stereo matching. Stereo matching of edge detection has a higher performance requirement. This article has been considered in the discussion of edge detection operator microscopic image edge extraction problems, a new approach. By refining and remove the burr on the edge of the non-generated comparison of a single pixel, the design of two different types of templates. Experiments show that using these templates can effectively solve the problem of refinement is not complete. Meanwhile, the paper presents an edge connection method, in the edge of the connection process and eliminates glitches. This paper makes a good foundation for the latter part of the stereo matching.

References

[1] Sonka, M., Hlavac, V., Boyle, R.: Image processing analysis and machine Vision, pp. 385–392. Posta & Telecom Press, Beijing (2003)
[2] Worthington, P.L.: Enhanced Canny edge detection using curvature consistency. In: Proceedings International Conference on Pattern Recognition, pp. 596–599. IEEE, Computer Soc. Press, Los Alamitos (2002)

[3]   Liu, Y., Gou, Q.: An Algorithm of Selecting Feature Points Indirectly in an Outline. Computer Engineering and Applications 16(1), 244–256 (2004)

[4]   Tao, J.: The Accurate 3D Object Reconstruction Based on the Camera-Projector System, 2004 Doctoral Forum of China(Beijing University), Beijing, China (2004)

[5]   Wang, Y.-Z.: Study on Edge Extraction of Objects in Microscopic Images. Minimicro Systems 24(12), 2177–2184 (2003)

[6]   Phillips, J.: Comments - The Thin Edge of the Wedge. European Intellectual Property Review 27(1), 31–34 (2005)

[7]   Wen, J.-G., Xiao, Q., Xu, H.-P.: Study on edge-extraction of remote sensing image. Uranium Geology 21(1), 59–65 (2005)

[8]   Espinosa-Duro, V.: Fingerprints thinning algorithm. IEEE Aerospace and Electronic Systems Magazine 18(9), 28–58 (2003)

[9]   Espinosa-Duro, V.: Fingerprint thinning algorithm. Aerospace and Electronic Systems Magazine 18(9) (September 2003)

[10] Tao, J., Zhang, J., Zhang, Z.: Calibration of a Projector with a Planar Grid. In: ISPRS, Istanbul, Turkey (2004)

A Certain Asymptotically Strict Pseudocontractive Mappings in the Intermediate Sense Semigroup

Wang Erli, Duan Qibin, Wu Dingping, and Zhao Hang

College of Mathematics, Chengdu University of Information Technology
Chengdu, Sichuan 610225, P.R. China
wdp68@163.com

Abstract. It is studied that the asymptotically kn-strict pseudocontractive mapping in the intermediate sense and proved the modified Mann iteration process with errors. Then the iteration converges strongly to a common fixed point p which is the nearest point to u in F. The results in this paper presented extend and improve the corresponding results of many authors.

Keywords: Asymptotically kn-strict pseudocontractive mapping in the intermediate sense, Mann iteration process with errors, Semigroup, Lipschitzian.

1 Introduction and Preliminaries

Throughout the paper, H be a real Hilbert space and R^+ denote the set of positive real number, with inner product $<\cdot,\cdot>$ and norm $\|\cdot\|$, respectively. C a closed convex subset of H. Let $T: C \rightarrow C$ be a mapping, we use $F(T)$ to denote the set of fixed point of the mapping T.

Definition 1.1 (1) The one-parameter family $\Im := \{T(t) : t \geq 0\}$ of mappings from C into itself is called a nonexpansive semigroup if the following conditions are satisfied:

(a) $T(0)x = x$ for each $x \in C$;

(b) $T(t+s)x = T(t)T(s)x$ for any $x \in C$ and $t, s \in R^+$;

(c) for any $x \in C$, the mapping $t \rightarrow T(t)x$ is continuous;

(d) for any $x, y \in C$, there exists $j(x+y) \in J(x+y)$ such that

$$\|T(t)x - T(t)y\| \leq \|x - y\| \tag{1.1}$$

for any $t \geq 0$.

(2) The one-parameter family $\Im$ of mappings from C into itself is called an asymptotically k_n – strict pseudocontractive mapping in the intermediate sense with sequence $\{\gamma_n\}$ semigroup and the conditions (a)-(c) and the following conditions (e) are satisfied:

(e) there exists a $\{k_n\} \subset [0,1)$ with $\lim_{n\to\infty} \sup k_n < 1$, $\{\gamma_n\}$, $\{c_n\} \subset [0,1)$ with $\{k_n\} \subset [0,1)$ $\lim_{n\to\infty} \gamma_n = \lim_{n\to\infty} c_n = 0$ such that

$$\left\| T^n(t)x - T^n(t)y \right\|^2 \le (1+\gamma_n)\left\| x-y \right\|^2 + k_n \left\| x - T^n(t)x - \left(y - T^n(t)y \right) \right\| + c_n \qquad (1.2)$$

for all $x, y \in C$, $t \ge 0$ and $n \in N$.

Remark 1.2([3]). Let C be a nonempty subset of a Hilbert space H, $T : C \to C$ asymptotically k_n – strict pseudocontractive mapping in the intermediate sense with sequence $\{\gamma_n\}$. Then

$$\left\| T^n(t)x - T^n(t)y \right\| \le \frac{1}{1-k_n}\left(k_n \left\| x-y \right\| + \sqrt{\left(1 + (1-k_n)\gamma_n\right)\left\| x-y \right\|^2 + (1-k_n)c_n} \right)$$

for all $x, y \in C$ and $n \in N$.

The convergence problems of the iterative sequences for nonexpansive semigroup to a common fixed point have been considered by some authors in the settings of Hilbert spaces [4, 7].

In 2009, Zhang, Yang, Li and Chen[7] introduce and study the strong theorem of nonexpansive semigroup the following iteration in Hilbert spaces:

$$\begin{cases} x_0 \in C, \\ x_{n+1} = \alpha_n u + (1-\alpha_n)T(t_n)x_n, & n \ge 0. \end{cases}$$

Recently, Zeng [7] study the strong convergence problem of the Ishikawa Iterative of error correction for the Lipschitzian and asymptotically pseudocontractive mappings in arbitrary real Banach space. Zhang [12], by using different methods, introduce and study the weak or strong convergence in general Banach spaces.

On the other hand, the class of asymptotically mappings in the intermediate sense was introduced by Bruck, Kuczumow and Reich[6] and iterative methods for the approximation of fixed points of such types of non-Lipschitzian mappings have been studied by Agarwal, O'Regan and Sahu[2] and many others. Recently, D. R. Sahu, Hong-Kun Xu, Jen-Chih Yao [1] introduced the concept of asymptotically k – strict pseudocontractive mapping in the intermediate sense, Wu [3] study the mapping with k_n to extend the mapping.

Inspired by the authors above, the purpose of this paper is to introduce and study the strong converges problem of the following modified Mann iteration process with errors:

$$\begin{cases} x_0 \in C \\ x_{n+1} = (1-\alpha_n)x_n + \alpha_n T^n(t_m)x_n + \varepsilon_n, & \forall n \in N \end{cases}$$

The following Lemmas are important to prove results in this paper.

Lemma 1.3([9]). Assume $\{a_n\}$ is a sequence of nonnegative real numbers such that

$$a_{n+1} \le (1-\gamma_n)a_n + \delta_n, \qquad n \ge 0.$$

Where $\{\gamma_n\}$ is a sequence in $(0,1)$ and $\{\delta_n\}$ is a sequence in R such that

(1) $\sum_{n=1}^{\infty}\gamma_n = \infty$;

(2) $\limsup_{n\to\infty}\delta_n/\gamma_n \le 0$ or $\sum_{n=1}^{\infty}|\delta_n| < \infty$.

Then $\lim_{n\to\infty}a_n = 0$.

Lemma 1.4([5]). Let $\{a_n\}$, $\{\beta_n\}$ and $\{\gamma_n\}$ be three sequences of nonnegative numbers satisfying the recursive inequality

$$a_{n+1} \le \beta_n a_n + \gamma_n, \qquad \forall n \in N.$$

If $\beta_n \ge 1$, $\sum_{n=1}^{\infty}(\beta_n - 1) < \infty$, and $\sum_{n=1}^{\infty}\gamma_n < \infty$. Then $\lim_{n\to\infty}a_n$ exists.

Lemma 1.5([3]). Let C be a nonempty subset of a Hilbert space H , $T:C\to C$ a uniformly continuous asymptotically k_n – strict pseudocontractive mapping in the intermediate sense with sequence $\{\gamma_n\}$. Let $\{x_n\}$ be a bounded sequence in C such that $\|x_n - x_{n+1}\| \to 0$ and $\|x_n - T^n x_{n+1}\| \to 0$ $(n\to\infty)$.
Then $\|x_n - T^n x_{n+1}\| \to 0$ as $n\to\infty$.

Lemma 1.6([7]). Let E be a arbitrary real Banach space, E^* is the dual space of E , $J:E\to 2^{E^*}$ is the normalized duality mapping defined by

$$J(x) = \left\{f \in E^* : \langle x, f\rangle = \|x\|\|f\|, \|x\| = \|f\|\right\}.$$

For all $x\in E$, we have

$$\|x+y\|^2 \le \|x\|^2 + 2\langle y, j(x+y)\rangle, \quad \forall j(x+y)\in J(x+y).$$

Especially, if E be a Hilbert space, then $J = I$. Thus we have

$$\|x+y\|^2 \le \|x\|^2 + 2\langle y, x+y\rangle$$

for all $x, y \in E$.

Lemma 1.7([7]). Let H be a real Hilbert space, C a closed convex subset of H , for any $x\in H$ and $u\in C$, we have

(1) $\langle z - P_c x, P_c x - x\rangle \ge 0$, $\forall z\in C$;

(2) $\langle z - u, u - x\rangle \ge 0$, $\forall z\in C$, then $u = P_c x$,

where $P_c x$ is the nearest projection from H to C .

2   Main Result

Now, we give our main results in this paper.

Theorem 2.1. Let C be a nonempty closed convex subset of a Hilbert space H, $T:C \to C$ a continuous asymptotically k_n – strict pseudocontractive mapping in the intermediate sense with sequence $\{\gamma_n\}$. Then $F(T)$ is closed and convex and $I-T$ is demiclosed at zero.

The proof can be found in [3].

By using Theorem 2.1, we have the following:

Theorem 2.2. Let C be a nonempty closed convex subset of a Hilbert space H, $T \in \mathfrak{I}$ of mapping from into C itself be a Lipschitzian and uniformly continuous asymptotically k_n – strict pseudocontractive mapping in the intermediate sense with sequence $\{\gamma_n\}$ semi-group. If $F(T(t)) \neq \varnothing$, u is a given point in C.

Assume that $u_n \subset C$, $\{\alpha_n\} \subset (1/2,1)$ and $\{t_m\}$ is an increasing sequence in $[0,\infty)$. $\{x_n\}$ be defined:

$$x_{n+1} = (1-\alpha_n)x_n + \alpha_n T^n(t_m)x_n + \varepsilon_n, \quad \forall n \in \mathbb{N}$$

If the following conditions are satisfied:

(ⅰ) $\sum_{n=1}^{\infty} \gamma_n < \infty$, $\sum_{n=1}^{\infty}\|\varepsilon_n\| < \infty$ and $\sum_{n=1}^{\infty} \alpha_n c_n < \infty$;

(ⅱ) $\lim_{m \to \infty} \sup_{x \in C, s \in R^+} \|T(s)T(t_m)x - T(t_m)x\| = 0$.

Then $\{x_n\}$ converges strongly to a common fixed point p which is the nearest point to u in F.

Proof: (1) Firstly, we prove that $\lim_{n \to \infty}\|x_n - p\|$ exists.

For any given $p \in F$, we have

$$\|x_{n+1} - p\|^2 \le (1-\alpha_n)\|x_n - p\|^2 + \alpha_n c_n + 2\langle (1-\alpha_n)(x_n - p) + \alpha_n(T^n(t_m)x_n - p), \varepsilon_n \rangle$$

$$+\alpha_n\left[(1+\gamma_n)\|x_n - p\|^2 + k_n\|x_n - T^n(t_m)x_n\|^2\right] - \alpha_n(1-\alpha_n)\|T^n(t_m)x_n - x_n\|^2 + \|\varepsilon_n\|^2. \tag{2.1}$$

And

$$2\langle (1-\alpha_n)(x_n - p) + \alpha_n(T^n(t_m)x_n - p), \varepsilon_n \rangle \le \|(1-\alpha_n)(x_n - p) + \alpha_n(T^n(t_m)x_n - p)\|^2 + \|\varepsilon_n\|^2$$

$$\le (1+\gamma_n\alpha_n)\|x_n - p\|^2 + \alpha_n c_n - \alpha_n(1-k_n-\alpha_n)\|T^n(t_m)x_n - x_n\|^2 + \|\varepsilon_n\|^2. \tag{2.2}$$

From (2.1) and (2.2), we have

$$\|x_{n+1} - p\|^2 \le 2(1+\gamma_n)\|x_n - p\|^2 + 2\left(\alpha_n c_n + \|\varepsilon_n\|^2\right). \tag{2.3}$$

By condition (ⅰ) and Lemma 1.4, we know $\sum_{n=1}^{\infty}\|\varepsilon_n\| < \infty$, $\sum_{n=1}^{\infty} \alpha_n c_n < \infty$.

Thus $\lim_{n \to \infty}\|x_n - p\|$ exists, so $\{x_n\}$ be bounded.

(2) Next we show

$$\lim_{n\to\infty,m\to\infty}\left\|T\left(t_m\right)x_n-x_n\right\|=0.$$

From Remark 1.2, we easy to know

$$\left\|T^n\left(t_m\right)x_n-p\right\|\le\frac{1}{1-k_n}\left(k_n\left\|x_n-p\right\|+\sqrt{\left(1+\left(1-k_n\right)\gamma_n\right)\left\|x-y\right\|^2+\left(1-k_n\right)c_n}\right)\le\mathsf{M}$$

for some certain M.

By (2.1), we have

$$\left\|x_{n+1}-p\right\|^2\le\left(1+\gamma_n\alpha_n\right)\left\|x_n-p\right\|^2+2\left\|\varepsilon_n\right\|\left[\left(1-\alpha_n\right)\left\|x_n-p\right\|+\alpha_n\left\|T^n\left(t_m\right)x_n-p\right\|\right]$$
$$-\left\|x_{n+1}-p\right\|^2+\alpha_nc_n+\left\|\varepsilon_n\right\|^2.$$

Since $\left\{\left\|T^n\left(t_m\right)x_n-p\right\|\right\}$ be bounded and $\lim_{n\to\infty}\left\|x_{n+1}-p\right\|=\lim_{n\to\infty}\left\|x_n-p\right\|$ exists, and $\lim_{n\to\infty}\left\|\varepsilon_n\right\|=\lim_{n\to\infty}\gamma_n=0$.

We observe that

$$\lim_{n\to\infty,m\to\infty}\left\|T^n\left(t_m\right)x_n-x_n\right\|=0. \tag{2.4}$$

This implies that

$$\left\|x_{n+1}-x_n\right\|=\left\|\left(1-\alpha_n\right)x_n+\alpha_nT^n\left(t_m\right)x_n+\varepsilon_n-x_n\right\|=\left\|\alpha_n\left(T^n\left(t_m\right)x_n-x_n\right)+\varepsilon_n\right\|$$
$$\le\alpha_n\left\|T^n\left(t_m\right)x_n-x_n\right\|+\left\|\varepsilon_n\right\|$$
$$\to0 \tag{2.5}$$

From (2.4), (2.5) and Lemma 1.5, we know that T is uniformly continuous, and by Theorem 2.1. Thus, we obtain that

$$\lim_{n\to\infty,m\to\infty}\left\|T\left(t_m\right)x_n-x_n\right\|=0.$$

(3) We prove that $\lim_{n\to\infty}\left\|T(t)x_n-x_n\right\|=0$, for all $t\ge0$.

In fact, it follows from the condition (ii) and step (2) that, for any $t\ge0$,

$$\left\|T(t)x_{n_k}-x_{n_k}\right\|\le(1+L)\left\|x_{n_k}-T\left(t_{n_k}\right)x_{n_k}\right\|+\left\|T\left(s+t_{n_k}\right)x_{n_k}-T\left(t_{n_k}\right)x_{n_k}\right\|$$
$$\le(1+L)\left\|x_{n_k}-T\left(t_{n_k}\right)x_{n_k}\right\|$$
$$+\sup_{z\in\{x_n\},s\in R^+}\left\|T\left(s+t_{n_k}\right)z-T\left(t_{n_k}\right)z\right\|\to0\ \ (as\ \ n_k\to\infty). \tag{2.6}$$

Since $x_{n_k}\to p$ as $n_k\to\infty$ and $\lim_{n\to\infty}\left\|x_n-p\right\|$ exists, which implies that $x_n\to p\in F$ as $n\to\infty$, where p is any given point in F.

(4) Now we prove

$$\limsup_{n\to\infty}\left\langle x_n-P_Fu,u-P_Fu\right\rangle\le0. \tag{2.7}$$

With the boundedness of $\left\{x_n\right\}$ and $\left\{T\left(t\right)x_n\right\}$, so $\left\langle T\left(t\right)x_n-P_Fu,u-P_Fu\right\rangle$ bounded.

Then we know $\limsup_{n\to\infty}\{\langle x_n - P_F u, u - P_F u\rangle\}$ exists. There exists a subsequence $\{x_{n_j}\} \subseteq \{x_n\}$, and for a certain $q \in C$,

$$T(t)x_{n_j} \to q. \tag{2.8}$$

From step (3), we know that

$$q \in F := \bigcap_{t\geq 0} F(T(t)).$$

By Lemma 1.7 and (2.8)

$$\limsup_{n\to\infty}\left\langle T(t)x_n - P_F u, u - P_F u\right\rangle = \lim_{j\to\infty}\left\langle T(t)x_{n_j} - P_F u, u - P_F u\right\rangle$$

$$=\left\langle q - P_F u, u - P_F u\right\rangle \leq 0. \tag{2.9}$$

And from step (3), we have

$$\limsup_{n\to\infty}\left\langle x_n - P_F u, u - P_F u\right\rangle \leq 0. \tag{2.10}$$

(5) Finally, we prove $x_n \to P_F u \in F := \bigcap_{t\in R^+} F(T(t))$, where $P_F u$ is the fixed point which is nearest to u in $F(T(t))$.

In fact, by Lemma 1.6, we have

$$\left\|x_{n+1} - P_F u\right\|^2 \leq \left(1-\alpha_n\right)^2 \left\|x_n - P_F u\right\|^2 + 2\alpha_n\left\langle T^n(t_m)x_n - P_F u, x_{n+1} - P_F u\right\rangle$$
$$+2\left\langle \varepsilon_n, x_{n+1} - P_F u\right\rangle$$

$$=\left(1-\alpha_n\right)^2\left\|x_n - P_F u\right\|^2 + 2\alpha_n\left[\left\langle T^n(t_m)x_n - p, x_{n+1} - P_F u\right\rangle + \left\langle p - x_{n+1}, x_{n+1} - P_F u\right\rangle + \left\langle x_{n+1} - P_F u, x_{n+1} - P_F u\right\rangle\right]$$
$$+2\left\langle \varepsilon_n, x_{n+1} - P_F u\right\rangle.$$

Simplify the formula above

$$\left\|x_{n+1} - P_F u\right\|^2 \leq \left(1-\frac{\alpha_n^2}{2\alpha_n - 1}\right)\left\|x_n - P_F u\right\|^2$$

$$+\frac{2\alpha_n}{1-2\alpha_n}\left[\left\langle T^n(t_m)x_n - x_n, x_{n+1} - P_F u\right\rangle + \left\langle x_n - x_{n+1}, x_{n+1} - P_F u\right\rangle\right] + \frac{2}{1-2\alpha_n}\left\langle \varepsilon_n, x_{n+1} - P_F u\right\rangle. \tag{2.11}$$

Let $\qquad a_{n+1} = \left\|x_{n+1} - P_F u\right\|^2 \qquad , \qquad \gamma_n = \alpha_n^2/2\alpha_n - 1 \qquad ,$

$$\delta_n = \frac{2\alpha_n}{1-2\alpha_n}\left[\left\langle T^n(t_m)x_n - x_n, x_{n+1} - P_F u\right\rangle + \left\langle x_n - x_{n+1}, x_{n+1} - P_F u\right\rangle\right] + \frac{2}{1-2\alpha_n}\left\langle \varepsilon_n, x_{n+1} - P_F u\right\rangle.$$

By Lemma 1.3, we can obtain $\{\gamma_n\}$ is a sequence in $(0,1)$ and $\{\delta_n\}$ is a sequence in R.

Also, one hand, we can easy to find that $\{\gamma_n\}$ satisfy the condition $\sum_{n=1}^{\infty}\gamma_n = \infty$.

On the other hand, from (2.4), (2.5) and condition (i), we have

$$\limsup_{n\to\infty} \frac{2\alpha_n[\langle T^n(t_m)x_n - x_n, x_{n+1} - P_{Fu}\rangle + \langle x_n - x_{n+1}, x_{n+1} - P_{Fu}\rangle] + 2\langle \varepsilon_n, x_{n+1} - P_{Fu}\rangle}{-\alpha_n^2} \le 0.$$

So we obtain $\|x_n - P_F u\| \to 0$, namely

$$x_n \to P_F u \in F := \bigcap_{t\in R^+} F(T(t)).$$

This completes the proof.

References

[1] Sahu, D.R., Xu, H.-K., Yao, J.-C.: Asymptotically strict pseudocontractive mappings in the intermediate sense. Nonlinear Anal. 70, 3502–3511 (2009)

[2] Agarwal, R.P., O'Regan, D., Sahu, D.R.: Iterative construction of fixed points of nearly asymptotically nonexpansive appings. J. Nonlinear Convex Anal. 8(1), 61–79 (2007)

[3] Wu, D.: Asymptotically k_n- strict pseudocontractive mappings in the intermediate sense (in press)

[4] Li, S., Li, L.H., Su, F.: General iterative methods for one-parameter nonexpansive semi-group in Hilbert space. Nonlinear Anal. (2008) doi:10.1016/j.na.2008.04.007

[5] Osilike, M.O., Aniagbosor, S.C.: Weak and strong convergence theorems for fixed points of asymptotically nonexpansve mappings. Math. Comput. Model. 32, 1181–1191 (2000)

[6] Goebel, K., Kirk, W.A.: A fixed point theorem for asymptotically noexpansive mappings. Proc. Amer. Math. Soc. 35(1), 171–174 (1972)

[7] Zhang, S.S., Yang, L., Lee, J., Chan, C.K.: On the strong convergence theorems for nonexpansive semi-group in Hilbert spaces. Acta Math., Sinica, Chinese Series 52, 337–342 (2009)

[8] Ming, X., et al.: Iterative approximation problem of fixed points for uniformly L-Lipschitzian and asymptotically pseudocontractive mappings. J. Sys. Sci. & Math. Scis. 30(9), 1206–1213 (2010)

[9] Xu, H.K.: Iterative algorithms for nonlinear operators. J. London Math. Soc. 66, 240–256 (2002)

[10] Suzuki, T.: On strong convergence to a common fixed point of nonexpansive semigroup in Hilbert spaces. Proc. Amer. Math. Soc. 131, 2133–2136 (2003)

[11] Zeng, L.: On the strong convergence of iterative method for non-lipschitzian asymptotically pseudocontractive mappings. Acta Math. Appl. Sinica 27(3), 230–239 (2004) (in Chinese)

[12] Zhang, S.S.: Weak convergence theorem for Lipschitzian pseudo-contraction semi-groups in Banach spaces. Acta Math. Sinica, Chinese Series 26, 337–344 (2010)

[13] Wu, D.: Strong convergence theorems for Lipschitzian demi-contraction semigroups in Banach spaces (in press)

[14] Wang, E., et al.: Strong convergence theorems for Lipschitzian demi-contraction semi-group in Banach space (in press)

A Multi-stage Stochastic Fuzzy Methodology for Credit Evaluation

Guo Minmin and Wang Li

School of Economics and Management Beihang University, Beijing 100191, China

Abstract. A credit evaluation methodology is developed to help decision makers identify the overall credit score which consists of financial score and non-financial score, according to the different qualifications of financial criteria system and non-financial criteria system. The proposed methodology above is able to search for dynamic evaluation about the hierarchical decision structure using stochastic AHP and fuzzy AHP. Finally, the proposed approach is verified through the application in construction industry. Based on the experimental results, the proposed approach performed well for uncertain datasets.

Keywords: Credit score, stochastic AHP, fuzzy AHP, credit evaluation.

1   Introduction

Recently serious credibility crisis exists in trades among construction companies with the mushroom of the economic environment, such as the low rate of the contract, compliance for the performance, poor quality, violations of law and discipline, and arrears of project payment etc. Credit plays an important role in business value chain for construction industry. As a kind of value, credit information can expand the trade quantity, improve success ratio, reduce trade cost and increase trade efficiency.

A great deal of literature used different methodology, for example, AHP, Fuzzy AHP plus TOPSIS, DEA, PROMETHEE etc, to solve the MCDM problem, with the result just being a best alternative or a ranking list. In such an application, Zhu (2000) used the DEA approach in the evaluation of Fortune 500 companies' financial performance and assigned the companies into one of the two groups of financially efficient or non-efficient. Babic and Plazibat (1998) ranked the firms with respect to business efficiency using PROMETHEE. Yurdakul and Tansel (2004) proposed a hierarchical decision structure, which consisted of three levels criteria, to calculate overall credibility scores based on financial criteria and non-financial criteria system in the process of AHP. Ertugrul and Karakasoglu (2009) introduced a methodology, combined Fuzzy AHP and TOPSIS, for evaluating the performance of fifteen Turkish cement firms by using their financial tables. In evaluating the credit score of enterprises along the supply chain of construction industry, quantitative/qualitative assessments are often required to deal with uncertainty, subjective, and imprecise data (Kuo et al. 2007). Secme et al. (2009) suggested that not only financial performance but also non-financial performance should be taken into care in a competitive environment.

The above-mentioned literatures had solved the qualitative data using Fuzzy AHP well. But they did not consider how to deal with quantitative data more adequately. The stochastic model is accurate and comprehensive while other techniques are approximations (R. Lahdelma and P. Salminen, 2001).

Given the assumption that the DMs were assumed to compare and submit a ranking of the entire criteria set, lots of literature did not consider a partial list of the set, So the paper focuses on two aspects: Partial list and methodology.

The structure of this paper is organized as followed. In the third section the partial list is explained in detail. In the fourth section, the new methodology which combines stochastic AHP and Fuzzy AHP is presented. In the fifth section, an application is introduced and results are presented. At last, we conclude with final remarks and future work for research.

2   Partial List

We assumed that the pair-wise process cannot provide full information extracted from the DMs regarding the financial criteria system and non-financial criteria system.

As stated by Hochbaum and Levin (2006) facing partial list, they addressed the problem by using acyclic graph G = (V, A) in order to consistent with the preference showed by DMs.

$$\begin{cases} \displaystyle\sum_{k=1}^{m}\sum_{j=0}^{n}\sum_{i=0}^{l} w_{ijk} = 1 \\ \qquad w_{ijk} \geq 0 \end{cases}$$

Once get the partial list from the DMs, we can use the proposed methodology to calculate the explicit credibility score.

3   The Multi-stage Methodology

Ni (2008) had designed a hierarchical structure specifically for enterprises of construction industry. Compared with the hierarchical decision structure (Yurdakul and Tansel, 2004), Ni's structure may be sort of embedding in the decision structure of Yurdakul and Tansel. In order to evaluate the credit score accurately, we would rather adopt the hierarchical decision structure for our research.

3.1   Evaluation Index in Hierarchical Structure

A hierarchical structure specifically for enterprises of construction industry has been designed by Ni (2008), this paper incite Ni (2008)'s evaluation index in Table 1.

3.2   Stochastic AHP Analysis

Eskandari and Rabelo (2006) introduced a stochastic approach. stochastic AHP methodology calculates the global weight variances accounting for judgmental errors resulting from inconsistent pair-wise comparisons (Rabelo et al., 2007).

Consider formula above, the DMs' preferences are represented by a weight distribution with density function f(w) in the set of feasible weights W defined as formula:

$$W = \{w_{ijk}, i = 0,...l; j = 0,...n; k = 1,...m\}$$

For being in the stochastic AHP, total knowledge about weights is represented in "Bayesian" spirit by a uniform weight distribution in W. So this distribution has density function:

Table 1. Credit evaluation index system for construction industry

The credit evaluation index for construction industry (A)	Financial index (B1)	Revenue type index(C1)	Rate of current from total capital(D1)
			Pure profit rate of enterprise (D2)
			Current profit rate of enterprise (D3)
		Expense index (C2)	Salary expense (D4)
			Administration expense (D5)
		Utility index (C3)	Total capital circulation rate (D6)
			Fixed asset circulation rate (D7)
			Visible fixed asset circulation rate (D8)
		Productivity index (C4)	Total capital investment efficiency(D9)
			Equipment investment efficiency (D10)
			Machinery investment efficiency (D11)
		Circulation on index(C5)	Circulation rate (D12)
			Current deposit rate (D13)
			Fixed rate(D14)
	Non-financial index (B2)	IT application (C6)	Information system (D15)
		Credit condition (C7)	A passing rate of handing over(Ratio of Project Award-winning works) (D16)
			Contract compliance rate(The number of court proceedings) (D17)
		Quality (C8)	Technical staff ratio(patent number) (D18)
			Enterprise performance of the past three years(leading working years) (D19)

$$f(w) = \frac{1}{vol(w)}, \qquad vol(w) = \frac{z^{1/2}}{(z-1)!}, z = l-k, m-k, n-k$$

k just means the missing arcs which represent the partial list number.

$$w_{ijk}^c = \int_X f(\varepsilon) \int_{w_{ijk}(\varepsilon)} \frac{f(w_{ijk})wdwd\varepsilon}{a_{ijk}}, \qquad a_{ijk} = \int_X f(\varepsilon) \int_{w_{ijk}(\varepsilon)} f(w_{ijk})wdwd\varepsilon$$

The confidence factor is then computed as an integral over the criteria distributions ε by:

$$p_{ijk}^c = \int_\varepsilon f(\varepsilon)d\varepsilon$$

In order to satisfy a consistent ranking matrix, each row and column should contain the full information of the entire matrix. So for every missing i, j entry, a value

generated from a sequence of entries, (i, k1), (k1, k2),...,(kp, j). If such a sequence exists, and let $i = k_0$ and $j = k_{p+1}$, the joint distribution value is:

$$f(\varepsilon) = \prod_{i,j,k} f_{ijk}(\varepsilon_{ijk})$$

Then utilize the global AHP weights and their corresponding estimated variances and expected value, and Monte Carlo simulation is employed for handling the related uncertainty in the global AHP weights.

3.3   Fuzzy AHP Analysis

In this paper, the ratings and variables of qualitative criteria, just non-financial criteria, are considered as linguistic variables, which can be expressed in positive triangular fuzzy members, (Chen, 2000) to evaluate the importance of the criteria, as in Table 2 and 3.

Table 2. Linguistic variable for the ratings

Linguistic variable (benefit criteria/cost criteria)	Triangular fuzzy members
Very poor/very low	(0,0,1)
Poor/low	(0,1,3)
Medium poor/ medium low	(1,3,5)
Fair/medium	(3,5,7)
Medium good/medium high	(5,7,9)
Good/high	(7,9,10)
Very good/very high	(9,10,10)

Table 3. Linguistic variables for the important weight of each criterion

Linguistic variable	Triangular fuzzy members
Very low	(0,0,0.1)
Low	(0,0.1,0.3)
Medium low	(0.1,0.3,0.5)
Medium	(0.3,0.5,0.7)
Medium high	(0.5,0.7,0.9)
High	(0.7,0.9,1.0)
Very high	(0.9,1.0,1.0)

We review the essentials of Fuzzy AHP, which was originally introduced by Chang (1999), briefly as follows. Triangular fuzzy members (TFN) are introduced to express m extent analysis values for each object, with the signs bellowed:

$$M_{gi}^1, M_{gi}^2, ..., M_{gi}^m, i = 1, 2, ..., n$$

The Chang's extent analysis consists of four steps to compute weight (Ertugrul and Karakasoglu, 2009).

After the weights for a number of criteria are determined based on the opinions of experts using FAHP method and Least Squares Support Vector Machine, these weights are inputs of TOPSIS method to evaluate the non-financial score (Mian, 2008). Then total credit score can be calculated as:

Total Score= $w_{k=1,i,j=0}$ * Financial Score + $w_{k=2,i,j=0}$ * Non-Financial Score.

4   Application

The hierarchical structure developed in this study is a four-level hierarchy in which the top level represents the total score of the problem. The second level of the hierarchy contains the financial and non-financial criteria.

According to the hierarchical, we designed three questionnaires in consistent with the three different entities, respectively the contractors, equipment suppliers and material suppliers. Three questionnaires studied the specific characteristic of different companies. Moreover, nowadays the companies grow and survive along the value chain. So the investigation participants should be 3×3. That means we can evaluate the particular company from three perspectives, so that the calculated total credit reflect the overall aspects in the way of scorecards.

In order to prove the proposed methodology proposed for the credit evaluation, we collected the data from financial reports (From Shanghai Stock Exchange) of five companies, which consist of Gezhouba Gorp, Wuhan Iron and Steel Corp, Green Town, Panzhihua Iron and Steel Group, China Building Materials Group, for three years, the partial list of preference showed by 4 experts, who had viewed the scorecards from the questionnaires about the five companies, and utilizing the methodology, we can get the credit score of the five companies as detailed in Table 4 and judgment error can be shown by using Bayesian Monte Carlo Simulation for 5500 times, 500 times of which are discarded because 500 times are within the simulation reaching the steady state .

From the results shown in Table 4 and Fig 1, the results show threefold feature with mean absolute deviation (MAD): 0.487, while without inconsistency for CI: 0.08456:

Table 4. The total credit score of the five companies based on the methodology

Company	Financial Score	Non-Financial Score	Financial Weight	Non-Financial Weight	Total Credit Score
1	0.873928	0.585253	0.649556	0.350444	0.772764
2	0.883389	0.490616	0.649556	0.350444	0.745744
3	0.780602	0.429029	0.649556	0.350444	0.657395
4	0.689077	0.409986	0.649556	0.350444	0.591271
5	0.665578	0.278085	0.649556	0.350444	0.529783

Note:1-Gezhouba Gorp   2-Wuhan Iron and Steel Corp   3-Green Town 4-Panzhihua Iron and Steel Group   5-China Building Materials Group

(1) The total credit score is almost precise when reflecting the true credibility of the company because MAD is so low that it can be neglected and CI has reached extent analysis for CI<0.1

(2) The scores, computed by the proposed methodology, are within the interval provided by the experts at first. The scores of each company are respectively 0.772764, 0.745744, 0.657395, 0.591271 and 0.529783.

(3) According to the preference of 4 experts, the financial weight, 0.649556, is higher than the non-financial weight, 0.350444. More or less, for each company, the final financial score is also higher than the final score of non-financial hierarchy decision structure. That is also true for the criteria be more important when the criteria weights more heavily. The financial score and the non-financial score is strongly positive described as Fig 2.
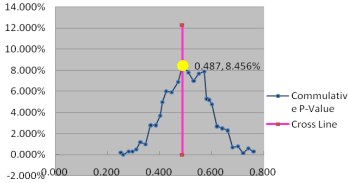
Fig. 1. AHP Monte Carlo Analysis
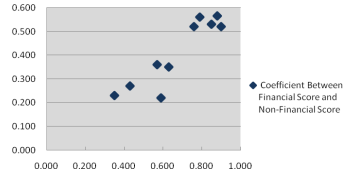
Fig. 2. Coefficient Between Financial Score and Non-Financial Score

(4) At last but also important, that is the bigger the financial score is, the bigger the non-financial score is, and then the according total credit score is big. That is also in consistent with the realistic financial world and the preference of the DMs.

According to the score and the expectation of the experts, we can conduct that the ranking is approximate for the real business world.

5   Summaries

In this paper, credit evaluation of companies along the supply chain is discussed, and a multi-stage stochastic fuzzy methodology is formulated. The application of this methodology is also illustrated. The AHP-based model is an excellent credit evaluation tool for companies. Since the data is uncertain and the DMs are unable to compare every criteria of the hierarchical decision structure, stochastic AHP and fuzzy AHP are introduced in the methodology in order to solve the problem.

In the further we will extend more rigorously to collect a lot of related data of companies to verify and evaluate the methodology.

References

[1]   Babic, Z., Plazibat, N.: Ranking of Enterprises Based on Multicriterial Analysis. International Journal of Production Economics 56(57), 29–35 (1998)
[2]   Chang, D.Y.: Applications of the Extent Analysis Method on Fuzzy Ahp. European Journal of Operational Research 95(3), 649–655 (1996)

[3] Chen, C.T.: Extensions of the Topsis for Group Decision Making under Fuzzy Enviornment. Fuzzy Sets and Systems 114, 1–9 (2000)

[4] Ertugrul, I., Karakasoglu, N.: Performance Evaluation of Turkish Cement Firms with Fuzzy Analytic Hierarchy Process and Topsis Methods. Expert Systems with Applications 36, 702–715 (2009)

[5] Eskandari, H., Rabelo, L.: Handling Uncertainties in the Analytic Hierarchy Process: A Stochastic Approach. International Journal of Information Technology and Decision Making 5(3) (2006)

[6] Hahn, E.D.: Decision Making with Uncertain Judgments: A Stochastic Formulation of the Analytic Hierarchy Process. Decision Sciences 34(3), 443–466 (2003)

[7] Hochbaum, D.S., Levin, A.: The k-Allocation Problem and Its Variants. In: Erlebach, T., Kaklamanis, C. (eds.) WAOA 2006. LNCS, vol. 4368, pp. 253–264. Springer, Heidelberg (2007)

[8] Kratschmer, V.: A United Approach to Fuzzy Random Variables. Fuzzy Sets and Systems 123, 1–9 (2001)

[9] Kuo, M.S., Tzeng, G.H., et al.: Group Decision Making Based on Concepts of Ideal and Anti-Ideal Points in Fuzzy Enviorment. Mathematical and Computer Modelling 45, 324–339 (2007)

[10] Lahdelma, R., Salminen, P.: Smaa-2: Stochastic Multicriteria Acceptability Analysis for Group Decision Making. Operations Research 49(3), 444–454 (2001)

[11] Ni, D.M., Xia, G.P., et al.: A Mulit-Structure Fuzzy Integrated Credit Evaluation Model Based on Entropy Weight for Construction Industry. In: Proc: 38th International Conference on Computer and Industiral Engineering, vol. 3(16), pp. 2364–2369 (2008)

[12] Rabelo, L., Eskandari, H., et al.: Value Chain Analysis Using Hybrid Simulation and Ahp. International Journal of Production Economics 105(536-547) (2007)

[13] Yurdakul, M., Tansel, Y.: Ahp Approach: in the Credit Evaluation of the Manufacturing Firms in Turkey. International Journal of Production Economics 88, 269–289 (2004)

Nonlinear Predictive Control
Based on Inverse System Method

Fuhua Song and Bin Lu

China Jiliang University,
Hangzhou 310018, China

Abstract. To deal with the difficulty of nonlinear inverse model identification of directive inverse control and improve the ability of robustness and anti-interference of the open system, the paper studied the realization of inverse system identification and control using least squares support vector machine (LS-SVM), and proposed a new nonlinear predictive control algorithm based on inverse system method. The method cascades the αth-order inverse model approximated by LS-SVM with the original system to get the composite pseudo-linear system. Then the predictive control method is introduced to the pseudo-linear system. The simulation results show that the combined method does not depend on the accurate mathematical model and has the characteristics of better robustness stability, simpler design process and high tracking accuracy. And this approach is one of the applicable methods for the control of nonlinear systems.

Keywords: least squares support vector machine (LS-SVM), nonlinear predictive control, inverse system method, robust stability.

1   Introduction

Predictive control is a model-based closed-loop optimal control strategy, which in a variety of predictive models based on rolling optimization targets and use of online feedback correction strategy, Strive to effectively overcome the uncertainty of the controlled object, hysteresis and time-varying factors such as the dynamic effects to achieve the desired control objectives - reference trajectory input, and make the system has good robustness and stability.

   With the development of intelligent control, In recent years, there have been some based on neural networks and fuzzy model predictive control algorithm. In these algorithms, neural networks are used to approximate nonlinear systems predictive models, while the neural network and fuzzy model of the learning algorithms are based on empirical risk minimization principle, this learning algorithm exists "over-fitting" problem Also, such a conventional predictive control algorithm structure is relatively complex, and because of the controlled system and prediction models are highly non-linear, therefore, closed-loop system robustness, stability and other features is often difficult to analyze.

   Support Vector Machine (Support Vector Machine SVM) for linear, nonlinear system identification and control has made great progress in recent years.Compared

with the neural network, SVM empirical risk minimization based on the same time, using structural risk minimization criterion, there is no curse of dimensionality and the local minimum problems, promote better performance, applications are increasingly being used. In 1999, Suykens JAK [4] proposed a new type of SVM: Least Squares Support Vector Machines (LS-SVM, Least Squares Support Vector Machines). LS-SVM linear system to the introduction of least squares support vector machine, instead of the traditional SVM uses quadratic programming method to solve the problem of estimating the function, the method simple operation, fast convergence and high accuracy.

This article has developed relatively rapidly in recent years, LS-SVM, inverse system method [9-11] and the predictive control method combining the three proposed a new approach based on inverse system predictive control algorithm, Namely, first using LS-SVM approximation of the original system order integral inverse system, and its thread in the original system prior to constitute the basic linear-oriented pseudo-linear composite system. The existence of pseudo-linear systems modeling error the introduction of predictive control strategy. Through a typical nonlinear system simulation to verify the validity of the method.

2   Least Squares Support Vector Machine Function Fitting Characteristics

The basic idea of Support vector machine regression is to map the data through a non-linear mapping to high-dimensional feature space, and then in this space to create a linear regression function, The decision function given by an SVM is:

$$f(x) = w^{\mathrm{T}} \varphi(x) + b \tag{1}$$

Where $\varphi(\cdot)$ is Non-linear function, ω is weights b is Threshold. Thus, in high-dimensional feature space of the linear regression on the low-dimensional input space corresponds to nonlinear regression. The decision function given by LS-SVM is as follows:

Given to the training sample collection $\{x_k, y_k\}_{k=1}^{d}$ of d, defined the optimization problems as follows:

$$\min_{w,b,e} J(w, \xi) = \frac{1}{2}\|w\|^2 + \frac{1}{2}C\sum_{k=1}^{d}\xi_i^2 \tag{2}$$

subject to $y_i = w^{\mathrm{T}}\Phi(x_i) + b + e_i, i = 1, \cdots, d$

The corresponding Lagrange function is defined as follows:

$$L(w, b, \xi, a) = J(w, \xi) - \sum_{i=1}^{d} a_i[w^{\mathrm{T}}\Phi(x_i) + b + \xi_i - y_i] \tag{3}$$

where $a_i, i = 1, \cdots, d$ is Lagrange multiplier, $a = [a_1, \cdots, a_d]^{\mathrm{T}} \in \mathsf{R}^d$, $\xi = [\xi_1, \cdots, \xi_d]^{\mathrm{T}} \in \mathsf{R}^d$. To optimize the solution:

$$\begin{cases} \dfrac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}} = 0 \Rightarrow \boldsymbol{w} = \sum_{i=1}^{d} a_i \, \Phi(\boldsymbol{x}_i), \\[2mm] \dfrac{\partial \boldsymbol{L}}{\partial b} = 0 \Rightarrow \sum_{i=1}^{d} a_i = 0 \\[2mm] \dfrac{\partial \boldsymbol{L}}{\partial \xi_i} = 0 \Rightarrow a_i = C\xi_i, \quad i = 1,2,\cdots,d \\[2mm] \dfrac{\partial \boldsymbol{L}}{\partial a_i} = 0 \Rightarrow \boldsymbol{w}^{\mathrm{T}} \, \Phi(\boldsymbol{x}_i) + b + \xi_i - y_i = 0, \quad i = 1,2,\cdots,d \end{cases} \tag{4}$$

So to solve the optimal problem is transformed into solving linear equations:

$$\begin{bmatrix} 0 & \boldsymbol{I}^{\mathrm{T}} \\ \boldsymbol{I} & \boldsymbol{\Omega} + \gamma^{-1}\boldsymbol{I} \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{5}$$

Where $\boldsymbol{y} = [y_1, \cdots, y_d]^{\mathrm{T}}$, $\boldsymbol{I} = [1, \cdots, 1]^{\mathrm{T}}$,
$\boldsymbol{\Omega} = \{\boldsymbol{\Omega}_{ij}\}_{d \times d}, \Omega_{ij} = \Phi(\boldsymbol{x}_i)^{\mathrm{T}} \cdot \Phi(\boldsymbol{x}_j), i, j = 1, \cdots d$.

According to the mercer conditions[3], There is a mapping function and nuclear function of 2.

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \Phi(\boldsymbol{x}_i)\Phi(\boldsymbol{x}_j) \tag{6}$$

Least squares support vector machine function approximation as follows:

$$y(\boldsymbol{x}) = \sum_{i=1}^{d} a_i K(\boldsymbol{x}, \boldsymbol{x}_i) + b \tag{7}$$

Where $\mathbf{a}, b$ can be calculated form Eq. (5)

The corresponding sample of a_i is support vector, which a_i is not zero.

3   Inverse System Control Method

We study the inverse system method through the BIBO stability of SISO with reversible non-linear system. The single variable discrete-time system is $\Sigma : u_k \rightarrow y_k$, It can be expressed as:

$$f[y(k+\alpha), \cdots, y(k+\alpha-n), u(k), \cdots, u(k-m)] = 0 \tag{8}$$

Where is the system output, u the system input, n is system order, m as the input delay, α is the system output to the input number of the total delay, And it is usually greater than or equal to 1.

If the equation (7) is reversible, then the inverse system can be express as:

$$u(k) = g[y(k+\alpha), \cdots, y(k+\alpha-n), u(k-1), \cdots, u(k-m)] \tag{9}$$

Then we defined α first-order inverse system for the input: $\varphi(k) = y(k+\alpha)$,as The adverse system input for $\varphi(k)$, output for $u(k)$, It is a regular system which can be achieved as follows:

$$u(k) = g\left[\varphi(k), y(k), \cdots, y(k-n+1), u(k-1), \cdots, u(k-m)\right] \tag{10}$$

It can be form of a pseudo-linear composite system befor first-order inverse system will be a thread in the original system, And must be meet the following transfer function.

$$y(z) = z^{-\alpha}\varphi(z) \tag{11}$$

4   Predictive Control Based on Inverse System Method

With the principle of inverse system there are two conditions: 1) mathematic model of the system; Accurate known. 2) From the original system can mathematical model of $u(k)$, however, In practical application, it is very difficult to do even less. While pseudo linear system is an open loop controller that derived from the inverse system method, there is no feedback, so it is difficult to guarantee robustness and anti-jamming performance. In order to solve the problem of modeling, we use LS-SVM Recognition BIBO stability of nonlinear system inverse model, The linear system of false, introducing the ideas to complete control of the pseudolites linear systems, take SISO system as an example, The control structure is shown in fig 1, For a given input in s.
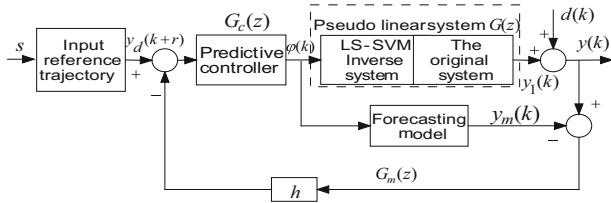
Fig. 1. Inverse system method based on internal model control structure

① the pseudo linear system prediction model
As the pseudo-linear system $G(z)$ has been basically linear into z^{-r}, therefore, simplify the prediction model as follows:

$$G_m(z) = z^{-r} \tag{12}$$

② reference trajectory
Control system of the expected output is a reference trajectory. The initial value for reference trajectory of the system is given before the change output value:

$$y_d(k+i) = y(k) \tag{13}$$

In a given reference trajectory, the future after changing the value of a time for i

$$y_d(k+i) = \beta_r y_d(k+i-1) + (1-\beta_r)s \tag{14}$$
$$(i = r, r+1, \cdots, 0 < \beta_r < 1)$$

where β_r is Soften coefficient, and it is the coefficient of the system robustness and control of the output.

③ feedback correction

There must be a certain margin of error in a nonlinear system of inverse modeling, Therefore, in predictive control feedback correction must be introduced. According to k times the output value of the object $y(k)$ and prediction error between the model output $y_m(k)$, correction predicted output, the output forecast to be revised:

$$y_p(k+r) = y_m(k+r) + h[y(k) - y_m(k)] = \varphi(k) + he(k) \qquad (15)$$

Where h is the error correction coefficient, $e(k)$ is the error that Composite controlled system and prediction model at the time of k.

④ performance index and optimal control law

For $\alpha = r$ step model prediction. Quadratic optimal objective function can be expressed as:

$$J(k) = q[y_p(k+r) - y_d(k+r)]^2 + \gamma \varphi(k)^2 \qquad (16)$$

minimize the objective function, To take $\varphi(k)$ bias and set as 0:

$$\frac{\partial J}{\partial \varphi(k)} = 2q[y_p(k+r) - y_d(k+r)]\frac{\partial y_p(k+r)}{\partial \varphi(k)} + 2\gamma \varphi(k) = 0 \qquad (17)$$

Get the current moment for the control of k:

$$\varphi(k) = \frac{q}{q+\gamma}[y_d(k+r) - he(k)] \qquad (18)$$

All of above, Predictive control algorithm which Based on the LS inverse system of SVM - first-order can be definded as follows:

(1) Select the LS-SVM kernel function and parameter C, σ, Given the input s, and the parameters β_r, q, γ, h.
(2) According to the historical data; identification of inverse model, and get LS - SVM inverse order when $\alpha = r$.
(3) To obtain the LS-SVM-order inverse system to link to the original nonlinear system prior to constitute a pseudo-linear composite system, and as a controlled object.
(4) To establish the prediction model by control system(11).
(5) At the moment of k, by the reference trajectory equation (12) Calculate the moment of hope that the next set value $y_d(k+1)$.
(6) We can get Output loop $y_p(k+1)$ from equation (13);
(7) By equation (16) to strike the optimal amount of control $\varphi(k)$, as the current state of the controlled object's input, so that $k = k+1$, transferred to subsection (5) step.

5   Simulation Study

Consider the following discrete nonlinear invertible systems:

$$y(k+3) = \frac{-0.55y(k+2) + 0.2y(k+1) + u(k)}{1 + y(k+2)y(k+1)u(k-1)} \qquad (19)$$

As we know $\alpha = 3$, $m = 1, n = 2$, In identifying the LS - SVM a = 3 rank inverse system, using RBF kernel function, regularization parameter $C = 5000$, nuclear parameter $\sigma = 4$. Identification, the input signal is used to take the mean of 0 and variance of 0.5 800 data is used to constitute the training sample set $\{\varphi_1(k), \varphi_2(k), y_1(k), y_2(k)\}_{k=1}^{800}$ training LS-SVM inverse model. In order to test identification obtained LS-SVM inverse model, using input signals to get it,and the input signal is :

$$\varphi(k) = 0.4\, sin(2\pi k / 200) + 0.1\, cos(2\pi k / 40)$$

As shown in Figure (2), as shown by the figure we can see,
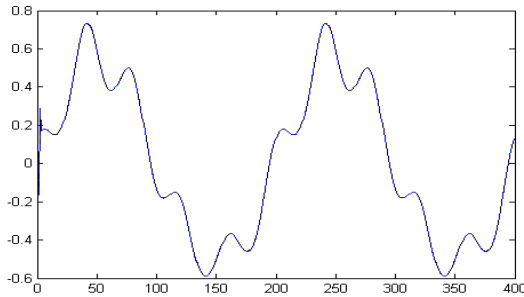
Fig. 2. LS-SVM inverse system of the tracking
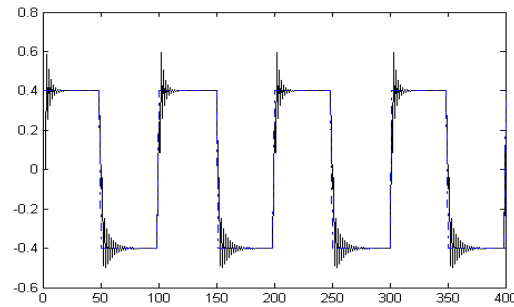
Fig. 3. Pseudo linear system of open loop control effect
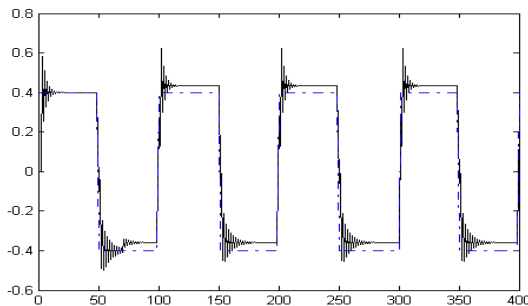
Fig. 4. The pictures which Pseudo linear system of open-loop
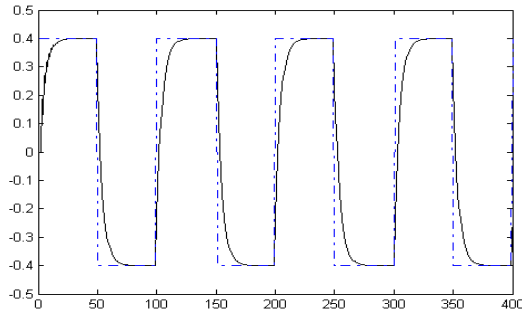
Fig. 5. Predictive control system of each wave signal output response control by constant value when disturbed

LS-SVM inverse model has been accurately received.

LS-SVM inverse model will be to link in the original system prior to constitute a pseudo-linear composite system, pseudo-linear composite system imposed after the amplitude of the square wave signal, Simulation results are shown in fig (3) ,it clear that the open-loop effects is not good. When disturbed open-loop system to the disturbance without any restrain, shown in fig(4).

By $\alpha = 3$, the prediction model can be expressed as: $y_m(k+3) = \varphi(k)$, reference trajectory using equation (12) as shown in the form of it. Soften coefficient for $\beta_r = 0.8$, $q = 1, r = 0$, Feedback coefficient for $h = 1$. After Predictive controller design, It applied to the entire square-wave signal, and the amplitude is ± 0.4. The output response curve is shown in fig(5).It is clear that that the system actual output can be a good reference trajectory tracking. Fig(6) is Simulation diagram that System into the steady-state after k> 70 when you join the disturbance d = 0.1 in case of simulation of Figure. From the figs we can know the system has a certain anti-interference ability.
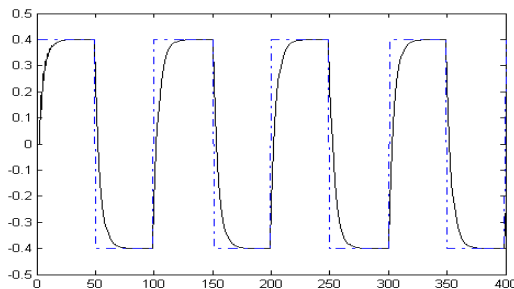
Fig. 6. Predictive control system into the steady-state response
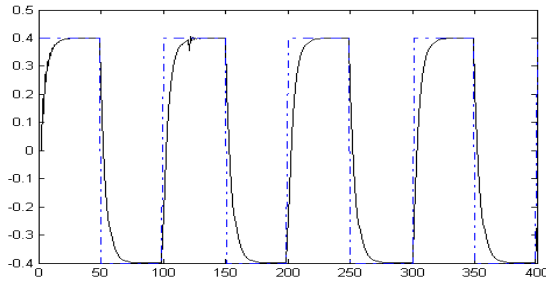
Fig. 7. After the steady-atate model into the responses of the changes after the output of interference

Fig(7)is Simulation diagram which after the system entering steady-state $k > 120$ and system model into $y(k+3) = \dfrac{-0.6y(k+2)+0.1y(k+1)+u(k)}{1+0.5y(k+2)y(k+1)u(k-1)}$ As shown in the figs System for non-linear parameter perturbations has a good robustness.

All of above, The LS-SVM-order inverse system and the original nonlinear system, composed of pseudo-linear system as a controlled object, the introduction of predictive control method to control to achieve good control effect, anti-jamming ability, the system non-linear parameter has a good robustness. This method is accused of nonlinear systems without needing to know the exact mathematical model, and the predictive control system is easy to simplify the design of conventional predictive controller design.

6   Summaries

In this paper, LS-SVM Recognition BIBO stability of nonlinear system inverse model, derived using the basic principle of inverse system method for linearization of the pseudo-linear composite system, and as a controlled object, combined with predictive control is proposed based on the idea inverse system method for predictive control new methods. From reversible simulation of nonlinear systems, usefull conclusion can be made:

(1) LS-SVM is a good inverse model of nonlinear systems identification, this method does not require a priori knowledge of system model, only need to use a small amount of input and output data as the sample set, we can identify the desired effect, to overcome the traditional method of two inverse system bottlenecks;

(2) predictive control method which based on SVM LS - a first-order of inverse system have high tracking accuracy, and when there is interference or parameter perturbation, the system can have very good performance, It proved robustness.

Therefore, Inverse system method based on predictive control program adapted to a more general system BIBO stability of predictive control problems, and its physical concept is clear to adapt to a wide range and convenient to use.

References

[1] Jing, Z.: Intellect predictive control and application. Zhejiang University Press, China (2002)

[2] Tan, Y., Canwenberghe, A.V.: Nonlinear one-step-ahead control using neural networks: control strategy and stability design. Automatica 32(12), 1701–1706 (1996)

[3] Vapnik, V.N.: Statistical Learning Theory. John Wiley, New York (1998)

[4] Suykens, J.A.K., Vandewalle, J.: Least squares support vector machine classifier. Neural Processing Letters 9(3), 293–300 (1999)

[5] Suykens, J.A.K., Lukas, L., Vandewalle, J.: Sparse approximation using least squares support vector machine. In: IEEE Int. Symposium on Circuits and Systems, Geneva, vol. (II), pp. 757–760 (2000)

[6] Suykens, J.A.K.: Support vector machines: A nonlinear modeling and control perspective. European Journal of Control 7(2-3), 311–327 (2001)

[7] Wang, Y.-H., Huang, D.-X., Gao, D.-J., Jin, Y.-H.: Nonlinear predictive control based on support vector machine. Information and Control 33(2), 133–140 (2004)

[8] Zhong, W.-M., Pi, D.-Y., Sun, Y.-X.: Support vector machine based direct inverse-model identification. Control Theory & Applications 22(2), 307–310 (2005)

[9] Li, C.-W., Feng, Y.-K.: Inverse system method of multi-variable nonlinear control. Thinghua University Press, Beijing (1991)

[10] He, D.: Nonlinear systems control based on artificial neural networks, Doctoral dissertations of Southeast University, Nanjing China (1999)

[11] Dai, X.-Z., He, D., Zhang, X., Zhang, T.: MIMO system invertible and decoupling control strategies based on ANN α th-order inversion. IEE Proc. Control Theory Appl. 148(2) (March 2001)

Research on Optimization Algorithm for Analog Circuit Simulation

Wang Jing-Li

Electronic and Information Engineering College,
Liaoning University of Technology,
Jinzhou, 121001, P.R. China

Abstract. Optimization design for analog circuit simulation based on the Monte Carlo, its goal wants to reduce test time and cost. This article researches the optimization algorithm of simulation parameters and the simulation precision. The two algorithms can reduce the function operational cost of optimization process and the SA's calculation times effectively. The simulation effect is improved further.

Keywords: Analog circuit simulation, Optimization algorithm, Macromodel algorithm, Precision optimization.

1   Simulation Parameters' Optimization Algorithm

The method of solving the output maximization is usually divided into two classes: the Monte Carlo method and geometric operation method. The Monte Carlo method calculates process outputs directly by using some estimated values. Instead, geometric method accepts process outputs indirectly by constructing approximate receivable domain of input space. But each method has the following two disadvantages:

(1) Precise outputs are not feasible, and the simulation cost is too much.
(2) Process input variables' dimensions are too big, the typical values can amount to hundreds of times, which increase the optimization algorithm's complexity.

This paper puts forward a method based on the macromodel method, it overcomes the shortcomings mentioned the above. This method uses a group of simple analysis functions which can replace the complicated relations between the input variables and output variables, and can reduce the function operational prices effectively.

1.1   The Problem's Description

A parametric process output function Y is defined, Y is a part of the equipment, and its features can satisfy the acceptable constraint conditions, Y can be expressed as follows:

$$Y = \oint_{a_y} f_y(y)dy \tag{1}$$

Among them: y is the characteristic vector of device, fy is the joint probability density function of y, a_y is the acceptable output domain of space y.

X is defined as process random variable, and is assumed to meet the normal distribution and is independent on statistics. X can be expressed as follows:

$$X = \begin{bmatrix} x_c \\ x_d \end{bmatrix} + \begin{bmatrix} \varsigma_c \\ \varsigma_d \end{bmatrix} \tag{2}$$

Among them: x_c is rated process control parameters vector, and its values can be designed and changed. x_d is rated physical parameters vector, and its values can not be changed in design. ς_c and ς_d are independent gaussian random variables, supposing the statistical properties can be predicted in advance.

Output maximization problem (YMP) is to optimize Y through choosing the ideal x_c, Namely:

$$\max_{x_c} Y(X) \quad (x_c \in D) \tag{3}$$

Among them, D is defined by user as the acceptable designed parameters scope.

1.2   Research on Algorithm

The input acceptable domain a_x is defined as the value domain of input process variables in X space, and decided by the acceptable output value domain, namely

$$a_x = \{x / y' \le y(x) \le y''\} \tag{4}$$

The above acceptable domain and the under acceptable domain is:

$$a_x = \{x / y' - y(x) = \theta_1(x) \le 0, and, y(x) - y'' = \theta_{m+i}(x) \le 0. \ i = 1, 2...m\} \tag{5}$$

Among them, m is the dimension of y.

After input fields are given, process output function can be defined as:

$$Y = \int_{a_x} f_x(x) dx \tag{6}$$

Here, f_x meets gaussian distribution, and a_x is gibbosity.

$$\max_{x_c} \min_{i \in Q} r_i(x_c) \tag{7}$$

Here, $x_c \in \{x_c / \theta_i(x_c) \le 0, i \in Q\} \cap D$, Q is non-redundant acceptable domain boundary index, r_i is the shortest distance of x_c to the boundary i, namely

$$r_i(x_c) = \min \|x - x_c\| \tag{8}$$

The main advantage of formula 8 is that this is a determinate optimal problem and it can be solved easily compared with the formula 3.

1.3 Macromodel Algorithm

In order to reduce the dimensions of tectonic approximate values of a_x, approximate values of Y are solved by means of using a group of macromodel. Macromodel algorithms are as follows: given output variables of number d, input variables of number n, a group of simple description formulas are as follows:

$$
\begin{aligned}
y_1 &= h_1\left(x_{11}, x_{12}, ..., x_{1n}\right) \\
y_2 &= h_2\left(x_{21}, x_{22}, ..., x_{2n}\right) \\
&\vdots \\
y_d &= h_d\left(x_{d1}, x_{d2}, ..., x_{dn1}\right)
\end{aligned}
\tag{9}
$$

Among them, y_i is the ith approximate output response, h_i is the ith macromodel,

x_{ij} is the jth input variable of the ith macromodel.

In order to get a proper macromodel and minimize the cost of modeling and keep the systems accuracy, this article adopts variable selection method to obtain effective input fields. The purpose of is to choose a group of meaningful macromodel input variables so as to reduce the number of model variables and the systems complexity. This method uses the subsection regression model, each item's step of model is the two steps. The meaningful parameters can be obtained through the low second square adapted method. Formulas are as follows:

$$
y = \beta_0 + \sum_{i=1}^{m} \beta_i x_i + \sum_{i=1}^{m} \sum_{j=1}^{n} \beta_{ij} x_i
\tag{10}
$$

Among them, y is a approximate value of output response, x_i is the ith important input variable, β is an estimated attenuation coefficient, m is an important input variable number.

2 Simulation Accuracy' Optimization Algorithm

This algorithm mainly uses the statistics theory and two-step CMOS circuit to explain its principles and applications. This algorithm may reduce the Monte Carlo analysis times of the phase and the gain. This algorithm can improve the SA's calculation times of obtaining optimized results and the systems' whole performance.

When the results of defective circuit and undefective one are too close, a mass of simulation calculations are necessary. So for many defects, especially disastrous ones, a mass of Monte Carlo analysis are unnecessary.

Let D_g and D_f to seperately express probability density of output values in the defective circuit and undefective one. The every distribution value is determined by average value μ and standard migration σ.

According to the required accuracy, we defined a threshold value S, namely, the defect $\left(\mu_g - \mu_f > S\right)$ is found, when

$$
S = k * \left(\sigma_g + \sigma_f\right) , \ k \in [1,3]
\tag{11}
$$

Variable k can decide the optimization test result. If probability of k is very low, then undefective circuit can pass test with high probability and defective one can pass test with very low probability, and vice versa.

For the average distribution, the probability values of continuous random variable x in a certain time intervals $[\mu - k\sigma, \mu + k\sigma]$ are in Tab.1.

Table 1. Probability values

k	Probability(%)
1	68.26
2	95.44
3	99.74

In this paper, the choice $k=2$, so $S = 2(\sigma_g + \sigma_f)$, when $|\mu_g - \mu_f| > 2(\sigma_g + \sigma_f)$, defect is found. If $\sigma_g \geq \sigma_f$, when $|\mu_g - \mu_f| > 4\sigma_g$, defect is found.

Assuming $V_f \in [\mu - 2\sigma_f, \mu + 2\sigma_f]$ are the simulation value of defective circuit, when $|\mu_g - \mu_f| > 2\sigma_g + 4\sigma_f$, defect is found. when $\sigma_g = \sigma_f$, namely, $|\mu_g - \mu_f| > 6\sigma_f$, defect is found.

If the simulation value V_f of defective circuit satisfied condition $|\mu_g - \mu_f| > 6\sigma_f$, we need not calculate the distribution of defective values and not to use Monte Carlo analysis and we can find defects too.

For each test T_i, μ and σ need to be calculated. For each defect F_j, we use the first simulation values to estimate it. If meet condition $|\mu_g - \mu_f| > 6\sigma_f$, defects are found and the next defective analysis may begin. Otherwise saving data, to choose the next simulation repeating the above process. If reaching maximum, formula $|\mu_g - \mu_f| > 6\sigma_f$ still don't dissatisfied, then μ and σ of defective circuit need to be calculated. Finally the following probability function which may find the defects can contrast two distribution:

$$PDF = 1 - \int_{\mu_G - 2\sigma_F}^{\mu_G + 2\sigma_G} D_f(x)dx \qquad (12)$$

If D_f closes to the normal distribution, formula 12 may change into:

$$PDF = 1 - \int_{\mu_G - 2\sigma_F}^{\mu_G + 2\sigma_G} \frac{1}{\sqrt{2\pi}\sigma_f} \exp\left(-\frac{(x - \mu_f)^2}{2\sigma_f^2}\right) dx \qquad (13)$$

Finally we can get a matrix, which is listed in Tab.2.

Table 2. Matrix values

Test times	Defects				
	2E0	2E1	2E2	2E3	2E4
1	1.00	1.00	1.00	1.00	0.00
2	1.00	0.96	0.06	0.00	0.40
3	1.00	0.96	0.00	0.00	0.14
4	0.57	0.69	0.00	0.00	0.08
5	0.02	0.01	0.00	0.00	0.00
FC	84%	85%	53%	53%	33%

Considering little independent parameters can be defined so as to meet requirements, other parameters can be derived from the independent parameters, simulation speed may be accelerated and cost may be reduced.

In each interval, systems' variables are produced with the following formula:

$$W_{k+1}[j] = W_K[j] + coef^* md_k[j] \qquad (14)$$

Among them, $W_K[j]$ is the right amount of the jth dimension system variable, $md_k[j]$ is random variable.

Coef is defined as follows:

$$coef = \begin{cases} 2coef, & objdf \ \text{f} \ \text{max\_var} \\ coef/8 \end{cases} \qquad (15)$$

Among them, max_var is the biggest acceptable maximum difference, acceptable *objdf* is to use the average of the previous analysis subtract the current absolute minimum value.

Due to existing shortcomings of large number of calculating times in the above method, this paper put forward to the following improving method. It could guarantee the good results in the premise of decreasing computing times. Formula 14 will be replaced as follows:

$$W_{k+1}[j] = W_k[j] + coef^* \left(tcoef^* md_k[j] - (1 - tcoef^*) grad_k[j] \right) \qquad (16)$$

Among them, k determines the starting point, $tcoef = \dfrac{temp}{k + temp}$, $grad_k[j] = \left(\dfrac{\partial_f}{\partial_x}\right)$ is tilted function.

Comparing this method with SA on execution time, this paper's method has good results, which can be shown in Tab.3.

Table 3. Execution time

Model	Two-step CMOS circuit	
	SA	Paper's method
Average interval	11834	4195
Average time	4.32	2.54
Minimum cost rate	0.786	0.784
Average cost rate	0.792	0.784

3   Summaries

This paper researches on the simulation parameters' optimization algorithm and simulation accuracy' optimization algorithm. Simulation parameters' optimization algorithm based on the macromodel method uses a group of simple analysis functions which can replace the complicated relations between the input variables and output variables, and can reduce the function operational cost effectively. Simulation accuracy' optimization algorithm mainly uses the statistics theory and two-step CMOS circuit to explain its principles and applications and may reduce the Monte Carlo analysis times of the phase and the gain and can improve the SA's calculation times of obtaining optimized results and the systems' whole performance.

References

[1] Di Yuan, X.: The Application of EDA in Teaching of Digital Circuit. Modern Electronics Technique 04 (2006)
[2] Ping, L.: Research on simulating techniques in analog electronic circuits. Huazhong University of Science and Technology 04 (2006)
[3] Min, D.: Universal analog circuit fault knowledge acquisition flat base on simulation. Computer Engineering and Design 27(1) (2006)
[4] Wang, Z.H.: Testing and Fault Detecting of Analog Integrated Circuit. Acta Electronica Sinica (10) (1995)
[5] Zhao, W.: Research on Simulation of Analog-circuit Fault Diagnosis Technology. Huazhong University of Science and Technology 05 (2006)

Shadow Removal Based on Physical Features and Morphological Methods

Zhang Chao, Liu Shao Jie, and Zhao Yu Ming

Department of Automation, Shanghai Jiao Tong University, and Key Laboratory of System
Control and Information Processing, Ministry of Education of China, Shanghai 200240
{nero614,arola_zym}@sjtu.edu.cn,
ziqidonglai03@163.com

Abstract. In this paper an algorithm has been presented to remove the shadow
from images, which includes two steps. First step is to detect shadow regions
using physical features of shadow for an initial result with less shadow. Two
properties of shadow, luminance and gray density are taken into account, to get
the primary result. The second step is based on morphological methods. Using
the similarity of gray density and the gradient value, a refined result can be
obtained with the help of real edge of the object and the direction to expand.

Keywords: Shadow removal, Object tracking, Intelligent surveillance system.

1   Introduction

Over the past decade, many researches have been focusing on the intelligent
surveillance system. This system can help people to deal with traffic problems, keep
residential area safe, etc. To detect and track moving object is key problem in the
surveillance system. Unfortunately, the shadow often influences the result of tracking
in some sunny days. So the shadow must be removed to get better detection and
tracking results.

Much work about shadow removal has been introduced. Otsu[1] proposed a
method using optimum threshold to remove shadows. This algorithm computes
inexpensively, but it can't deal with dark objects well. K. Onoguchi[2] proposed a
method using two cameras and examining the images produced to remove the shadow
on the plane. This algorithm isn't influenced by the object type and plane color, but
the object must be visible to the two cameras and the road plane must be flat. I.
Mikic[3] proposed a method using three sources of information (local, spatial and
temporal) to classify pixels into three classes: shadow, object and background. This
algorithm isn't influenced by the number of objects and can deal with outdoor scenes,
but it assumes the background to be flat road surface and needs position of the sun
with respect to camera.

In this paper, we will present our shadow removal method consisting of two parts.
One is shadow detection using the physical features of shadow to get a result without
shadow. It makes use of the two properties of shadow, luminance and gray density, to
realize this procedure. But there is a problem that the object region may be smaller
than the real object. The second part employs morphological methods. Using the

similarity of gray density and the gradient value, a better result can be obtained with the help of real edge of the object and the direction to expand. We will explain these later in details.

2   Shadow Detection Using Physical Features of Shadow

In the First step, we use the Gaussian Mixture Model (GMM) [4] to learn   the background model and detect the primary foreground object with shadow, depicted in Fig.1, Fig.2, Fig.3. GMM is an popular and robust method. Unfortunately, it just detects the moving objects, and cannot distinguish the shadow from object.

2.1   Using Shadow's Luminance Property

According to the shadow's physical property, it can be known that the difference between the luminance value of the object and that of the background should be greater than a threshold named L_1 and less than another threshold named L_2.

Thus, we transform the current image M_c and the background image M_b from the RGB color space to YUV color space, so the current image M_{cyuv} and background image M_{byuv} in YUV color space can be obtained. Then, they are split to get the first plane of them and name them as M_{cy} and M_{by} respectively, i.e., the luminance's image of the current image and background image.

Then we judge the difference between M_{cy} and M_{by} where every pixel's value of M_f is not equal to zero. According to (1), if the difference is greater than L_1 or is less than L_2, we set this corresponding pixel of M_l to 255, on other conditions we set this pixel to 0.

$$M_1 = \begin{cases} 255 & L_o - L_b \succ L_1 \\ 255 & L_o - L_b \prec L_2 \\ 0 & else \end{cases} \tag{1}$$

where M_l is the result of luminance's operation, L_o is the luminance of M_{cy} in one point, and L_b is the luminance of M_{by} in this point. L_1 and L_2 are two thresholds.

2.2   Using Shadow's Gray Density Property

Similar to the shadow's physical property, it also known that the difference value of the gray density of the object and that of the background should be greater than a threshold G_1 and less than another threshold G_2.

Another operation is used to transform the current image M_c and the background image M_b from the RGB color space to gray space, so the current image M_{cg} and background image M_{bg} in gray space can be obtained.

Then we calculate the difference between M_{cg} and M_{bg} where every pixel's value of M_f. If the difference is greater than G_1 or less than G_2, we set this corresponding pixel of M_g to 255, on other conditions we set this pixel to 0. It can be written as

$$M_g = \begin{cases} 255 & G_o - G_b \succ G_1 \\ 255 & G_o - G_b \prec G_2 \\ 0 & else \end{cases} \quad (2)$$

where, M_g is the result of gray density's operation. G_o is the gray density of M_{cg} in one point, G_b is the gray density of M_{bg} in this point, G_1 and G_2 are two thresholds.

At last we add M_l and M_g to get the coarse object regions named of M_{io} according to (3). Except for explaining, the objects talked later is M_{io}. You can see the result in Fig.4.

$$M_{io} = \begin{cases} 255 & P_l = 255 \, or \, P_g = 255 \\ 0 & otherwise \end{cases} \quad (3)$$

where P_l is the value of M_l in one point, P_g is the value of M_g in the same point.

Fig. 1. Current image, M_c

Fig. 2. Background image, M_b
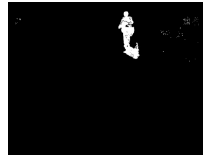
Fig. 3. Foreground image, M_f

Fig. 4. M_{io}

3   Using the Morphological Method to Improve the Result

From above mentioned, we have got some parts of the moving objects, but they may be the incomplete ones which should be expanded for a satisfactory result. Therefore, some principles must be followed to achieve that; otherwise more mistakes will be made. To expand this coarse result, some restricts on the direction and scope should be given. Next detailed explanation will be showed.

At first some work about gradient need to be done, which will be used later. First we transfer the current image M_c from RGB color space to gray space. Therefore two images are gotten, named M_{cgx} and M_{cgy} respectively. Then we compute them with

two different kernels of operator, $\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$ and $\begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix}$, to get the first x-image

derivate M_{cdx} and y-image derivate M_{cdy}. The reason we use those two kernels is they are Sobel operators with Gaussian smoothing and differentiation, so the result is less affected by the noise.

At last according to (4), we plus the absolute value of Mcdx and absolute value of Mcdy to get the gradient value of Mc, named of Mcgv. Similarly we get the first x-image derivate Mbdx, first y-image derivate Mbdy and gradient value Mbgv from background image Mb. You can see the result in Fig.5 and Fig.6.

$$M_{cgv} = \left|M_{cdx}\right| + \left|M_{cdy}\right| \tag{4}$$

$$M_{bgv} = \left|M_{bdx}\right| + \left|M_{bdy}\right| \tag{5}$$

3.1   Getting the Real Edge and the Expanding Directions of the Objects

The gradient value of edges in the current image is greater than that of background image, so we compute the differences between gradient value of current image and background image and compare them with a threshold named G. by this meaning, the edges of objects are gotten.

First we get M_{cgv} and M_{bgv} as mentioned before, then judge the difference between M_{cgv} and M_{bgv} where every pixel's value of M_f is not equal to zero. According to (6), if the difference is greater than G, this corresponding pixel of M_{fe} is set to 255, other's condition this pixel is set to 0. You can see the result in Fig.7.

$$M_{fe} = \begin{cases} 255, & G_{cgv} - G_{bgv} \succ G \\ 0, & else \end{cases} \tag{6}$$

where M_{fe} is the result of edges of objects, G_{cgv} is the gradient value of M_{cgv} in one point, G_{bgv} is the gradient value of M_{bgv} in this point, and G is a threshold.

This result not only has the edges of the objects but also has the edges of shadow. To get the real edges of the objects, first we define nine directions, depicted in Fig. 8, those directions will be used to expand the area of objects later, and how to get them will be discussed below.
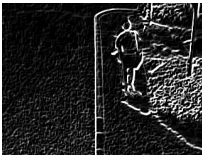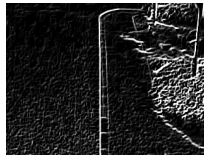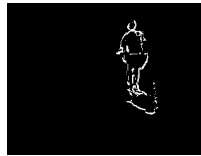
Fig. 5. M_{cgv} Fig. 6. M_{bgv} Fig. 7. M_{fe} Fig. 8. Nine directions are defined
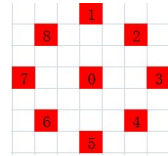
A mask of 7×7 is used to scan the full edges pixel by pixel, you can see Fig.9. We set the center of mask to one point of the full edges and count the number of points which belong to both objects and mask, besides we choose the direction by judging whether this direction have the most numbers. For example you can see Fig. 10, the black point is to be determined whether it belongs to the real edges, the blue points do not belong to the real objects and the purple points belong to both the real objects and the mask. We compare the number of purple points with a threshold to determine if it is one of points of the real edges and this point is a point of real edges. At last we choose the direction of this point, and in the direction of 7 there are most points. So the direction of this point is 7. The yellow points are the points of real edges. In the
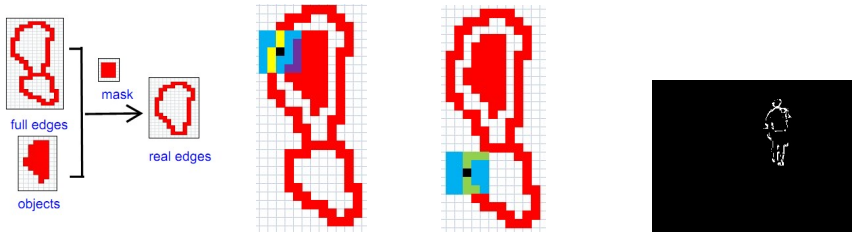
Fig. 9. The process to get the real edges
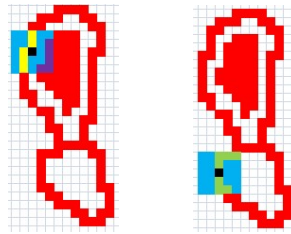
Fig. 10. This point belongs to real edges
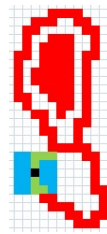
Fig. 11. This point does not belong to the real edges
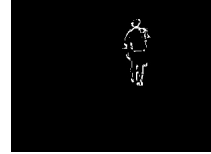
Fig. 12. Real edges

Fig. 11, you can see there are not repeat points at the side of the black point. So this point does not belong to the real edges. The green points in the figure are not the points of real edges. You can see the real edges in Fig.12.

After the direction of the real edges is ascertained, the direction of objects is defined as direction 0.

At last we plus the real edges and the objects to get the objects with real edges named of M_{eo}. Therefore every point of M_{eo} has a direction. You can see the result in Fig.13.

3.2   Using Features of Gray Density and Gradient Density to Expand the Result

As talked before, we have got some part of the objects (M_{eo}). You can regard them as seed points. So what we will do next is to expand the seed points. To accomplish this aim, we search the area around the seed points and get the points like the seed points. The search direction is important. If we get the accurate direction, we can save computing time, computer's memory and get precise results. The method to get the direction has been explained before, now you can see the concrete search's area in Fig.14 and Fig.15.

Then we scan every point of M_{eo}, without loss of generality named P_c and do some operations in this point's searching area. If those points in the searching area satisfy three expanding principles, they will be regarded as part of real objects. Firstly this point named P_s must belong to M_f. This principle can avoid some mistakes, such as overexpansion. Secondly we will judge the difference between the gray density of P_c and that of P_s, and the difference must be less than a threshold named T_g. At last we will judge the difference between the gradient value of P_c and that of P_s, the difference must be less than a threshold named of T_d. The last two principles are based on the same object have the similarity. After the expanding process, we finally get the result objects without shadows. You can see the Fig.16.
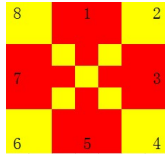
Fig. 13. M_{eo} **Fig. 14.** The searching area   **Fig. 15.** The searching area of 0   **Fig. 16.** Results after
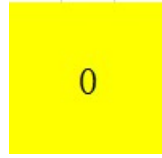
4   Experiments

You can see the flow chart of all the process in Fig.17.
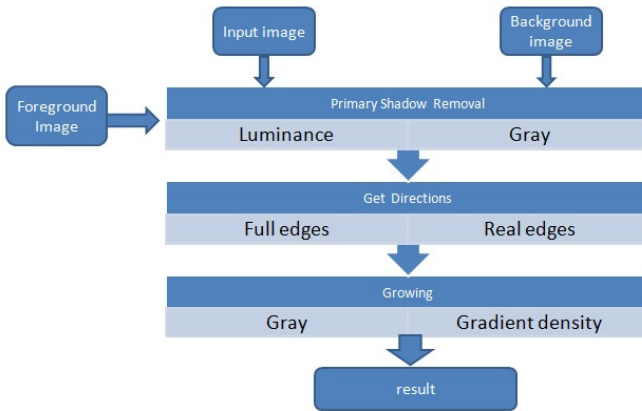
Fig. 17. Flow chart

In this part, we would demonstrate the results of our proposed shadow removal algorithm. We implemented our method on the platform of PC with Intel Core 2 CPU T5500 @ 1.66GHz 1.67GHz and 2 GB RAM. All the testing inputs are uncompressed AVI video files. The resolution of each video frame is 640×480. The average processing speed is 8 frames per second.

You can see from examples, depicted in Fig.18 and Fig.19, our method not only can remove the shadow of white objects but also can remove that of black objects. And our methods can deal with both cars and people.

Fig. 18. Examples of person, first row is current image, second row is foreground image with shadow, third row is result

Fig. 19. Examples of car, first row is current image, second row is foreground image with shadow, third row is result

References

[1] Ostu, N.: A Threshold Selection Method from Gray-Level Histograms. IEEE Transactions on Systems, Man and Cybernetics 9(1), 62–66 (1979)
[2] Onoguchi, K.: Shadow Elimination Method for Moving Object Detection. In: Pro. Int'l. Conf. Pattern Recognition, vol. 1(2), pp. 583–587 (1998)
[3] Mikic, I., Cosman, P.C., Kogut, G.T., Trivedi, M.M.: Moving Shadow and Object Detection in Traffic Scenes. In: Proc. Int'l Conf. Pattern Recognition, vol. 1(1), pp. 321–324 (1998)
[4] Lee, D.-S.: Effective Gaussian Mixture Learning for Video Background Subtraction. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(5), 827–832 (2005)

An United Security Model for Multi-Class-IS

Wang Chao and Chen Xingyuan

ZhengZhou Information Science and Technology Institute, Zhengzhou, China
Wangchao302@sina.com

Abstract. Different information systems (IS) usually do access control independently, but different systems cannot cooperate. Therefore, in the multi class information system (Multi-Class-IS), the privilege will be out of control and the security policies will not be coincident. Furthermore, the available information system only analyzes the security of the access control between subject and object and controls it. But it cannot protect the indirect information caused by access cross class and domain from leaking. This article proposes an united access control model, which can be used to analyze the security of information flow in the access cross class and domain, then control it, and finally resolve the above problem of security risk effectively.

Keywords: United access control model, Multi-Class-IS, Information flow graph, BLP.

1 Risk of Information Leakage in Multi-Class-IS

According to Classified Security Protection, Different IS should take different security protection. Our Classified Security Protection standard[1] requires that IS beyond class 3 implement s Mandatory Access Control (MAC), while Class 2 implements Discretionary Access Control (DAC). Because of different access control mechanism and model, if we share data between class 2 and class 3, information leakage rises.
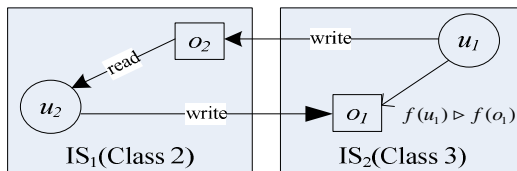
Fig. 1. Information leakage rises in Multi-Class-IS

As Fig.1 illustrates, there are two informatrion systems. IS_1 is Class 2, and IS_2 is Class 3. In IS_2 there have one user u_1 and one resourse o_1, and $f(u_1) \rhd f(o_1)$ .In IS_1 there have one user u_2 and one resourse o_2. Because IS_1 implements MAC, $u_1$can not write o_1.But if u_1 read o_2、 u_2 writes o_1、 u_2 read o_1, u_1 will send information indirectly to o_1 through o_2 and u_2, which is called "indirectly access". In fact, The "indirectly access" shown in Figure 1 would probably occur and will disobey MAC of IA_2.

The reason is that there is no united model for Multi-Class-IS. For the available access is independent, Eithor class 2 IS or class 3 IS controls user's accessing according to its own access control model.

The previous solution[2][3] is realized through unified policy inter-system. But the point is that it diverts security of access cross Multi-Class-IS to correctness of unified policy inter-system. The security depends on the coherence of access control policy of Multi-Class-Iss. And an error of access control policy will still lead to the information leakage and un-authorization access problem caused by indirectly access.   Therefore, the advantage of MAC in high-Class IS (prefer the ISs   whose class ≥ 3)would not be brought into full play, and protect ability of high-Class IS is reduced.

BLP[4] is widely used security model in High-Class IS. In BLP, (b,M,f,H)denotes a security state. b donets set of system's current access behavior, which is constituted by (subject×sbject×action). BLP model restricts the subject's access through S-property、 *-property and D-property. If access meet these properties, the system is safe. But BLP only records ervery direct-access, not the   chansfer relation of accesses, which causes indirect-access and brings the risk of information leakage. Some literature[5,6] improve on BLP, which proposes a Multi-Level model with memory capability. This model expands the security level, which is separated into independence read-write sections. They aime to prevent information flow from the object a subject have read to the object the subject will write. But in fact, most of subject's security level are invariable. Under this circumstance, the same problem does exist. Therefore, the "indirect access" problem is not solved in this literature.

The paper first consults Denning's information flow model[7], then puts forward a union security model for cross multi-Class –IS which is based on recording and analysing the informarion flow caused by operation between subject and object, and finally tries to realize the protection for cross Multi-Class-IS.

2   Description of the Model

2.1   Basic Elements

$U=\{u_1,u_2,...\}$, represents user. In Multi-Class-IS, U includes all users in different IS Class.

$O=\{o_1,o_2,...\}$, represents applied resource to protect. O includes all resources in different IS Class

$L=\{C,K\}$, represents security level. Security level is the foundation of MAC, which falls into two parts:

$C=\{C_1,C_2,…,C_n\}$, $C_1>C_2>…>C_n$. C is the classification or clearance, such as unclassified、 confidential and secret.

$K=\{K_1,K_2,…,K_m\}$，K is a set of category, such as {finance, production}.

$f(x)=L$，$x \in U$ or O, f is called level function. Given an user or resource, f will get it's security level.

$G=\{1,2,3,4,5\}$, represents class Class of IS.

2.2   Definition of Information Flow Graph

Definition 1: Information flow $\varphi = \{x_1,x_2,…,x_t\}$,$x \in U$ or O, denotes a flow of $x_1 \rightarrow x_2 \rightarrow … \rightarrow x_t$. Symbol $\rightarrow$ denotes a stream. A flow consists of many streams. The

two conjoint elements form the direct stream, while two non-conjoint elements form indirect stream. For example: $\varphi_1 = \{x_1, x_2, x_3, x_4\}$, $x_1 \rightarrow x_2$ is a direct stream, $x_1 \rightarrow x_3$ is indirect stream.

It should be noted here that a stream may be produced by read, and it may also be produced by write. For example, u_1 reads o_1, and products a stream of $o_1 \rightarrow u_1$; u_1 writes o_1, and products a stream of $u_1 \rightarrow o_1$.

Definition 2: flow policy $\#x_1 \rightarrow \#x_2$, denotes a flow policy which allow $x_1 \rightarrow x_2$. flow policy is one of the foundations of flow security. Any element in access control matrix can constitute a flow policy.

Definition 3: Flow graph T, shows the total access relation of IS. In the graph, a node represents an user or resource, a line a stream$\rightarrow$. In this paper, flow graph T is used to describe information flow relation of Multi-Class-IS, where a route denotes a flow φ.

2.3   Definition of the Security State

Definition 4: The state V of Multi-Class-IS is showed below:

$V = (U, O, T, M, f)$

Flow graph T is a set of information flows, which records the one caused by the current access, including the information flows produced by Cross-Class access. $M = \{M_1, M_2, ..., M_n\}$, is access control matrixs, which is constituted by access control matrix of different information system.

Definition 5: Rule $\rho : (R \times V) \rightarrow D \times V$, in which R is access request, $R = (S \times O \times op)$, $op \in \{r, w, ...\}$. In Multi-Class-IS, R contains not only local access requests, but also Cross-Class access requests. D is the set of respond, $D \in \{yes, no, ?, error\}$.

The interpretation for rule is: in the current state, if given a request, a respond and a next state occurs on effect of the rule. such as $(d_n, v^*) = \rho_i(r_j, v_c)$ indicates that rule ρ_i produces a responded d_n and a next state v^*, when a request r_j is asked in the current state v_c.

Theorem 1: as for $(d_n, v^*) = \rho_i(r_j, v_c)$, ρ_i is security role, only if the current state v_c is safe, and the next state v^* is also safe.

Definition 6: Automator $W = \rho_1, \rho_2, ... \circ \ \forall r_j \in R$, $v_m \in V$, $\exists$ only $\rho_i \in W$, $(d_n, v^*) = \rho_i(r_j, v_c)$. In any state, as long as there is a r_j, there is only one rule ρ_i, which lead to a respond d_n and enter the next state v^*.

Theorem 2: Automator W is safe, iff $\forall \rho_i \in W$ is safe.

Definition 7: Multi-Class-IS is denoted by (R, D, W, v_0).

3   Safe Property and Safe Axiom

Direct safe property(D-property) : V satisfies D-property，$iff \ \forall (x, y) \in \psi, \exists \#x \rightarrow \#y$
   Mandatory safe property(M-property): V satisfies M-property，
   $iff \ \forall \varphi \in T, \forall x, y \in \varphi, G(y), G(x) > 2, y = after(x), \exists f(y) \rhd f(x)$
   Transfer safe property (T-property): V satisfies T-property,
$iff \ \forall \varphi_m \subset T, A(\varphi_m) > 1, \forall x_i, x_j \in \varphi_m, i > j, \exists \#x_i \rightarrow \#x_j$
and $[f(x_j) \rhd f(x_i), if \ G(x_i), G(x_j) > 2]$

Axiom 1: $V=(U,O,T,M,f)$ refers to the secrity state , only when V meets D-property、M-property and T-property.

Axiom 2: ML-S refers to secrity system, only when initial state v_0 is safe, and so is every next state v_i.

Dedution 1: ML-S is secrity system, only when initial state v_0 is safe, and W is safe automator.

4   Rule

4.1   Read Rule

Rule 1: $dom\ R_1=(u_i,o_j,read)\in R\circ$

$$\rho_1(R_k,v_c) = \begin{cases} (?,v_c) & if\ R_k \notin dom(R_i) \\ (yes,(O,U,T\|(o_j \to u_i),M,f)) & if\ R_k \in dom(R_i)\wedge cond1\wedge cond2\wedge cond2 \\ (yes,v_c) & if\ R_k \in dom(R_i)\wedge \exists(o_j \to u_i) \\ (no,v_c) & other \end{cases}$$

$cond1$: $r\in M_{ij}$

$cond1$: $\forall \varphi_k \supset (o_j \to u_i), A(\varphi_k) > 1, \forall o_f \in forward_o(\varphi_k,u_i), o_b \in behind_o(\varphi_k,o_j),$
$u_f \in forward_u(\varphi_k,u_i), u_b \in behind_u(\varphi_k,o_j), \exists r\in M_{fb}, w\in M_{bf}$

$cond1$: $\forall \varphi_f \supset (o_j \to u_i), \forall x_b, x_f \in \varphi_k, G(x) > 2, G(x) > 2, x_f \in behind(\varphi_k,x_b)$
$\exists f(x_f) \rhd= f(x_b)$

Interpretation:

(1) if the request is no in the format of definition domain, ρ_1 dose not work.

(2) if the request is in the format of definition domain, and meanwhile matches the three conditions below, it is allowed to perform. The current state changes.

$cond1$： r is in the access control matrix.

$cond2$： In the flow graph T, as figure 2-(a) shows, in all information flows where $(o_j \to u_i)$ is used as the path and nodes belong to different IS, any forward subject node u_f of u_i has "read" competence for any behind object nodes o_b of $o_j(r\in M_{fb}$),and any behind subject node u_b of o_j has "write" competence for any forward object nodes o_f of u_i ($w\in M_{bf}$).
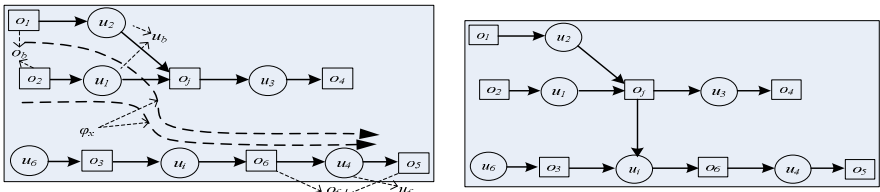
(a) Information Flow graph in current state   (b) Information Flow graph in next state

Fig. 2. Change of Information Flow graph

$Cond3$: according to the flow graph T, in all information flows φ_k where $(o_j \to u_i)$ is used as the path, , as for those nodes whose system Class is greater than 2, the security

level of any forward node x_f dominates or equals to the behind object nodes x_b's $f(x_f) \vartriangleright= f(x_b)$.

The next state is $(O,U,T \parallel (o_j \rightarrow u_i), M, f)$. Compared with the current state，the flow graph T changes and adds a line of $o_j \rightarrow u_i$,as figure 2-(b) shows.

(3) If there is the line of $o_j \rightarrow u_i$ in the flow graph, the request is allowed, and the current state does not change.

(4) In other conditions, the system will give a "no" respond, which means refusing to access, and the current state remains.

4.2 Write Rule

Rule 2：$dom\ R_2 = (u_i, o_j, \underline{write}) \in R。$

$$\rho_1(R_k, v_c) = \begin{cases} (?, v_c) & if\ R_k \notin dom(R_2) \\ (yes, (O,U,T \parallel (u_i \rightarrow o_j), M, f)) & if\ R_k \in dom(R_2) \wedge cond1 \wedge cond2 \wedge cond2 \\ (yes, v_c) & if\ R_k \in dom(R_2) \wedge \exists(u_i \rightarrow o_j) \\ (no, v_c) & other \end{cases}$$

> cond1：$w \in M_{ij}$
> cond2   ：
> $\forall \varphi_k \supset (u_i \rightarrow o_j), A(\varphi_k) > 1, \forall o_f \in forward_o(\varphi_k, o_j), o_b \in behind_o(\varphi_k, u_i),$
> $u_f \in forward_u(\varphi_k, o_j), u_b \in behind_u(\varphi_k, u_i), \exists r \in M_{fb}, w \in M_{bf}$
> cond3：$\forall \varphi_f \supset (u_i \rightarrow o_j), \forall x_b, x_f \in \varphi_k, G(x) > 2, G(x) > 2, x_f \in behind(\varphi_k, x_b),$
> $\exists f(x_f) \vartriangleright= f(x_b)$

Interpretation:

(1) if the request is no in the format of Definition domain, ρ_2 dose not work.

(2) if the request is in the format of Definition domain, and meanwhile matches the three conditions below, it is allowed to perform. The current state changes.

cond1："w" is in the access control matrix.

cond2：In the flow graph T, as figure 2-(a) shows, in all information flows where $(u_i \rightarrow o_j)$ is used as the path and nodes belong to different IS, any forward subject node u_f of o_j has "read" competence for any behind object nodes o_b of u_i ($r \in M_{fb}$),and any behind subject node u_b of u_i has "write" competence for any forward object nodes o_f of o_j ($w \in M_{bf}$).
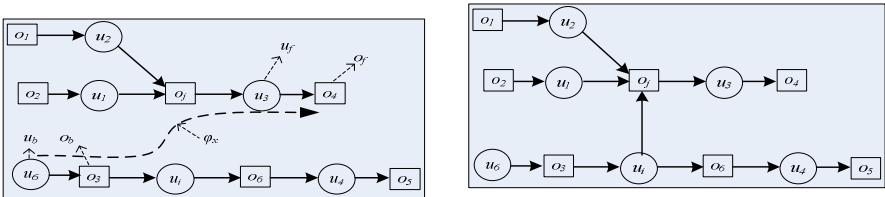
(a) Information Flow graph in current state   (b) Information Flow graph in next state

Fig. 3. Change of Information Flow graph

Cond3: according to the flow graph T, in all information flows φ_k where ($u_i \rightarrow o_j$) is used as the path, as for those nodes whose system Class is greater than 2， the security level of any forward node x_f dominates or equals to the behind object nodes x_b's $f(x_f) \vartriangleright= f(x_b)$.

The next state is $(O, U, T \parallel (u_i \rightarrow o_j), M, f)$ .compared with the current state，the flow graph T adds a line of $u_i \rightarrow o_j$, as figure 2-(b) shows.

(3) If there is the line of $u_i \rightarrow o_j$ in the flow graph, the request is allowed, and the current state remains.

(4) In other conditions, the system will give a "no" respond, which means refusing to access, and the current state remains.

5   Analysis for the Model

Through constructing the information flow graph T, the model records local and Cross-Class access, and effectively avoids security problems illustrated in figure 1. If the case that u_1 write o_2 and u_2 write o_1 happens, there will be an information flow of $u_1 \rightarrow o_2 \rightarrow u_2$ in the graph. At this time, if u_2 want to read o_1 further , *cond3* in the rule ρ_1 cannot be matched, and ρ_1 will not allow the request. After the analysis, we can see that there will never be an information flow of $u_1 \rightarrow o_2 \rightarrow u_2 \rightarrow o_1$. The model can be applied to Multi-Class_IS, with the help of which we can effectively avoid the risk of information leakage and un-authorization access cause by independency access control.

References

[1] GB 17859-1999. Classified criteria for security protection of Computer information system. Chinese Standard Press, Beijing (1999)
[2] Jiang, Y.-X., Lin, C., Tan, Z.: An Authentication Model for Multilevel Security Domains. In: IEEE International Conference on Systems, Man and Cybernetics (2003)
[3] Duan, S.-J., Hong, F., Luo, T.: Secure interoperable authorization model of multi-domain application. Journal of HuaZhong University of Science &Technology (Nature Science Edition) 11, 4–6 (2003)
[4] Bell, D.E., LaPadula, L.J.: Secure computer system: unified exposition and multics interpretation. Mitre Report, MTR-2997 Rev 1 (1976)
[5] Cai, Y., Zheng, Z., et al.: A planar attributes model based on multi level security policy. Chinese Journal of Computers 27(5), 619–624 (2004)
[6] Li, R.-X., Zhao, Z.-X., Wang, Z.G., Lu, Z.-D.: A BLP Model Based on Access History. Computer Science 7, 286–289 (2006)
[7] Denning, D.: A lattice model of secure information flow. Communications of the ACM 19(5), 236–243 (1976)

An Improved Method for Close Approach Prediction between Space Objects

Guo Xiaozhong, Yang Datao, Shen Ming, and Gao Pengqi

National Astronomical Observatories, CAS

Abstract. Based on several typical close approach prediction analysis methods, an improved method, which aimed for an simplified data-processing algorithm also easy to programming, easy to parallelization, was presented in this paper.

Keywords: Space debris, Close Approach Prediction Analysis.

1   Introduction

The space exploration cause a large number of satellites and payloads have been set into the near earth orbit, so as a lot of direct or indirect by-products of space activities, such space debris or space junks consist of everything from spent rocket stages and defunct satellites to explosion and collision fragments. The collection of objects in orbit around Earth have always grown larger and larger since the beginning of Sputnik-1, and the orbits of these objects often overlap each other's, potentially the probability of collision risks the safety of manned or robotic spacecrafts. So far a few of the collision events have been identified, one of them is the first catastrophic collision between two intact satellites(Iridium-33&Cosmos-2251) that happened in February 11th 2009 (UTC Time 16:56, February 10th 2009), which is a big challenge to space surveillance technology and collision avoidance research.

Theoretically all the close approaches between each two space objects could be predicted for a future short period of time, if the catalogue of all space objects, which were detectable for most bigger size debris, is updated continually. The Close Approach Prediction Analysis is the pre-process for Collision Probability Analysis of spacecraft, and it can provide suggestions and guides to necessary maneuver to protect manned spacecrafts or robotic satellites, also have an insight into the potential hazard during operation life time. US DoD has established a tracking network consisted of radars and telescopes to survey all the space objects. According to the data published by SSN, there are around 16,000 objects in orbit, in which only 6% are operating objects such as satellites. A "catastrophic" collision — one that causes the objects to completely fragment into debris —was estimated to take place every 12-14 years for the come decades due to the deteriorating space debris environment[1].

The problem is stated most generally by specifying a time span of interest, say $[t_B, t_E]$, and a separation distance D. As two satellites move around their respective orbits, the relative separation distance between the two is given by

$$r_{\text{REL}} = r_{\text{P}} - r_{\text{S}} \tag{1}$$

where r denotes the position vector of the satellite and where the sub-scripts P and S denote the primary and secondary satellite. Over the span of interest, a graph of r_{rel} might appear as in Figure 1. We seek all points of closest approach which are less than the threshold D(e.g., times t_1 and t_2 in Figure 1).
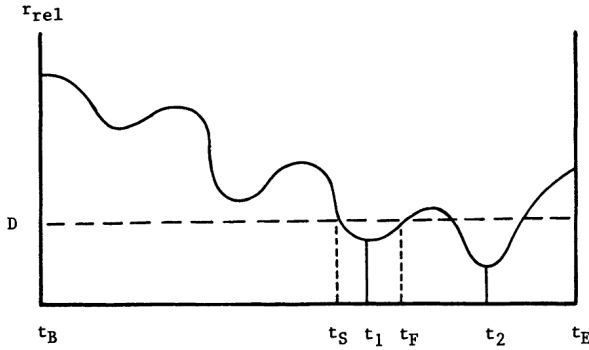
Fig. 1. Relative separation distance during some time span

Based on several typical close approach prediction analysis methods, an improved method, which aimed for a simplified data-processing algorithm also easy to programming, easy to parallelization, was presented in this paper. Then the Iridium-33&Cosmos-2251 collision event was analysis using this method as a demonstration case to show the validity.

2   Typical Close Approach Prediction Analysis Methods

The Close Approach Prediction Analysis method for space objects became a basic routine requirement for safety of manned spaceship, also the protection of robotic operational satellites. To protect or warn a satellite (called the primary) of a future collision, its orbit must be compared with every other object (called the secondary) in orbit about the Earth. The prediction time span could be several days or a week or more. Actually, the daily routine analysis may focus on many a primary object (tens of, hundreds of or even more) versus entire space debris environment which is available as the catalogue data. It is obvious that the calculation time cost on orbital data processing and close approach analysis gets much longer as the primary's number increases or for a longer time span. So it is important, as the need of space debris surveillance and warning mission to lower the computation consumption of Close Approach Prediction Analysis, speed up its prediction procedure and improve its accuracy, to use every different optimizing method to solve the same problem.

Several typical close approach prediction analysis methods have been given out by different people. Hoots[2] has firstly formulated a method using analytical techniques

to calculate the times of future close approaches between pairs of satellites based on usage of a series of pre-filters which eliminate many cases from further consideration. His method basically was consisted of the perigee-apogee pre-filter, the orbital geometrical pre-filter and the time pre-filter, and the interval of interest is usually shortened to a small subset of candidate times, of which the exactly closest approach point could be determined iteratively. ZHEN[3] have also given a similar method, which includes perigee-apogee pre-filter, geocentric distance pre-filter and time pre-filter. The test cases shown that this method can rule 99% background objects out. Klinkrad[4] applied the Close Approach Prediction Analysis on ESR-1 and ESR-2, in which a method of epoch filter, altitude filter, plane&geometry filter and orbit phase filter was proposed.

Alfano[5] gave out a new Close Approach Prediction Analysis method based on the relative distance function, which is different from the geometrical filter based methods. An algorithm for close approach determination was developed based on polynomial spline techniques. LI[6] has made an improvement upon the A-N algorithm. A self-adaptive method based on polynomial splining error theory for choosing interpolation step is presented. Comparison shows that the result of the improved A-N algorithm is more reliable than the original one.

WANG[7] adopted an improved geometrical filter to develop an algorithm, which combined a dynamic sieve and relative motion function, and use it to detect close conjunctions between a specific spacecraft and the object catalog. The results of several numerical tests indicated the new methods reduced compute time by about 10 times.
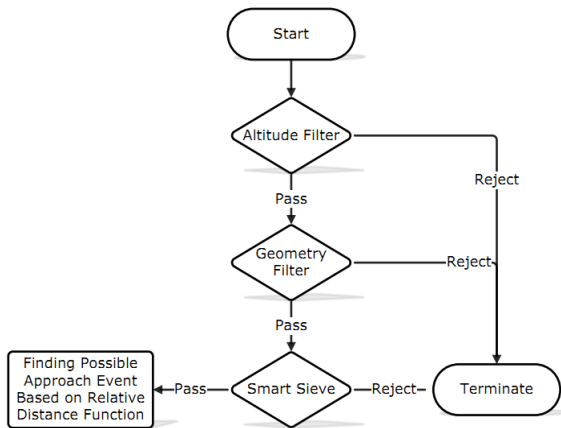
Fig. 2. One procedure of the typical methods[7]

From the above, all the methods are different from each other in the combination of different pre-filters or data process algorithm. One procedure of the typical Close Approach Prediction Analysis method was illustrated in the Fig.2.

3   An Improved Method Based on Adopted Filter Function

It is concluded that only geometry filter based methods may miss close approach events, and the relative distance function based method is more applicable for full event searching. So an improved method based on adopted filter function is presented here in this paper.

Start with the geocentric equatorial (ECI) position vectors of the primary, r_P, and secondary, r_S, satellites at time t. the relative-distance vector, r_D, become

$$r_D = r_S - r_P \tag{2}$$

With any orbit propagator, the ephemeris table for both primary and secondary satellites can been produced by a fixed time step for days long or longer. We define the adopted filter function using dot product as below:

$$f(t) = \cos \alpha = \frac{r_S \cdot r_P}{|r_S||r_P|} \tag{3}$$

We define the adopted filter function to be the cosine value so we do not need to evaluate the inverse cosine function.

Then we evaluate all the adopted filter function values for a sequence of epoch to generate a series of data. Because the cosine indicated the angle between the two positions vectors, it obvious that if the cosine is bigger than a chosen threshold, the primary and secondary satellite could be never close enough to each other in the moment. So the series of data is sequentially checked to rule the unnecessary data points out, and subsets of candidate time span are determined by the first epoch time, where cosine is under the threshold, and the last epoch time, where cosine is up the threshold.
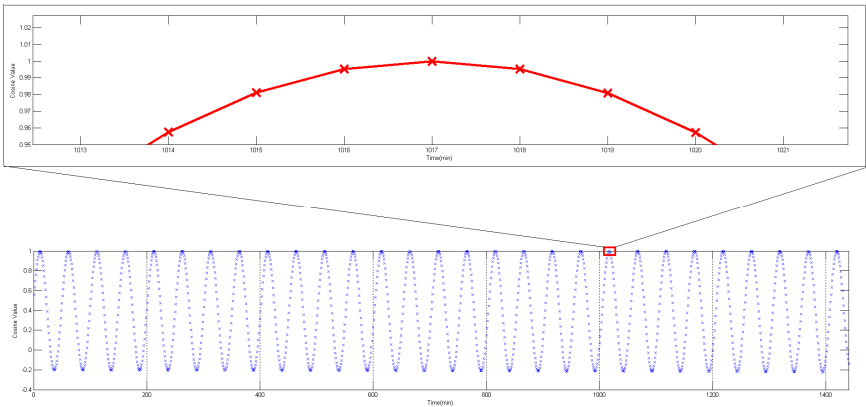
Fig. 3. The subset time span after adopted filter function processing

For an actually Close Approach Prediction Analysis method, the finally output should be the exactly epoch of the closest relative distance. So the above subsets must be passing to numerical iteration algorithm, or quick search algorithm, or spline interpolation algorithm to get the results. A fourth-order polynomial curve fitting method is suggested here to solve the searching problem; also it serves as a double filter to enhance the adopted filter function. As it is shown in the Fig.3, the whole method is really straightforward versus the others complex mathematical requirement, so it simplified the data-processing, also is easy to programming and more reliable. The series data processing is comparing the every cosine to the threshold, which is logically independent of each other, so the candidate subsets could be parallelized for a long period analysis.

4   The Preliminary Iridium-33&Cosmos-2251 Analysis

An improved method based on adopted filter function for Close Approach Prediction Analysis was presented in this paper. And we conducted a preliminary Iridium-33&Cosmos-2251 analysis by using this method. We validated our method to a brute-force step-forward prediction searching results, and the comparison is listed in the Tab.1. The orbit propagator SGP4 was used with TLE data as basic input, and the time span is 24h long begin at UTC 2009 Feb 10th. The selected TLEs were the latest available pair before the actually collision.

Table 1. Comparison of the prediction results by different method

	Brute-force method	Our method
Predicted Closest distance epoch	Feb 10th, 2009 16 : 55 : 59.795	Feb 10th, 2009 16 : 55 : 59.799

5   Summaries

An improved method for Close Approach Prediction Analysis was presented in this paper. The preliminary Iridium-33&Cosmos-2251analysis shown its prediction is acceptable as engineering application. The proposed method will be a suggest way to component of an collision risk assessment tool. But before the real application, the method also needs a more wildly case test for its validation to deal with some critical situations.

References

[1] Wright, D.: Space Debris. Physics Today, 35–40 (2007)
[2] Hoots, F.R., Crawford, L.L., et al.: An analytic method to determine future close approaches between satellites. Celestial Mechanics and Dynamical Astronomy 33(2), 143–158 (1984)

[3] Zhen, Y., Wu, L.: A Computation Method to Warn the Collision Event between Space Probe and Debris. Acta Astronomica Sinica 45(4), 422–427 (2004) (in Chinese)

[4] Klinkrad, H., Alby, F., et al.: Space Debris Activities in Europe. In: Fifth European Conference on Space Debris, Darmstadt, Germany (2009)

[5] Vallado, D.A.: Fundamentals of Astrodynamics and applications, 3rd edn., pp. 910–912. Springer, NY (2007)

[6] Li, J., Xiao, Y.: An Improved Efficient Algorithm for Close Approach Determination of Space Objects. Acta Aeronautica Et Astronaut Ica Sinica 28(sup.), S42–S48 (2007) (in Chinese)

[7] Wang, T., Huang, H.: Methods to Determine Close Approaches Among Orbiting Objects. Journal of Astronautics 29(6), 1747–1751 (2008) (in Chinese)

Construction of Viral Soft Handover Nonlinear System Model in Frequency Domain Based on Cross Entropy

Xia Wang[1], Zhiqiang Zhu[1], Shunxia Cao[2], and Fangfang Hao[3]

[1] Electrical Information Department, Shandong University of Science and Technology, Jinan, China, 250031
[2] Electronic Engineering Department, Shandong College of Electronic Profession and Technology, Jinan, China, 250000
[3] College of IEE, Shandong University of Science and Technology Qingdao, China, 266510

Abstract. Aiming at broadband CDMA mobile communication systems, around mobile users' nonlinear transmission power identification and tracking, introducing the concept of cross entropy, this paper comes up with a construction method of viral voice business soft handoff communication model in frequency domain based on cross entropy. In the point of frequency-domain model, we simulate and analyze the impact of corresponding parameters on the performance of viral mobile communication network. The simulation result shows that the normalized probability of viral mobile users MSB transmission power can reach to 37%. Meanwhile it offers a method for the application of viral voice business soft handover communication model in frequency domain in 3G and 4G CDMA mobile communication system, as well as the performance optimization.

Keywords: Viral mobile communication, Cross entropy, Soft handover.

1 Introduction

The problems about nonlinear tracking control, which supplies people abundant and effective applications, are studied extensively. For example, in the field of space technology and communication, the trajectory control of spacecrafts and attitude tracking now have become a major research direction [1]. In the field of robots many important engineering problems need to be studied furtherly, such as trajectory tracking of robots, trajectory instruction trace of homework mechanical arms and so on[2]. At present most actual mobile communication systems are of nonlinear characteristics, such as the identification and tracking for mobile users' transmission power. Therefore nonlinear control theory has always been a fairly hot research topic, but many problems still need to be solved urgently, when concrete research methods are concerned.

For the further maturity and perfection of 3G mobile communication system performance, Some new standards and technology are required to emerge and develop constantly. Due to the application of key technology such as soft handover and so

on[3][4][5], 3G voice business performance receives more mature development. Viral communication theory[6][7] will also be devoted to solving point-to-point information transmission between wireless LAN mobile terminals, trying to realize the point-to-point information transmission and exchange between mobile terminals without the participation and support of the upper devices of mobile terminals. Combining with soft handover technology, Literature[8]studies the performance of video business correspondingly, applying viral communication theory to 3G and 4G mobile communication systems. However, in the point of frequency domain, this paper will verify the feasibility of point-to-point data transmission and exchange among the moile users in viral mobile communication network, by constructing viral voice business soft handover communication model and analysing its performance, verifying the effectiveness of mobile users in viral mobile communication network carrying on point-to-point data transmission and exchange.

Generally, in the process of system performance analysis, the signal treatment method for time domain and transform domain will be involved, when transform domain consists of fourier transform and Laplace transform, which conduct signal process in the frequency domain and S domain respectively. According to the features of viral mobile communication, only focusing on Fourier transform, this paper mainly processes the relavent parameters of fast Fourier transform (FFT) in discrete systems and studies its impact on system performance.

2   Viral Soft Handover Communication Models in Frequency Domain

When analyzing the performance of CDMA viral mobile communication system, according to literature[9] ,we normalize MS_A transmission power P_{cA} in order to reflect the features of discrete signal, where P_{cA}=1.Because the transmission power of MS_B is smaller than that of MS_A, the ratio of each other can be considered a probability and the probability of normalized P_{cA} can be taken as 1.Therefore the normalized probability of viral mobile users MS_B transmission power P_{mB} in discrete systems can be expressed by the following equation:

$$p(n) = 1 - \frac{1}{P_{cA}} \{ B_{p1} - [SIR(r)_{VMC} - (E_c/I_t)_{req}(n)] -$$

$$[(N_0 \bullet B)_A - (N_0 \bullet B)_B] - 10\log[\frac{\phi_p - (E_c/I_t)_{req}(n)(1+\varepsilon)}{1 - X}]\} \tag{1}$$

In the formula (1)，B_{p1} is the power balance factor between links of MS_A and MS_B, $SIR(r)_{VMC}$ is the signal-to-noise ratio required by viral mobile communication system, $(E_c/I_t)_{req}(n)$ is the minimum required by pilot frequency, $(N_0B)_A$ is the thermal noise of MS_A, $(N_0B)_B$ is the thermal noise of MS_B, Φ_p is the proportion assigned to pilot frequency power which is part of cell power, X is   the load factor, ε is the ratio of the interference from other cells to that from the same cell, $p(n)$ is the ratio of transmission power of mobile users MS_B to that of MS_A, and the ratio is taken as a probability.

According to the formula (1), the normalized probability $p(n)$ of viral mobile users MS_B transmission power P_{mB} in frequency domain may be written simply as:

$$p(n) = f((E_c / I_t)_{req}(n)) \tag{2}$$

Taking fast Fourier transform for the formula (2) gives:

$$p(k) = FFT[\ p(n)\ ] = \sum_{n=0}^{N-1} p(n)W_N^{nk} \tag{3}$$

Where $0 \leq k \leq N - 1$.

Considering that $p(n)$ and P_{cA} are both the functions of mobile users in the formula (1), there we introduce the concept of cross entropy[10][11].As the definition of cross entropy is   the measure between two different probability distributions, according to the formula (2) cross entropy under discrete environment can be expressed as:

$$H(n) = \log[\ p(n) \log \frac{p(n)}{p_{cA}}] \tag{4}$$

Written simply as:

$$H(n) = f(\ p(n),(E_c / I_t)_{req}(n)) \tag{5}$$

Assuming $H(n)$ is the limited long sequence in CDMA discrete system, on the basis of the fast Fourier transform algorithm that the cross entropy of viral mobile users in discrete system frequency domain can be expressed as:

$$H(k) = FFT[\ H(n)\ ] = \sum_{n=0}^{N-1} H(n)W_N^{nk} \tag{6}$$

3   Performance Analysis of Video Business in Frequency Domain

According to the performance of time domain system, in order to reflect the parameters characteristics of relevant indexes in frequency domain, we have simulated the relevant performance of voice business in microcell. In order to embody this purpose, we should take fast Fourier transform for relevant parameters, simulating and analysising the relevant features when $p_{sho} = 0.400$, $p_{vmc} = 0.147$ are given.

3.1   Simulation Performance of $p(n)$ in Frequency Domain

Formula (3) is the expression for FFT transform of $p(n)$, which reflects the relations between $p(k)$ and (Ec/It)req(n). The simulation results after fitting process will be given below.
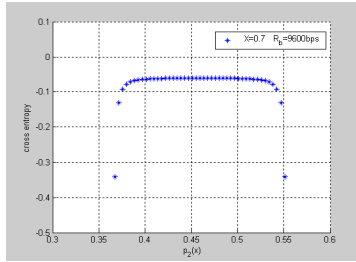
Fig. 1. The relations between *p(n)* and cross entropy in frequency domain when X=0.7

The simulation results after fitting process for the relations between the cross entropy results in frequency domain and *p(n)* are shown in fig.1 when X equals to 0.7.

From fig.1, we can see the cross entropy changes along with p(n).The curve changes gently when the cross entropy equals to -0.05,and apparently when *p(n)* is less than 0.37 or greater than 0.54.At the same time the curve changes when *p(n)* ranges from 0.37 to 0.55.Since the system is in the state of heavy load when X equals to 0.7,the result shows the corresponding relations between the cross entropy of viral mobile users in frequency-domain microcell and *p(n)* and its influence on the system performance when X is 0.7.

The simulation results after fitting process for the relations between the frequency-domain cross entropy results and *p(n)* are shown in fig. 2 when X equals to 0.5.

From fig.2, we can see the cross entropy changes along with *p(n)*.The curve changes gently when the cross entropy is 0, and apparently when *p(n)* is less than 0.28 or greater than 0.44.At the same time the curve changes when *p(n)* ranges from 0.27 to 0.46 .Since the system is in the balance state of light load when X is 0.5,the result shows that the corresponding relations between the cross entropy of viral mobile users in frequency-domain microcell and *p(n)* and its influence on the system performance when X is 0.5.
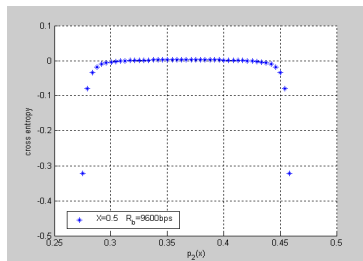
Fig. 2. The relations between *p(n)* and cross entropy in frequency domain when X=0.5

Combing picture 1 with picture 2, we can see the curve changes gently when X equals to 0.7 and cross entropy is -0.05 or when X equals to 0.5 and cross entropy is 0. The result just indicates the variations of system load. It also shows the effect of

corresponding parameters on the system performance in time domain, which is determined mainly by the changing relations between time domain and frequency domain. For example, bandwidth of the signal is wider in frequency domain when the pulse width of pulse signal is narrower in time domain.

Additionally, combing picture 1 with picture 2, we also can see the curve of frequency-domain cross entropy ranges from 0.37 to 0.55 when X is 0.7 and ranges from 0.27 to 0.46 when X is 0.5. Considering the variation of the load, we can get the range of $p(n)$ based on the values in the above two intervals. According to the interval intersection characteristics, $p(n)$ ranges from 0.37 to 0.46. The result just shows the range of the transmission power normalized probability $p(n)$, and also indicates the effect of corresponding parameters on the system performance is uniform when in time domain and frequency domain.

3.2 Simulation Performance of Cross Entropy in Frequency Domain

Formula (6) is the FFT expression for cross entropy of viral mobile users which reflects the relations between $p(n)$ and cross entropy. The simulation result after fitting process in frequency domain is shown in fig.3.
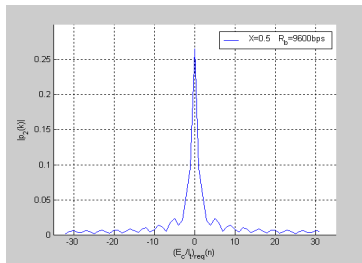
Fig. 3. The amplitude spectrum of $p(n)$ in frequency domain when X=0.5
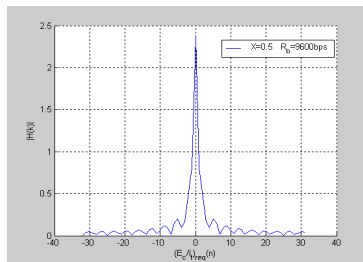
Fig. 4. The amplitude spectrum of cross entropy in frequency domain when X=0.5

According to the analysis for the performance of viral mobile communication system as well as the expression of transformation in frequency domain, we give out the simulation result of amplitude spectrum for Formula (6) shown in fig.3,when $(E_c/I_t)_{req}(n)$ ranges from -35 to 35. From the fig, we can see the normalized probability $p(k)$ of transmission power P_{mB} in microcell MS_B occuring with a sampling function change rule whose energy mainly concentrating on the main lobe .We can also see the right and left half axle shaft of the curve is symmetrical.

Similarly, due to Formula (6) the simulation result for amplitude spectrum of cross entropy in frequency domain shown in fig.4   can also be given. From the fig., we can see the cross entropy $H(k)$ in microcell occuring with a sampling function change rule whose energy mainly concentrateing on the main lobe. We also can see the right and left half axle shaft of the curve is symmetrical.

4   Summaries

Aiming at broadband CDMA mobile communication system, introducing the concept of cross entropy, this paper conductes quantitative analysis and simulation of the relations between cross entropy in frequency domain and $p(n)$ when the former reverse link balance is considered. From the result we can see that $p(n)$ may reach to 0.37, which reflects the optimizing degree of viral broadband CDMA mobile communication system MS_A and MS_B's transmission power. So supposing that we can choose the performance parameters of viral mobile communication network reasonably, broadband CDMA mobile communication can be received correspondingly, conducting point-to-point voice business data transmission and exchange between mobile users. The analysis result for the applying of viral soft handoff voice business communication models to the multimedia service of 3G and 4G CDMA mobile communication system and performance optimization has important reference value. Additionally according to voice and video business, around relay collaboration diversity and soft handover technology, how to make viral mobile communication better applied to 3G and 4G mobile communication system will be the focus of research in this field in the future.

References

[1] Marino, R., Tomei, P.: Adaptive tracking and disturbance rejection for uncertain nonlinear systems. IEEE Transactions on Automatic Control 50(1), 90–95 (2005)
[2] Liao, F., Wang, J.L., Yang, G.H.: Reliable robust flight tracking control: an LMI approach. IEEE Transactions on Control Systems Technology 10, 76–89 (2002)
[3] Gillhousen Klein, S., Jacobs Irwin, M.: On the capacity of a cellular CDMA systems. IEEE Transactions on Vehicular Technology 40(2), 303–311 (1991)

[4] Chang, S.K., Cho, H.-S., Dan, K.S.: Capacity analysis of spectrally overlaid macro/microcellular CDMA systems supporting multiple types of traffic. IEEE Transactions on Vehicular Technology 52(2), 333–346 (2003)

[5] Zhu, Z.-Q., Xu, G.-Y., Guo, J.-Q.: Soft handover algorithm analysis of overlaid cellular in CDMA system based on speed and direction. Dynamics of Continuous, Discrete and Impusive Systems, Series B: Applications and Algorithms 11, 753–756 (2006)

[6] Andrew, L., Reed, D.P.: Viral communications,
`http://dl.media.mit.edu/viral/viral.pdf`

[7] Aggelos, B., Andrew, L.: Efficient collaborative (viral) communication in OFDM based WLANs,
`http://web.media.mit.edu/~aggelos/papers/bletsas_isart03.pdf`

[8] Zhu, Z.-Q., Xu, G.-Y., Xu, L., Guo, J.-Q.: Performance Analysis of Viral CDMA Mobile Communication Based on Cross Entropy. In: The 5th International Conference on Wireless Communications, Networking and Mobile Computing, Beijing, China, vol. 1 (2009)

[9] Zhu, Z.-Q., Xu, G.-Y., Xu, L.: Performance Analysis of Viral Soft Handover Communication Based on Cross Entropy. In: The 2010 International Conference on Future Computer and Communication, Wuhan, China, vol. 1, pp. 681–685 (2010)

[10] Shore, J.E., Johnson, R.W.: Properties of cross entropy minimization. IEEE Transactions on Information Theory 27(4), 472–482 (1981)

[11] De Boer, P.-T.J., Kroese, D.P., Mannor, S., et al.: A tutorial on the cross-entropy method,
`http://web.mit.edu/6.454/www/www_fall_2003/gew/CEtutorial.pdf`

The Research of Methodology of Multi-body Parameterized Modeling for Vehicle Body Crashworthiness

Xueqin Dong and Xichan Zhu

Tongji University

Abstract. Side Crashworthiness is a main research point of vehicle passive safety. A finite model of vehicle is crashed by Ls-dyna, and body structure deformation, energy absorb, load path and velocity history are analyzed, which help to determine the main factors of the side-crashworthiness. The stiffness and inertia of the deformable barrier and the vehicle is abstracted with Ls-dyna. The corresponding multi-body system is built with MADYMO, and verified with the finite model. The results show that the side multi-body model has good correlation with the prototype finite model and can be used for further parameter study and optimization.

Keywords: Body Structure, Crashworthiness, Conceptual Design, Simulation.

1 Introduction

Now Vehicle's Crashworthiness on the increasingly high demand, but engineer had a very short design cycle under pressure in the market. So engineer should propose new design method more effective, resulting to design high Vehicle's Crashworthiness in limited time [1]. This paper presents a multi-body dynamics in the MADYMO software environment, the establishment of vehicle side collision of parametric modeling methods. MADYMO (Mathematical Dynamic Model) is a computer package that is used to model rigid multi-body dynamics and finds application in simulating various crash simulations and to assess occupant injuries. Input for the MADYMO model consisting of rigid body joint stiffness was obtained from an existing Finite Element (FE) model [2]. Model validation was done by comparing the finite model and the results correlated very well with FE data. This model demonstrates the utility of rigid body base full car models for crashworthiness analysis. Such models result in significant saving in computational time and resources.

2 Methods of Multi-body Parameterized Model

Multi-body Parameterized(MBP) model is a kind of lumped-mass spring model which is composed of rigid bodies connected by joints, So the Structural body deformation transfer to various types of joints movement [3]. Such as, the translational joint can represent Axial crushing and revolute joint can represent bending deformation etc. The rigid bodies represent undeformed areas of vehicle. The body masses and inertia data

are calculated using geometrical data from the full vehicle finite element model, Point restraints and Cardan restraints of MADYMO in 3 dimensional coordinates [4]. The shear effect of panels e.g. roof panel and floor are replaced with Kelvin restraints [5]. The joint positions and connections are chosen at the locations where structural deformation can be known crash modes shown by simulation of FE model.

   The Restraint characteristics of joints are extracted by means of finite element calculation. At present, the extraction of stiffness classified up in three ways: (1) experiment. The components to install special test machine, its quasi-static and dynamic load applied to produce the axial crushing or bending deformation, to obtain the stiffness curve of axial and bending; (2) with professional calculation software, such as DEEPCOLLAPSE; (3) extract from FE model. In the finite element software, simulation by finite element software in actual conditions same as experiment, with post-processing to obtain the corresponding stiffness curve. In this paper, the stiffness extraction method is   the last one.

   The area in which deformation occurs is cut out from full vehicle finite element vehicle model and modified to be a component non-linear finite element model. Joint characteristics are calculated from these component models. Axial crushing stiffness extraction method is bound the six degrees freedom of one component's end part, rigid Wall crushing components with uniform velocity, then we can get the data of rigid wall displacement and section force. As an example, a component model and extracted axial stiffness curves are shown in Fig.1.
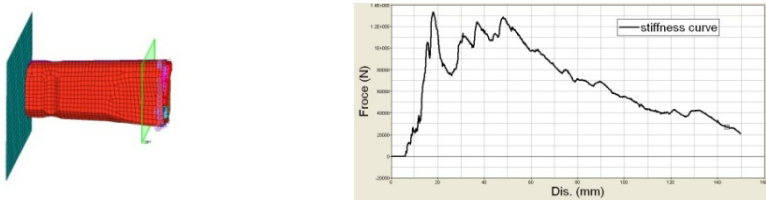
Fig. 1. Extract axial stiffness in FE model

   For calculating the bending moment, one end is simply supported and the other section is welded with a linear beam . Then the linear beam is loaded by rotating motion and the torque and rotation angle of the beam are extracted, as show in Fig.2.

Fig. 2. Extract bending stiffness in FE model

   The various stiffness curves can be described by a few important points. In the case of axial Crushing, the material yield strength determines the deformation force when structure began to crush at point A (Fig.3), and the force began to decline after crushing, until deformation displacement to point B, then nearing the limits of

compression reaction force increased rapidly to point C. So FA-DA FB-DB FC-AC (F represent Force, D represent Displacement) can get the original the stiffness curve parametric. Similarly, for bending deformation, the deformation curve is mainly as a high moment of the beginning (point D) and subsequent gradual decline (point E), and MD-RD, ME-RE are parameters of the bending stiffness curve (M represent Moment, R represent Rotation).
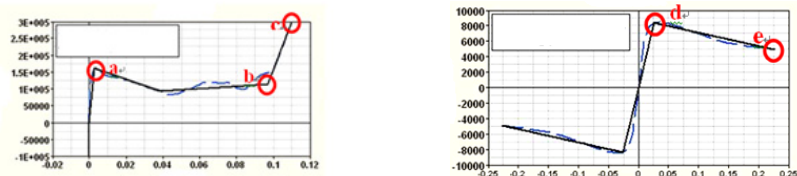
Fig. 3. Simplicity stiffness curve

The body masses and inertia data are calculated using geometrical data from the full vehicle finite element model.

3 Analyze Side Crashworthiness of FE Model

From the simulation result of FE model, we can see the deformation of B pillar and the threshold beam is largest, and the deformation can be found mainly in the following three deformation modes: (1) bending deformation, such as the B pillar bending around the x-axis, the floor beams slightly bending around z-axis; (2) torsional deformation, such as the threshold beam in bending occur simultaneously slightly torsion; (3) translation, such as the distant parts almost have no deformation, can be classified as translational motion.
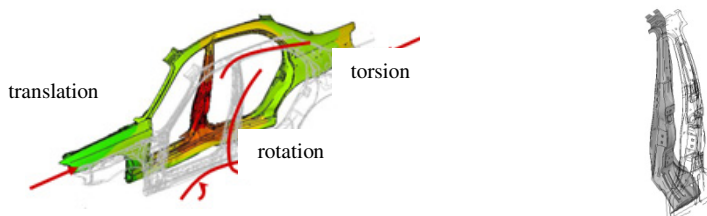
Fig.4. Deformation of side part

Near the threshold beam and middle of B pillar have taken place the plastic bending(Fig.4), this region corresponding to dummy near the chest, so the plastic joint in the middle of B pillar is harmful for side protection.

By observing the deformation and stress response changes in the simulation of FE model, can be found in the vehicle side collision, the impact force directly on the door, B pillar and other parts of the threshold (Fig.5 Fig.6), and pass to the body mainly in three directions: (1) upwards, press the A pillar and C pillar of the roof beam through the B pillar

bending; (2) downward, Moving Deformation Barrier (MDB) directly on the threshold beam in the bottom of B pillar, the impact force pass to floor beams, floor, rear seats and the central channel through the threshold beam bending, and this force transmission paths accounted for a considerable proportion of the process in the side impact; (3) impact force pass to A pillar through front door, Anti-collision beam and threshold beam, and the remaining deformation energy pass to other side of the vehicle body by IP beam.
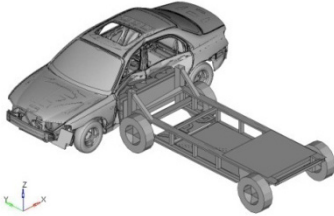
Fig. 5. Simulation of FE model
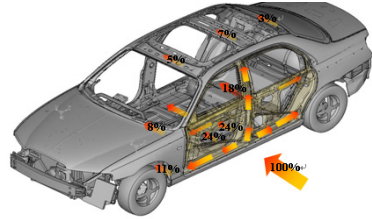
Fig. 6. Force transfer path

Vehicle body side structure and the floor is the most important part in the side impact energy absorption, so the side impact bearing structure, floor, beams and other supporting transmission part is the focus of Multi-body Parameterized Model, and the rest can be simplified as rigid. The B pillar deformation displacement, deformation rate, force pass through floor, beam and other structural as reference a for evaluating the reasonableness of the Multi-body Parameterized Model

4   Build Multi-body Parameterized Model

MBP side model contain A pillar, B pillar, C pillar, some columns, floor, threshold beam and roof beams which constitute the entire side force support structure, as show in Fig.7. It should be noted that, as is the use of multi-body modeling, therefore, If you do not set the translation joint to simulate movement of tension and compression between the two body, then it could not be demonstrated tensile and elastic character similar the simulation of FE model. So it is often have unexpected interference in the movement of whole model. Such as, a lot of compression or tension of the structure is not obvious, so it is difficult to cause plastic joint, if you don't give the appropriate degree of freedom corresponding direction of motion, then deformation and force of some structure would cause disorder.
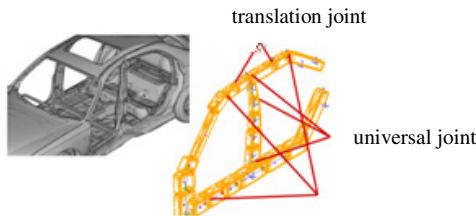
Fig. 7. Multi-body model of beam on side part

In the vehicle side impact process, the impact force of front door transfer forward to A pillar through two hinges of door, and backward to B pillar by lock. Force of rear door transfer forward to the lower B pillar through two hinges of rear door, and backward to C pillar by lock. Meanwhile the lower parts of two doors have a strong role in promoting threshold beam. The force transmission between door and other body is very complex, so how to effectively describe the deformation of door, and make a reasonable allocation of door force to the various body parts, which is the major difficulty. In this paper, the door is divided into structure and coverage of two major parts. Structure part mainly contain anti-impact beam of door and the other various beam which have a strong anti-bending ability. Coverage part mainly include door outside panel of finite element, which can provide cover plate-shaped surface, and the barrier's impact force accurate transmission to the other part. As show in Fig.8

Spherical joint
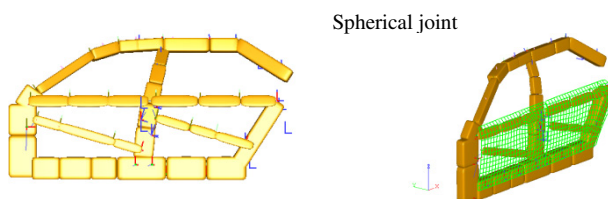
Fig. 8. Multi-body model of door

For some parts which don't deformation in the impact process, we can make it be simplified into two mass, and connected to the passenger cabin. As previously completed the frontal impact multi-body model, so the various movements of front part joints are locked, then side impact Multi-body Parameterized model in this paper can be found in Fig.9. A lot of ellipsoid, Point Restraints, Kelvin elements are composed to be the model.
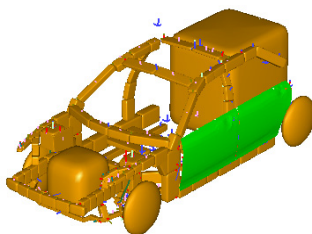
Fig. 9. Multi-body Parameterized model of side impact

5   Model Verification

The model was correlated with a full vehicle finite element model for the same crash conditions. Model validation, mainly from the following aspects: the body as a whole the results of crash deformation, B pillar deformation of shape, B pillar intrusion displacement and intrusion velocity. The deformation and vehicle velocity time history curves in Fig.10, Fig.11 and Fig.12 respectively.
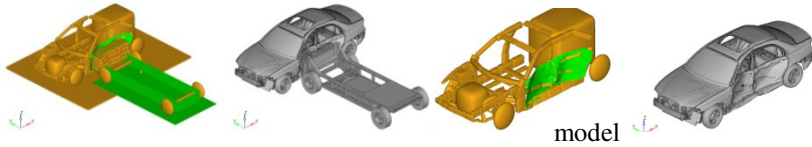
Fig. 10. Deformation trends of all vehicle comparison between Finite element and Multi-body Parameterized model
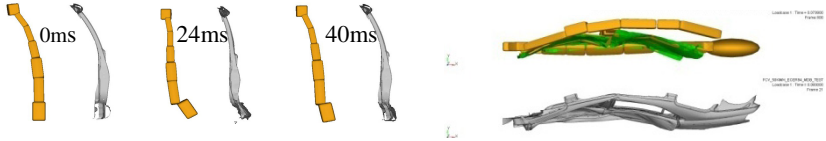
Fig. 11. Deformation trends of B pillar and door comparison between Finite element and Multi-body Parameterized model
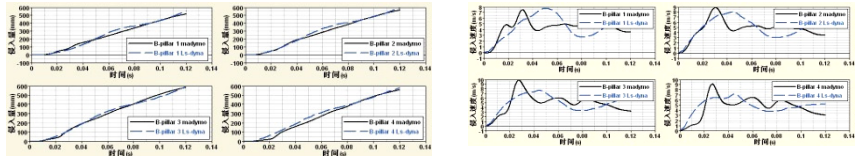
Fig. 12. B pillar velocity time history comparison between Finite element and Multi-body Parameterized model

The results can be seen from the various comparison, both of model show the consistent trends in deformation shape, intrusion displacement and velocity of some key parts, which can be identified with the validity of this multi-body model. This model can provide support for further research. Meanwhile, the establishment of such parameters of the model designer can also enhance the understanding of vehicle side crashworthiness

6   Summaries

Crashworthiness conceptual model is concise and able to show the crashworthiness essence, thus appropriate understanding of the crashworthiness performance is the necessary premise and base of building an excellent conceptual model. In this paper, we built multi-body parameterized model by MYDYMO base on the depth analysis of LS-Dyna finite element side impact model. The method of modeling process has universal. Using this method to establish the side impact parameter model can quickly analyze the key factors of the side crashworthiness for vehicle safety.

References

[1] Carrera, A.C., Mentzer, S.G., RadwanSamaha, R.: Lumped-Parameter Modelling of Frontal Offset Impacts, SAE950651

[2] Ls-Dyna Keyword User'S Manul. Livermore Software Technology Corporation, 20.119–20.224 (2003)

[3] Huibers, J., Nieboer, J.J., De Coo, P.: Design Tools for Front and Side Impact Protection. In: 5th International Madymo User's Meeting (1994)

[4] Madymo User'S Manul V6.2.2

[5] Pant, R., Cheng, J., O'Connor, C., Jackson, D., Mellireri, A.: Light Truck Concept Models and Their Applications. ESV96-S1-W-19

Active Phased Array Antenna Beam of the Integrated in Design Optimization Algorithms and Simulation

Guilin Lu, Xun Lei Wu, Shaohong Wang and Ji Hai Yang

[1] Guangxi University of Technology (GXUT),
No. 268 Donghuan Road, Liuzhou, 545006, Guangxi, P.R. China
`mathsphysics@yahoo.cn`
[2] Chinese Academy of Sciences, Institute of Electronics
North Fourth Ring Road, Haidian District, No. 19, Beijing, China
`xlwu@mail.ie.ac.cn`
[3] Ari Force No.95275 LiuZhou, China, 545006
[4] Department of Electronic Science and Technology, USTC
Hefei 230026, China

Abstract. Integrated design algorithm of quantum Control theory information and Wavelet neural network are proposed to the Loss and radar sidelobe cancellation rate, it is applicated to PDW flow simulation Of the output beam planar array and Parameters Estimation, its Improve the signal resolution, and Real-time processing speed.

Keywords: quantum Control theory, Wavelet neural network, PDW flow simulation, signal resolution, Real-time processing speed.

1 Introduction

Digital filter structure of Phased array radar system is completed by CORDIC operation mode at this stage, its Frequency conversion is very Stable, but that Reflection coefficient of Predictions of the basic standard components is completed by Lattice Prediction Filters on Standard components of the predicted reflection coefficient, it is very difficulty that Predict the reflection coefficient is completed, Feeder is not conducive to the formation of the control beam, and no Changes in the rate of scanning.in order that More reliable data are provided by Command and control systems in time, The rise of multi-target tracking algorithm that is adaptive algorithm, Multi-target tracking algorithm is risen ,it is instead of Track While Scan algorithm, its characteristic of Digital beam is adoptive by Multiple independent method of beam control, The structure of parallel processing methods can be taked by independent structures beam of the array number, Low side lobe of the beam's amplitude and phase to achieve continuously adjustable, it can reduce disturb of Specific sources of interference and sidelobe multipath main lobe.DBF technology launched can Implementation delay by clock control, The key technology can allows t o use effectively all the features of the radar .

Cancellation of the Improvement of Active Phased Array Radar on Array sidelobe need not DOA Estimation, system can automatically adjust with the output signal

structure, and need not prejudge knowledge of the exact waveform, synthesize adaptive wavelet neural network algorithm, etc, and Calculate the antenna.

2   Integrated Design Method of Cancellation of the Array Beam Sidelobe in Active Phased Array Radar

Second order cone programming approach is proposed by Yan Shefeng to optimized FIR broadband beam forming array in this year. Array beam array is be improved by the robustness of the error, But the main downside is computation of large amount, Differentiate between the high noise threshold, target azimuth resolution performance are to be improved.so we We propose using a sampling rate than maneuvering target goal of flight, it can extract the signal in all directions, and to collect all kinds of information .the theory is based on the existence of electromagnetic waves, it is always occupy a certain space, The point is combination of the smallest vector vector and panel. a Describe is necessarily that the generation of electromagnetic waves and the reasons for the nature of the charge, Magnetic field B and E, and the information into a single particle can be Captured by Divergence and curl into a physical quantity.three quantum numbers are introduced by n,l,m, Electromagnetic spectrum is used to describe wave function of the terahertz THZ material, That is

$$\psi_{n,l,m}$$

each wave function taken as wave function of zero level, Combined to get more accurate Energy expression of wave function for the spectral term,

$$E = \frac{1}{2P} \sum_{P=1}^{P} d_n^p$$

In Where, P is the number of multi-electron system of electronic, d_n^p are Sum of The first n-electron kinetic energy and potential energy, Minimum of energy function is be determined by the wavelet neural network algorithm, the energy gradient Formula is :

$$\nabla .E = i\frac{\partial v}{\partial x} + j\frac{\partial v}{\partial y} + k\frac{\partial v}{\partial z}$$

Output layer is:

$$V_k = \sum_{k=1}^{k} d_n \sum_{t=1}^{T} S_n(t)\psi(w)$$

Where the wavelet neural network output layer is equivalent to the expression (traveling salesman) the shortest path problem state function [1],

$$\psi(w) = \sin(0.75\frac{x-b}{a})\exp(-0.5(\frac{x-b}{a})^2)$$

In This time, Schrodinger equation is:

$$-\frac{\hbar}{2m}\nabla^2\psi + V(r)\psi = E\psi$$

Order

$$\psi(r,t) = \psi(r)e(-w^2)[1 - \exp(-jwr)]$$

In where, $w = \dfrac{E}{\hbar}$

The possible solution to meet the light of all the other pieces is uniform linear array of conditions, construction algorithm for the approximate solution:

$$U(n) = 2^{-1} + \sum_{i=0}^{n} 2^{-1} x(n - m1) +$$

$$\sum_{m1=0}^{p}\sum_{m2=0}^{p} 2^{-2} x(n - m1)x(n - m2) + \ldots\ldots$$

When m_1, m_2 achieve threshold set

$$V_{\varphi,m} = \sum 2^{-1} x(n)\Psi_{m1,m2}(k)$$

Approximate solution form [2]:

$$U(n) = \sum_{k} \hat{h}(0.5n - k)a_{j,n} + \sum_{k} \hat{g}[0.5n - k]d_{j,k}$$

In where,

$\hat{g}(n)$ and $\hat{h}(n)$ Known as the reconstruction filter,

It can be achieved By the digital beam (DBF) technology, the advantage is high reliability constitute active phased array surface element, Deduced the nature of the filter,

$$\Phi(w) = \prod_{j=1}^{\infty} x(\frac{w}{2^{-j}})$$

$$\varphi(w) = G(\frac{w}{2}) \prod_{j=2}^{+\infty} x(\frac{w}{2^{-j}})$$

Wavelet function can be used Neural network output layer of Structure of Basis function neural network, then [3]:

$$D = \sum_{p=1}^{p} D^p - \frac{1}{2p}\sum_{p=1}^{p}\sum_{n=1}^{p}(x_n^{2n} - y_n^{2n})$$

In where, y_n^{2n} is The actual output of the network.

3   Simulation and Experimental

For example of the ridge waveguid, Conductor transmission line is equal to the wavelength cut-off wavelength, the boundary conditions of the wave is TE wave and TM wave,

$E_z = 0$ and $E_y = 0$

Derived the relationship with the cutoff frequency, electromagnetic spectrum

$$E(w) = \sum_p E_P 2^{-\frac{j}{2}} \Psi(2^{-j}t - k)|_{j\in z, k\in z}$$

It can satisfice Maxwell equations

$$\nabla \times E = \frac{\partial}{\partial t}(\Phi(w)) = \frac{\partial}{\partial t}(x(\frac{w}{2^{-j}}))$$

$$\nabla \times H = \frac{\partial}{\partial w}G(\frac{w}{2})\prod_{j=2}^{+\infty}(\frac{w}{2^{-j}})$$

RLT filtering is be used different sampling time signal processing, active phase control radar signal by mixing said, as shown fig.1 [4]:
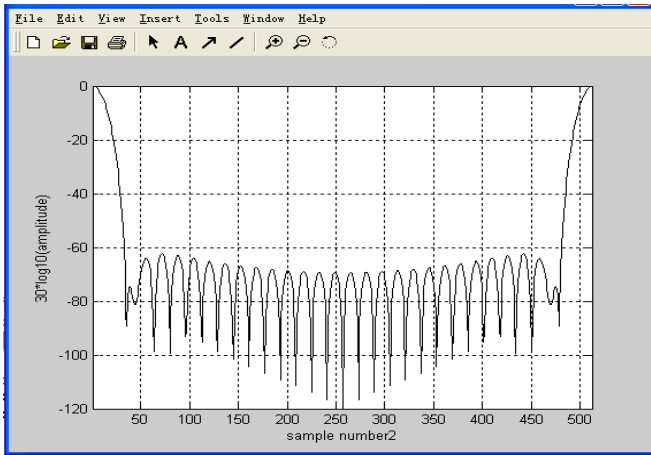
Fig. 1. Array beam

Kalman filtering method is used to establish the relationship between the current and time to design tracking filters,

$$Z(k) = H(w)X(w) + E(w)$$

In where, H , E is known matrix, $n \times n$ Victoria, $n \times p$ Victoria, $n \times q$ Victoria, When the received signal, assuming that the signal from the 100 sent to the

receiver sub-array beamforming filtering neural network algorithm module, define a target error function, we get P×q Real Matrix A, Singular value decomposition [5]:

$$A = S \sum Q^t = \sum_{i=1}^{r} \sigma_i \, sq^T$$

For real matrices, the existing CORDIC processor has been used to calculate the singular value and eigenvalue decomposition, and find the estimation accuracy of target position and velocity, we get the coordinate, transformation under the radar derived

$$x = r \sin \beta \cos \theta$$
$$y = r \sin \beta \sin \theta$$
$$z = r \cos \beta$$
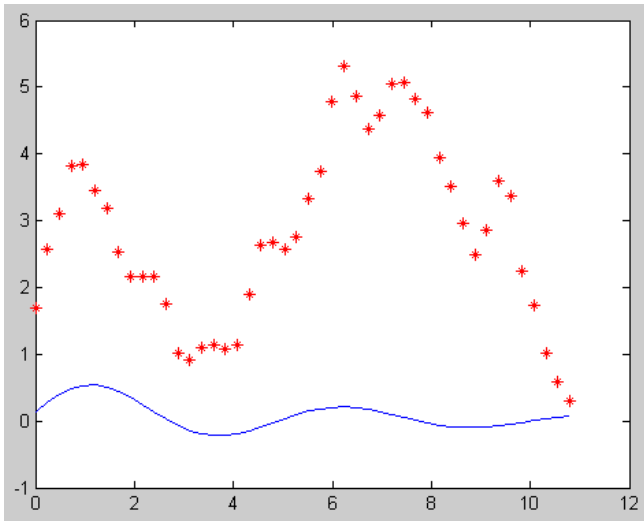
Fig. 2. Target prediction

For example figure 2, synthesize above the algorithm described ,we get derived forecast Of Objective observation, The red line shows the integrated design method and the actual target trajectory, while the Blue Line is not using an integrated algorithm, Kalman filtering results of method using a simple line drawn track ,we get Construction of wavelet basis functions,its are given neural network by selecting the differential form of Gaussian pulse, and measuring the output of the PDW stream parameter identification and parameter estimation, including Various parameters of direction of arrival, pulse differential form, it is superior than a single CORDIC processor algorithm, can resist the noise which is not interested at this time, the beam can effectively suppress the side lobe, robustness can be improved, it is clear that the proposed the algorithm is clearly very accurate, to reduce the trajectory error of the require.

4   Summary: Simulation and Experimental

Active Phased Array Radar minimum sidelobe canceller design involves the directional antenna array , and antenna array aperture distribution function, Fourier transform can be set to the antenna main beam antenna array and zero by the weighted value of each unit to determine , in order that filters have the spatial filtering characteristics, we consider that array sensitivity is a function of position and frequency, the formation of the spherical coordinate system to determine the direction of the base array receiving system, Array beam are dealt with quantum information control by wavelet neural network algorithm theory and the integrated design method, Simulation results show the Sidelobe robustness are Increase that CORDIC processors, compared with the method ,and that Flap to the sidelines to enter the interference signal is effectively suppressed.

References

[1] Zhang, M.Y.: Digital Array Radar and Software Defined Radar. Publishing House of Electronics Industry (February 2008)
[2] Yan, S.: Sensor Array Beampattern Optimization Theory with Applications. Science Press (June 2009)
[3] Ge, J.: Realization of neural network theory and MATLAB7. Publishing House of Electronics Industry (March 2005)
[4] Mahafza, B.R.: Radar systems Analysis and Design Using MATLAB, 2nd edn. Publishing House of Electroncs Industry, Beijing (2008)
[5] De Fong, Z.: MATLAB Neural Network Application Design. China Machine Press (September 2006)

A Hyrid Tabu Search for the Vehicle Routing Problem with Soft Time Windows

Liang Nai-Wen[1] and Liu Chang-Shi[2]

[1] Probability Theory and Mathematical Statistics Research Laboratory of Mathematics Science and Computing Technical College, Central South University, China 410083
[2] Management School, Hunan University of Commerce, China 410205

Abstract. The vehicle routing problem with soft time windows is a well known NP-hard problem and instances with 100 customers or more are very hard to solve optimally. A hybrid tabu search is designed in this paper for the proposed problem to minimize the total travel distance and the number of used vehicle. The proposed hybrid tabu search is tested on the data sets of VRPSTW benchmarks, the results show the high performance and effectiveness of the proposed approach.

Keywords: Vehicle routing problem, Soft time windows, Tabu search.

1   Introduction

The vehicle routing problem (VRP) is a well known NP-hard problem. The vehicle routing problem with time windows (VRPTW) is an extension of the VRP, VRPTW includes vehicle routing problem with hard time windows (VRPHTW) and vehicle routing problem with soft time windows (VRPSTW). VRPHTW deliverying goods outside the time window are not allowed at all. On the other hand, VRPSTW deliveries are still possible outside the time windows with some penalty costs [1].

Many researchers have presented lots of useful and interesting literatures for VRPHTW. In comparison, VRPSTW has received meager attention. The solution of VRPSTW has valuable practical applications [3]. In addition, VRPSTW solutions provide a workable alternative plan of action when the problem with hard time windows is infeasible [2].

Min considered a single-vehicle library distribution system with a single depot, and incorporated customer's specific time window preferences into the modeling process and provided a goal approach for the problem [8]. Koskosidis et al. formulated VRPSTW as a mixed integer program, presented a new formulation based on the treatment of the time window constraints as soft constraints [6]. Balakrishnan described three simple heuristics for VRPSTW, presented the computational results of the proposed methods on the example sets proposed by Solomon [5]. Fu et al. analysed different types of VRPSTW, presented a unified penalty function and a unified tabu search algorithm for the main types of VRPSTW [11]. Qureshi et al. (2008) presented an exact solution approach for VRPSTW, and they solved elementary shortest path problem with resource constraints and late arrival penalties as a subproblem [12].

Indeed, the objective function often drives the search towards solutions with low travel costs that may require more vehicles [4,7]. To overcome this limitation, this paper proposes a hybrid tabu search algorithm divides the search in two steps: (1) the minimization of the number of used vehicles and (2) the minimization of the total travel cost. This two-step approach makes it possible to design algorithms tailored to each sub-optimization.

The remainder of this paper is organized as follows. Section 2 gives problem assumption. Section 3 describes the proposed hyrid tabu search. Section 4 presents the computational results for our algorithm. Finally, Section 5 gives conclusions and recommendations for future research.

2   Problem Assumptions

VRPSTW can be described as follows: Let $G = (V, A)$ be a graph, the vertex set V includes the depot vertex 0 and set of customers $C = \{1, 2, ..., n\}$, The arc set A consists of all feasible arcs $(i, j), i, j \in V$. Both cost c_{ij} as well as time t_{ij} are associated with each arc $(i, j) \in A$. Time t_{ij} includes the travel time on arc (i, j) and the service time at vertex i. Every vertex of V there is an associated demand d_i, with $d_0 = 0$ [1]. There is an allowable violation of time windows in VRPSTW denoted $p_{\max} \succ 0$, The time window of each customer i can be enlarged to $(a_i - p_{\max}, b_i + p_{\max})$. In addition, an early penalty is applied if service time starts early a_i, Similarly, a late penalty is applied if service starts late b_i. The primary objective function for the VRPSTW is the minimization of the number of routes. The secondary objective is the minimization of the number of time window violations. The third objective is the minimization of total time or distance plus penalties for early or late deliveries [2].

In a broader sense, a solution to the VRPSTW is a routing plan $\sigma = \{r_1, ..., r_m\}$ $(m \leq n)$ satisfying the capacity constraints and time window constraints. The size of a routing plan σ, denoted by $|\sigma|$, is the number of non-empty routes in σ, i.e., the number of used vehicles because each route is served by only one vehicle in this paper. Let $t(r)$ denote the travel cost of a route, the VRPSTW problem consists of finding a solution σ which minimizes the number of used vehicles and, in case of ties, the total travel cost, i.e., a solution σ minimizing the objective function specified by the lexicographic order

$$f(\sigma) = \left\langle |\sigma|, \sum_{r \in \sigma} t(r) \right\rangle \qquad (1)$$

3   The Tabu Search

Our hybrid tabu search (TS) uses and explores most of the more important concepts in TS, namely, the following: initial solution, minimizing vehicle number, neighbourhood structure, tabu list, moves and evaluation of the moves, fre- quency-based memory, stopping criterion.

3.1   Initial Solution

The set of customers to be served by each vehicle is found by using the sweep algorithm. It is assumed that each customer i is located with angle θ_i from the horizontal axis originating from the depot to the right. Starting from the customer arbitrarily chosen, for example, the customer with the smallest value of θ_i, customers are assigned to each vehicle in increasing order of hi as many as possible until the capacity is allowed. Depending on the starting point, and the rotating direction, clockwise and counter-clockwise. By this approach, $2(m-1)$ kinds of clustering can be generated at most.

3.2   Minimizing Vehicle Number

As mentioned by Bent and Van Hentenryck in 2006, the objective function $f(\sigma) = \left\langle |\sigma|, \sum_{r \in \sigma} t(r) \right\rangle$ is not always appropriate, since it may lead the search to solutions with a small travel cost and makes it impossible to remove routes. To overcome this limitation, our simulated annealing algorithm uses a more complex lexicographic ordering

$$e(\sigma) = \left\langle |\sigma|, -\sum_{r \in \sigma} |r|^2, \sum_{r \in \sigma} t(r) \right\rangle \tag{2}$$

especially tailored to minimize the number of used vehicles [9]. The first component is, of course, the number of the vehicles used. The second component maximizes $\sum_{r \in \sigma} |r|^2$ which means that it favors solutions containing vehicles with many customers and vehicles with few customers over solutions where customers are distributed more evenly among the vehicles. The intuition is to guide the algorithm into removing customers from some "small" vehicles and adding them to "large" vehicles (see insertion heuristic). Components of this type are used on many problems, a typical example being graph coloring proposed by Tang and Galvao in 2002[10]. The third component minimizes the travel cost of the solution.

3.3   Neighbourhood Structure

Neighbours of the current solution are generated by modes exchange and relocation.

3.3.1   Nodes Exchange

Exchange the execution mode of current route with every other feasible mode in the other routes to create a set of neighbouring solutions, each of them differing from the current solution.

3.3.2   Relocation

This movement consists of removing a node from one of the routes and including it in another route. The movement is in principle attempted for all possible pairs of routes and for each node in each of the two routes.

3.4   Recording in Tabu List

Whether two customers had performed exchange or not can be held in the tabu list. For example, a move $v = (i, j)$ is added to tabu list T in the following standard way. The tabu list T is shifted one position forward and put v in the last position in the list, that is, $T_q = T_q + 1$, $q = 1, 2, ..., \max q - 1$ and $T_{\max q} = v$. In this paper, the length of tabu list is set to eighteen and first-in-first-out rule is adopted; that is, when the tabu list is full, the new move replaces the earliest one entering tabu list and adds the maximal searching times by one.

3.5   Short-Term Memory

A recency-based memory structure was implemented to avoid cycling. Every movement is characterized by two sets of edges, which define its attributes: ① edges to be inserted into the solution; ② edges to be removed from the solution. The short-term memory maintains a list of the edges that were used (inserted or removed) in the recent past. The same tabu list is used to save both inserted and removed edges, since it is always possible to identify if an edge was inserted or removed by observing the current solution.

   A movement is considered tabu when all the edges involved are tabu in the current iteration. This is not so restrictive, in the sense that movements can be made even when some (but not all) of the associated edges are classified as tabu.

3.6   Stopping Rules

   ① No feasible movement exists.
   ② Maximum number of iterations for the corresponding phase was reached.

4   Computational Results

The hybrid tabu search (HTS) has been implemented in Matlab and run on a personel computer with 1 G memory and 2.26 GHz processor. The parameter of the HTS is maxiteration=2000. We now present the test results conducted on instances of VRPSTW benchmarks was proposed by Fu et al. (2008) [11] assuming that the late

delivery time window can be violated with a penalty coefficient for late time windows that is equal to 100, i.e. a unit of time of time window violation is equal to 100 units of distance traveled. The soft time window for late deliveries can be extended up to the depot closing time. Our test instances named R101, R102, R103, R104, R105, R106, R107, R108, R109, R110.

After concluding the computer experiments, the first question to be answered was the efficiency of the proposed algorithm.

We compare our HTS with an iterative route construction and improvement algorithm (IRCIA) presented by Figliozzi et al. [2] for VRPSTW, the results are shown in table 1. "AVTTD" is the average value of the total travel distance over the ten runs; "VN" is the vehicle number.

Table 1. Results of HTS and IRCIA

problem	HTS		IRCIA	
	AVTTD	VN	AVTTD	VN
R101	1482.4	13	1535.2	14
R102	1397.2	13	1416.8	13
R103	1246.3	11	1267.3	11
R104	963.8	9	983.5	9
R105	1406.9	12	1441.2	13
R106	1326.7	11	1355.3	11
R107	1128.4	10	1147.6	10
R108	964.5	9	978.7	9
R109	1234.1	11	1264.2	11
R110	1067.3	11	1084.0	11

From tab.1 we know that our HTS is very effective to deduce the total travel distance for all test instances. On the other hand, it is not easy to deduce the vehicle number. Only the vehicle number of instance R101 and R105 is decreased.

5   Summaries

VRPSTW is a well known NP-hard problem, but has many practical applications. In this paper, a hybrid tabu search is designed to deduce the vehicle number and total travel distance, the computational results show that the proposed approaches is very effective and competitive. It is easy to deduce total travel distance, but not easy to deduce the vehicle number.

The real world is dynamic, thus, effective and effcient decision support tools are needed to address real word vehicle routing problems. So, in the future, we can exploit the solutions provided by fast heuristics to help logistics enterprises to cut down cost.

References

[1] Qureshi, A.G., Taniguchi, E., Yamada, T.: An exact solution approach for vehicle routing and scheduling problems with soft time windows. Transportation Research Part E, 960–977 (2009)

[2] Figliozzi, M.A.: An iterative route construction and improvement algorithm for the vehicle routing problem with soft time windows. Transportation Research Part C, 668–679 (2010)

[3] Chiang, W.C., Russell, R.A.: A metaheuristic for the vehicle-routing problem with soft time windows. Journal of the Operational Research Society 55, 1298–1310 (2004)

[4] Homberger, J., Hermann, G.: A two-phase hybrid metaheuristic for the vehicle routing problem with time windows. European Journal of Operational Research, 220–238 (2005)

[5] Solomon, M.M.: Algorithms for the vehicle routing and scheduling problems with time windows constraints. Operations Research 35, 254–265 (1987)

[6] Koskosidis, Y.A., Powell, W.B., Solomon, M.M.: An optimisation-based heuristic for vehicle routing and scheduling withsoft time window constraints. Transportation Science, 69–85 (1992)

[7] Bent, R., Hentenryck, P.V.: A two-stage hybrid local search for the vehicle routing problem with time windows. Transportation Science, 515–530 (2004)

[8] Min, H.: A multiobjective vehicle routing problem with soft time windows: The case of a public library distribution system. Socio-Economic Planning Science, 179–188 (1991)

[9] Bent, R., Hentenryck, P.V.: A two-stage hybrid algorithm for pickup and delivery vehicle routing problems with time windows. Computers & Operations Research 33, 875–893 (2006)

[10] Tang, F.A., Galvao, R.D.: Vehicle routing problems with simultaneous pick-up and delivery service. Journal of the Operational Research Society of India 39, 19–33 (2002)

[11] Fu, Z., Eglese, R., Li, L.Y.O.: A unified tabu search algorithm for vehicle routing problems with soft time windows. Journal of the Operational Research Society, 663–673 (2008)

[12] Qureshi, A.G., Taniguchi, E., Yamada, T.: An exact solution approach for vehicle routing and scheduling problems with soft time windows. Transportation Research Part E, 960–977 (2009)

Transistor Level Modeling for Circuit Simulation

Gao Jin, Duan Zhe-min, Yin Xi-peng, and Gao Wei

School of Electronics and Information, Northwestern Polytechnical University, Xi'an, China

Abstract. During the development of servo system, there are some problems in circuit simulation, particularly at the transistor level. One way of solving this problem is to use a consistent modeling and simulation framework that allows for the description of systems from different disciplines and for the description of interactions between these systems. To find the advantage of the VHDL-AMS language for analog system, this modeling method is also effective to describe the mixed-signal system's structure and action.

Keywords: Transistor Level Modeling, VHDL-AMS, Circuit Simulation.

1 Introduction

According to the signal flow, the structure of servo control circuit can be divided into three parts: Proportional-Integral-Derivative (PID), Pulse-Width Modulation (PWM) and power output.

The accuracy of the circuit simulation performed in contemporary design flows is directly correlated to the quality of its fundamental components-the models. To ensure on-time delivery of simulation, these model's characterization and generation must be rapid and precise.

In this paper, a 'top-down' methodology will take advantage of servo system design. Throughout the servo system circuit design, how VHDL-AMS can be effectively used for the top-down design methodology has been shown.

2 Process Simulation

In EDA language, 'top-down' design means engineer can divide the top level design into a number of models, without go through all details of each model. Then engineer can design each model, which may involve defining more underlying sub-modules. The process should stop when engineer have designed all the underlying sub-modules. The basic idea is to build a hierarchy which starts with fewer details on the top and more details as engineer goes down the hierarchy. Because simulation is always accompanying with design, engineer can finish his work rapidly and efficiently.

The behavior and properties of all semiconductor devices are defined by their three geometrical dimensions and concentration profile of impurities [1].The main goal of process simulation is to model a virtual device with geometry and properties identical with the real structure. The advanced physical models with calibrated parameters characterizing individual fabrication steps are integrated into the process simulators.

To a servo system, there are some non-liner parts such as valve in the whole configuration. These non-liner functions can be done with machine. If non-liner parts are ignored here, the control circuit is shown as Fig.1.
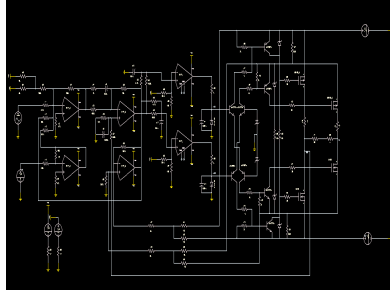
Fig. 1. The Control Circuit

After the control circuit has been designed, this model can be transferred from MATLAB to SPICE. Then it will be performed as wish for alternating current, direct current, Fourier and Monte Carlo analyses.

3   Problems in Circuit Simulation

Though SPICE is very useful for simulation, it still has a problem in the simulation progress. Because the device's SPICE model highly relates to the device's technique, the industry may only declare its ordinary and old device's SPICE model. But in a high-tech product, modern device can be found on its PCB everywhere. So the absence of device's SPICE model is a bottleneck with simulation [2].

VHDL-AMS is a language to describe digital, analog and mixed-signal systems. It is an extension of VHDL. It is designed to fill a number of needs in the design progress [3]. In the first, it allows description of the structure of a system, that is, how it is decomposed into subsystems and how those subsystems are interconnected. In the second, it allows the specification of the function of a system using familiar programming language and equation forms. In the third, as a result it allows the design of a system to be simulated before being manufactured, so that designers can quickly compare alternatives and test for correctness without the delay and expense of hardware prototyping. At last, it allows the detailed structure of a design to be synthesized from a more abstract specification, allowing designers to concentrate on more strategic design decisions and reducing time.

A model of an analog system consists of the circuit nodes, analog unknowns to be calculated and the characteristic equations that specify analog behavior. In VHDL-AMS, designers use terminals to represent the circuit nodes, quantities for the analog unknowns and simultaneous statements for the characteristic equations. So it can describe an analog system perfectly.

4   Transistor Level Modeling with VHDL-AMS

Saber is the multi-language simulation software of Synopsys. It allows SPICE circuits and models to be integrated into the VHDL-AMS environment in two ways. At the First, a VHDL-AMS model may include SPICE sub-circuit as a part of components; second, a SPICE net-list may include VHDL-AMS design entities as a part of components.

In Fig.1, the typical PID and PWM circuit can run by Saber because their SPICE model can be found to perform circuit simulation. But the MOSFET model of IRF150 in power output circuit doesn't exist in the library. So we try to model at the transistor level with VHDL-AMS. The complete information of large-signal vector measurements allows the direct extraction of the large-signal equivalent scheme [4]. The IRF150 Characteristics can be understood from the $I_{ds}(t)$ versus $V_{ds}(t)$ trajectory in Fig.2.
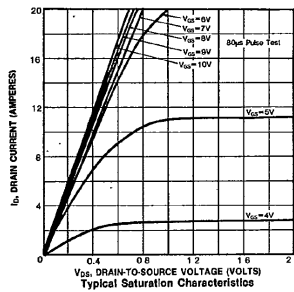
Fig. 2. IRF150 Characteristics

It can be expressed as

$$I_{ds} = I_{dsi}(V_{gsi}, V_{dsi}) + C_{21}(V_{gsi}, V_{dsi})\frac{dV_{gsi}}{dt} + C_{22}(V_{gsi}, V_{dsi})\frac{dV_{dsi}}{dt} \tag{1}$$

The following simulation model of IRF150 is developed by VHDL-AMS, and it passed the compiling in Saber:

```
library ieee;
    use ieee.math_real.all;
library ieee_proposed;
    use ieee_proposed.energy_systems.all;
    use ieee_proposed.electrical_systems.all;
    use ieee_proposed.thermal_systems.all;
entity irf150 is
 generic (
   MICRO:real:=1.5e-7;
   MILLI:real:=1.1e-7;
   WEF:real:=1.0*MICRO;   --effective channel width
   LEF:real:=0.15*MICRO; --effective channel length
   VT0:real:=0.5;                --long channel threshold voltage
   RDS:real:=0.04;           --drain source resistance
   PHI:real:=0.97;               --bulk Fermi potential
```

```
  GMA:real:=0.71;            --body effect parameter
  KP:real:=487.0*MICRO; --transconductance parameter
  THETA:real:=50.0*MILLI; --mobility coefficient
  TCV:real:=1.5*MILLI; --temp.coef.of threshold voltage
  BEX:real:=-1.5;    --temp.coef.of transcond parameter
  );
port (
    terminal td,tg,ts,tb:electrical;
    terminal tj:thermal);
end entity irf150;
architecture ideal of irf150 is
  constant kOQ:real:=K/Q;
  constant TEMPREF:real:= 269.3;
  quantity vg across tg to tb;
  quantity vd across td to tb;
  quantity vs across ts to tb;
  quantity ids through td to ts;
  quantity igs through td to ts;
  quantity gpower through thermal_ref to tj;
  quantity temp across tj to thermal_ref;
begin
  vt:=KOQ*temp+1.0e-6;
  ratio:=abs(temp/TEMPREF+1.0e-6);
  vds:=ids*RDS;
  vto_th:=VT0-TCV*(temp-TEMPREF);
  kp_th:=KP*(ratio**BEX);
  vgp_0:=vg-vto_th+PHI+GMA*sqrt(PHI);
  vgp:=0.5*(vgp_0+sqrt(vgp_0*vgp_0+1.0e-3));
  vp:= vgp-PHI-GMA*(sqrt(vgp+0.25*GMA*GMA)-0.5* GMA);
  iff:=i_v((vp-vs)/vt);
  irr:=i_v((vp-vd)/vt);
  beta:=kp_th*(WEF/LEF)*(1.0/1.0+THETA*vp));
  n:=1.0;
end architecture ideal;
```

After VHDL-AMS modeling, the simulation wave of IRF150 of power output circuit in Saber is shown as Fig.3.
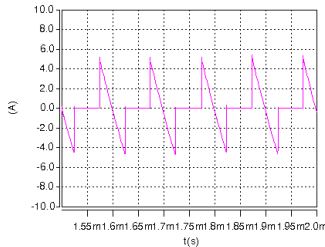
Fig. 3. Simulation wave of IRF150

5   Circuit Verify

After simulation, corrective actions are added into the actual servo circuit. The actual output wave is shown as Fig.4, exactly suited to simulation results (because the different testing terminal, the scope is shown as mirror style).
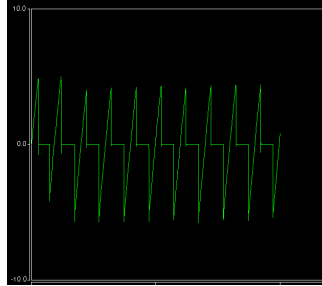
Fig. 4. Actual output wave

6   Summaries

After simulation in analog circuit can be described by VHDL-AMS language in Saber, simulation result will be perfect showed. Because it is easy to use mode in VHDL-AMS than SPICE, it is very helpful to find a solution, which simulation model is not enough for the transistor level in the analog circuit, but for the mixed-signal circuit.

Because the absence of simulation models, the design and verification of analog and mixed-signal systems are difficult and complex. The progress of top-down design for these types of systems is less straightforward than for digital systems. VHDL-AMS model can play an important role in it.

References

[1] Attia, J.O.: PSPICE and MATLAB for electronics, pp. 127–128. CRC Press, London (2002)
[2] Ashenden, P.J.: The System Designer's Guide To VHDL-AMS, pp. 534–536. Morgan Kaufmann Publishers, San Francisco (2003)
[3] Williams, T.: The Circuit Designer's Companion, pp. 76–79. Publishing House of Electronics Industry, Beijing (2006)
[4] Scheffer, L.: EDA for IC System Design, Verification and Testing, pp. 313–315. CRC Press, London (2006)

Expert System for Forest Type Interpretation on Aerial Photographs

Yin Ying-ji and Wang Ni-hong

Information and Computer Engineering College
Northeast Forestry University
Harbin, China
`yinyingji2005@126.com, wnh@mail.nefu.edu.cn`

Abstract. How to transit from visual interpretation for forest type to computer interpretation is an important and meaningful question. Through making an intensive study of operating mechanism, system structure and core components of the system: knowledge base, inference engine, this study came up with the basic framework of forest type interpretation expert system, gave the method of designing the expert system and implemented the initial function according to user-oriented design.

Keywords: remote sensing, interpretation for forest type, expert system, database.

1   Introduction

With the development of science and technology, whether forest resources dynamic monitoring or forest management work, is dependent on remote sensing data to interpret and analyze forest type and land type. In all the interpretations for forest classification, interpretation for forest type is the most difficult, but it is very important. Master a variety of tree forms and imaging features on aerial photographs is the key of species interpretation. Visual interpretation for image data has long been the main technology of remote sensing. Visual interpretation is the observation by the eyes and through the brain to make a quantitative evaluation for the spatial images of the scene. However, visual interpretation technique has some shortcoming. It requires extensive training and is meticulous work. In addition, some spectral features can't be determined by visual interpretation alone. In such a case, the best method is using image data analysis and image brightness value matrix, qualitatively analyzing these data that can be applied to the computer of the land type through a quantitative method. Therefore, at this stage, we should seek out a comprehensive interpretation approach for remote sensing data based on the two methods. The system combines partial function and principle of the visual interpretation and computer interpretation system. It relies on visual observation method to extract image features and quantify these characteristics into the computer-compatible digital form, which can be sensed by the system, then process the quantified image data to complete the identification process [1].

The existing methods of remote sensing image interpretation system can be classified into three types of interpretation systems:

- Visual interpretation system based on people themselves (interpretation people). Visual interpretation system is through the eyes and brain

understanding and processing the information on remote sensing image. Through getting the image features by visual observation and establishing thinking, interpretation people link the image features observed with the stored memory of the relevant interpretation knowledge, concepts and interpretation rules to synthesize, analyze and complete the identification process.

- Interpretation system based on computer and assisted equipments. The interpretation system can obtain the perception of the remote sensing image through determinating the optical density of remote sensing image by optical densitometer, camera, or directly inputting digital image data from the data file. The system has artificial intelligence to perform and complete the identification task after owing the knowledge base or making the basic structured model constituted with the interpretation rules into computer software.

- Comprehensive interpretation system. This interpretation system consists of some structure features of visual interpretation system and computer interpretation system. It relies on visual method to observe the image features and to quantify these features into the computer-compatible digital form, making the system perceived. The computer runs the artificial intelligence software and processes the quantified image data, at last completes the interpretation process.

Here, I adopted the third interpretation system. Visual interpretation factors are defined as the ground-object morphology features identified from the images when we do the visual interpretation. In other words, interpretation people should focus on the interpretation factors which are obvious enough to analyze. The main interpretation factors are: crown shape, crown size, tree elevation difference, tone, crown density uniformity, slope and aspect.

In addition, the shape and the angle of the objects observed on the ground are different with when we observed on the aerial photographs. On the ground it is based on the side shape of trees to identify the forest type, while the morphological characteristics of the upper part of trees are reflected on the aerial photographs. Interpretation people must adapt themselves to this observation way and establish the thinking associated with the form of the actual objects.

2   Study Area Selection and Data Acquisition

The study area is Laoshan circle unit of Maoer Mountain forest farm attached to Northeast Forestry University. I adopted the Digital Ortho photo Map (DOM) after being corrected and mosaic and eleven aerial photos were taken in September 2003 at Laoshan circle unit distributing in three strips. The scale of the photographs was approximately 1:15000. Sub-Compartment information is acquisitioned from the 2004 forest management survey data. The source data of retrieval samples information table which will be used in the later experiment is acquisitioned from Sub-Compartment information.

3   Research Steps

3.1   The Overall Structure and Function of the Expert System

The expert system consists of 5 parts: knowledge base, inference engine, integrated database, interpreter and interface. Integrated database stores initial data from specific domains or questions and middle data from reasoning process, which are some current facts of being processed objects. Interpreter can explain the behavior of expert system to
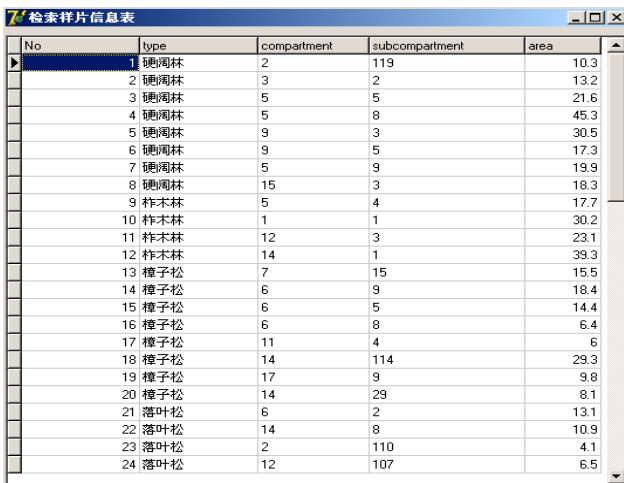
the users, including the correctness of the conclusions and the reason of outputting other candidate solutions. Interface is capable of establishing a dialogue between system and users, allowing users to input the necessary data, ask questions and understand the reasoning process and results. System asks the users to answer questions through the interface or answers the questions by giving some necessary explanations which are asked by the users. As for the establishing process of knowledge base and inference engine, will be explained in detail in the next section [2].

3.2   Establish Knowledge Base

In order to build knowledge base, we must solve the problems of knowledge acquisition and knowledge representation. Knowledge acquisition is to collect and assort the specialized knowledge of solving problems. Knowledge acquisition is instead of knowledge engineers doing automatic knowledge acquisition to achieve self-learning of expert system and constantly improves the knowledge base in expert system knowledge base construction, including sample plots selection, discriminating factors selection, making discrimination reaction table. Knowledge is stored, used and modified in appropriate representation. Production rule representation is used most widely, which is satisfying the following condition statement as "if this condition is satisfied (provided), then to take the action (conclusion)". The expert system uses production rule to represent knowledge [3].

1)   Knowledge Acquisition

a) Sample plots selection: The representative forest types in Laoshan circle unit of Maoer Mountain Forest Farm were divided into eight categories: hard broad-leaved forest, Oak lumber, Pinus sylvestris, Larch, soft broad-leaved forest, rare broad-leaved forest, artificial and natural mixed forest, Birch wood. Select 10% of each category as sample plots at random, and make them as interpretation samples, then store the information about the interpretation samples into the database as retrieval samples information table shown in the figure 1 below.

No	type	compartment	subcompartment	area
1	硬阔林	2	119	10.3
2	硬阔林	3	2	13.2
3	硬阔林	5	5	21.6
4	硬阔林	5	8	45.3
5	硬阔林	9	3	30.5
6	硬阔林	9	5	17.3
7	硬阔林	5	9	19.9
8	硬阔林	15	3	18.3
9	柞木林	5	4	17.7
10	柞木林	1	1	30.2
11	柞木林	12	3	23.1
12	柞木林	14	1	39.3
13	樟子松	7	15	15.5
14	樟子松	6	9	18.4
15	樟子松	6	5	14.4
16	樟子松	6	8	6.4
17	樟子松	11	4	6
18	樟子松	14	114	29.3
19	樟子松	17	9	9.8
20	樟子松	14	29	8.1
21	落叶松	6	2	13.1
22	落叶松	14	8	10.9
23	落叶松	2	110	4.1
24	落叶松	12	107	6.5

Fig. 1. Retrieval samples information

b) Discriminating factors selection: Select seven factors to reflect the differences between various types. The seven factors are crown size, crown shape, tone, tree elevation difference, crown density uniformity, slope and aspect. In the further study I found:

The crown of Oak lumber, Pinus sylvestris is small or medium occasionally, while the crown of some kinds of broad-leaved tree such as Ash, Manchurian walnut is bigger. Therefore, the crown size is divided into three levels as large, medium and small, which can be identified by eyes and have a small error rate. Crown shape can be divided into round, oval, star, irregular shape. For example, crown shape of Larch and Pinus sylvestris is round, Oak trees irregular because of closely arranging, broad-leaved forest oval.

Tone is a difficult factor to grasp and many factors can affect it. Overall, the tone of the forest consisting of many tree species is gray and white, because the reflectivity from the tree species is different. In addition, the tone is divided into dark gray, gray and light gray, which is relatively easy for human eyes to identify.

Different forest types have different internal structure laws. Therefore, we must choose some factors which can reflect the internal structure as the interpretation factors. The crown density uniformity of different forest types is divided into tufted, tight, porosity with blankets and porosity without blankets. Another factor is the tree elevation difference, which can show the composition of forest type to some extent. If one tree species only, tree elevation difference is small, otherwise it is large. This factor is also divided into three categories as large, medium and small.

As for slope and aspect, they are also affecting the distribution of forest types to some extent. Slope is divided into four categories as upper, middle, lower and valley. Aspect is divided into three categories as large, medium and small.

c) Make interpretation samples: The mosaic image of Laoshan circle unit was coincided with sub-compartment vector image in ERDAS remote sensing image processing software. Then, search the region corresponding to the forest type, according to the vector attribute information table. Using ✍ in the AOI (Area Of Interest) menu to draw sub-compartment and save the picture as .IMG. At last, by using picture format conversion function, the image is converted to .JPG and intercepted as the picture shown in the figure 2 below.
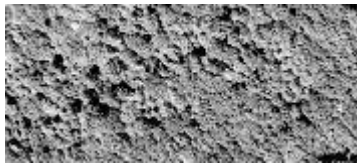
Fig. 2. Sample of hard broad-leaved forest

From the graph, we can see the crown size, crown shape, tone, tree elevation difference and other information. The collected samples had also been stored into database, convenient for beginners to do massed learning.

One problem here is that regardless of how random the interpretation samples are selected there is still great uncertainty. Because I could choose the samples from any sub-compartment in any compartment if it meets the forest type and entry the samples into

sample library for learning. In addition, even if the samples are intercepted different parts from the same interpretation sample, it could also express different information. This shows, interpretation samples could only interpretive the general information about the forest type. The general information might not be suitable for individual situation.

Fig. 3. Interpretation sample

d) *Make discrimination reaction table:* I picked out eight typical species, including 43 sample plots. Enter the interpretation results as reaction table into the database for inquiries and classification. The study of the interpretation samples later is based on this discrimination reaction table.

Fig. 4. Discrimination reaction table

2) Knowledge Representation

Make the corresponding relations between the forest type interpretation factors information and the forest types into rules. Make these rules as the basis for reasoning and enter these rules into computer program in the form of "production rules". The statement of production rule representation is: IF (Condition is satisfied) THEN (action). That is, "If this condition occurs, then perform this function." Between these rules are "and" or "or" relationships, after random combinations, then form multiple branches which could be used when we are reasoning along the chain of reasoning. Partial production rules are as shown in below.

IF   crown size is large or medium
    AND   crown shape is oval
    AND   crown density uniformity is tufted
    THEN   guess the forest type is one kind of hard broad-leaved forest, soft broad-leaved forest and rare broad-leaved forest.

IF   crown size is small
    AND   crown shape is irregular
    THEN   Oak lumber

IF   crown size is small
    AND   crown shape is round
    THEN   Pinus sylvestris

IF   crown size is large, medium
    AND   crown shape is round
    AND   tree elevation difference is small
    THEN   Larch

The table 1 lists the structure of rule base and partial rules.

Table 1. Structure of rule base

Rule NO	Condition	Next rule NO
1	Crown size is large or medium	2
2	Crown shape is oval	3
3	Crown density uniformity is tufted	void
4	Crown size is small	5
5	Crown shape is irregular	void
6	Crown size is small	7
7	Crown shape is round	void
8	Crown size is large or medium	9
9	Crown shape is round	10
10	Tree elevation difference is small	void

In the meanwhile, the type of the rule base makes the reasoning procedure reasons in turn according to the structure of list.

Knowledge base includes rule base and fact base. Rule base stored rule set and fact base stored fact set which is including the inputted facts or the intermediate results (facts) and the final inference facts. Among them, the first two belongs to the fact set and the third one belongs to the rule set.

- species database (stores 43 sub-compartment sample plots information)
- picture bank (stores 43 sub-compartment sample plots aero photos)
- rule base (stores the eight interpretation factors in the form of "production rules")
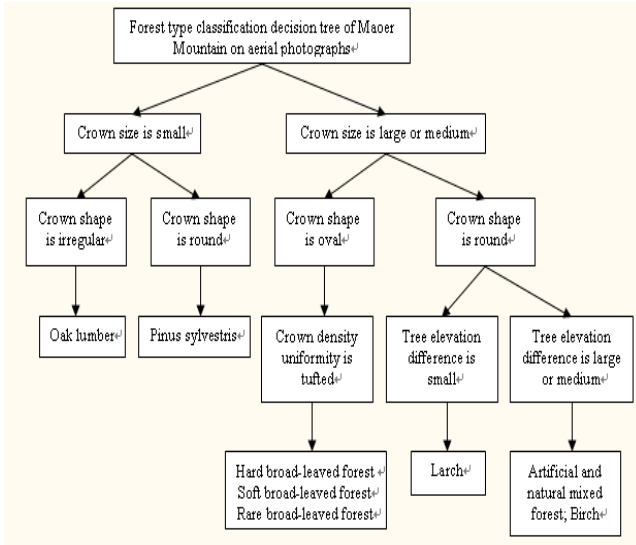
Fig. 5. Forest type classification decision tree

3.3   Establish Inference Engine

Inference engine is the procedures which are used for remembering the rules and control strategies, so that the whole expert system is able to work coordinately and logically. Forward inference, backward inference and bidirectional inference are three kinds of methods of rule-based reasoning. The expert system in this article adopted the rule-based forward inference.

When the users use the interpretation program, they just follow the tips and enter the known interpretation factors information, such as: crown size, crown shape, tone, tree elevation difference and other information, the inference engine finds the corresponding factors by the rules according to the known factors which the users submitted, then match the factors to the information in the knowledge base to get the forest type results and display them to the users. For example: When a user input his visual interpretation factors basing on an image to be interpreted: the crown size: small; crown shape: irregular; It is easy to know the forest type is Oak lumber from the forest type classification decision tree.

4   Discussion

The expert system for forest type interpretation had been established, which achieved the desired functionality initially. Due to the limitation of the designer's level, data source and time, this article only established prototype system. There are still many imperfections. The inference rules can also continue to expand in many directions. In addition, this system is stand-alone operation, so it could be developed further into a network system, and the system's intelligence is not high, because learning function needs further development. In this study, there are several factors which determined the uncertainty of the interpretation results and possibility of the misjudge situation.

- The interpretation samples were selected can only on behalf of the feature of the species to some extent.
- The given characteristic information of the forest type, can only determine the type to some extent. Because the impurity level of the node exists, "Misclassification impurity level" can be defined as: $i(N) = 1 - \max_j P(\omega_j)$, $P(\omega_j)$ is the frequency of the node N belonging to ω_j type. It can be used to measure the minimum probability of node N training samples classification error rate. That is the minimum probability of the training samples classification error rate of an interpretation factor. Take Pinus sylvestris and Larch as examples: there are 8 Pinus sylvestris plots. The number of the plots which crown size is small is eight and the number of which crown shape is round is eight. There are 13 Larch plots. The number of the plots which crown size is medium or large is 11, crown shape is round is 13 and tree elevation difference is small is 10, the other is 3. That is to say, "Misclassification impurity level" of Pinus sylvestris is 0 while Larch's "Misclassification impurity level" of crown size is 0.15; crown shape is 0; tree elevation difference is 0.23. So we can figure out that the correct interpretation possibility of Pinus sylvestris is high, while the possibility of Larch is low. Through getting the statistical information of the interpretation factors about the eight categories, we find out that some interpretation factors are great uncertain, can not be determined. So these interpretation factors are placed in the back, while the interpretation factors which can be easy to distinguish species are placed in the front [4].
- The location I picked to intercept the interpretation samples will also affect the production of discrimination reaction table, such as the information from the middle of the region and the marginal area is different.
- This study made the interpretation samples taking from the corrected Digital Ortho photo Map (DOM), the scale is decreased compared with the original uncorrected aerial photographs. If the study uses the uncorrected aerial photographs, the accuracy of interpretation will be increased.
- In addition, due to the limitation of the aero photographs, only through visual inspection to extract the interpretation factors information from the aero photographs, it is hard to distinguish some factors clearly. For example: From the aero photographs, we can distinguish the type of the species as hard broad-leaved forest, soft broad-leaved forest, rare broad-leaved forest if the crown size is large, shape is oval, crown density is tufted, but these three   species are difficult to distinguish. Mostly because these three species are all mixed forest and exist the composition of the same stand, which is hard to distinguish. This article provides a kind of thinking and method. Therefore, the final results of the experiment are limited to a qualitative interpretation of which types of species.

References

[1] Valérie, T., Marie-Pierre, J.: Tree species identification on large-scale aerial photographs in a tropical rain forest, French Guiana—application for management and conservation. Forest Ecology and Management 225, 51–61 (2006)

[2] Yang, X., Skidmore, A.K., Melick, D.R., Zhou, Z., Xu, J.: Mapping non-wood forest product (matsutake mushrooms) using logistic regression and a GIS expert system. Ecological Modelling 198, 208–218 (2006)
[3] Hatzilygeroudis, I., Prentzas, J.: Integrating (rules, neural networks) and cases for knowledge representation and reasoning in expert systems. Expert Systems with Applications 27, 63–75 (2004)
[4] Li, M., Liu, L., Li, C.-B.: An approach to expert recommendation based on fuzzy linguistic method and fuzzy text classification in knowledge management systems. Expert Systems with Applications 38, 8586–8596 (2011)

Software Designing and Simulation of GPS Receiver

Liang Wei-Tai, Wang Jun, and Yang Jin-Pei

Science and Technology on Information System Engineering Key Laboratory

Abstract. This paper studies the methods of GPS Receiver designing based on computer simulation. The mathematical models of GPS receiver are firstly developed, then GPS navigation information demodulation are given, and finally GPS Receiver are simulated and validated. The simulation results reveal that the methods have an application widely.

Keywords: Delay Lock Loop (DLL), Phase Lock Loop (PLL), Intermediate Frequency (IF).

1   Introduction

According to the typical GPS receiver, firstly the paper analyze GPS receiver architecture, secondly the mathematical model of GPS receiver are established, finally the GPS receiver design method are given based on computer simulation. It is support GPS signal research.

2   The GPS Receiver Basic Architecture and Principle

The basic architecture of typical GPS receiver is as show Fig 1[1][2][3].
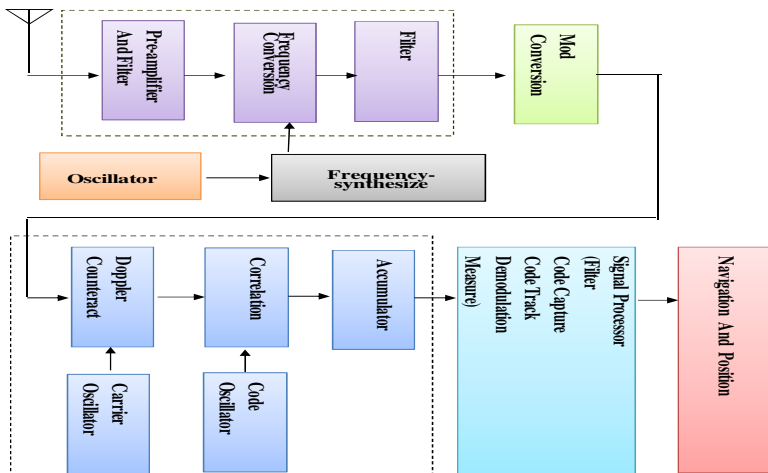
Fig. 1. The architecture of a typical GPS receiver

It can be know from Fig.1, the GPS receiver architecture are composed of receiving antenna, and preamplifier and low-noise filter, and transfers cable, and frequency transform circuit, and modulus circuit and signal processor and so on. The GPS signal that from satellite are processed by the GPS receiver, the distance of GPS to satellite and receiver self-situation can be resolved, so that achieve to navigation and position.

3   If Receiver Channel Mathematical Model

The model of intermediate frequency receiver channel are composed Phase Lock Loop (PLL), Delay Lock Loop (DLL) and central processing unit and so on, the function is code capture and to achieve code and carrier synchronization track.

3.1   Code Capture Arithmetic

The objective of code capture is to estimate visible satellite and its carrier frequency, and code, and code phase. The GPS system has twenty-four different pseudo-random codes. Besides, satellite perturbation factor shall arouse signal carrier frequency to occur Doppler shift, so it is need good code search method to ensure local-code and carrier synchronization. The method of Code capture has serial search method, and parallel frequency search method and parallel code phase search method.

I) Serial search
The serial search method is one of prime employ in CDMA system, and is very intuitionistic, as show in Fig.2.
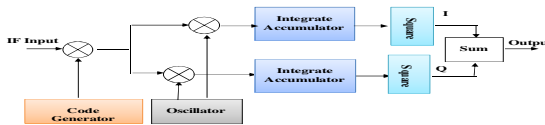
Fig. 2. The framework of serial search method

The serial search method can be divided into two step : first step is carrier search, namely it is searched in bound of carrier frequency shift, second step is code phase search, namely it is searched in bound of code periods.

II) Parallel frequency search
The serial search method is one step and one step to search code and carrier, so that search time is very long, and parallel frequency search method is only to search code, so that search time is shot, as show in Fig.3.

Fig. 3. The framework of parallel frequency search method

III) Parallel code search

The method of Parallel code search is very novel, as show in Fig.4.
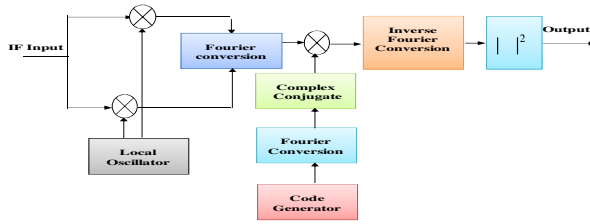
Fig. 4. The framework of parallel code phase search method

The mathematical model of parallel code phase search method is expressed as flow.

If input numeric intermediate frequency signal is $s(n)$, that code signal $y(n)$ are given by.

$$y(n)=s(n)\cos(\omega_0 n)+js(n)\sin(\omega_0 n)=I(n)+jQ(n)$$ (1)

Where ω_0 is local carrier angle Frequency. The $y(n)$ is Fourier transformed and inverse Fourier transformed, the code signal $z(n)$ are given by.

$$z(n)=IFFT\left(FFT(x(n))(FFT(y(n)))^*\right)=\sum_{m=0}^{N-1} x(m)y(n+m)$$ (2)

Therefore, parallel code phase search method is two signal correlation essentially.

3.2   Code and Carrier Track

The core module of carrier and code track circuit are PLL and DLL. Especially, the DLL has lot of types, such as one rank, two ranks and three ranks. The paper use one rank as example to deduced the mathematical model of code and carrier track circuit.

I) Code track circuit modeling

A typical code track circuit as show Fig 5[4], The six output ($I_E, Q_E, I_P, Q_P, I_L, Q_L$) were obtained by filtering and correlating GPS digital intermediate frequency signal, and it was processed by distinguish discriminator, and it has four differentiation character.
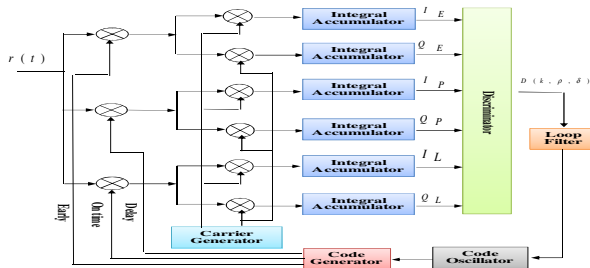
Fig. 5. The framework of code track circuit

Coherent
$$D = I_E - I_L$$
$$D = \left(I_E^2 + Q_E^2\right) - \left(I_L^2 + Q_L^2\right)$$

The mathematical model of coherent code track circuit is expressed as flow.

If i th GPS digital intermediate frequency signal is $r(i)$, it is expressed as follow.

$$r(i) = AC[(1+\varsigma)iT_s - \xi T_c]\cos[(\omega_b + \omega_d)i + \phi_0)] + n(i) \tag{3}$$

Where A is signal amplitude, $C(\cdot)$ value is ± 1, T_c is code width, $\tau = \xi T_c$ is time delay, T_s is sample interval, $\omega_b = 2\pi f_b T_c$ is intermediate frequency signal angle frequency, $\omega_d = 2\pi f_d T_s$ is Doppler angle frequency, $\varsigma = f_d/f_{L1}$ is frequency shift rate.

Code generator generate i th sample :

$$I_{I\pm}(i) = C[(1+\hat{\varsigma})iT_s - \hat{\xi}T_c \pm \delta T_c]\cos[(\omega_b + \omega_d)i] \quad I_{Q\pm}(i) = C[(1+\hat{\varsigma})iT_s - \hat{\xi}T_c \pm \delta T_c]\sin[(\omega_b + \omega_d)i]$$

The output of DLL discriminator is expressed as.

$$D(k,\rho,\delta) = \left(I_E^2 + Q_E^2\right) - \left(I_L^2 + Q_L^2\right) = L(k,\rho,\delta) - E(k,\rho,\delta) = K_0 D_\Delta(\rho,\delta) + N_D(k,\rho,\delta) \tag{4}$$

Where: $D_\Delta(\rho,\delta) = R^2(x)\big|_{x=\rho-\delta} - R^2(x)\big|_{x=\rho+\delta}$, $R(x) = \overline{C[(1+\varsigma)iT_s + \tau_0]C[(1+\hat{\varsigma})iT_s + \tau_0 + xT_c]}$,

$N_D = 2(\frac{N_0}{2})^2 B_L[1 - R^2(2\delta)] + \frac{A^2}{4}[\frac{1}{2}(N_0)]\sin c^2[(\Delta\omega_d)N/2]f(\rho,\delta)$, $K_0 = \frac{A^2}{4}\sin c^2[(\Delta\omega_d)N/2]$, $B_L = \pi/N$,

$f(\rho,\delta) = R^2(\rho-\delta) + R^2(\rho+\delta) - 2R(\rho-\delta)R(\rho+\delta)R(2\delta)$, $F(z)$ is loop filter transfers function,

$H_N(z) = A^2(1-\delta)F(z)/(A^2(1-\delta)F(z) + z - 1)$, $F(z) = G_1$, $B_{cl} = 1/(2\pi j)\oint_{|z|=1} H_N(z)H_N(z^{-1})z^{-1}dz$, $\xi(k)$ is signal phase(phase stepped is $aU(k)$, frequency stepped is $akU(k)$, frequency slanting stepped is $ak^2U(k)$).

II)   Carrier track circuit modeling

Carrier track circuit is to capture GPS signal carrier frequency, and typical carrier circuit as show Fig 6.
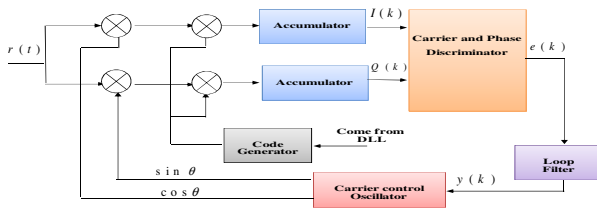
Fig. 6. The framework of carrier capture and track circuit

It can be know from Fig.6, the mathematical model of the carrier track circuit is expressed as.

$$e(k) = \arctan[Q(k)/I(k)] = g[\psi(k) + n_\theta \ e(k) \in [-\pi, +\pi]$$

$$g[\psi(k)] = \psi(k)\bmod[-\pi, +\pi], \quad \psi(k) = \theta(k) - \hat{\theta}(k), \quad y(k) = F(z)e(k), \quad \hat{\theta}(k) = D(z)y(k) = \sum_{i=0}^{k-1} y(i) \tag{5}$$

Where $D(z)$ is local oscillation transform function ($D(z) = z^{-1}/(1-z^{-1})$), $F(z)$ is loop circuit low filter, D is discriminator character, $I(k)$ and $Q(k)$ are expressed as.

$$I(k)=\frac{A}{2}R(\tau-\hat{\tau})\operatorname{sinc}[(\Delta\omega_d)_k\,N/2]\cos[\theta(k)-\hat{\theta}(k)]+N_I(k)\quad Q(k)=\frac{A}{2}R(\tau-\hat{\tau})\operatorname{sinc}[(\Delta\omega_d)_k\,N/2]\sin[\theta(k)-\hat{\theta}(k)]+N_Q(k)\quad .$$

If $N_I(k)$ and $N_Q(k)$ is white gauss noise, the performance parameter of carrier phase track circuit can be deduced as.

$$\operatorname{var}[\psi_{ss}]=\lim_{k\to\infty}\psi(k)=\frac{G_1}{2-G_1}\operatorname{var}[n_\theta] \tag{6}$$

Where n_θ is random noise, and the probability is expressed as.

$$p(n_\theta(k))=\frac{1}{2\pi}\exp(-\alpha)+\frac{\sqrt{\alpha}\cos n_\theta(k)}{\sqrt{\pi}}\exp[-\alpha\sin^2 n_\theta(k)]\times\{\frac{1}{2}+erf[\sqrt{2\alpha}\cos n_\theta(k)]\} \tag{7}$$

Where $\alpha=A^2/(4\sigma_n^2)NR^2(\tau-\hat{\tau})\operatorname{sinc}^2[(\Delta\omega_d)_k\,N/2]$, N is sample number, σ_n is white noise squared error.

3.3 Satellite Navigation Information Demodulation

It can be know from reference [4], the GPS navigation information are deduced as follow.

Step 1: GPS signal is expressed as.

$$s_d^k(t)=\sqrt{2P_{Cd}}C^k\left(t-\tau_d-K/f_{L1}^2\right)D^k\left(t-\tau_d-K/f_{L1}^2\right)\times\cos\left(2\pi f_{L1}t-2\pi f_{L1}\tau_r-K/f_{L1}^2+\phi_1\right)+\sqrt{2P_{PL1d}}P^k\left(t-\tau_d-K/f_{L1}^2\right)D^k\left(t-\tau_d-K/f_{L1}^2\right)$$
$$\times\sin\left(2\pi f_{L1}t-2\pi f_{L1}\tau_d-K/f_{L1}^2+\phi_1\right)+\sqrt{2P_{PL2d}}P^k\left(t-\tau_d-K/f_{L2}^2\right)D^k\left(t-\tau_d-K/f_{L2}^2\right)\times\cos\left(2\pi f_{L2}t-2\pi f_{L2}\tau_d-K/f_{L2}^2+\phi_2\right) \tag{8}$$

Where P_{PL1d}、P_{PL2d} is signal P code power, P_{Cd} is signal C/A code power, τ_d is transmit time from satellite to receiver, K/f_{L1}^2 K/f_{L2}^2 is signal transmit error.

Step 2: $s_d^k(t)$ is processed by band-pass filter, and GPS intermediate frequency output is expressed as.

$$s_{dIF}^k(t)=\sqrt{2P_{CdIF}}C^k\left(t-\tau_d-K/f_{L1}^2\right)D^k\left(t-\tau_d-K/f_{L1}^2\right)\times\cos\left(2\pi(f_{IF})t-2\pi f_{L1}\tau_d-K/f_{L1}^2+\phi_{dIF}\right)$$
$$+\sqrt{2P_{PL1dIF}}P^k\left(t-\tau_d-K/f_{L1}^2\right)D^k\left(t-\tau_d-K/f_{L1}^2\right)\times\sin\left(2\pi f_{IF}t-2\pi f_{L1}\tau_d-K/f_{L1}^2+\phi_{dIF}\right) \tag{9}$$

Where f_{IF} is intermediate frequency, P_{CdIF} is C/A code power, P_{PdIF} is P code power.

Step 3: $s_{dIF}^k(t)$ are processed by code capture and track circuit, and output is as.

$$s_{dIF}^k(n)\cos\left(2\pi f_{IF}n\right)=\sqrt{2P_{CdIF}}C^k(n-m)D^k(n-m)\cos\left(2\pi f_{IF}n-\phi_n\right)\cos\left(2\pi f_{IF}n\right)+e_3(n)$$
$$=\frac{\sqrt{2P_{CdIF}}}{2}C^k(n-m)D^k(n-m)\cos(\phi_n)-\frac{\sqrt{2P_{CdIF}}}{2}C^k(n-m)D^k(n-m)\cos(4\pi f_{IF}n-\phi_n)+e_3(n) \tag{10}$$

Formula (10) are processed by low pass filter, and it is expressed as.

$$D^k(n)=\sum_{n=0}^{N}\frac{\sqrt{2P_{CdIF}}}{2}C^k(n-m)\left(C^k(n-m)D^k(n-m)\cos(\phi_n)\right)=N_1D^k(n-m) \tag{11}$$

Where N_1 is constant.

3.4 Receiver Self-position Computation

Satellite ephemeris is refers to satellite Kepler orbit parameters. According to satellite ephemeris, satellite position and velocity can be resolved at any time.

I) Kepler orbit parameter

The position of S satellite can be deduced from Kepler equation[1] [5] :

$$S = \begin{bmatrix} \xi \\ \eta \\ \varsigma \end{bmatrix} = \begin{bmatrix} r\cos f \\ r\sin f \\ 0 \end{bmatrix} = \begin{bmatrix} a(\cos E - e) \\ a\sqrt{1-e^2}\sin E \\ 0 \end{bmatrix} \tag{12}$$

Where i is orbit surface slant angle, Ω is ascending node right ascension, ω is perigee angular distance, e is ellipse excentricity of satellite orbit, a is ellipse major axis, n_0 is average velocity, M_0 is flat anomalistic angle, E is partial anomalistic angle, r is radius, (ξ, η) is orbit plane, ς is perpendicularity orbit plane, $f = \arctan(\sqrt{1-e^2}\sin E/(\cos E - e))$ is really anomalistic angle.

Passing coordinate transform, the space orbit plane position of the k satellite at t_j time is expressed as.:

$$\begin{bmatrix} x^k(t_j) & y^k(t_j) & z^k(t_j) \end{bmatrix}^T = R_3\left(-\Omega_j^k\right)R_1\left(-i_j^k\right)R_3\left(-\omega_j^k\right)\begin{bmatrix} r_j^k\cos f_j^k & r_j^k\sin f_j^k & 0 \end{bmatrix}^T \tag{13}$$

$$R_3\left(-\Omega_j^k\right) = \begin{bmatrix} \cos\Omega_j^k & -\sin\Omega_j^k & 0 \\ \sin\Omega_j^k & \cos\Omega_j^k & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad R_1\left(-i_j^k\right) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos i_j^k & -\sin i_j^k \\ 0 & \sin i_j^k & \cos i_j^k \end{bmatrix} \quad R_3\left(-\omega_j^k\right) = \begin{bmatrix} \cos\omega_j^k & -\sin\omega_j^k & 0 \\ \sin\omega_j^k & \cos\omega_j^k & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{14}$$

II) Satellite position computation

According to ephemeris parameter, the satellite position can be computed accurately, the step is follow.

（1） Normalization time t_j : $t_j = t - t_{oe}$
（2） Average angle velocity n : $n = n_0 + \Delta n = \sqrt{GM/(a^3)} + \Delta n$, where Δn is perturbation correction of average angle velocity in telegraphese.
（3） Satellite flat anomalistic angle M_j : $M_j = M_0 + nt_j$
（4） Partial anomalistic angle E_j : $E_j = M_0 + e\sin E_j$
（5） Ascending node longitude Ω : $\Omega = \Omega_0 + (\dot\Omega - \omega_e)t_j - \omega t_{oe}$, where ω_e is earth rotation velocity.
（6） ascending node distance angle u_j , satellite radius vector r_j and orbit obliquity i_j :

$$u_j = \omega + f_j + C_{uc}\cos\left(2(\omega + f_j)\right) + C_{us}\sin\left(2(\omega + f_j)\right)$$

$$r_j = a\left(1 - e\cos E_j\right) + C_{rc}\cos\left(2(\omega + f_j)\right) + C_{rs}\sin\left(2(\omega + f_j)\right) \tag{15}$$

$$i_j = i_0 + it_j + C_{ic}\cos\left(2(\omega + f_j)\right) + C_{is}\sin\left(2(\omega + f_j)\right)$$

(7) satellite position in WGS-84 coordinate system is .

$$\begin{bmatrix} x_j(t_j) \\ y_j(t_j) \\ z_j(t_j) \end{bmatrix} = \begin{bmatrix} x_j'\cos\Omega_j - y_j'\cos i_j\sin\Omega_j \\ x_j'\cos\Omega_j + y_j'\cos i_j\sin\Omega_j \\ y_j'\sin i_j \end{bmatrix} \quad \begin{bmatrix} x_j' \\ y_j' \end{bmatrix} = \begin{bmatrix} r_j\cos u_j \\ r_j\sin u_j \end{bmatrix} \tag{16}$$

III) Receiver self-position computation

The pseudo-distance equation of satellite k from to receiver i is :

$$P_i^k = \rho_i^k + c\left(dt_i - dt^k\right) + T_i^k + I_i^k + \epsilon_i^k \tag{17}$$

where ρ_i^k is distance of from satellite k to receiver i , c is velocity of light, dt_i is clock excursion of receiver, dt^k is clock excursion of satellite, T_i^k is troposphere delay, I_i^k is ionosphere delay, ϵ_i^k is append error of pseudo-distance.

The geometry distance ρ_i^k of satellite k from to receiver i is follow :

$$\rho_i^k = \sqrt{\left(x^k - x_i\right)^2 + \left(y^k - y_i\right)^2 + \left(z^k - z_i\right)^2} \tag{18}$$

Where $\left(x^k, y^k, z^k\right)$ is k position, (x_i, y_i, z_i) is receiver i position. formula （18） into formula （17）, it is obtained .

$$P_i^k = \sqrt{\left(x^k - x_i\right)^2 + \left(y^k - y_i\right)^2 + \left(z^k - z_i\right)^2} + c\left(dt_i - dt^k\right) + T_i^k + I_i^k + e_i^k \tag{19}$$

Use least square rule to linearize formula (19), the formula (19) changed as.

$$
\begin{bmatrix}
-\dfrac{x^1 - x_{i,0}}{\rho_i^1} & -\dfrac{y^1 - y_{i,0}}{\rho_i^1} & -\dfrac{z^1 - z_{i,0}}{\rho_i^1} & 1 \\
-\dfrac{x^2 - x_{i,0}}{\rho_i^2} & -\dfrac{y^2 - y_{i,0}}{\rho_i^2} & -\dfrac{z^2 - z_{i,0}}{\rho_i^2} & 1 \\
-\dfrac{x^3 - x_{i,0}}{\rho_i^3} & -\dfrac{y^3 - y_{i,0}}{\rho_i^3} & -\dfrac{z^3 - z_{i,0}}{\rho_i^3} & 1 \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
\cdot & \cdot & \cdot & \cdot \\
-\dfrac{x^m - x_{i,0}}{\rho_i^k} & -\dfrac{y^m - y_{i,0}}{\rho_i^k} & -\dfrac{z^m - z_{i,0}}{\rho_i^k} & 1
\end{bmatrix}
\begin{bmatrix}
\Delta x_{i,1} \\
\Delta y_{i,1} \\
\Delta z_{i,1} \\
cdt_{i,1}
\end{bmatrix}
= b \tag{20}
$$

$$b = P_i^k - \rho_i^k + cdt_i + T_i^k - I_i^k - e_i^k$$

If $m \geq 4$, the formula （20） can be resolved, this solution is error between receiver begin position and true position, next approximately position is.

$$x_{i,1} = x_{i,0} + \Delta x_{i,1}, \quad y_{i,1} = y_{i,0} + \Delta y_{i,1}, \quad z_{i,1} = z_{i,0} + \Delta z_{i,1} \tag{21}$$

If receiver begin position is located earth center, namely $(x_{i,0}, y_{i,0}, z_{i,0}) = (0,0,0)$, the formula (20) and (21) were iterated for six time, the receiver position solution error is centimeter rank, thereby the distance from satellite to receiver and receiver position are resolved.

4   Receiver Channel Software Simulation

The GPS receiver are designed by computer simulation., and the program flow framework of software are show Fig 7.
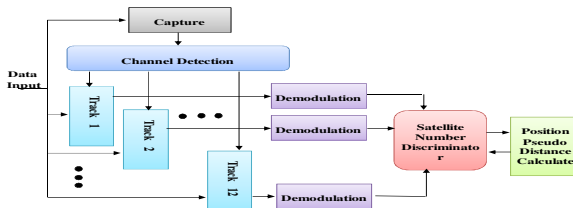
Fig. 7. The framework of GPS receiver software designing

Example: Suppose GPS signal contained two satellite signal add noise, signal intermediate frequency is 3.563MHz, navigation information is 111110000011111 00000111110000011111000001111100000, signal length is 1s, sampling rate is 12, one GPS satellite signal code shift and Doppler shift are separately 其中5$^{\mu s}$ and -1200Hz, another GPS satellite signal code shift and Doppler shift are separately 10$^{\mu s}$ and 800Hz. Signal-to-Noise is -15Db. According to module that from formula (2),(4),(5) and (11), and navigation information decompose from code and carrier capture by relevant receiver software. The result as show Fig.8.

```
. . ***************************************************************
**  Channel   Sat      Carrier    Doppler      C/A Code      State   **
              Number              Shift        Phase
****************************************************************************
**     1        3      3561800     -1200          60         Capture  **
**     2       10      3562000     -1000          84         Capture  **
**     3        0         0          0             0            -     **
**     4        0         0          0             0            -     **
**     5        0         0          0             0            -     **
**     6        0         0          0             0            -     **
**     7        0         0          0             0            -     **
**     8        0         0          0             0            -     **
**     9        0         0          0             0            -     **
**    10        0         0          0             0            -     **
**    11        0         0          0             0            -     **
**    12        0         0          0             0            -     **
****************************************************************************


   navigationdecode =

   Columns 1 through 14

       1    1    1    1    1   -1   -1   -1   -1   -1    1    1    1    1

   Columns 15 through 28

       1   -1   -1   -1   -1   -1    1    1    1    1    1   -1   -1   -1

   Columns 29 through 42

      -1   -1    1    1    1    1    1   -1   -1   -1   -1   -1    1    1

   Columns 43 through 50

       1    1    1   -1   -1   -1   -1   -1
```

Fig. 8. Code, carrier and navigation information demodulation

The resulting of computer simulation indicate the mathematical module of GPS receiver software that is correcting and validating.

5 Summaries

The paper actually is realization of high dynamic GPS receiver software designing. The method of GPS receiver software designing is adapt to similarity navigation and position system, therefore it has some reference value in mutuality domain.

References

[1] Liu, J.: Principle and method of GPS satellite navigation and position. Science Publishing Company, Beijing (2003)
[2] Guo, J., Hu, X., Vucetic, B.: Performance of a digital code tracking loop for DSSS systems in the presence of doppler shift. IEE Proc. Commun. 150(3), 202–206 (2003)

[3] Kim, S.J., Iltis, R.A.: Performance comparison of particle and extended kalman filter algorithms for GPS C/A code tracking and interference rejection. In: Conference on Information Sciences and Systems, pp. 20–22. Princeton University Press (2002)

[4] Jwo, D.J.: Optimisation and sensitivity analysis of GPS receiver tracking loops in dynamic environments. IEEE Proc.-Radar, Sonar Navig. 148(4), 241–250 (2001)

[5] Kamata, M., Shono, T.: Third-order phase locked loops using dual loops inserting an active filter in the second loop with improved stability. Electronics and Communications 84(6), Part 3, 75–84 (2001) (in Japan)

Feasibility of the Virtualization Based on OpenIMSCore

Shan Chuan[1], Han Xiaoyong[2], and Duan Xiaodong[2]

[1] Beijing University of Posts and Telecommunications
[2] China Mobile Research Institute
Beijng, China

Abstract. When Enterprise-level data center using virtualization technology for server consolidation is becoming more and more general, discusses and researches on virtualization for the telecommunications industry tend much fiercer. With virtualization of the system architecture, it can be saved mass time and costs, and receive better flexibility. Under this background, how to afford expected performance for each service becomes a new challenge. Combined with the ideas of virtualization and telecommunication service, this paper tests the performance of OpenIMSCore based on virtual platform to show feasibility of the virtualization for telecommunication service.

Keywords: Virtualization, OpenIMSCore, SIPp.

1 Introduction

As IT technology developing with each passing day, basic hardware with high performance is merging in large numbers. Virtualization,as an exciting architecture, which could create dynamic resource pool with limitless capacity to help users to access resource anywhere or anytime. And discussion between telecom industory becomes much fierce.

Taking advantage of virtualization,we can integrate several Operation Systems into one physical server,which can maximize the utilizaion of the hardware resource. While it can improve the flexibility and security of basic architecture, it will also reduce the cost of device, power supply of numerous servers, and cooling systems. However, overhead of virtualization must occupy some physical resource, which may have some effect on efficiency of CPU and memory. So we have to face this challenge whether virtualization could meet the demands of telcoms service system.

2 Related Technologies

2.1 Virtual Platform Vmware ESX

Many companies such as VMware, Microsoft and IBM launch virtualized products. This paper aims mainly at challenge that core network facing, and focuses on the feasibility of the virtualization based on VMware vShpere.

VMware vShpere is composed of virtual platform ESX and management tool called vCenter Server. This virtualized technology encapsulates the whole x86 server (including driver, OS, application and configuration) into a virtual machine,and transplant it on virtual platform ESX. VMware ESX uses disk partition on virtual machine and resource subdivision to simplify the base frame, which can make certain of remote management and standardization on resource. This virtual schema provides load isolation, that means system and applications are distributed in several virtual machines which run on single physical server together. That is to say only one server can run various OS and applications independently. What's more,system resource could be assigned automatically to users on demand. In the mean time, vCenter server provides windows service to manage numbers of hosts, which use VMware HA to realize high availability.

2.2   OPENIMSCORE

The Open Source OpenIMSCore is the specific application of 3GPP, 3GPP2, ETSI TISPAN and IMS/NGN,which realize the fundamental function of IMS core network. It is an open source implementation of the Call-Session Control Functions (CSCFs) and a lightweight Home Subscriber Server (HSS) developed in compliance to the IMS architecture standards given by the 3GPP. The implementation of the CSCFs is based on SIP Express Router written in C, while the HSS developed in Java. Open IMS Core support access of kinds of UE to hold VoIP communication.

2.3   IMS BENCH SIPP

With the aim of evaluating the performance of the core components of an IMS network, the European Telecommunications Standards Institute (ETSI) developed the IMS/NGN Performance Benchmark Specification, consisting of guidelines for applying a set of tests to determine how the system behaves when the load on the system is increased. This benchmarking standard makes the benchmarking results comparable, which is an important step in taking decisions regarding the deployment of IMS systems.

The IMS Bench SIPp is a modified version of SIPp, a free open source traffic generator for the SIP protocol. The test environment meet a criterion of IMS Performance Benchmark specificaion,ETSI TS 186.008.The test system consists of one manager instance controlling the whole benchmark run, a fixed number of SIPp load generators, and a monitoring tool for the SUT collecting information about CPU load and memory consumption. At the configuration of the test, the manager assigns each SIPp instance a fixed number of users generating a predefined database containing user data as well as configuration scripts.It also generates a deployment script to deploy all needed files to each SIPp instance. Each SIPp instance then generates SIP traffic towards the SUT in accordance with the statistical distribution of each scenario ordered by the manager instance and reports the total number of generated scenarios, the number of IHS and system information of the host machine.

3   Performance Benchmark

3.1   Delay Contrast

ESX can be seen as a software tire bewteen hardware and operation system, which runs in privilege to take charge of management and isolution of virtual machine. It virtualize a hardware environment independent of physical hardware for each virtual machines and provides secure and independent condition for these virtual machines. However, virtual platform adds ESX to the physical platform, which may take some effect on IMS performance.

This experiment placed all elements of Open IMS Core including PCSCF, ICSCF, SCSCF and HSS together on one physical machine or a virtual one. The virtual machine occupies all the resource(CPU and Memory) of its physical host. To avoid numerous CPS(Call per Second) leading to signaling queuing delay,we used only 20 CPS in the test. The result of delay is shown in Tab.1, which demonstrates the introduction of ESX cause five times diference between virtual platform and physical platform.

Table 1. Delay contrast

	Session setup (ms)	INVITE arrive (ms)	Session release (ms)
Virtual platform	15.45	8.94	8.95
Physical platform	3.24	1.21	1.96

For ESX catches sensitive privilege instruction between guest OS and hardware, which can run guset OS without modification. This virtual technology use Ring Compression and Binary Translation. The former make ESX and guest OS run in different privilege, which means ESX runs in the highest privilege Ring 0, the guest OS kernel code runs in Ring 1, the application code runs in Ring 3.With this method ESX will intercept and capture some privilege instruction of guest OS.However,some instruction is not suitable for ring compress, it must scan and modify binary code of guest OS to make these instructions support virtualization. The advantage of the virtual technology is no modification of guest OS, meanwhile speed and function meet users' demand. The most disadvantage is its performance especially for I/O. So Open IMS Core running on virtual platform brings more signaling delay.

When all elements run on one virtual machine,the next experiment replaced 4 element of Open IMS Core onto the physical machine in turns, and the others still run on virtual machine. So we can know in which condition the delay is close to all elements run in one physical machine. This experiment is to find which element is the most sensitive for virtualization, and when deploy IMS core network in reality, it can choose right element runing on virtual platform.

Table 2. Delay contrast of network elements replacement

	Session setup (ms)	INVITE arrive (ms)	Session release (ms)
All on Virtual platform	15.45	8.94	8.95
All on Physical platform	3.24	1.21	1.96
Replace PCSCF on Physical platform	3.74	1.41	2.12
Replace ICSCF on Physical platform	13.87	5.51	8.52
Replace SCSCF on Physical platform	12.24	4.97	6.76
Replace HSS on Physical platform	12.9	5.35	7.62

As is shown in Tab.2 above, the delay is close to all components on physical machine when replace PCSCF onto physical machine. So PCSCF may be a bottleneck of IMS virtualization. For PCSCF is the first interface between IMS and outside environment, every service request from users must pass through PCSCF.

3.2 Maximum Call Attempts Contrast

Because of the introduction of virtual platform, there may be more difference between physical and virtual platform. This experiment tested the max number of concurrent calls on both platform, and the results can be shown in Fig.1 below. When CPS increase to 500, Open IMS Core on virtual platform can not handle these call attempts.
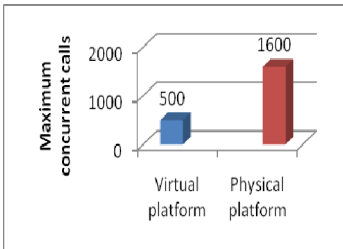
Fig. 1. Maximum concurrent calls contrast   **Fig. 2.** Delay contrast
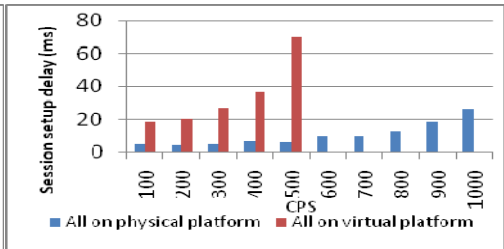
To find reason why different maximum calls exist in virtual and physical platform, this experiment also tested the delay. And then this experiments collected delay results using one IMS Bench SIPp test system, because more test system will cause larger delay when test same CPS. To exclude the effect as told above, we choose only one test system. The results are depicted in Fig.2. Signaling delay on virtual platform is much larger than delay on physical platform, which is the reason that call attempts failed when CPS are 500.This result also demonstrates that real-time service has high

demand on virtualization, because signaling delay leads to failure of service requests, which takes effects on the whole system performance.

3.3 CPU Utilization of ESX

As for the maximum number of concurrent calls on virtual and physical platform existing huge difference, this experiment need to test how much resource ESX used. Besides signaling delay, we tested whether hypervisor was another reason because it would take use of much CPU resource when Open IMS Core handles much call attempts. This experiment used VMware vCenter Server to real-time monitor virtual machine and physical server, and results can be seem in Fig. 3 and Fig. 4.
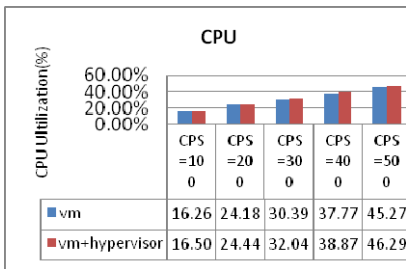
Fig. 3. CPU utilization of virtual machine and physical machine
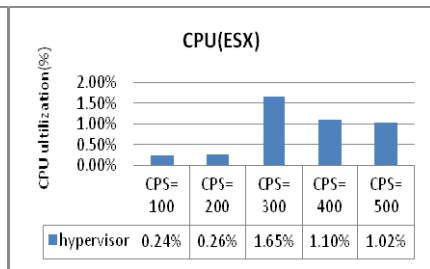
Fig. 4. CPU utilization of ESX

In Fig.3, the blue instance stands for the CPU utilization of virtual machine, the red one stands for physical server CPU utilization. Because CPU of physical server is shared by ESX and virtual machine, the red minus the blue is the ESX utilization. As Figure 4 depicted, with CPS increasing from 100 to 500, the utilization of CPU is always below 2%, which means ESX does not take huge resource that may influence the performance of Open IMS Core on virtual platform.

4   Summaries

The test results based IMS Bench SIPp demonstrate that ESX does not use huge physical resource, and when CPS increasing the resource utilization is within a fixed renge. Virtualization technology would not modify guest OS on its virtual machine, but may take some effect on performance, especially for I/O. So there may be larger signaling delay when Open IMS Core runs on virtual platform.

However the introduction of ESX takes some effect on delay and the maximum number of concurrent calls, it can integrat several physical server into only one physical server, that is to say server virtual machine runs on one physical server.If these virtual machine can handle more service than one physical server,this can increase the utilization of servers. With consolidation of server,it will reduce the number of physical servers,and increase the resource utilization,which will cut costs

on hardware and management. Some statistics shows that virtualization technology increases the average utilization of single server.After using virtualization technology,the average utilization of each server improve from 5-15% to 60-85%,meanwhile it will lower the cost on hardware and software from 30-60%.

This is the original study, and there will be more follow-up studies. As for real-time demand on telecommunication service,we will choose para-virtualization to test the delay,which may have less delay costs than full-virtualization.

References

[1] vmware official website, http://www.vmware.com/cn/
[2] OpenIMSCore Project official website, http://www.openimscore.org
[3] IMS bench SIPp official website,
     http://sipp.sourceforge.net/imsbench/intro.html
[4] Herpertz, R., Carlin, J.M.E.: A Performance Benchmark of a Multimedia Service Delivery Framework. In: 2009 Mexican International Conference on Computer Science (2009)
[5] Din, G.: An IMS Performance Benchmark Implementation based on the TTCN-3 Language. International Journal on Software Tools for Technology Transfer (STTT) 10(4), 359–370 (2008)
[6] European Telecommunications Standards Institute. Telecommunications and Internet converged Services and Protocols for Advanced Networking (TISPAN); IMS/NGN Performance Benchmark. ETSI TS 186 008-1 (October 2007)

The Application of the Genetic Anneal Simulation Support Vector Machine on the Predicting of the Consumer Price Index

Luo Fangqiong and Huang Shengzhong

Department of Mathematics and Computer Science Liuzhou Teachers College
Liuzhou, Guangxi 545004, China
{gxlzhsz,Luofangqiong123}@163.com

Abstract. In order to have the predicting of the consumer price index, the application of the support vector machine on the predicting of the consumer price index carried out. Firstly, the basic theory of the least square support vector machine was put forward, and then the genetic simulated annealing algorithm was put forward, and finally the case study was carried out, and the results showed that this new method had a higher ability of predicting.

Keywords: support vector machine, consumer price index, predicting.

1   Introduction

Consumption index in the internationally accepted, the significance of the data indicators that reflect the degree of deflation or inflation, for pricing policy, wages policy, monetary policy and consumer policies favorable theoretical basis, but also conducive to national Accounting of the economy, has been, and many scholars since the semi-phase regression model, ARIMA prediction model, the function coefficient autoregressive model, time series and the residents of mixed regression model for prediction of the price the consumer price index, but the prediction accuracy is not high, And the consumer index does not reflect the complex variation of the predicted results often in a distorted state.

Support vector machine with a simple structure and generalization ability advantage. Standard support vector machine algorithm gradually evolved out of the least squares support vector machines, least squares support vector machine parameters are adjusted for the two nuclear width and the regularization parameter, least squares support vector machine training and generalization Completely determined by these two parameters, but the choice of adjusting the parameters of these two is difficult, no effective selection methods, usually selected based on experience, time and effort on the one hand, on the other predicted effects Accuracy is not high. Therefore, we should find a way to adjust the parameters optimized to determine the means. Many pattern classification will require taking a large number of potential features, therefore, must be taken in the underlying characteristics of the child to identify significant features, and then, based on characteristics of the beginning sub-classification. Thus, different characteristics should be clear of its

contribution to the classification, in this way the method of feature subset found is Optimization. Consumer price index in accordance with the residents to predict the optimal encountered during the feature selection and least square support vector machine to adjust parameter optimization problem, a simulated annealing algorithm based on least squares support vector machine resident consumption price index Prediction methods, and ultimately to achieve the purpose of improving prediction accuracy.

2   Support Vector Machines

Suykens and Vandewalle first proposed the concept of least squares support vector machines, and its expression is as follows:

(x_i, y_i), $i = 1,2,\cdots,n$ set of input and output for the nonlinear systems sample data set, which indicates that the variable $x_i \in \vec{R}^d$, $y_i \in \vec{R}$, d dimension. Least squares support vector machines using these data samples back into the nonlinear system model structure. Main steps are as follows:

(1) First, a nonlinear transform the original input to the system $x \rightarrow \varphi(x)$ data sample point x_i to switch to another high-dimensional space Xi and Gilbert, the original problem is converted to high-dimensional space in a linear problem.

(2) Next, create a linear model of high-dimensional space.

In high-dimensional feature space corresponding to the original input space sample points classification function can be expressed as equation:

$$f(x) = < \omega, \varphi(x) > +a \tag{1}$$

Where $<,>$ used to represent the inner product of two vectors, ω representative of the weight vector, $\varphi(x)$ on behalf of non-linear mapping function, a said threshold.

In accordance with the method of Vapnik's VC dimension and structural risk minimization theory, while using the least squares support vector machine to complete the system optimization, therefore, equation (1) the function of $f(x)$ should meet the following conditions:

$$\min J = \frac{1}{2} \| \omega \|^2 + \lambda \sum_{i=1}^{n} \theta_i^2 \tag{2}$$

$$s.t. \ y_i = < \omega \cdot \varphi(x_i) > +a + \theta_i \tag{3}$$

Where, λ behalf of the regularization parameter.

The introduction of the Lagrangian operator equation (2) and formula (3), we get:

$$L = \frac{1}{2} \| \omega \|^2 + \frac{1}{2} \lambda \sum_{i=1}^{n} \theta_i^2 - \sum_{i=1}^{n} \beta_i [< \omega \cdot f(x_i) > +a + \theta_i - y_i] \tag{4}$$

Where: $\beta_i \geq 0$ on behalf of the Lagrange operator. Find the formula (4) stationary, so that a ω 、 a 、 θ_i 、 β_i order partial derivatives to 0, we get:

$$\omega = \sum_{i=1}^{n} \beta_i \varphi(x_i) \tag{5}$$

$$\sum_{i=1}^{n} \beta_i = 0 \tag{6}$$

$$\beta_i = \lambda \theta_i \tag{7}$$

$$< \omega \cdot \varphi(x_i) > +a + \theta_i - y_i = 0 \tag{8}$$

Solving equation (5) - (8) can be ω and θ to remove these two parameters, and then can be:

$$\begin{bmatrix} 0 & q^T \\ q & \Omega + \lambda^{-1} I \end{bmatrix} \begin{bmatrix} A \\ \beta \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{9}$$

Which, $\vec{y} = [y_1, y_2, \cdots y_n]^T$; $\vec{q} = [1,1,\cdots,1]^T$; $\vec{\beta} = [\beta_1, \beta_2, \cdots, \beta_n]^T$; $\Omega_{kh} = \varphi(\vec{x}_k)^T \varphi(\vec{x}_h)$ 。

Equation using the least squares method (9), therefore, in training speed, the least squares support vector machine than the traditional support vector machine has a faster speed, $\vec{\beta} = [\beta_1, \beta_2, \cdots, \beta_n]^T$ and a can be calculated using the least squares method, which can be non-Linear system model:

$$f(x) = \sum_{i=1}^{n} \beta_i < \omega(x) \cdot \varphi(x_i) > +a \tag{10}$$

Kernel function corresponding nonlinear system model can be expressed as the following form:

$$f(x) = \sum_{i=1}^{n} \beta_i h(x, x_i) + a \tag{11}$$

3   Improved Genetic Simulated Annealing Algorithm

3.1   Nested-Based Algorithm Framework Hybrid Strategy RCPSPGSA

The problem of fault diagnosis of CNC machine tools, the hybrid strategy is usually divided into three categories, namely, the embedded strategy, a two-stage strategy and nesting strategy. For embedded strategy, will inspire his element as a means of evolutionary function. Two-stage strategy inspired by other element method in the first phase of the initial GA population initialization, and GA in the second phase of its traditional way of working. In the nested strategy, inspired by the other element method is nested in the GA in order to improve local search ability of GA. In the improved genetic simulated annealing algorithm is selected nesting strategy.

Genetic simulated annealing algorithm to improve the operation flow is as follows:

1) the initial parameter set;
2) generate an initial population.
3) assessment of the initial population, and to make Best equal to the current optimal solution, k = 0.
4) If the termination conditions (evolution of the number of iterations reaches the preset value or the optimal solution $\mathbb{IV}$ iterations continue unchanged) satisfied, go to step 11.
5) The selection operation.
6) Cross operation.
7) the mutation.
8) form a PopSize (population number of individuals) on behalf of individuals under the temporary population, the population of individual solutions as the initial SA. Calculate the optimal solution of the population, and update the Best.
9) The use of a given interim step to improve the next generation of SA population in the ability of each individual, the probability of AA to a new individual to accept, then update the Best.
10) Order $k = k + 1$, $t_k = \mathcal{E}_{k+1}$, go to step 4 ($\mathcal{E}$ temperature is the retreat rate).
11) Output the optimal solution Best.

3.2   Encoding and Decoding Rules

(1) parameters: the use of genetic simulated annealing algorithm to adjust the parameters of SVM are optimized for least squares support vector machine has two tuning parameters, the width of a nuclear σ, the other is the regularization parameter λ. Features and two different values corresponding to a 0, the other is 1.0 indicates that the feature is not taken, one that recognized the features. Therefore, the feasible solutions can be expressed as $(\lambda, \sigma, F_1, F_2, \cdots, F_n)$.

(2) classification accuracy: the use of ten fold cross-validation method the data set is divided into ten, of which 1 cycle to the data set as test data, and the remaining nine as training samples to test the resulting ten Times the average of the diagnostic accuracy of classification on the classification accuracy.

(3) With the idea of genetic variation: the formation of new parameters of the order to be able to form a combination of parameters in the parameter space is full. Genetic algorithms and simulated annealing hybrid algorithm can improve the diagnostic ability of simulated annealing, to form a new feasible parameters.

4   Residents Consumer Price Index Forecast

In order to verify the effectiveness and practicality of the method, select the years 2005 to 2010 the consumer price index data as the input support vector machine samples, the least squares support vector machine as a resident consumer price index forecast Classification into genetic simulated annealing algorithm, first, the training sample data using this algorithm, which can get the best support vector machine parameters and the optimal combinations of features. Then, using least squares support vector machine to predict the residents consumer price index price.

Table 1. 2004 - 2010 in the consumer price index table

Year	2004	2005	2006	2007	2008	2009	2010
CPI	455.8	464.0	471.0	493.6	522.7	532.9	550.4

First, the use of genetic simulated annealing support vector machines in China 2005-2010, the consumer price index for the simulation results shown in Tab.2:

Table 2. Based on genetic simulated annealing CPI simulation results

Year	Actual value	Predictive value	Error (%)
2005	464.0	464.9	0.19
2006	471.0	476.6	1.19
2007	493.6	496.2	0.53
2008	522.7	528.1	1.03
2009	532.9	536.3	0.64
2010	550.4	563.2	2.33
Average relative error			0.98

By Tab.2 the following conclusions can be drawn, the use of genetic simulated annealing support vector machines can accurately predict the consumer price index. Genetic simulated annealing corresponding support vector machine to predict the consumer price index of the error curve shown in Fig.1, support vector machines over 750 iterations the average error reduced to 0.12456, indicating that the method has good predictive ability.
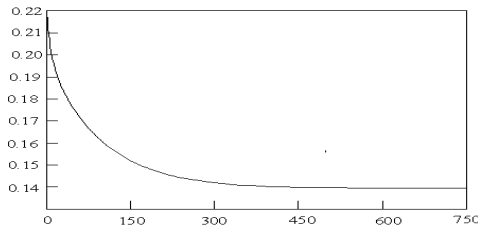
Fig. 1. Support vector machine error of the consumer price index curves

Then, using genetic simulated annealing support vector machines for the 2010 - 2012 China's consumer price index for the predicted results in Tab.3. Thus, according to forecasts of monetary policy, wage policy.

Table 3. 2010-2012 predicted the consumer price index

Years	Predictive value
2010	563.2
2011	597.5
2012	601.6

Parameters on genetic simulated annealing algorithm the predictive ability of support vector machines is relatively large, the genetic simulated annealing algorithm of support vector machines and comparison grid search method, its performance is better than the grid search method to compare the results in Tab.4.

Table 4. Genetic simulated annealing algorithm of support vector machines and network search method comparison table

Support Vector Machine			Network search method		
λ	σ^2	Accuracy (%)	λ	σ^2	Accuracy (%)
103. 24	0. 14	95	45. 23	0. 25	87

5   Summaries

Genetic Simulated Annealing Algorithm for SVM parameter selection and optimization characteristics, and the method used in the consumer price index forecast to use 2005 to 2010 the consumer price index data for 2011 in China Residents consumer price index is predicted, results showed that the genetic simulated annealing algorithm of support vector machine has better prediction accuracy, can successfully eliminate adverse interference.

References

[1] Zhang, H., Xu, H.: China's CPI fluctuations in factor analysis. Statistics and Decision (19), 86–88 (2009)
[2] Wang family planning, Huang, Y.X. provided the public. Wavelet Packet Analysis and Support Vector Machine tool fault diagnosis. Vibration, Testing and Diagnosis 28(3), 273–276 (2008)

Research on Model of Command Signal and Instruction of Some Single Channel Control-and-Guide System

Wei Xian-Jie, Zhang Yan-Pu, and Yuan Li-Zhe

Department 3, Artillery Commanding Institute, Langfang, 065000, China

Abstract. The mathematic model of command signal and instruction modulus is the key of simulation of single channel control-and-guide system. A model is established for most common single channel control-and-guide system in the paper. The command signal of some single channel control-and-guide system contains Blank Block, so it is difficult to simulate the process in actual. In the paper a logic relation model is based on command signal and instruction modulus by analyzing single channel control-and-guide system, cycle average control power and research on difficult command instruction classification. This model is successfully applied in simulation.

Keywords: Single Channel Control-and-guide, Command Signal, Cycle Average Control Power, Logic Model.

1   Introduction

Some one system is one of single channel control-and-guide systems, which are applied to vector-controlled. It is cycle average control power, not instantaneous power that decides its movement. So, research on relations to command signal and instruction modulus is the key to study dynamic flight performance, the linkage to its control-and-guide simulation and movement.

2   The Mathematic Model of Command Signal and Instruction Modulus

There are four times directional conversion of a command signal modulation in common single channel control-and-guide systems. A command signal comprises orientation of pitching's signal and yawing's signal, so corresponding instruction coefficients are complex. Consequently，there are three steps to realize the single channel control-and-guide systems simulation[1].

Step 1: to calculate instruction coefficients basis on command signals.
Step 2: to calculate cycle average control power basis on instruction coefficients.
Step 3: to establish the mathematic model of dynamic flight performance using cycle average control power.

Now, to deduce the general cycle average control power formulation, the pitching's cycle average control power F_{kpy3} and the yawing's cycle average control power F_{kpz3} follow.
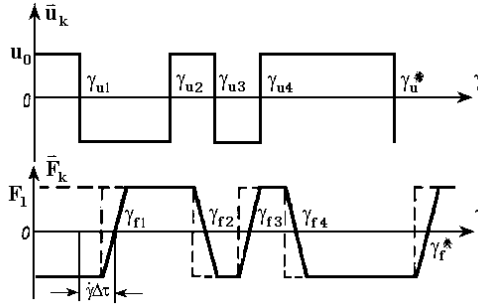
Fig. 1. Relation to command signal and control power

As Fig.1 showing, calculate the pitching's cycle average control power F_{kpy3}.

$$F_{kpy3} = \frac{1}{2\pi} \int_0^{2\pi} - F_k \sin\gamma d\gamma = -\frac{1}{2\pi} \left(\int_{\gamma_{f1}}^{\gamma_{f2}} F_k \sin\gamma d\gamma + \int_{\gamma_{f2}}^{\gamma_{f3}} F_k \sin\gamma d\gamma + \int_{\gamma_{f3}}^{\gamma_{f4}} F_k \sin\gamma d\gamma \right.$$

$$\left. + \int_{\gamma_{f4}}^{\gamma_f^*} F_k \sin\gamma d\gamma \right) = -\frac{2F_0 \sin\frac{a}{2}}{\pi a} (\cos\gamma_{f1} - \cos\gamma_{f2} + \cos\gamma_{f3} - \cos\gamma_{f4}) \quad (1)$$

Similarly, calculate the yawing's cycle average control power F_{kpz3}.

$$F_{kpz3} = \frac{1}{2\pi} \int_0^{2\pi} F_k \cos\gamma d\gamma = -\frac{2F_0 \sin\frac{a}{2}}{\pi a} (\sin\gamma_{f1} - \sin\gamma_{f2} + \sin\gamma_{f3} - \sin\gamma_{f4}) \quad (2)$$

In the process of control and guide simulation, formulation (1) (2) is not easy to direct computer calculation. To predigest formulation, and at the same time to extend all common instances, we review control power trapezia cycle. Provided that there was n times zero crossings point from $\gamma_f^* - 2\pi$ to γ_f^* in a whole period, n would be a even number, and phase shifting angle of every zero crossings point would be γ_{f1}、 $\gamma_{f2} \cdots \gamma_{fn}$. F_{kpy3} and F_{kpz3} can be expressed as following.

$$F_{kpy3} = -\frac{1}{2\pi} \int_{\gamma_f^* - 2\pi}^{\gamma_f^*} F_k \sin\gamma d\gamma = \frac{2F_0 \sin\frac{\dot\gamma\tau_2}{2}}{\pi\dot\gamma\tau_2} \sum_{i=1}^n (-1)^{i+1} \cos\gamma_{fi}$$

$$F_{kpz3} = \frac{1}{2\pi} \int_{\gamma_f^* - 2\pi}^{\gamma_f^*} F_k \cos\gamma d\gamma = \mp \frac{2F_0 \sin\frac{\dot\gamma\tau_2}{2}}{\pi\gamma} \sum_{i=1}^n (-1)^{i+1} \sin\gamma_{fi}$$

In this way, we can calculate the pitching's cycle average control power F_{kpy3} and the yawing's cycle average control power F_{kpz3} conveniently, and programmed algorithm is greatly simplified and easily to be realize in its control and guide simulation[2] [3] [4].

3   Analysis of Some System Control and Guide Process Simulation

Comparison of common single channel control and guide systems, this system have two differences.

(1) To adopt double-core lead, time-sharing transmitting information signal. specifically, in the system circumvolution period, there are two areas, one area only uses to transfer of control signal,　the another one uses to the system info signal, namely Blank Block($\pm \lambda$), exclusive from transfer of command signal. Because of existence of Blank Block, unavoidably impacts orientation and modulus of the cycle average control power.

(2) It is not four-time directional conversion of　command signal modulation in the system. Influencing to formation of command signals is the instruction unlock angle(α) and the instruction unlock central angle (θ)of a command signal.

Therefore, ignoring time-retarded function of signals transfering and rudder working, the orientation of the cycle average control power is decided by the instruction unlock central angle (θ)，and the magnitude of the cycle average control power by the instruction unlock angle(α).

4   Mathematic Model of Command Signal and Instruction of Some System for Simulation

In the control-and-guide simulation, to calculate instruction coefficient, we need to analyse cautiously the influence of Blank Block($\pm \lambda$). Because of complexity of this influence, classification analysis method is introduced to establish logic mathematic model of command signal and instruction of some system for simulation.

(1) The instruction unlock central angle (θ) belongs to ($360^{\circ} - \lambda$, $360^{\circ} + \lambda$).

Case one: $0 < \alpha < 2\lambda$, then command instruction, $K_Z = 0$

Case two: $\theta + \dfrac{\alpha}{2} > 360^{\circ} + \lambda$, $\theta - \dfrac{\alpha}{2} < 360^{\circ} - \lambda$ and $2\lambda < \alpha < 180^{\circ}$, then the command instruction spans Blank Block ($\pm \lambda$).

$$\text{If } \alpha_{s1} = a_1 \cdot (\frac{\alpha}{2} - \theta + (360^{\circ} - \lambda)), \quad \theta_{s1} = a_1 \cdot (\theta - \frac{\alpha}{2}) + \frac{\alpha_{s1}}{2},$$

$$\alpha_{s2} = a_1 \cdot (\frac{\alpha}{2} + \theta - (360^{\circ} + \lambda)), \quad \theta_{s2} = a_1 \cdot (360^{\circ} + \lambda) + \frac{\alpha_{s2}}{2},$$

$$\text{then } K_{py31} = -\sin \frac{\alpha_{s1}}{2} \cos \theta_{s1}, K_{pz31} = -\sin \frac{\alpha_{s1}}{2} \sin \theta_{s1},$$

$$K_{py32} = -\sin \frac{\alpha_{s2}}{2} \cos \theta_{s2}, \quad K_{pz32} = -\sin \frac{\alpha_{s2}}{2} \sin \theta_{s2}.$$

Case three: $\theta+\dfrac{\alpha}{2}<360°+\lambda$, $\theta-\dfrac{\alpha}{2}<360°-\lambda$ and $2\lambda<\alpha<4\lambda$, then the command instruction spans Blank Block($\pm\lambda$).

If $\alpha_{s1}=(\dfrac{\alpha}{2}-(\theta-(360°-\lambda)))\cdot a_1$ $\theta_{s1}=(\theta-\dfrac{\alpha}{2})\cdot a_1+\dfrac{\alpha_{s1}}{2}$,   then

$K_Y=-\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$, $K_Z=-\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$.

Case four: $\theta+\dfrac{\alpha}{2}>360°+\lambda$, $\theta-\dfrac{\alpha}{2}>360°-\lambda$, and $2\lambda<\alpha<4\lambda$, then the command instruction don't span Blank Block ($\pm\lambda$).

If $\alpha_{s1}=(\dfrac{\alpha}{2}+\theta-(360°+\lambda))\cdot a_1$, $\theta_{s1}=(360°+\lambda)\cdot a_1+\dfrac{\alpha_{s1}}{2}$,

then $K_Y=-\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$, $K_Z=-\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$.

(2) The instruction unlock central angle (θ) belongs to $90°+\lambda<\theta<270°-\lambda$, $\theta-\dfrac{\alpha}{2}>\lambda$, $\theta+\dfrac{\alpha}{2}<360°-\lambda$ and $0°<\alpha<180°$, then the command instruction don't span Blank Block ($\pm\lambda$).

If $\alpha_{s1}=\alpha\cdot a_1$ and $\theta_{s1}=\theta\cdot a_1$,   then $K_Y=-\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$,

$K_Z=-\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$.

(3) The instruction unlock central angle (θ) belongs to $\lambda<\theta<90°+\lambda$.

Case one: $\theta-\dfrac{\alpha}{2}>\lambda$ and $0°<\alpha<180°$, then the command instruction don't span Blank Block ($\pm\lambda$). If $\theta_{s1}=\alpha\cdot a_1$, $\theta_{s1}=\theta\cdot a_1$, then

$K_Y=-\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$, $K_Z=-\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$.

Case two: $-\lambda<\theta-\dfrac{\alpha}{2}<\lambda$ and $4\lambda<\alpha<180°$, then the command instruction doesn't span

Blank Block ($\pm\lambda$). $K_Y=-\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$ $K_Z=-\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$.

Case three: $\theta-\dfrac{\alpha}{2}<-\lambda$ and $4\lambda<\alpha<180°$, then the command instruction spans Blank

Block   ($\pm\lambda$). $K_{py31}=-\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$, $K_{pz31}=-\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$,

$K_{py32}=-\sin\dfrac{\alpha_{s2}}{2}\cos\theta_{s2}$, $K_{pz32}=-\sin\dfrac{\alpha_{s2}}{2}\sin\theta_{s2}$.

(4) The instruction unlock central angle (θ) belongs to $\theta>360°$.

Case one: $0°<\theta<\lambda$, $\theta+\dfrac{\alpha}{2}<\lambda$, $\theta-\dfrac{\alpha}{2}>-\lambda$ and $2\lambda<\alpha<4\lambda$, then the command instruction is inside Blank Block($\pm\lambda$). $K_Y=0$, $K_Z=0$

Case two: $0° < \theta < \lambda$, $\theta + \dfrac{\alpha}{2} > \lambda$, $\theta - \dfrac{\alpha}{2} > -\lambda$ and $2\lambda < \alpha < 4\lambda$, then the command instruction spans Blank Block ($\pm \lambda$). $K_Y = -\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$,

$K_Z = -\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$.

Case three: $0° < \theta < \lambda$, $\theta + \dfrac{\alpha}{2} > \lambda$, $\theta - \dfrac{\alpha}{2} < -\lambda$ and $2\lambda < \alpha < 4\lambda$, then the command instruction spans Blank Block ($\pm \lambda$). If $\alpha_{s1} = (\dfrac{\alpha}{2} - \theta - \lambda) * a_1$,

$\theta_{s1} = (\theta + 360° - \dfrac{\alpha}{2}) * a_1 + \dfrac{\alpha_{s1}}{2}$, $\alpha_{s2} = (\dfrac{\alpha}{2} + \theta - \lambda) * a_1$, $\theta_{s2} = (\theta + \dfrac{\alpha}{2}) * a_1 - \dfrac{\alpha_{s2}}{2}$, then

$K_{py\,31} = -\sin\dfrac{\alpha_{s1}}{2}\cos\theta_{s1}$, $K_{pz\,31} = -\sin\dfrac{\alpha_{s1}}{2}\sin\theta_{s1}$, $K_{py\,32} = -\sin\dfrac{\alpha_{s2}}{2}\cos\theta_{s2}$,

$K_{pz\,32} = -\sin\dfrac{\alpha_{s2}}{2}\sin\theta_{s2}$.

5   Summaries

In the paper, the rules of forming cycle average control power be studied on common single channel control-and-guide systems. by analysis of characteristics of some system control-and-guide system and rensearch on its Blank Block's impact on control power, the calculation of cycle average control power be analyzed cautiously, and there are four kinds and eleven cases in this calculation. Then mathematic model of command signal and instruction be educed and established according to every case.

This model is considerably simplified the process of control-and-guide system simulation, and convenient for computer calculation. It has been validated by computer simulation on this system control-and-guide closed loop calculation, and proved reliably and accurately. Results of computer calculation is fully consistent to practical testing.

References

[1] Zhang, Y.-P.: Guided System. Artillery Commanding Institute, pp. 184–186 (2005)
[2] Sun, W.-J.: Research Report On Some System Control-And-Guided System Simulation. Artillery Commanding Institute (2003)
[3] Sun, W.-J.: Control and guide theory. Artillery Commanding Institute (2003)
[4] Yu, C.-Z.: System generalities. Peking Industry College Publishing Company (1985)

An Ontology Cloud-Shadow Model Based Knowledge Service Framework

Feng Wang[1], Jiang Zhu[1], Maoxiang Chen[2], and Shoulin Shen[1]

[1] Department of Military Operational
Nanjing Army Command College
Nanjing, China
[2] 61541 Troop
Beijing, China

Abstract. In this article, recent research of knowledge management is reviewed and a cloud-shadow model based knowledge service framework(CBKSF) is proposed. This framework use the definition and operation from cloud-shadow model theory which can reflect new-knowledge-view and knowledge's characters such as uncertainty, inconsistency, time-varying and sociality, regularity. By this means, it has some special advantages and can solve the problem exists in other frameworks. It is more suitable for Semantic Web.

Keywords: ontology, Cloud-Shadow model, ontology cloud, knowledge service.

1 Introduction

Current challenges of applications for semantic web had been discussed in SVWIC (Silicon Valley World Internet Center Symposium). As was concluded, most of these applications require a flexible, credible, collaborative architecture supplying deep knowledge-sharing and utilization in an open and dynamic environment.

In traditional architecture, knowledge must be objective, consistent. it is extracted, purified, stored and acquired in centre market, eliminating the subjective aspects of knowledge and leaving aside user context of using knowledge, it led to depriving the knowledge of its creativity and variety.[1][2]. Different from traditional architecture, distributed knowledge management architecture (DKM) regards that viewpoints from different domain are not an obstacle for using knowledge, but quite a opportunity to pregnant with creativity and vitality of knowledge [3]. So more and more people focus on DKM. By the way, P2P, Semantic Web, MAS (multi-agent) technology is widely used in DKM.

Kex[3] has proposed putting same nodes in a group, aggregating knowledge bottom-up at different layer. yet the classfication of nodes is from the P2P viewpoint leaving aside knowledge structure. FP5[6] project put same concepts in the same category, without taking into account the perspective of users context. (under different context, one concept may be put in different group, and with different understanding). In addition, it generates new knowledge only according to statistics, ignoreing the ambiguity of knowledge, uncertainty, disharmony and other properties. SWAP[4],

InfoQuilt[5], FP5 involved knowledge discovering, ontology creating, and ontology evolutionary process, but the process is not automatic enough, and still very vague.

The rest of paper is structured as follow, In Section 2 an ontology-cloud model is put forward, CBKFM is proposed in section 3. Relative operation and function of framework is illustrated in section 4. In the last section, some overall conclusions along with the future work are presented.

2   Ontology Cloud-Shadow Theory

Ontology Cloud-Shadow Theory is put forward to reflect the new-knowledge-view and pragmatic which leads to virginity and creativity of knowledge, and the knowledge's characters including uncertainty, inconsistency, time-varying and sociality, regularity is concerned too. As has been described in detail[7][8], here is only a brief introduction. Ontology Cloud-Shadow Theory consists of the following five components:

1) Philosophical basis--new knowledge view In the classical knowledge view, to be knowledge at least three criteria must be fulfilled: justified, true and believed. Knowledge, therefore is some entity existing independent of the mind of individuals and we can say things about it that are objectively, absolutely and unconditionally true or false

Since the twentieth century, there is growing criticism and doubts that the scientific rationality and the absolute objectivity of knowledge. People put forward new knowledge view which includes several characters: 1) Construction: knowledge is constructed by the knower based on mental activity instead of cognition of absolute subjectivism. 2) Individuality: knowledge is created and accepted by individual one, if that experience changes, the knowledge may need to be modified. 3) Environment sensitivity: under different circumstance knowledge has his personal meaning. 4).Sociality: all knowledge are created under the guidance of society's value, knowledge is a product of social construction.

In past, many knowledge and intelligent system and method are based on traditional knowledge view, problems and questions are caused by its absolute objectivity of "knowledge". So, new knowledge view is adopted consciously or unconsciously in recent years. New knowledge view have an important instructive meaning to nature science methodology, that is, knowledge has the character of uncertainty, inconsistency, time-varying and abnormality, it also has the character of sociality and regularity, all above led to the creativity and vitality of our world.

2) The basic concepts of model

Ontology cloud model theory involves several entities: personal ontology, common ontology, group ontology and ontology cloud. As Figure1 shows, ontology cloud is made up of common ontology and a set of personal ontologies. Moreover, fc function is defined to map from common ontology to personal ontology, fc is very like the "viewpoint" of common KADS[9], independent variables of fc is related to the context of using personal ontology.
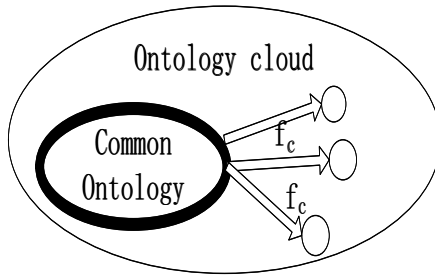
Fig. 1. Ontology cloud model

3) Semi-formalized description

Ontology cloud model needs clearly expressed using predicate logic and set theory. Some definition about ontology-cloud model is given as semi-formalized description is list below: Ontology cloud (personal ontology, common ontology), Ontology relation (divergence-relation, same-as-relation, overlap-relation, high-universal-relation, low-universal-relation), Ontology operation (integration-operation, combination-operation, division-operation), knowledge granularity.

3   CBKMF

Ontology cloud has sociality, ontology cloud may be seemed as a social group for individual ontology to join in and individual ontology may join different group according to context. Therefore, it is natural of combing with distributed architecture such as P2P and social networks. Seen from Fig 2. CBKSF is based on distributed architecture, using CORBA，EJBs，COM/DCOM or JINI technology. it provides service such as ontology discorvery, Resource Allocation, Agent Management.

Ontology cloud were used to construct distributed ontology system in which each ontology agent is an ontology   provider, those agents form self-organized ontology-cloud scheduling topology structure.

CBKSF not only fufill the goal of knowledge-sharing and reusing, but also try to solve the shortage of framework mentioned above. In this framework, the node sturcture and knowledge structure are combined together, knowledge can be expressed in a an appropriate manner which fully reflect the knowledge characteristic, and provide interaction, management, discovery, evolution function and operation of knowledge.

In CBKSF, we use clustering method on a set of similar personal ontology to create an ontology cloud containing a common ontology. Common ontology in the low level is the personal ontology of the high level ontology cloud. Then recursively use clustering method, we can get highest level ontology cloud (the Global cloud) at last. A lower level in the hierarchy corresponds to a more specific cloud (Local Cloud), less abstract representation of the domain. Moreover by ontology cloud combination and division operation, different granularity knowledge is built.
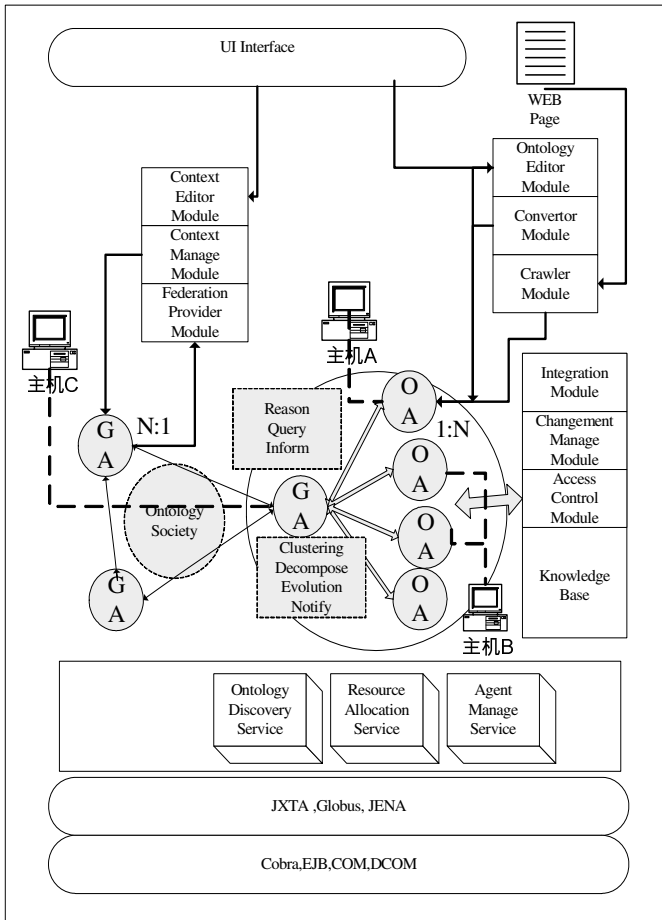
Fig. 2. Cloud Based Knowledge Management Frame (CBKMF)

There are several import functions and operations in framework are list below:

1) Knowledge Discovery Operation: Every OA(Ontology Agent) can provide his own knowledge. Ontology Editor Module gives interface to add, delete, change the relation between concepts, which help knowledge engineer intervening into the progress of ontology creation. Ontology Converter Module provide automatic and semi-automatic ontology creation capability. Crawler Module can draw knowledge from web, and this knowledge will be marked and drawn into ontology later.

2) Context Handle Operation: In ontology cloud model, Mapping function fc is related with context. Context Editor Module gives interface to edit context by Q/A(Question/answer) or Office assistant manner. Context Manager Module provides context to GA, by context, GA can be mapped into one OA in group, and this OA can be provided to application layer for solving special task. Federation Provider Module put ontology into group by context.

3) Knowledge Management Operation: Changement Manager Module propagates all change of knowledge souce to related OA. Integration Module links OA into different

group to serve application's need. Access Control Model controls the combination or division of ontology cloud according to knowledge granularity, deciding which level of ontology cloud are published in public and user can access then.

4   Functions in CBKMF

Some functions which is key to CBKSF are list below. Other functions such as semantic quering, knowledge space constructing, knowledge discoverying have been discussed in other framework, it's same with our framework, It's beyond our talk here.

1) Ontology Registration Funtion

In ontology clouds model, group ontology and common ontology at all level are registered and published in public. Personal ontology registered at group ontology, and only save the difference compared with group ontology.

GA(Group ontology agent) have a few copies, it checks out outdated and useless connections and removes them. When a GA fail, GA generate new group and register it.

2) Ontology Join / Leave

Ontology can join or leave ontology cloud according to the context, GA will check the status of OA while joining GA. If OA is in the constraints of GA cloud , accept it, otherwise reject it.

3) Ontology Notify

GA notify his changes to the OA by flooding approach. when   update notification from GA is obtained by OA,   OA marks it and request at runtime to cache a copy of GA, but not apply for new copy immediately. That avoids bandwidth consumption for least-used-service, on the other hand, frequently-used-OA is in the cache so as to reduce overall system cost.

OA notify his changes to the GA by unicast, GA counts the change, when it is accumulated to a threshold, GA calculates the same concepts and the dissimilar concepts to form new group.

4) Ontology Querying

Cloud model makes ontology sharing and querying quite wasy. Each OA have a complete copy of the GA, and application can access to the latest copy of GA. If nothing changed in GA, getting it from the cache. In additional, GA has a certain hierarchy, low-level GA can be used as a OA in high-level clouds, it is very convenient for manage ontology, different levels of knowledge of cloud reflecting different granularity. According to application needs, while quering, the scope is getting smaller and smaller, from the top to the low level, from the coarse to fine. As a result, ontology queries therefore are limit to a size, and will not grow unlimitedly.

5) Ontology Integration

Ontology Integration is a important process of ontology cloud model. There are two different reuse processes which can be used according to task in distributed architecture: merge and integration. Merge is the process of building an ontology in one subject reusing two or more different ontologies on that subject Integration is the process of buildingan ontology in one subject reusing one or more ontologies in different subjects[9].

6) Ontology Division

There are two type of inconsistency, that is syntax inconsitency and semantic inconsistency[10]. Most systems provide a simple way to treat inconsistency. that is

deleting all sub concepts when deleting a super concept. but it makes against maintaining the diversity of knowledge.

But in Cloud shadow, inconsistency are allowed in sub-ontologys. When inconsistency is beyond threshold, the ontology cloud will divide into several clouds at divergence point. Later, each ontology cloud contact members by broadcast.

5   Summaries

Our framework is not limited to the knowledge service discovery and distribution, but also providing the definition and operation of interaction, management, discovery of knowledge, such as ontology registration, ontology join/leave, ontology notify, ontology searching, ontology integration, ontology division.

The framework uses the definition and operation from cloud-shadow model theory which can reflect new-knowledge-view and knowledge's characters such as uncertainty, inconsistency, time-varying and sociality, regularity. By ontology cloud model, knowledge can be expressed, stored, maintained, shared in a uniform manner. what's more, each operation in ontology cloud model has algorithm in detail, which can heighten automatic ability of knowledge service.

References

[1] Bonifacio, M., Bouquet, P., Traverso, P.: Enabling Distributed Knowledge Management: Managerial and Technological Implications. Novatica and Informatik/ Informatique 3(1) (2002)

[2] Zaihrayeu, I., Bonifacio, M.: KEEx: A Peer-to-Peer Solution for Distributed Knowledge Management. In: Proc. of the MobiQuitous 2004 Workshop on Peer-to-Peer Knowledge Management, P2PKM 2004, Boston, MA, USA, August 22 (2004)

[3] Sheth, A.P., Thacker, S., Patel, S.: Complex relationships and knowledge discovery support in the InfoQuilt system. VLDB Journal 12(1), 2–27 (2003)

[4] Ehrig, M., Tempich, C., Staab, S.: SWAP:ontology-based knowledge management with peet-to-peer. In: Proc. of the 4th European Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS 2003, London (2003)

[5] Arumugam, M., Sheth, A., Arpinar, I.B.: Towards Peer-to-Peer Semantic Web:A Distributed Environment for Sharing Semantic Knowledge on the Web. In: Proc. of WWW- 2002 Workshop on Real World RDF and Semantic Web Application, Honolulu, Hawaii (2002)

[6] Aldea, A., Bañares-Alcántara, R., Bocio, J., et al.: An Ontology-Based Knowledge Management Platform. In: Proc. of IJCAI, Acapulco, México (August 2003)

[7] Zhu, J.: A ontology cloud based parameter expressing method for content adaptatiom. In: WCSE 2008, Moscow, pp. 143–147 (2008)

[8] Zhu, J., Wang, W.: New-Knowledge-view based ontology cloud model. In: CSSE 2008, Dalian, pp. 1140–1144 (2008)

[9] Benjamin, A., Ehrig, M., Euzenat, J., et al.: Integrating Ontologies. In: Proc. of K-CAP, Banff, Canada (October 2005)

[10] Zhu, X.: Inconsistency Measurement of Software Requirements Specifications: An Ontology-Based Approach. In: Proc. of the 10th IEEE International Conference on Engineering, ICECCS 2005, pp. 402–410. IEEE Computer Society, Washington, DC (2005)

Correlation Analysis between Stand Growth Factors and Volume of Stand Using Generalized Linear Models

Lei Kong, Hua Yang, Xin Gang Kang, and Shuo Cai

The Key Laboratory for Silviculture and Conservation of Ministry of Education
Beijing Forestry University
Beijing, P.R. China, 100083

Abstract. As for simulating stand growth Model, the classical linear models often failed to explain the discrete component of data. However, some discrete variables really influence the stand growth. In this research, this paper describes the use of the generalized linear model for analyzing linear and non-linear effects of continuous and discrete predictor variables on a continuous dependent variable. The relationship between stand growth characters and volume of stands was discussed as a case study. Through analyzing attribute of the 1046 stander samples, generalized linear models to the analysis of stand growth data was proposed. A class of estimating equations was introduced, and consistent estimates of the regression parameters were given. All the results of tests were significance.

Keywords: Generalized Linear Models, discrete variables, stand growth.

1   Introduction

The modeling of natural forests growth and yield has received increasing attention in recent years, as they play a more important role than planted forests in ecological services. And the natural forests resource is one of the important indicators for characterizing local ecological environment status. It is clear that, because of the long production cycles and of the numerous management alternatives, the possibilities of learning from direct experimentation or through trial and error are limited. Consequently, mathematical growth models capable of predicting treatment effects are essential for rational forest management. [1]

At stand scales, modeling tree growth has gained momentum with recent developments in the fields of statistical techniques applied to forestry management. Indeed, many regression models have been fitted to the data with a variety of continuous factors, including the average height of stand, Canopy Density or the average age of stand, but without concern for discrete factors like the Origin of stand, slope position. One of the problems with these analyses is that Model results are dependent on the adequacy of cases and continuance of variables. To provide good predictions of stand growth, caution must be given to the discrete variables for analysis. Thus, generalized linear models have been applied in modeling stand growth.

The developments leading up to the general overview of statistical modelling, known as generalized linear models (GLMs), extended over more than a century. In

the years since the term was first introduced by Nelder and Wedderburn in 1972, generalized linear models have slowly become more well known and more widely used.[2] A thorough account is given in McCullagh and Nelder(1989).[3]

The methodology has already been applied in modeling characters of forest, and will get a future development. Wilfried et al [4] applies three statistical methods (generalized linear models, generalized additive models and classification tree analysis) to predict spatial distribution of plant species at different scales. Gretchen et al [5] uses generalized linear models to construct approximately unbiased and efficient estimates of population totals.

2   Methods

2.1   Gerneralized Linear Mordels

In general, GLMs are an extension of the linear regression methods, which allow spatial and temporal nonstationarities to be incorporated and quantified for variables of interest. [6] Comparing classic multiple regression, GLMs provide a less restrictive form. They provide error distributions for the dependent variable other than normal and non-constant variance functions. [3] A typical assumption in GLMs is that all of the observations are drawn from the same family of distributions, for example, normal, Poisson, Logistic, or gamma. A GLM is importantly a multiple regression model for the chosen family of distributions; the regression-like approach makes the individual distributions change with time, site and external factors. Using likelihood-based methods, it can be judged if a possible factor has a genuine effect on the studied phenomenon.

There are 3parts in the generalized linear models: the random component, the systematic component and the link between the random and systematic components.

1. The random component: As an illustration, define Y to be the response variable of interest with components of Y being independent, coming from an exponential family, not just the normal. E(Y) =μ and constant variance σ2;

2. The systematic component: The x may be discrete variables, or continuous variables. The resulting model can be used to predict the response variable over any surface;

$$Y = b_0 + b_1 X_1 + b_2 X_2 + ... + b_k X_k$$

In this equation b0 is the regression coefficient for the intercept and the bi values are the regression coefficients (for variables 1 through k) computed from the data.

3. The link between the random and systematic components: the linear (multiple regression) model might be inadequate to describe a particular relationship is that the effect of the predictors on the dependent variable may not be linear in nature. To overcome this problem, the link function is used. In this paper, the response variable is modeled as a Poisson random variable with the log link function.

$$g(\mu_i) = \log(\mu_i)$$

However, one could assume that the response variable is Binomial or Multinomial, but the results would not differ from those obtained assuming the response variable to be Poisson distributed.

2.2   Akaike Information Criterion

To select for the most parsimonious model, the Akaike Information Criterion (AIC) is used as an automatic stepwise model selection. [7] Generally speaking, with the lower AICc value, the predicted value of the model is closer to the observed value. The model can be selected with the smallest AICc.

$$AIC_C = 2n \log e(\hat{\sigma}) + n \log e(2\pi) + n \left\{ \frac{n + tr(S)}{n - 2 - tr(S)} \right\}$$

Where n is the sample size, and   is the estimated standard deviation of the error term, and tr(S) denotes the trace of the hat matrix.

2.3   Spearman's Rank Correlation

Spearman's rank correlation provides a distribution free test of independence between two variables. It is, however, insensitive to some types of dependence. Kendall's rank correlation gives a better measure of correlation and is also a better two sided test for independence. Spearman's rank correlation coefficient is calculated as:

$$\rho = \frac{\sum_{i=1}^{n} R(x_i)R(y_i) - n\left(\frac{n+1}{2}\right)^2}{\left[\sum_{i=1}^{n} R(x_i)^2 - n\left(\frac{n+1}{2}\right)^2\right]^{0.5} \left[\sum_{j=1}^{n} R(y_j)^2 - n\left(\frac{n+1}{2}\right)^2\right]^{0.5}}$$

where R(x) and R(y) are the ranks of a pair of variables (x and y) each containing n observations.

2.4   Likelihood Ratio Test

In statistics, a likelihood ratio test is a statistical test used to compare the fit of two models, one of which (the null model) is a special case of the other (the alternative model). The test is based on the likelihood ratio, which expresses how many times more likely the data are under one model than the other. This likelihood ratio, or equivalently its logarithm, can be used to compute a p-value, or compared to a critical value to decide whether to reject the null model in favour of the alternative model.

3   Case Study: Application of a Generalized Linear Models to Establish Stand Growth Model

The data of case study was forest management inventory data which was conducted at the Jin Gouling forest farm, Wangqing country, Jilin Province of china. There were 1064 standard sample stands in the survey data. The investigable variables were Canopy Density, dominant species, species composition, age, Origin, slope gradient and so on. From the standard sample stands, 7 variables were selected to study the effect factors of stand growth.

3.1   Data Sets

- Origin: it means the origin of stands, and it includes 0=natural forest and 1=plantation;
- Slopegrad: it means the slope gradient of stands, and it includes 1=level ground, 2= gentle slope, 3=descent slope and 4=steep slope;
- Slopeposi: it means the slope position of stands, and it includes 1=valley, 2=flats, 3=down slope, 4=mid slope and 5=up slope;
- Average: it means the average age of stands;
- No: it means the number of trees per hectare, and it represents the density of stand;
- Averdbh: it means the average DBH (the diameter at breast height) of stands.
- Avervolume: it means the average volume of stands;

3.2   Correlation Analysis

We rejected the null hypothesis of mutual independence among Origin, Slopegrad, Slopeposi, Average, NO, Averdbh and Avervolume. However, we could conclude that there was a statistically significant independence among Origin, Average, Averdbh and Avervolume. So Slopegrad, Average and AverDBH were selected to be the independent variables.

Table 1. The results of Spearman's Rank Correlation

Table Head	Origin	Slope grad	Slop eposi	Ave rage	No	Aver DBH
spearman`s rho	-0.16	0.31	0.26	0.51	0.09	0.60
Sig.	0.00	0.00	0.00	0.00	0.00	0.00

From the tab.1, it was clear to know which variable is independent to Avervolume. There were negative relationship between the variable of Origin and the variable of Avervolume. If the origin of the stand was plantation, the average volume of stand was smaller than the natural forest.

There were positive relationship between the variable of Slopegrad and the variable of Avervolume. If the slope gradient of stands was bigger, the average volume of stand was bigger. There were positive relationship between the variable of Slopeposi and the variable of Avervolume. If the slope position of stands was upper, the average volume of stand was bigger. There were positive relationship between the variable of Average and the variable of Avervolume. If the average age of stands was older, the average volume of stand was bigger. There were positive relationship between the variable of No and the variable of Avervolume. If the number of trees per hectare was more, the average volume of stand was smaller. There were positive relationship between the variable of AverDBH and the variable of Avervolume. If the average DBH of stands was bigger, the average volume of stand was bigger.

3.3 The Distribution of Dependent and Independent Variables

As the figure1 showed, variable Slopegrad was discrete variable, not normally distributed; the classical linear model was not fit for the data. Inspection of the distributions of variables is critically important when using the generalized linear model. The stander deviation and mean of Avervolume were approximately equivalent. GLMs were considered to fit the data, and Poisson regression model was chose as the link function. Moreover, the linear regression model could be also chose, for the dependent variable Avervolume distribution was approximately normal.
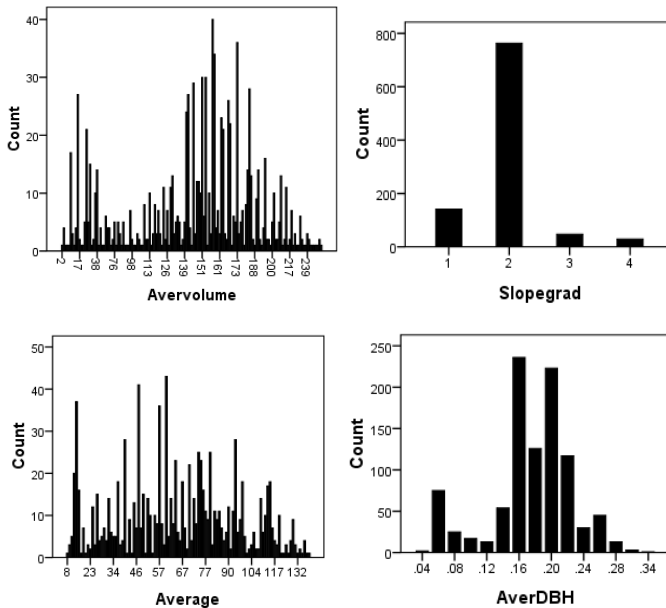
Fig. 1. Bar of count by variables

Totally, four variables were used in the stand growth models. However, it was hard to decide which variable should be taken, or all the variables should be taken. Thus, in order to obtain the best model, different Combinations of these variables were calculated in the stand growth models.

3.4 Statistical Models

Statistical analyses were performed using SPSS, with generalized linear models. In order to evaluate the quality of predictions, two different link functions, the linear regression model and the passion regression model were taken in the models. Each model was tested by the method of goodness of fit, and the results were showed as table 2.

Table 2. AIC, DF and Log Likelihood for stand growth models with different link functions

models	link	AIC	DF	Log Likelihood
1	possion	20432	974	-10210.2
	linear	1076020	974	-538004
2	possion	27292	975	-13641.2
	linear	1856004	975	-927997.3
3	possion	20708	975	-10349
	linear	1113300	975	-556645.2

As mentioned before, the lower the AIC, the closer the approximation of the model to reality. Thus the best model was the one with the smallest AIC. In Model 1, the model using link function which was passion regression model had the smallest AIC, AIC=20432, and its Log Likelihood was the biggest, Log Likelihood=-10210.2. The model using link function which was passion regression model was the best among these models.

Table 3. Coefficient, Stander error and 95% Wald Confidence Interval for the model 1

Parameter	coefficient	Std.error	95% Wald Confidence Interval		Sig.
			Lower	Upper	
Intercept	1.249	0.0312	1.188	1.31	0
Slopegrad=4	0.128	0.0657	-0.001	0.256	0.052
Slopegrad=3	0.23	0.0554	0.121	0.339	0
Slopegrad=2	0.292	0.0355	0.222	0.361	0
Slopegrad=1	0a				
average	0.002	0.0004	0.001	0.003	0
averDBH	6.15	0.263	5.636	6.665	0

a.   Set to zero because this parameter is redundant.

Through calculated the data by computer, it was easy to get the coefficients, stander error, 95% Wald Confidence Interval and significance of coefficients. From the table 3, it was clear that all coefficients of the model were significant (significant level < 0.1). The final model was the following:

$$Avervolume = 1.249 + 0.002 Average +$$
$$6.150 AverDBH, (Slopegrad = 1)$$
$$Avervolume = 1.249 + 0.002 Average + 0.292 Slopegrad$$
$$+6.150 AverDBH, (Slopegrad = 2)$$

$$Avervolume = 1.249 + 0.002 Average + 0.23 Slopegrad$$
$$+ 6.150 AverDBH, (Slopegrad = 3)$$
$$Avervolume = 1.249 + 0.002 Average + 0.128 Slopegrad$$
$$+ 6.150 AverDBH, (Slopegrad = 4)$$

3.5   Wald Test

The Wald test is to test the significance of particular explanatory variables in a statistical model. In possion regression we have a binary outcome variable and more explanatory variables. For each explanatory variable in the model there will be an associated parameter. The Wald test, described by Polit [8]and Agresti, [9]is able to test whether the parameters associated with a group of explanatory variables are zero. If for a particular explanatory variable, or group of explanatory variables, the Wald test is significant, then the parameters associated with these variables are not zero, so that the variables should be included in the model. If the Wald test is not significant then these explanatory variables can be omitted from the model.

Table 4. The Wald test of the model1

source	wald chi-square	df	sig.
intercept	5982.563	1	0.000
Slopegrad	76.517	3	0.000
average	21.659	1	0.000
AverDBH	548.743	1	0.000

From the tab.4, the Wald test was significant, the parameters associated with these variable Slopegrad, average and AverDBH were not zero. so that the variables should be included in the model.

4   Discussion and Conclusion

The paper illustrates how generalized linear models can describe changes in the stand growth. In particular, the methodology has great potential to identify weak relationships between variable Slopegrad and variable avervolume. Although the present paper does not exhaust all possible factors influencing average height of stand, it reveals some important features of the stand growth.

Generalized linear models relax many restrictions, which are often violated in practice. For example, binary (yes/no or 0/1) responses do not have same variance across classes. Furthermore, the sum of terms in a linear model typically can have very large ranges encompassing negative and positive values. For the binary response example, we would like the response to be a probability in the range [0, 1]. However, some limitations of the generalized linear modeling should be paid more attention.

The approach includes the need for the iterated weighted least squares (IWLS) procedure for estimation and deviances for inferences; these restrict the class of models that can be used and do not allow direct comparisons among models from different distributions. [10] The generalized linear modeling methodology will be gradually outdated, for more powerful methods, involving wider classes of distributions, non-linear regression, censoring and dependence among responses, are required. These methods would be got further application in modeling stand growth.

References

[1] Garcıa, O., Forestal, I.: Stand growth models: Theory and practice. In: Proceedings of the IUFRO Seoul Conference on Advancement in Forest Inventory and Forest Management Sciences. Forestry Research Institute of the Republic of Korea, pp. 22–45 (1993)

[2] Nelder, J.A., Wedderburn, R.W.: Generalized linear models. Roy. Statist. Soc. Ser. A 135, 370–384 (1972)

[3] McCullagh, P., Nelder, J.A.: Generalized Linear Models, 2nd edn., p. 532. Chapman and Hall, New York (1989)

[4] Thuiller, W., Araújo, M.B., Sandra, L.: Generalized models versus classification tree analysis: Predicting spatial distributions of plant species at different scales. Veg. Sci. 14, 669–680 (2003)

[5] Moisen, G.G., Edwards, T.C.: Use of generalized linear models and digital data in a forest inventory of northern Utah. Agric. Biol. Environ. Statist. 4, 164–182 (1999)

[6] Dobson, A.: An Introduction to Generalized Linear Models, 2nd edn., pp. 145–147. CRC Press, Boca Raton (2001)

[7] Chambers, J.M., Hastie, T.J.: Statistical models in S, pp. 23–25. Chapman&Hall, London (1997)

[8] Polit, D.: Data Analysis and Statistics for Nursing Research, pp. 45–46. Appleton&Lange, Stamford (1996)

[9] Agresti, A.: Categorical Data Analysis, 2nd edn. The Wiley Series in Probability and Statistics, pp. 24–26. John Wiley and Sons, New York (1990)

[10] Lindsey, J.K.: A review of some extensions to generalized linear models. Statistics in Medicien, Statist. Med. 18, 2223–2236 (1999)

Research on Biotic Complex System Based on Multi-Agents Technology

Hongliang Gao, Xianhe Zhang, and Jinhua Liu

College of Mechatronics and Control Engineering,
Hubei Normal University
Huangshi, P.R. China

Abstract. In recent years, much attention has been focused on the research of multi-agent system (MAS), since it is widely applicable in many areas. Model building and simulation of MAS is very important for the research of MAS. In this paper, the function structure and a multi-agent system architecture of biotic population is researched. Methods of how to model and simulate a MAS are discussed. Modeling and simulation for a complex system—epidemic model based on StarLogo TNG is presented in this paper. From the simulation results, the certain regularity of a biotic community is got.

Keywords: Multi-agent system, StarLogo TNG, Complex system, Epidemic model.

1 Introduction

Multi-agent system theory has been a research focus in recent years. The main problems of multi-agent systems include cooperative control[1], formation control [2~4], flocking[5], consensus[6~8] and so on. Jadbabaie, Lin, and Morse first gave a theoretical analysis of the consensus of Vicsek model [7]. Using some basic knowledge of algebraic graph theory, Olfati Saber and Murray [8] gave some distinguished theoretic results on the consensus problem of multi-agent systems with single integrator dynamics, in which the strong connectivity condition played a key role in achieving the consensus. Ren and Beard [9] extended the results of reference [8] and obtained a weaker condition.

Lately, great progresses have been made in communication, computing technology and relevant issues of biology, social behavioral science, statistical physics and computer graphics, so the research in this area has been supported and stimulated greatly. Researchers want to make explicit that how to possess certain regularity without point coordination to autonomous movement of creatures (clusters of birds, fish, and crowd of people) or autonomous artificial agents [10].

In this paper, we analyzed the function structure of agent in part 2, and then discussed the design and implementation of multi-agent system in part 3, in part 4 the detailed simulation results on epidemic problem are presented to show the effect of the total number of a colony on biotic population.

2   Function Structure of Agent

Considering that there are N+1 agents, their dynamic behavior is given by

$$b_i = v_i \ , v_i = u_i, \ i{=}0, 1, 2... \ N$$

where $b_i = (x_i, y_i)^T$ is agent i's position vector, $v_i = (\dot{x}_i, \dot{y}_i)$ is velocity vector, $u_i = (u_{xi}, u_{yi})^T$ is the control input of acceleration.

If there is a leader in a MAS, suppose the leader moves at a speed of $v_0(t)$, the relative position between agent i and the leader is r_i, the relative speed is $\dot{r}_i = v_i - v_0$, the kinematical equation can be expressed as follows:

$$\frac{d}{dt}\begin{pmatrix} r_i \\ \dot{r}_i \end{pmatrix} = \begin{pmatrix} \dot{r}_i \\ u_i - \dot{v}_0 \end{pmatrix}, \quad i = 1,2,...,N \tag{1}$$

The relative position between agent i and agent j is given by $r_{ij} = r_i - r_j$. The control input of agent i is given by

$$u_i = -\sum_{j \in N_i} \Delta r_i U_{ij} - \gamma \dot{r}_i + \dot{v}_0 , i = 1,2,...,N \tag{2}$$

where γ is a positive constant, U_{ij} is a potential field function which depends on the relative distance between agent i and agent j, an example of potential field function is given by

$$U_{ij}(\|r_{ij}\|) = \frac{1}{\|r_{ij}\|^2} + \log\|r_{ij}\|^2 \tag{3}$$

The idea behind any multi-agent system is to break down a complex problem handled by a single entity – a centralized system – into smaller simpler problems handled by several entities – a distributed system. The multi-agent system architecture of biotic population is presented in Fig. 1.
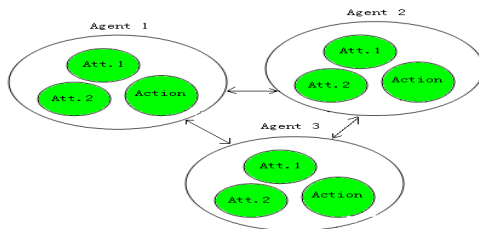
Fig. 1. The architecture of biotic population multi-agent system

As shown in Fig.1, each agent has an action definition, and agents interact with each other based on the rules defined by the action, also the inter-relationship between agents and the environment, so as to indicate the behaviors of complex system. Each agent has different attribute which represents individual's characteristics, that is Att.1, Att.2…, and presented in Fig.1. In fact, an agent is a controllable computer process which has these characteristics. The control model of agents in MAS theory is shown in Fig.1.

3   Design and Implementation

To implement a multi-agent system, there are a number of open-source agent platforms available in the literature that aid developers to build a complex agent system in a simplified fashion. There are some open-source agent platforms. Since performance of these agent development toolkits has been extensively studied previously, comparison of these agent platforms will not be discussed in this paper. In this paper, due to the object we researched is a complex system without a leader, we choose StarLogo TNG as the platform.

StarLogo TNG (The Next Generation) is developed at Media Laboratory, MIT. StarLogo is a programmable modeling environment for exploring the workings of MAS, especially for systems that are organized without an organizer, coordinated without a coordinator. By using StarLogo, we can model (and gain insights into) many real-life phenomena, such as bird flocks, traffic jams, ant colonies, and market economies.

StarLogo is a specialized version of the Logo programming language. With traditional versions of Logo, you can create drawings and animations by giving commands to graphic "agents" on the computer screen. StarLogo extends this idea by allowing you to control thousands of graphic agents in parallel. In addition, StarLogo makes the agents' world computationally active: you can write programs for thousands of "patches" that make up the agents' environment. Agents and patches can interact with one another [11, 12].

First, we created some agents (such as fish), set the number as 1000 and put them on the screen, let's leave their color as white (healthy), and then some of them should be infected, we can infect individuals with a certain chance (e.g., the probability is 10%) at the beginning, then the biotic population is set up (Fig.2).
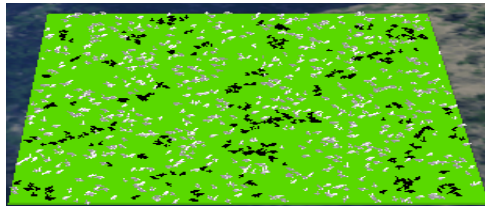
Fig. 2. The biotic population multi-agent model—fish flock

As shown in Fig.2, black agents represent sick individuals and white agents represent healthy individuals. Every agent moves randomly (e.g. every agent moves forward 1step and then turn a little to the left or right randomly), when a black individual touches a white one, the white one will be infected and get sick. We add a recovery variable that can be adjusted by the user and graph the changing number of sick and healthy agents. In many biotic communities some individuals start out with immunity to a certain disease. We can easily model that here, by giving some of the agents "immunity".

What is the result of the fish flock，will all the fish be sick or healthy, how does the probability of recovery (POR) affect the biotic population, and the total number of the fish effects the number of the healthy/sick fish? We have researched on the problems above in the next part—simulation analysis.

4   Simulation Analysis

In a previous paper [13], we have studied the third problem and got the relationship between the probability of recovery (POR) and the biotic population, now we focus on the next issue—how does the total number of the fish effects the number of the healthy/sick agents? In order to help to the analysis, we set the POR as a constant—6.364.

The simulation results are presented in Fig.3-5. In Fig.3, the total number of the agents is 100. After operating for a while, the colony became stable, and the number of healthy individuals is 100, the number of infected agents is 0. The result means the effect of infection is finite to the biotic population with undersized total number of individuals.

As presented in Fig.4, when the total number is set to 200, the number of sick (healthy) agents fluctuates as time goes by. After operating for a while, the number of healthy agents is 80-140, and the sick is 120-60, that means the colony keeps a balance with periodicity.

If the total number of the colony is bigger, such as 1000, the result is different. The curve of the number of individuals is presented in Fig.5. As presented in Fig.5, the number of healthy agents decreased dramatically, accordingly the number of sick agents increased quickly. After running for a short time, the number of the sick agents is 800-900, and then keeps stable. If the number of agents is big, then the density of the fish becomes big, the probability of touching each other also became big. That is the reason why we should keep sparse when a disease attacks people, even though some of individuals have immunity to the disease.
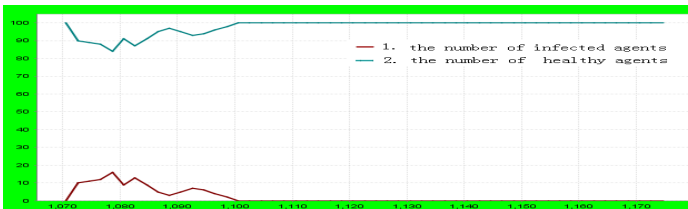
Fig. 3. The instance of the total number is 100
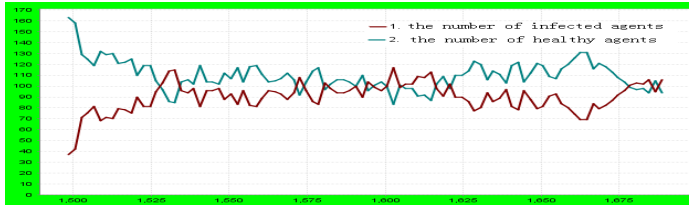
Fig. 4. The instance when the total number is 200
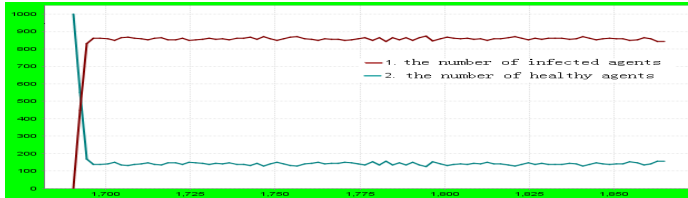
Fig. 5. The instance when the total number is 1000

5   Summaries

Complex system can be described naturally by using methods of modeling and simulation based on agents, and it makes the modeling of complex system more conveniently. The modeling and simulation of initiative behaviors is supported by using the method. Through expressing individuals in complex system by agents, microscopic behaviors of a complex system can be represented well, at the same time behaviors in highly dynamic environment also can be simulated, which is unavailable by traditional methods. In this paper, we have researched on a complex system-fish flock, which is a discrete multi-agent system with no leaders. From the simulation results of the model, we can get the law of the practical colony—the relationship between the total number of individuals and the last state of the colony. That is a decided rule in a complex system (biotic population). The research method can also be used in other areas, such as crowd of people, traffic and economy.

References

[1]   Wang, P.K.C., Hadaegh, F.Y.: Coordination and control of multiple microspacecraft moving in formation. Journal of the Astronautical Sciences 44(3), 315–355 (1994)

[2]   Balch, T., Arkin, R.C.: Behavior-based formation control for multirobot terms. IEEE Transactions on Robotics and Automation 14(6), 926–939 (1998)

[3]   Hu, X.: Formation control with virtual leaders and reduced Communications. IEEE Transactions on Robotics and Automation 17(6), 947–951 (2001)

[4]   Fax, A., Murray, R.M.: Information flow and cooperative control of vehicle formations. IEEE Transactions on Automatic Control 49(9), 1465–1476 (2004)

[5]   Olfati-Saber, R., Murray, R.M.: Flocking for multi-agent dynamic systems: algorithms and theory. IEEE Transactions on Automatic Control 51(3), 401–420 (2006)

[6]   Vicsek, T., Czirok, A., Ben-Jacob, E., et al.: Novel type of phase transitions in a system of self-driven particles. Physical Review Letters 75(6), 1226–1229 (1995)

[7]   Jadbabaie, A., Lin, J., Morse, A.S.: Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Transactions on Automatic Control 48(6), 988–1001 (2003)

[8]   Olfati-Saber, R., Murray, R.M.: Consensus problems in networks of agents with switching topology and time-delays. IEEE Transactions on Automatic Control 49(9), 1520–1533 (2004)

[9]   Ren, W., Beard, R.W.: Consensus seeking in multi-agent systems under dynamically changing interaction topologies. IEEE Transactions on Automatic Control 50(5), 655–661 (2005)

[10]   Low, D.J.: Following the Crowd. Nature 407, 465–466 (2000)

[11]   http://education.mit.edu/starlogo

[12]   Uhrmacher, A.M., Tyschler, P.: Modeling and Simulation of Mobile Agents. Future Generation Computer Systems, 118–124 (2000)

[13]   Gao, H.: Research on Modeling and Simulation for Multi-Agent System. In: The 3nd International Conference on Computer and Network Technology, pp. V9–V70 (2011)

Self-Growing Regularized Gaussian Mixture Models for Image Segmentation

Tao Guan, Hongxia Wang, and Yan Wang

Department of Computer Science and Application, Zhengzhou Institute of Aeronautic Industry Management, Zhengzhou, Henan, China
`timm.guan@gmail.com`

Abstract. Mixture Models(MMs) are a typical class of statistical models and have been applied to image processing in many situations, among which Gaussian MM (GMMs) are widely adopted. Main drawbacks of classical models involve that they need presetting the number of clusters, have not considered the influence of outliers. They will lead to unreasonable image segmentation results. This paper proposes the Self-Growing Regularized GMMs(SGRGMMs), which generalizes the classical GMMs, for image segmentation. We compute the unknown parameters using the self-branching competitive leaning and a new generalized EM algorithm, Regularized EM(REM). We carried out experiments on the segmentation of some images and our approach can automatically determine the number of clusters and efficiently erase the influence of outliers.

Keywords: Gaussian mixture models, Expectation Maximum(EM), Self-branching Competitive Learning.

1   Introduction

Image segmentation can be viewed as a process to partition an image into a series of non-overlapping regions. Currently, there have existed some approaches including histogram thresholding[1] and classification based methods. in which the pixels with same gray or color are grouped to one class. Many classification approaches are based on statistical learning models and the frequently used ones involve vector quantization, mixed models and Markov random fields. Mixture Models(MMs) are established on statistical theory and have received wide attention in image segmentation[2-9]. Gaussian mixture models(GMMs) are a kind of MMs whose components are Gaussian distributions. In image segmentation using GMMs, one assumes that data comply with Gaussian mixture distribution with unknown parameters and then determines the optimal parameters by EM algorithm. In segmentation of remotely sensed images, P. Masson and W. Pieczynski proposed a stochastic variant of EM to segment satellite images[9]. They pointed out that EM related algorithms had large dependence on the initialization. Recently, some researchers further develop GMMs and combine GMMs with Markov Random Fields for the spatial constraints of pixels[5,6]. However, there are still three drawbacks in GMMs and EM algorithm. First, one must preset the initial number of clusters, which

is almost unreasonable in many practical cases. Second, the unsuitable initialization of centers may lead to the local converge of EM algorithm. Third, in general, the GMMs incline to be affected by outliers.

In this paper we propose the SGRGMM model and the REM algorithm by imposing a regularized item on the MAP function to obtain parameter estimation. The Self-Branching Competitive Learning[10] is used to determine the initial number and locations of clusters. We carry out experiments on remotely sensed images and standard test images and get better results compared to GMMs+EM.

2   Mixture Models and EM Algorithm

In the probability space with d dimension, given sample X and probability density function $p_i(X|\theta_i)$, $i \le r$, the mixture model has the following form

$$p\left(X|\Theta\right) = \sum_{i=1}^{M} \alpha_i p_i\left(X|\theta_i\right),\tag{1}$$

where $\Theta = \left(\alpha_1,\cdots,\alpha_M,\theta_1,\cdots,\theta_M\right)$, $\sum_{i=1}^{M}\alpha_i = 1$.

Here, $p_i\left(X|\theta_i\right)$ can be any probability density function. In particular, if $p_i(\cdot)$ is the Gaussian distribution $G_i(\cdot)$, then we get GMMs that are frequently used in image segmentation. The coefficients α_i, which may be distorted by outliers, affect the performance of the estimators.

In common, it is assumed that we know the number of clusters and then compute the maximal likelihood estimation(MLE) of parameters by EM algorithm. EM is an iterative optimal method proposed by A. P. Dempster et al in 1977 and can be used to calculate the optimal parameters of mixture models when sample are incomplete[3]. In sequent years, EM is developed and many variants of it are proposed, such as Stochastic EM(SEM)[12]. However, it is not convenient to directly apply EM to (1) when estimating parameters.

3   Regularized Mixture Models and Generalized EM Algorithm

The conventional GMMs and EM algorithm classify the pixels of an image into several groups, each of which denotes a region with specified feature values. In common, the irregular parts in images, such as isolated points, will affect the convergence of EM algorithm and the segmentation quality of images. To erase the influence, we propose a regularized EM algorithm(REM) that defines a regularized $Q(\cdot|\cdot)$ function for M-step written as

$$\begin{aligned} Q\left(\Theta|\Theta^{(k)}\right) &= E\left[l(\Theta)|X,\Theta^{(k)}\right] - \beta\sum_{j=1}^{N}\sum_{i=1}^{M}\left(\alpha_i^j\right)^2 \\ &= \sum_{j=1}^{N}\sum_{i=1}^{M}E\left[z_i^j|x_j,\Theta^{(k)}\right]\ln\left[\alpha_i^j p_i\left(x_j|\theta_i\right)\right] - \beta\sum_{j=1}^{N}\sum_{i=1}^{M}\left(\alpha_i^j\right)^2, \end{aligned}\tag{2}$$

where $\beta \geq 0$ is the regularized factor , $l(\Theta)$ is the likelihood function and $\sum_{i=1}^{M} \alpha_i = 1$.

When $p_i(x_j | \theta_i)$ satisfies Gaussian distribution (3), $\theta_i = (\mu_i, \Sigma_i)$ with $d=2$, then (13) becomes

$$Q(\Theta | \Theta^{(k)}) = \sum_{j=1}^{N} \sum_{i=1}^{M} \left\{ (\omega_i^j)^{(k)} \ln \alpha_i^j - \beta \sum_{j=1}^{N} \sum_{i=1}^{M} (\alpha_i^j)^2 \right.$$

$$\left. - (\omega_i^j)^{(k)} \left(\ln 2\pi + \frac{1}{2} \ln |\Sigma_i| + \frac{1}{2} (x_j - \mu_i) \Sigma_i^{-1} (x_j - \mu_i) \right) \right\}, \tag{3}$$

where

$$E\left[z_i^j | x_j, \Theta^{(k)} \right] = (\omega_i^j)^{(k)} = \frac{\alpha_i^j p_i(x_j | \Theta_i^{(k)})}{\sum_{l=1}^{M} \alpha_l^j p_l(x_j | \Theta_l^{(k)})} . \tag{4}$$

To compute the unknown parameter α_i^j it needs to solve the following Lagrange equation. From (3) we get

$$\frac{\partial}{\partial \alpha_i^j} \left[Q(\Theta | \Theta^{(k)}) - \left(\sum_{j=1}^{N} \lambda_j \left(\sum_{i=1}^{M} \alpha_i^j - 1 \right) \right) \right] = 0 ,$$

where λ_j is the Lagrange multiplier. Solve above equation and we have

$$2\beta (\alpha_i^j)^2 + \lambda_j \alpha_i^j - (\omega_i^j)^{(k)} = 0 ,$$

Since $\sum_{i=1}^{M} \omega_i^j = 1$ and $\sum_{i=1}^{M} \alpha_i^j = 1$, then we get the iteration at kth step

$$\lambda_j^{(k)} = 1 - 2\beta \sum_i \left((\alpha_i^j)^{(k)} \right)^2 . \tag{5}$$

Obviously, $1 \geq \lambda_j^{(k)} \geq 1 - 2\beta$. Then we obtain the iterative equation at kth step as

$$(\alpha_i^j)^{(k+1)} = \frac{1}{4\beta} \left(\sqrt{(\lambda_j^{(k)})^2 + 8\beta (\omega_i^j)^{(k)}} - \lambda_j^{(k)} \right). \tag{6}$$

To find the priors of each component we take the mean of all α_i^j about j and then get

$$\alpha_i = \frac{1}{N} \sum_{j \leq N} \alpha_i^j . \tag{7}$$

The α_i need normalizing about i so as to satisfy $\sum_i \alpha_i^j = 1$ for j. The prior probabilities of α_i will be used to calculate ω_i^j in every iteration. Let

$$\alpha_i = \frac{\alpha_i}{\sum_{i=1}^{M} \alpha_i} . \tag{8}$$

In the iteration process, the normalized α_i is used to update ω_i^j in (4). For μ_i, we have

$$\mu_i^{(k+1)} = \frac{\sum_{j=1}^{N} \left(\omega_i^j\right)^{(k)} x_j}{\sum_{j=1}^{N} \left(\omega_i^j\right)^{(k)}} .$$ (9)

From (3), we get the covariance matrix for multivariate data as

$$\Sigma_i^{(k+1)} = \frac{\sum_{j=1}^{N} \left(\omega_i^j\right)^{(k)} \left(x_j - \mu_i\right)\left(x_j - \mu_i\right)^T}{\sum_{j=1}^{N} \left(\omega_i^j\right)^{(k)}} .$$ (10)

4   Experiment Results on Remotely Sensed Images

Remotely sensed images are widely adopted in city planning, Agriculture and resource survey. Here, we apply our algorithm to segmentation of color remotely sensed images. Figure 1 (a)   show a remotely sensed image and its segmented image (b) by our approach with $\sigma = 10$, K=10. The results present a favorable segmentation result.

(a)   origin   (b) segmented image
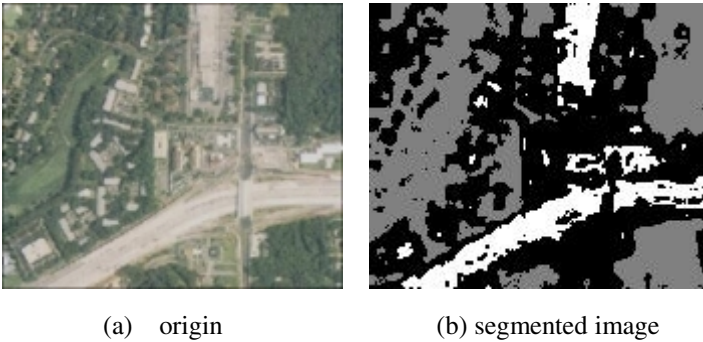
Fig. 1. The segmentation results by our approach

5   Experiment Results on Standard Test Images

We used two standard figures to compare the performance of classical EM and REM in segmentation quality. With the same set of preset parameters, we got the final variances and priors. From the data distribution in feature space as shown in Figure 2, we knew that the data distribution of Donna was more irregular than that of Lena, which made the final priors for Donna had more adjustment. So the difference in segmentation results between (c) and (d) in Figure 2 was less than that between (a) and (b). Since classical EM algorithm was sensitive to initialization and noise, thus the change of priors improved the segmentation quality.

(a) EM   (b) REM

(c) EM   (d) REM
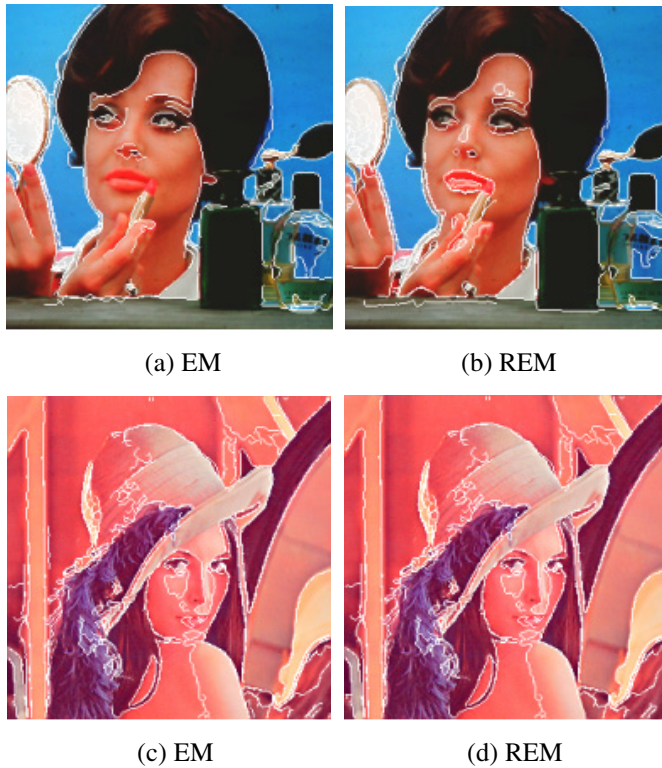
Fig. 2. The segmentation results by our approach and GMM+EM

References

[1] Gonzalez, R.C., Woods, R.E., Eddins, S.L.: Digital Image Processing Using MATLAB, 2nd edn. Prentice Hall, New Jersey (2010)

[2] Mclachlan, G., Peel, D.: Finite Mixture Models. John Wiley & Sons, New York (2000)

[3] Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., Ser. B 39, 618–633 (1977)

[4] Bilmes, J.A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, U. C. Berkeley, CA, TR-97-021 (1998)

[5] Sanjay-Gopal, S., Hebert, T.J.: Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. IEEE Transactions on Image Processing 7(7), 1014–1028 (1998)

[6]   Nikou, C., Likas, A.C., Galatsanos, N.P.: A Bayesian framework for image segmentation with spatially varying mixtures. IEEE Transactions on Image Processing 19(9), 2278–2289 (2010)

[7]   Sfikas, G., Nikou, C., Galatsanos, N., Heinrich, C.: Spatially varying mixtures incorporating line processes for image segmentation. Journal of Mathematical Imaging and Vision 36(2) (2010)

[8]   Blekas, K., Likas, A., Galatsanos, A.P., Lagaris, I.E.: A spatially constrained mixture model for image segmentation. IEEE Transactions on Neural Network 16(2), 494–498 (2005)

[9]   Masson, P., Pieczynski, W.: SEM algorithm and unsupervised statistical segmentation of satellite images. IEEE Transactions on Geoscience and Remote Sensing 31(3), 618–633 (1993)

[10] Guan, T., Li, L.-L.: Self-Branching Competitive Learning for image segmentation. In: International Conference on Bio-Inspire Computing: Theories and Applications, China, pp. 652–656 (2010)

Simulation Research on Ore Block Model Building Based on Mechanical Analysis of FLAC^{3D}

Sun Guang-Hua[1], Li Fu-Ping[2], and Wang Wen-Bin[3]

[1] College of Mining Engineering, Hebei United University, Tangshan, Hebei 063009, China
[2] HeBei Province Key Laboratory of Mining Development and Safety Technique, Tangshan, Hebei 063009, China
[3] Jizhong Energy FengFeng Group Wuan City Nanminghe Mining Co., Ltd., Wuan, Hebei 056300, China

Abstract. Based on powerful simulation ability and calculation function of FLAC3D, in the background of a specific mining engineering project, importance of the preparative work before modeling was studied. The principle and necessity of dividing finite unit grid was analyzed from the point of improving the numerical simulation results. On the basis of analyzing the interface problem between different ore-body and rock-mass, numerical model the mine was built using FLAC3D, and the mechanical simulation was carried out. According to the simulation results, feasibility and accuracy of the numerical model was analyzed.

Keywords: FLAC3D, Numerical model, Simulation, Mining stope stability.

1 Introduction

With the exploitation of underground mine, it inevitably leads to variation of stress in the surrounding rock and rock deformation, such as collapse, fracture, abscission layer, bending, and so on. The analysis of underground mining stope's mechanical stability has never been stopped. Due to the complicated variable orebody shape and occurrence state of metal mine, conventional methods are difficult to do further research. With the development of computer technology, based on numerical modeling, mechanical stability analysis is widely used in geotechnical engineering, geological engineering, mining engineering, and other fields [1-4]. Using computer numerical modeling technology, three-dimensional visualization of mine could be achieved in complex geological conditions. Based on building a platform for numerical computation and analysis, size, speed and reliability of the stope mechanical stability analysis can be improve.

In summary, numerical mining model which is based on mechanical analysis has many advantages, such as fast and accurate calculation, outstanding ability of three-dimensional visualization, and powerful mechanical analysis. In this paper, FLAC^{3D} numerical analysis software was used to build three-dimensional model, and then underground stope's stability was analyzed combined with mining engineering practice.

2   Engineering Situation of Mine

Liu Dian Iron Mine of Hebei Province is buried below the quaternary surface and weathered rock. The thickness of quaternary surface is about 40m and the thickness of weathered rock mass is about 160m. The ground surface is flat and   open, which is the local main agricultural producing areas. The orebody length along strike is about 200m, possessing from 24m to -248m altitude. The average thickness of ore body is 17~20m and the maximum thickness is 46.64m. The strike of this orebody is 40° to the North East, and the tendency is north-westward, and   the inclination angle is 50°.

   To prevent the surface subsidence which was caused by mining, and to ensure normal production of the farmland, subsequent filling phase stope mining method was used. The ore block was layed out vertically, with the strike and the structure parameters of ore block shown in Table 1. Considering filling cost, the room was cementedly filled, and the spacer pillar is uncementedly filled with unclassified tailings.

Table 1. Structure parameters of subsequent filling phase stope mining method (unit: m)

Stage height	Ore block length	Room width	Spacer pillar width	bottom pillar thickness	top pillar thickness
40	17~20	14	8	6	4

3   Three-Dimensional Numerical Model of Mining

Three-dimensional numerical model of mining is a comprehensive reflection on the basis of research fields of mining geology, mining engineering, mechanical analysis, and so on. To ensure the accuracy of three-dimensional numerical mechanical model, the preliminary work of modeling is essential [5]. According to Liu Dian Iron Mine, there is main preliminary work as follows:

   (1) To collect the whole site original data of mining and to ensure that the original data is reliable and true.
   (2) To select the 3D numerical analysis software.
   (3) The original data's extraction and entry must be accuracy.
   (4)The numerical model should be comprehensively and thoroughly examined , and the deviation of the results should be prevented.

In the progress of numerical simulation modeling for mining using FLAC^{3D}, the division of finite difference grid was essential, which directly affected the accuracy of subsequent analysis of numerical results [6]. To reduce the scale of the subsequent simulation process and to improve the accuracy, as well as distribution calculation data to adapt to the characteristics of grid, it was taken to divid the unit grid with density combination form. To better reflect the varivation of stress and strain in the stress concentrated area, the gradient calculations should be appropriately large area dense grid layout, and vice versa. While modeling, it can be divided into the relevant sections of the actual situation. For example, the external section lay large units, inside decreased, and there is only subdivision in shape of the innermost of the grid. It will not only ensure the accuracy, but also that the unit number is not too large. And

there will not be many virtual nodes that will be completely found. Thus, the analysis efficiency is improved and the meshing and its optimization are realized.

In the process of establishing mining numerical mechanical model, it involves issues of surface contact between different mining rocks, such as the Quaternary alluvium and weathered rock contact, weathered rock and the ore body contact, weathered rock and rock on the plate contact, weathered rock and the next disk surrounding rock, ore excavation and filling of the upper and lower plate rock Body contact, and so on. Thus, while building the model, the problem of contact surface's modeling method, parameter selection and some other issues should be fully considered. In the model building process, the length ratio between a unit side and its adjacent cell should be less than 1:1 or 1:2 [7], which will improve the approximation ability to complicated form boundaries. It is better in the effect.

According to the local engineering geological conditions, existing condition of orebody, and the structural parameters of mining method, there are the No.1 stope, the No.2 stope, and the No.3 stope, the frist excavation stage, as the numerical object, using FLAC^{3D} numerical simulation software in mining model.

According to excavation region's geological conditions, three-dimensional geological model size is determined as: length × width × height = 100m (X-axis: vertical to ore-body strike) × 206m (Y-axis: along the ore body strike) × 245m (Z-axis: vertical direction). From top to bottom of the model, it is as follows: the Quaternary alluvium (+50 m ~ +10 m), overlying weathered rock cover (+10 m ~-150m), excavated ore-body (-150m ~-190m). Generated grid is divided into 350,475 units, 374,713 nodes, as shown in Fig.1. Mining stope numerical model is shown in Fig. 2.
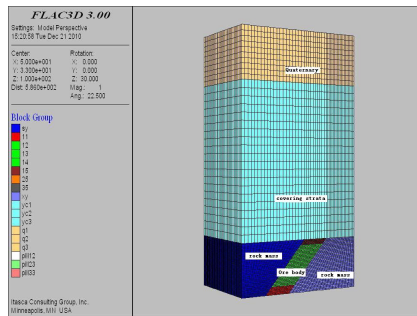
Fig. 1. Geological model
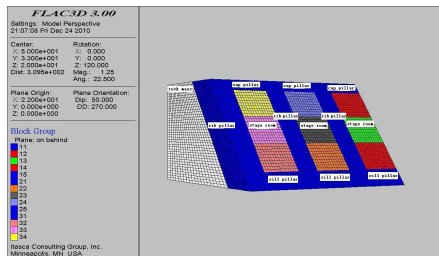
Fig. 2. Numerical model of mining stope

4   Numerical Analysis

Mechanical analysis is made based on geological numerical model. Calculation research scope of materials involves cement mortar filling, ore body, and rock-mass. They are elastic-plastic materials, which are suitable for the Coulomb-Mohr failure criteria. The mechanical model is as follows.

$$\frac{\sigma_1 - \sigma_3}{2} = \frac{\sigma_1 + \sigma_3}{2} \sin \phi + C \cdot \cos \phi \tag{1}$$

$$f_s = \sigma_1 - \sigma_3 \frac{1 + \sin \phi}{1 - \sin \phi} - 2 C \sqrt{\frac{1 + \sin \phi}{1 - \sin \phi}} \tag{2}$$

In the formula:

> σ_1——the maximum principal stress, MPa;
> σ_3——the minimum principal stress, MPa;
> C—— Cohesion;
> φ——Internal friction angle;
> f_s—— Failure coefficient.

The horizontal displacement, around the model, is restricted. The vertical displacement, of the model bottom, is restricted. The vertical load is set to be the weight of overlying rock mass. Three-dimensional stress of corresponding depth to each node of the model is imposed. The model deformation is set to be large deformations, and the model is valued according to relevant mechanical parameters. Then, subsequent stress distribution stope filling is assigned to.

1[#] ore block mining stope is immediately filled with cemented filling body after mining. When the FLAC[3D] program runs 500 steps (step500), the maximum principal stress is mainly concentrated in the filled mining stope's contact zone of hanging wall and footwall. At this point, the performance of the maximum principal stress presented as compressive stress, and the maximum stress values less than 10.1MPa, as shown in Fig.3. After mining the pillar between 1[#] stope and 2[#] stope, the non cemented filling body is immediately filled. It is shown that the maximum stress area mainly focuses on the left lower and right upper corner of the stope, the maximum stress is about 11.59MPa, as shown in Fig.4.
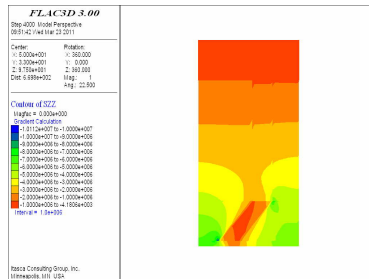
Fig. 3. Maximum principal stress contour of room 1[#]
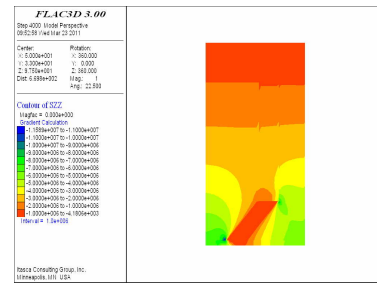
Fig. 4. Maximum principal stress contour of spacer pillar between rooms 1 -2

The ground surface settlement is relatively uniform after filled, as shown in Fig.5 and Fig.6. The maximum displacement ranges from 4.7cm to 4.8cm. Maximum displacement occurs in the central area of mining, and with the distance further away from the central region, the surface displacement decreases. According to stope mining sequence, the stress and displacement changes of the entire mining process can be simulated and analyzed, and reliable mining numerical simulation is achieved, and the safe production in mine is guided.
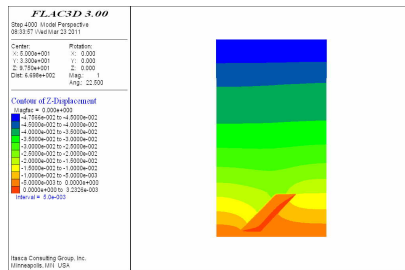
Fig. 5. Displacement contour along the ore-body strike
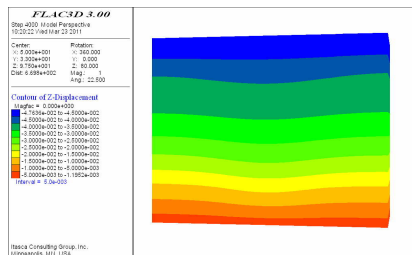
Fig. 6. Displacement contour of room 1 after filled

5   Summaries

According to practical mining example, by using Flac^{3D} numerical analysis software, this paper built mechanical numerical model and conducted numerical simulation, which was important and valuable to engineering projects.

(1) According to a practical mine, the paper discussed the importance of preparatory work in the process of building model. At the same time, the element mesh division principle and key problems of structural plane connection were analyzed.

(2) According to present situation of the mine and structural parameter characteristics of the mining method, numerical model of the mine was built. Based on that, the model was evaluated according to relevant mechanical parameters. And then, some mechanical calculational analysis were done.

(3) The results of numerical analysis could verify the accuracy and effectiveness of numerical model. The results also verified the necessity and feasibility of pre-preparation work in the process of model building.

References

[1] Pinto, V., Font, X., Salgot, M., Tapias, J.C., Mana, T.: Using 3D structures and their virtual representation as tool for restoring opencast mines and quarries. Engineering Geology 63(1/2), 121–129 (2002)

[2] Zhang, Y., Wen, G.-Q., Wang, X.-H.: Application of 3D volume visualization in geology of civil engineering. Chinese Journal of Rock Mechanics and Engineering 21(4), 563–567 (2002) (in Chinese)

[3] Xu, N.-X.: 3D engineering geological modeling method suitable for numerical simulation. Chinese Journal of Geotechnical Engineering 31(11), 1710–1715 (2009) (in Chinese)

[4] Huang, M., Li, X.-B.: Determination of Mine Movement Scope Based on 3D Modeling Technology. Mining and Metallurgical Engineering 29(4), 5–9 (2009) (in Chinese)

[5] Liu, A.-H., Zhao, G.-Y., Zeng, L.-F., Li, X.-B.: Application of 3D Mine Model to Stability Analysis of Landside. Chinese Journal of Rock Mechanics and Engineering 27(6), 1236–1242 (2008) (in Chinese)

[6] Feng, X.-L., Wang, L.-G., Bi, L., Shang, X.-M., Gong, Y.-X.: Cavability of rock mass based on 3 D simulation technology. Journal of China Coal Society 33(9), 971–976 (2008) (in Chinese)

[7] Kou, X.-Y., Jia, M.-T., Wang, L.-G., Wu, X.: Research on optimization ore block model technology for mechanics analysis of FLAC3D. China Molybdenum Industry 34(3), 10–13 (2010) (in Chinese)

Numerical Simulation on Deformation and Stress Distribution around Spraying Hole in Outburst Coal Seam

Yun Bao-Ju[1,2], Cheng Yuan-Ping[1], and Zhou Hong-Xing[1]

[1] Faculty of Safety Engineering, China University of Mining and Technology, Xuzhou, China
[2] College of Mining Engineering, Hebei United University, Tangshan, China
`cumtljz@163.com`

Abstract. Outburst elimination holes through beds are widely used in Chinese low permeability coal seam. The phenomena of hole spraying is helpful to increase the gas extraction volume. The rock mechanical mechanism of the holes is the core issue preventing spraying. Based on the finite difference software FLAC3D, we get appealing simulating results about the coal seam's deformation, stress distribution and displacement distribution around the spraying hole. Therefore, the correlative discipline is gradually cognized and a new reasonable method of penetration improvement is found. Furthermore, to a certain extent the studies we have performed perfects the theory of preventing outbursts and mine gas disasters.

Keywords: Outburst coal seam, Hole spraying, Numerical simulation, Stress and strain law.

1 Introduction

Occasional in outburst coal seam, hole spraying is the result of combined action by rock pressure and gas pressure[1,2]. The hole spraying is helpful to improve the permeability and efficiency, to increase the drainage volume [3,4]. Therefore, the rock's rheology, dilation, deformation, and fracture development should be an important research object on preventing and eliminating outburst. FLAC3D is a three-dimensional explicit finite-difference program for engineering mechanics computation, widely used in geotechnics, mining technology, tunnuel technology, etc [5]. The software provide users with abundant built-in constitutive models such as null, elastic, Mohr-Coulomb plasticity, strain-hardening/softening Mohr-Coulomb plasticity, etc. Some researchers have managed to utilize it in modeling the rock deformation round a cylindrical hole. So we attempt to modeling the coal deformation around spraying hole with FLAC3D. The appealing results help us to recognize the rock mechanism and defoming laws of spraying hole. Further, it will offer instruction to optimization design of hole preventing outburst.

2 Model Construction

2.1 Model Simplifying and the Rock Parameters

Because the main research scope is the surrounding coal and rock near the hole, we simplify our FLAC3D-model with three seams, whose mechanical parameters are shown in Table 1. The spraying hole can be modeled by excavating an proportional cylinder in diameter and height(model null FLAC3D order).

Table 1. Physical and mechanical property parameters of the three rock seams

Lithology	Young's modulus /GPa	Poisson's ratio	Friction angle /°	Tensile strength /MPa	Density /kg·m^{-3}	Thickness /m
Sandy Mudstone	27.0	0.4	30	1.0	2200	15
Coal	0.71	0.3	17	0.8	1400	2
Sandy Mudstone	27.0	0.4	30	1.0	2200	15

2.2 Model Size and Boundary Conditions

Studies have shown the great stress and strain influence range is confined to about four to five times of the excavating hole's diameter. In model constructing, the objecting hole's diameter is set 4 meters, which is accordance with our on-site observation in coal mines of Huainan Group. In order to eliminate the bad influence of boundary condition, we expand the model's horizontal length to 40 meters. So we get a model with 40 meters length, 40 meters width, and 32 meters height. Because of the symmetry of the research scope, we mesh it according to the requirement of the manual. The model consists of 62400 zones and 67229 gridpoints, as shown in Fig. 1 with quarter part. Based on the on-site actual mechanics, the models are confined with roller boundary conditions on five sides of the model, and the exceptional side on the upper model is applied different rock load.
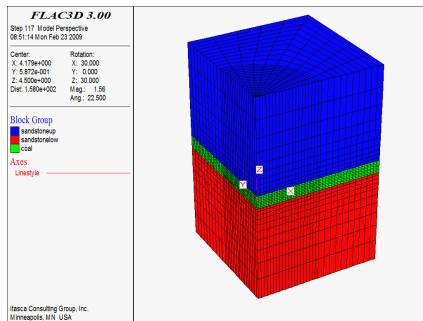
Fig. 1. FLAC3D model about spraying hole rock mechanism research

3   Simulation Computing Results

3.1   Stress Distribution Law around the Hole

In three directional uniform loads, the Mohr-Coulomb elastic-plastic model is applied with 15.0MPa(about 600 metres rock gravity stress) stress on the upper plane to solve the initial balance state. And then the middle cylinder with 2 metres diameter and 2 metres height, is trenched by transforming to null model, which simulates the spraying hole. There are five key monitoring zones near points with coordinate (2,0,0), (3,0,0), (4,0,0), (6,0,0), and (8,0,0). The distances from the points to the hole neutral line are 2.0 metres, 3.0 metres, 4.0 metres, 6 metres and 8 metres. In computing, the vertical component stress(Z direction) and the horizontal part(X direction) are tracked and recorded, as shown in Fig. 2 (left indicates the history near the hole point).

a- Horizontal l component stress   b- Vertical component stress
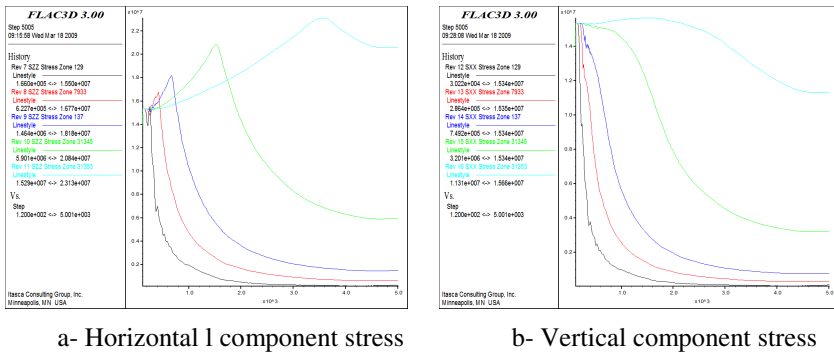
Fig. 2. History plot of stress on five points vs step

The Figures above showed the zones is subjected step by step to initial stress, concentration stress, peak stress, and residual stress. In the early spraying stage, the coal body zones are gradually loaded additional stress induced by drilling. The coal body realizes pressure relief as soon as its stress arrives at the plastic yielding point.The closer zones are relieved more pressure in the process. We can find clearly three districts called stress-relaxation zone, stress-increase zone, and stress-initial balance zone in Fig. 3.

(a) Horizontal l component stress   (b) Vertical component stress

Fig. 3. Stress distribution in the model
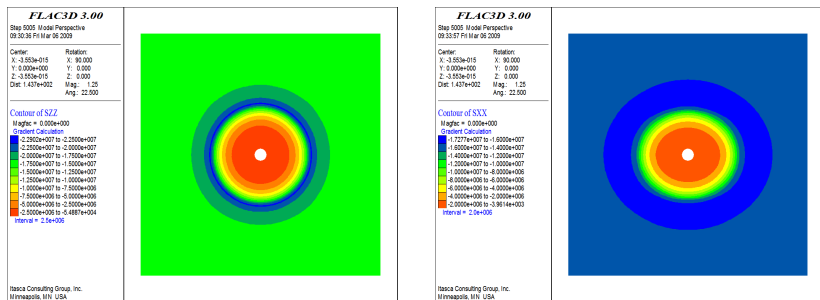
The stress nephogram indicates the vertical concentrating stress is bigger than the horizontal ones. The factor is a good indicator of pressure relief or concentration in equation:

$$k = \left(\sigma_{z0} - \sigma_z\right)/\sigma_{z0} \qquad (1)$$

Where σ_{z0} is the initial horizontal balance press and σ_z is the vertical stress after drilling. With the stress histories in the modeling, we plot the k distribution condition in Fig 4. The Fig shows the relief range confined to 7.8 meters to the hole central line. Moreover, the zones 8.5 meters to the center have the biggest vertical component stress 23.3 MPa with factor 1.52. In addition, the biggest horizontal ones is -17.2MPa with factor 1.12, which is lower than the former.
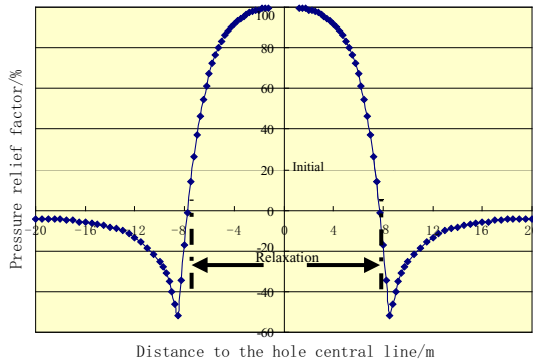
Fig. 4. Pressure relief factor *k* distribution around the hole

3.2   Displacement Distribution Law around the Hole

The key points' displacement are also monitored in the simulating. The history curves are characterized with bigger slope on two borders than the middle. The plot can be explained by three rock deforming stages as followed: Firstly, the displacement accumulates gradually with low speed. second, big concentrated stress applies the zones and surpasses their bearing capacity, which is followed by the coal body yielding, cracks forming, volume expansion. The process is associated with rapid increasing displacement. Finally, yielding zones suffered from peak stress are destoried and realize unloading, with the flat displacement curve.

In the static state, the displacement distribution is illustrated in Fig. 5. The figure indicates the radical displacement is damped speedy from 226 centimeters on the boundary to 60 centimeters on the 3 meters depth zones. The range from 3.0 to 7.8 meters depth is relaxation change district, the displacement from 60 centimeters to 1.5 centimeters. The number on the peak stress is 0.7 centimeter. The diaplacement on the outer zones tends to be zero gradually. Consequently, the effectual radical displacement range can be thought as crack area, from the boundary of the hole to the peak stress piont.
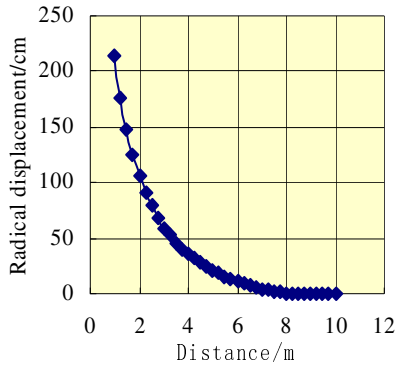
Fig. 5. Radical displacement around the spraying hole

4 Main Factors Influncing on Results

In according to the forementioned method in part 2 and part 3, we have simulated the problem with different hole radius, mining depth, and coal thickness. The simulating results are shown in Table 2, Fig. 6 and Fig. 7. Therefore, we get some valuable rules as follows:

(1)The influence range of the spraying hole is enlarged with the increasing drilling hole radius, and the stress-relaxation radius and crack radius has the same changing tendency. It is noticeable that the amplitude will become smaller and smaller.

(2)The influence range also changes incrementally with the mining depth. However, the crack radius is not sensitive to the depth. With deeper rock seams and bigger rock stress, there is greater extreme radical displacement around the hole, 800 meters depth to 350 centimeters and 500 meters depth to 152 centimeters.

(3)The coal thickness has some influence on the range of stress-relaxation district and crack disrict. The inducing spraying hole contributes immensely to improve the penetration in outburst coal.

Table 2. Comparison of stress distribution parameters in different hole radius

Hole radius /m	Stress-relaxation radius /m	Crack radius /m	Stress-relaxation mass-ratio /%	Biggest concentrated stress /MPa	stress concentration factor /%
0.5	6.1	6.8	0.68	-21.4	171
1.0	7.8	8.5	1.66	-22.9	166
1.5	8.7	9.8	3.03	-24.2	163
2.0	9.6	11.0	4.55	-24.6	162
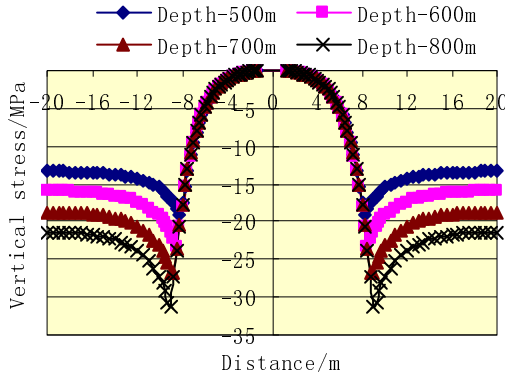
Fig. 6. Vertical stress distribution in different mine depth(Radius=1.0m)
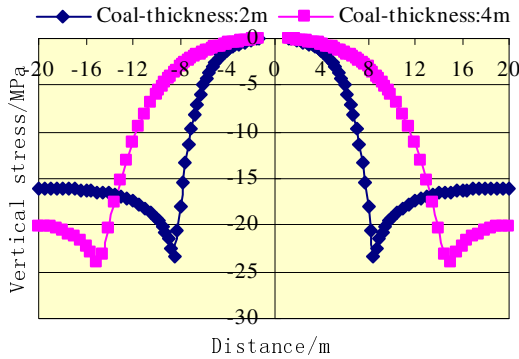
Fig. 7. Vertical stress distribution in different coal thickness(Radius=1.0m)

5   Summaries

(1) FLAC3D is an effective simulating software in research on outburst coal seam stress.

(2)The stress will resize to maintain a new mechanical balance in arround zones as the inducing hole preventing outburst is trenched. In this course, the near zones experience rheology, dilation, and cracking. Therefore it forms three basic districts. This is the mechnism of hole spraying which can conduct optimization design of hole preventing outburst.

(3)The influence range of hole spraying changes consistently with the hole radium, mining depth, and coal thickness. However, their effects are not same on present on-site condition.

References

[1] He, X.-Q.: Rheological Dynamics of Coal Rock Containing Methane. China University of Mining & Technology Press, Xuzhou (1995)

[2] Zhou, S.-N., Lin, B.-Q.: Theory of Gas Flow and Storage in Coal Seams. China Coal Industry Publishing House, Beijing (1992)

[3] Yang, Y., Cheng, Y.-P., Wang, L.: Research on Coal and Gas Outburst Prevention Technology During Mine Rock Roadway Heading with Short Distance to Seam. Coal Science and Technology, 31–34 (October 2009)

[4] Wu, X., Cheng, Y.-P., Hou, S.-J.: Coal and Gas Outburst Accident Key Chains Theory. Coal Technology, 86–89 (July 2010)

[5] Zhao, Y.-Q., Qi, L.-M.: Research on Avoiding Outburst by Gas Dialysis of Coal Seam. Journal of Mining & Safety Engineering, 110–113 (March 2009)

The Promotion of Model Research Given
by the Robot Visual Images Information Bunkers

Song Yang[1], Zhang Zhiyong[2], and Liu Yueqi[2]

[1] Jilin Teachers' Institute of Engineering & Technology
Changchun 130052, China
`manloveapple@163.com`
[2] Changchun Teachers College, The Computer Science and Technology Institute, Changchun,
Jilin Province 130032, China

Abstract. Today, with the aid of a model of the matrix, visual matrix estimates, visual elements and visual projective space transformation, we can adjust the robot's ability of recognize visual information for images. On the other hands, this visual for robots also improve and develop the visual images and information models.

Keywords: Promotion, robot matrix, visual images bunkers.

1   Introduction

After the robot visual image recognition and image preprocessing, you can identify the object and its orientation [1]. Therefore, after processing using the corresponding algorithm, the robot can make a difference on matters related to reflection. Solving the robot recognition processing problems, improving the level of awareness of the robot to reduce the identification error is about a very key issue. Visual robot on image pre-processing capability advantages and disadvantages requires a reasonable image quality evaluation algorithm rapidly as the support, but to build a robot model of the visual image information data base, can improve the robot ability to identify, solve the image Identification of Error.

2   The Basic Model of Visual Images

Robot Visual image acquisition to establish a basic model, can meet the complex and dynamic environment, when the new three-dimensional visual model of the planning requirements[2], the basic imaging model for robot visual, often referred to as the basic pinhole model, the three-dimensional space to the plane given central project transformation. Order spatial point O_c Is the project center to the plane it π's distance is f. Spatial point X_c In the plane π The project m The project O_c And through the points for the endpoint X_c The ray and plane π the point of

intersection, shown in Fig.1 flat. π The image plane as robot visual, point $\boldsymbol{O}_c$ as robot visual center f , called the focal length of robot visual in order to point $\boldsymbol{O}_c$, called the focal length of robot visual in order to point $\boldsymbol{p}$, As the main point of robot vision.
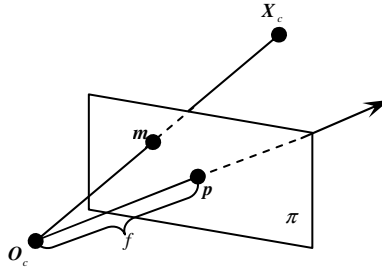
Fig. 1. The basic pinhole model

In order to describe the projection from the algebra, the need for the establishment of robot visual and image plane coordinate system coordinates. In the image plane to the main point $\boldsymbol{p}$, For the origin of the image plane coordinates $\boldsymbol{o}$ and lead to horizontal lines were x Axis and y Axis, the establishment of the image coordination system $\boldsymbol{o\text{-}xy}$ .In space, the visual center of the robot $\boldsymbol{O}_c$ Coordinate system for the robot visual coordination origin to spindle as z_c Axis, equal to x Axis and the center of the robot visual $\boldsymbol{O}_c$ the straight line x_c Axis parallel to y Axis and the center of the robot visual $\boldsymbol{O}_c$ Linear axis y_c The establishment of robot visual coordinate system $\boldsymbol{O}_c - x_c y_c z_c$ Shown in Fig.2.
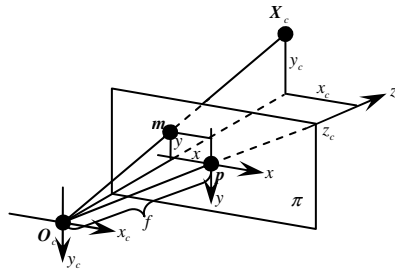
Fig. 2. Robot visual coordinates

Spatial point $\boldsymbol{X}_c$ Coordinate system in robot visual sitting marked the Euclidean $\tilde{\boldsymbol{X}}_c = (x_c, y_c, z_c)^T$, It' s like a point $\boldsymbol{m}$ The image coordinate system in the seat labeled $\tilde{\boldsymbol{m}} = (x, y)^T$. According to the triangle similar to the principle of Spatial point can be inferred $\boldsymbol{X}_c$ And its image point $\boldsymbol{m}$ Satisfy the following relations

$$\begin{cases} x = \dfrac{fx_c}{z_c} \\ y = \dfrac{fy_c}{z_c} \end{cases}$$

The equation can be expressed as the following matrix form

$$z_c \boldsymbol{m} = \begin{pmatrix} fx_c \\ fy_c \\ z_c \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \boldsymbol{X}_c$$

Among them, $\boldsymbol{X}_c = (x_c, y_c, z_c, 1)^T$, $\boldsymbol{m} = (x, y, 1)^T$, Space point and were homogeneous coordinates of image points. It is like a plane from space to a linear transformation. If you remember

$$P = diag(f, f, 1)(I, 0)$$

Then this linear transformation can be expressed as a more concise form

$$\boldsymbol{m} = P\boldsymbol{X}_c$$

Among them, the matrix is a matrix, usually called matrix of robot visual. This is the basic imaging model of algebra.

3   Visual Elements of the Matrix

Established, the robot visual servo system can achieve fast and accurate [3].

(1) To consider the visual center of the visual centers in the world coordinate system of coordinates. From the following formula

$$PC = KR(I, -\tilde{C}) \begin{bmatrix} \tilde{C} \\ 1 \end{bmatrix} = KR(\tilde{C}, -\tilde{C}) = 0$$

Known, the visual centers $C = (\tilde{C}^T, 1)^T$ is the equation $PC = 0$ A solution. On the other hands, the right only one-dimensional null space, because its rank is equal to 3. Thus, the visual centers constitute the homogeneous coordinates P right null space. Visual centers in the known matrix P circumstances can solve the equation PX=0 by visual 1 centers in the world coordinate system coordinates. In fact, if so $P = (H, p_4)$, among H this P's three columns posed 3X3 matrix p_4 is P The fourth column vector, from equation PX=0. Available visual centers in the world coordinate system of homogeneous coordinates

$$C = \begin{bmatrix} -H^{-1}p_4 \\ 1 \end{bmatrix}$$

2) the direction of the axis coordinate origin

Matrix Visual Center $P = (p_1, p_2, p_3, p_4)$ is among PJ to P's J Column vector.

The world coordinate system origin coordinates $X = (0,0,0,1)^T$, so image coordinates

$$s_0 m_0 = PX = (p_1, p_2, p_3, p_4)\begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = p_4$$

That is, the visual center of the first four vector matrix is the world coordinates origin homogeneous image coordinates. Considering the three coordinates of the world coordinate system, the direction of the image is three axis intersections with the image plane at infinity. Clearly, 3 axes, respectively with the plane at infinity $X = (1,0,0,0)^T, Y = (0,1,0,0)^T, Z = (0,0,1,0)^T$

Therefore, their image coordinates

were $s_1 m_1 = PX = p_1, s_2 m_2 = PY = p_2, s_3 m_3 = PZ = p_3$

Therefore, the visual center of the first three column matrix vector of the world coordinate system the direction of 3 axis image point of homogeneous coordinates

(3) The main plane and axial plane

Visual center matrix $P = \begin{bmatrix} p^{1T} \\ p^{2T} \\ p^{3T} \end{bmatrix}$,

among p^{1T}, p^{2T}, p^{3T} were P The 3 row vector.

① principal plane visual center coordinate plane $O_c - x_c y_c$, if the coordinates of the image plane parallel to the plane, it often referred as the main plane. Principal plane in the world coordinate system is available in the visual center line 3 vector matrixes to represent. Because the image plane parallel to the main plane, so their intersection line is a straight line at infinity, which is like the main plane of infinity as the plane straight. Order p^{3T} X The main any point on the plane, then its image will be under the visual center for the

$$\begin{bmatrix} u \\ v \\ 0 \end{bmatrix} = PX = \begin{bmatrix} p^{1T} X \\ p^{2T} X \\ p^{3T} X \end{bmatrix}$$

The main any point on the plane, then its image will be under the visual center for the so $p^{3T}X = 0$. Thus, the main plane in the world coordinate system of coordinates for the visual center line 3 matrix p^{3T}.

② axis plane

By the equation $p^{1T}X = 0$ Plane identified, that coordinates the world coordinate system p^{1T} Plane. It is the image plane v Axis and the plane determined by the visual center, it often referred to as axial plane. Order X For this axis at any point on the plane, then its image will be under the visual center for the

$$\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = PX = \begin{bmatrix} p^{1T}X \\ p^{2T}X \\ p^{3T}X \end{bmatrix} = \begin{bmatrix} 0 \\ p^{2T}X \\ p^{3T}X \end{bmatrix}$$

Thus, the axis of the image plane $\{(0, v, 1)^T : v \in R\}$, That is like a plane v.

By the $p^{2T}X = 0$ Identified another axis plane is the image plane u Axis and forms the visual center of the plane. Axis plane and the main plane of the difference is that it depends on the image coordinate system of choice, that is, different images coordinate system; the corresponding axis plane is different.

(4) principal axis and the main points① principal axis

Principal axis and the principal plane are orthogonal, so will the main plane of the normal spindle. Flat $\pi = (\pi_1, \pi_2, \pi_3, \pi_4)^T$, there are two directions the normal They are $\hat{\pi} = \pm(\pi_1, \pi_2, \pi_3)^T$, Planes at infinity they represent the same point, namely, the normal point of intersection with the plane at infinity. As the principal plane is $p^{3T} = (p_{31}, p_{32}, p_{33}, p_{34})^T$, therefore, the two axis direction $\hat{p}^3 = \pm(p_{31}, p_{32}, p_{33})^T$. The principal axis is usually talking about the direction of its positive direction, that point to the direction of the visual center of the front, because the visual center of a constant factor matrix can vary, so $\hat{p}^3 = \pm(p_{31}, p_{32}, p_{33})^T$ does not represent a positive number in the positive axis. If the visual center matrix $P = (H, p_4)$, matrix with the standard visual center $K(R, t)$ Difference between a normal number, there must be $\det(H) > 0$, otherwise $\det(H) < 0$. Therefore, the spindle is positive

$$v = \det(H)h^3$$

Among h^{3T} is the matrix of the first 3 Row vector.

② Main points

The main point is the main axis and the intersection of the image plane, due to the excessive visual center spindle, so the main point will be as the main direction of the image point, so the main point of coordinates $p = P(h^{3T}, 0)^T = Hh^3$

4   Visual Estimate of Matrix

Under normal circumstances, the robot can use the code, and inertial positioning plate method, but in the classical three-dimensional visual, the need to estimate the Euclidean coordinate system under visual matrices, be able to complete the stereo visual system calibration [4]. Chang Yong method is based on some spatial point coordinate system in Europe under the coordinate Yu Qing correspondence between image coordinates, Juan Li visual Zhongxinjuzhen the Yueshufangcheng, Congerqueding visual Zhongxin matrix. In practice, in order to get the Euclidean coordinates, need to make a calibration reference material, reference material on the calibration feature points through the accurate measurement of visual as the estimated center of the space required when matrix point[5].

Visual center matrix:

$$P = \begin{bmatrix} p^{1T} \\ p^{2T} \\ p^{3T} \end{bmatrix}$$

Among p^{jT} the first Row vector. Order $X_j = (x_j, y_j, z_j, 1)^T$ Is a feature point in world coordinate system of coordinates, the corresponding image point coordinates $m_j = (u_j, v_j, 1)^T$, So according to the projection of the visual center, we get

$$s_j m_j = PX_j = \begin{bmatrix} p^{1T} X_j \\ p^{2T} X_j \\ p^{3T} X_j \end{bmatrix}$$

Therefore, the elimination of the constant factor on the type, the get the following

equation $\begin{cases} p^{2T} X_j - v_j p^{3T} X_j = 0 \\ p^{1T} X_j - u_j p^{3T} X_j = 0 \\ v_j p^{1T} X_j - u_j p^{2T} X_j = 0 \end{cases}$

In this equation, the third equation by the linear representation of the first two equations, only two equations are linearly independent. Given $N \geq 6$ Feature points more than their corresponding image points can be solved linearly, the visual center of matrix P.

When the image data is measured error, the above equation is generally no nonzero solution. At this point, usually it's least square solution as the visual center of the estimated matrix. Corresponding to each point, it signs

$$
A_j = \begin{bmatrix} 0 & X_j^T & -v_j X_j^T \\ X_j^T & 0 & -u_j X_j^T \\ v_j X_j^T & -u_j X_j^T & 0 \end{bmatrix}
$$

It is a 3×12 Matrix A_j. Given N correspond, get N a matrix form as such, Then this N Matrices together to get a $3n \times 12$ Matrix $A = (A_1^T, \cdots, A_N^T)^T$, to A as Singular value decomposition(SVD), A=UDV^T so V The last column vector $p = v_{12}$ Is the equation $Ap = 0$ the Least squares solution, Then P Written in matrix form to get the visual center of matrix.

5   Euclidean Space and the Visual Center of the Space Transformation

In real life, the huge video data and the system processing power, and complex visual information redundant line data [6]. Visual center coordinates and projective coordinate transformation between the still can be 4X4 Homogeneous invertible H Express, so $X' = HX$, and from three-dimensional projective space to the image plane of the map can still use a rank to 3 order 3X4 Matrix P express[7]. For the most general case, the visual center model can be seen from the three-dimensional projective space to projective plane two-dimensional map, and this map can use the following matrix synthesis approach to expression

$$
P = K \begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix} H
$$

① if $H = \begin{pmatrix} A & b \\ 0^T & 1 \end{pmatrix}$, among $rank(A) = 3$, it is said the three-dimensional affine

transformation, the style becomes

It is the world coordinate system for the affine coordinate system of the visual center of matrix, called the visual center of affine space matrix [7]. ② if

$H = \begin{pmatrix} sR & t \\ 0^T & 1 \end{pmatrix}$, among R is Rotation matrix S   to Non-zero constant, it is said that

three-dimensional similarity transformation, into

$$
P = K \begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix} \begin{pmatrix} sR & t \\ 0^T & 1 \end{pmatrix}
$$

It is the world coordinate system for the Euclidean coordinate system, but the measurement unit S Matrix of the visual center, call center similar to the visual space in the matrix

③ if $H = \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix}$, among R is Rotation matrix that represents the

three-dimensional Euclidean space transform, the type change to

$$P = K \begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix} \begin{pmatrix} R & t \\ 0^T & 1 \end{pmatrix}$$

It is the world coordinate system for the Euclidean coordinate system, but the measure is an absolute measure of the visual center matrix, Euclidean space called the visual center in the matrix.④if $H = (I, 0)$,

Transform into

$$P = K \begin{bmatrix} 1000 \\ 0100 \\ 0010 \end{bmatrix}$$

It is based on the visual center of coordinate system for the visual center of the world coordinate system matrix.

References

[1] Cong, K., Han, J., Chang, F.L.: Visual robot contour extraction and location of goods. Shandong University (Engineering Science) (2010)
[2] Fu, W., Gu, X., Wang, Y.: Based on human visual characteristics of color image quality assessment. Microelectronics and Computer (2010)
[3] Wei, T., Li, H.: The new stereo visual model based on velocity vector obstacle avoidance planning. Computer (2009)
[4] Jiang, X., Li, Z.: Based DSP Robot Visual Servo System. Computer Information (2007)
[5] Chen, W., Wang, X., Wu, T., Xu, X.: Total angular correction of visual localization algorithm and error analysis. System Simulation (2007)
[6] Zheng, Y., Liu, D.: Foveated visual based monitoring image processing method of production. Tsinghua University (Natural Science) (2005)
[7] Hu, Z.Y., Wu, F.C., Wang, G.H.: Impossibility of affine reconstruction by a translation camera from two perspective images. Pattern Recognition Letters 24(16) (2003)

Study on Regularity of Stress Change for Quarry Wall Rock in Mining of Glacis Thin Orebody with the Long Wall Mining Based on FLAC

Zhao Shuguo[1,2], Song Weidong[1], and Xu Wenbin[1]

[1] Civil & Environment Engineering School, University of Science & Technology Beijing,
Beijing 100083 China
[2] Hebei United University, Tangshan 063009 China

Abstract. Targeted at the orebody section from Guandian Mining to Dazhuang Mining between line -8' and -16', the study focuses on the mechanism of wall rock deformation and failure and the deformation of ground surface led by the mining of stratified ore bodies with the long wall mining. FLAC software is adopted to analyze and study the simulation of the stress change of quarry wall rock without support on the working surface. Conclusion is drawn on the regularity for stress change of quarry wall rock to provide scientific and rational guide for the mining engineering practice.

Keywords: Long wall mining, FLAC, Quarry wall rock, Regularity of stress change.

1   Introduction

The software, FLAC (Fast Lagrangian Analysis of Continua) is mainly applied in simulation of geo-material calculation and mechanical behavior in geotechnical engineering, especially in the plastic flow when the material has reached its yield limit. The material is expressed by unit and area,.restricted by load and boundaries, each unit gives mechanical response based on the linear or nonlinear stress strain relation. It includes the special calculation function that reflects the geo-material mechanical effect; it can calculate high nonlinearity of the geo-material,irreversible shearing failure and Consolidation, viscoelastic (creep), and mechnical behaviors. Targeted at the ore body of Guandian Ore Mining from Liangshuijing to Dazhuang Mining in 8'to -16'section, the paper focuses on the mechanism of wall rock deformation and failure and the deformation of ground surface led by the mining of stratified ore bodies with the long wall mining. FLAC software is adopted to analyze and study the simulation of the stress change of quarry wall rock without support on the working surface. Conclusion is drawn on the regularity for stress change of quarry wall rock to provide scientific and rational guide for production.

2   The Geological Features in Model Mining Area and Physical Mechanics Parameters

2.1   The Geological Features in Model Mining Area

Guandian Iron Mine of Jianshi County belongs to the Devonian "Ningxiang" pelagic sediment hematite ore deposit. The main industrial deposit is the third ledge (Fe_3) in the middle with the layered and lensoid ore body.The ore body Guandian Mining from Liangshuijing to Dazhuang Mining -8' to -16' section is about 890m; the ore body beats at the F_1 fault and stops at hole CK12 on line -8' to hole CK168 on line -12'; the thickness of the ore body is stable, with the maximal thickness of 4.74m （TC9）, the minimal thickness of 2.10m(CK168)，and the average of it is 3.74m. The trend of it is east--northeast to west--southwest；the dip direction is north--northwest, the inclination 8°~20°, with the average of it, 10°~11°. The ledge is in stratiform--bedded production. The intercalation larger than 0.3 is never found there. The mining design elevation of the ore body is +1430~+1380m; the vertical height is 50m; the largest distance from the deepest place to the surface of the ground is 250m. The research approach is on the study the simulation of the stress change of quarry wall rock without support on the working surface and the deformation law at the excavation of 8m intervals.

2.2   Physical Mechanics Parameters for the Ore Body Wall Rock

The ore body wall rocks are mainly the shale and limestone with weak tensile strength. The physical mechanics parameters for the main terrane are shown in tab.1.

Table 1. Rock mechanics calculation parameters

Materials	Deformation modulus（Gpa）	volume weight （kN·m^{-3}）	Poisson ratio（μ）	Lap Shear Strength	
				φ	c/MPa
The surface layer calcareous rock	18~23	26	0.24	50.20	1.50
Limestone in Maokou Group	43.96	27.10	0.23	44.10	6.72
Marlite in Huanglong Group	8.00	25.00	0.24	36.00	6.00
Quartz arenite in Xiejingsi Group	44.00	25.87	0.11	42.05	10.10
Sand shale in Huangjiadeng Group	10.00	25.00	0.30	30.00	14.00
Quartzite in Yuntaiguan Group	68.00	30.00	0.17	50.00	19.60
Iron ledge	14.96	37.20	0.21	33.60	8.32

3   Modelling and the Analysis of the Results

3.1   Build Numerical Calculation Model

The mining area model is made (as shown in Fig.1) on the procedures of numerical calculation software after meshing, setting model criteria and defining the boundary conditions and initial conditions. The excavation is carried down the trend, at the same time, monitoring points are set up on the mining area wall rocks to monitor the stress and deformation change. The changes of the wall rocks can be obtained through the record of stress and deformation change and area of plastic zone of the wall rocks of their special monitoring points after the step excavation.
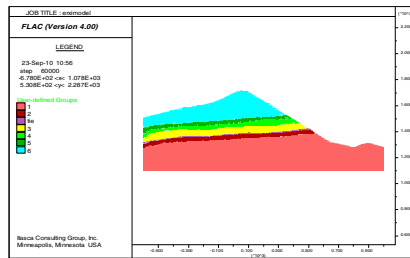
Fig. 1. Numerical simulation diagram

3.2   Analysis of the Model Calculation Results

3.2.1   Analysis of the Regularity of Stress Change

As can be shown in the stress distribution of the steps(shown in Fig.2), stress of the wall rocks changes with the steps and times of the excavation. The bigger the goaf, the more significant is the concentration of the stress from the front of the ore body to the back of it.

As shown in the stress distribution of the wall rocks when the excavation is 8m deep(shown in Fig.2(a)), the front and back of the working face have the stress concentration, with the maximum stress value of 6Mpa. The top of the goaf has the minimum stress, as can be seen in the arched stress cloud chart, therefore, it relieves the stress. At 16m excavation section, both the front and back stresses increase with the maximum at 7Mpa; the front of the working face is the wavy peak value zone. The broken rock zone of the top wall rocks increases, too(shown in Fig.2(b)). The above tendancy is the same at the excavation section of 32m, 48m and 96m(shown in Fig.2(c)、Fig.2(d)). The only difference is in the peak value at each step, with the peak value of 12.5MPa at 32m excavation and the maximum of 20Mpa when all the ore body is excavated. The concentration of front stress is more significant than that of the back.
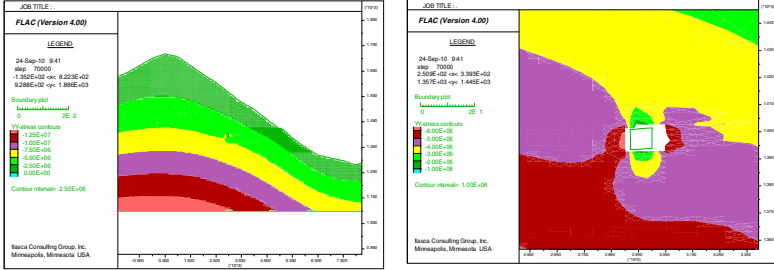
(a)   Stress cloud chart of Y Position wall rocks at 8m excavation section
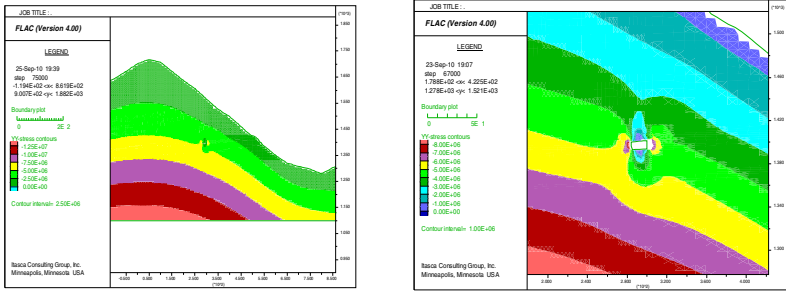
(b)   Stress cloud chart of Y Position wall rocks at 16m excavation section
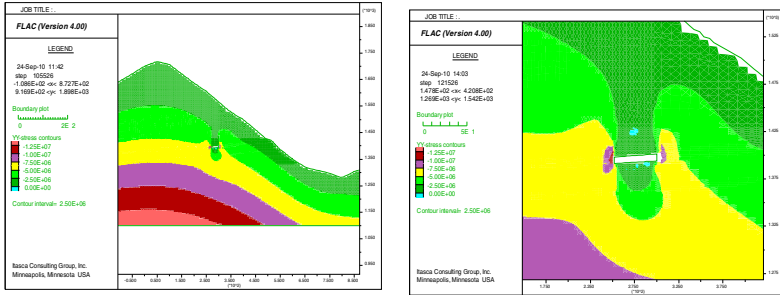
(c)   Stress cloud chart of Y Position wall rocks at 48 m excavation section
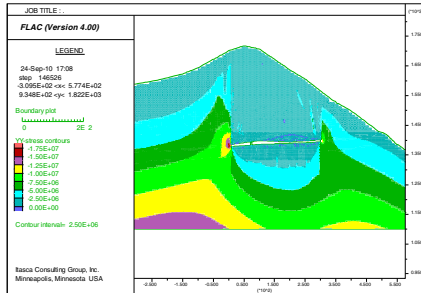
(d)   Stress change of wall rocks when the ore body is excavated

Fig. 2. Stress change of ore body wall rocks at excavations of different steps

On the whole, in excavation, the concentration of stress the result of stress transfer and the reflection of self-stability of the wall rocks so that the stress state will be different accordingly; the difference of the peak value at different steps is mainly influenced by the increase of the area of goaf. The deeper the excavation is, the more weight for the rocks on the front working face to bear the cover rocks and the more factors will be involved to influence the stress concentration.

3.2.2   Analysis of the Regularity of Plastic Zone Change

The plastic zones led by excavation are shown in the respective figures. At 8m excavation section(shown in Fig.3(a)), the wall rocks can manage to be self-stable with elastic small plastic zone area. At 16m excavation section, the tensile zone of the wall rocks becomes larger. From 16m to 48m excavation section(shown in Fig.3(b)), the front working face is mainly in elastic state with small shearing area; as the tensile zone on the top of goaf becomes larger, the tensile zone expands upward at an angle in trapezium; the stress disturbed zone has reached +1460m, with 60 m vertical height from the ore body. The plastic zone shape is similar to the modelling with similar materials in the experiment.

At 96m excavation section, the rocks on the working face top are mainly influenced by shearing effect; the tensile area becomes larger accordingly and the plastic zone is so large that it reaches the surface of the ground..At 144m, 192 m excavation section(shown in Fig.3(c))and when the ore body is excavated out(shown in Fig.3(d)), the rocks above produce the more significant shearing area in similar sile slope direction, which shows that the top rocks slide to the goaf along the side slope after the tensile area reaches the surface of the ground.

The larger the excavation area, accordingly the larger the plastic zone produced. The stress state changes with different position of the excavation. Influenced by the side slope of the ground surface, the rocks of the top working face is influenced by the tensile stress at the beginning and then at more than 96m excavation section, influenced by shear yield; in vertical direction, stress plasticity of the rocks above +1460 is different from the lower part but the shearing area direction is similar to that of the side slope. But in the section below 1460m, the shape of the plastic zone is the upward expansion of the trapezoid; the yield range of the rocks is close to the working face direction.
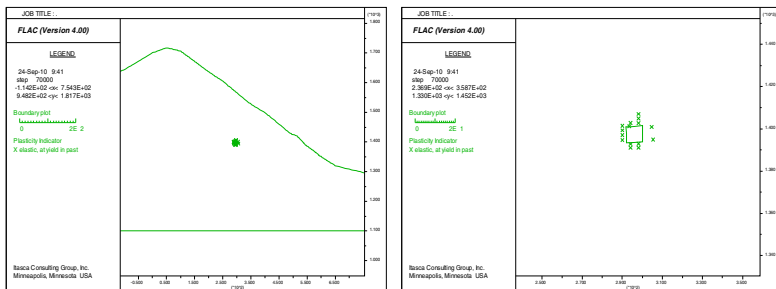
(a)   The wall rock changes of the plastic zone at 8m excavation section

Fig. 3. The wall rock changes of the plastic zone at excavations of different steps
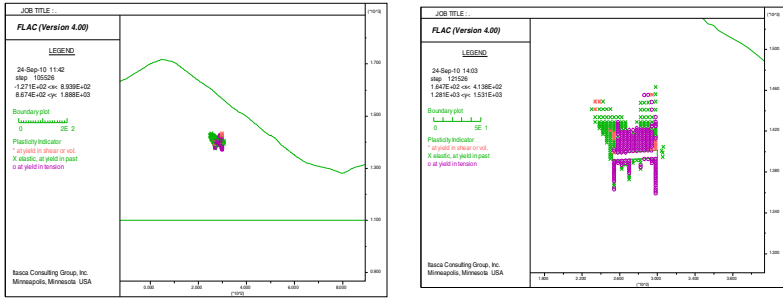
(b) The wall rock changes of the plastic zone at 48m excavation section
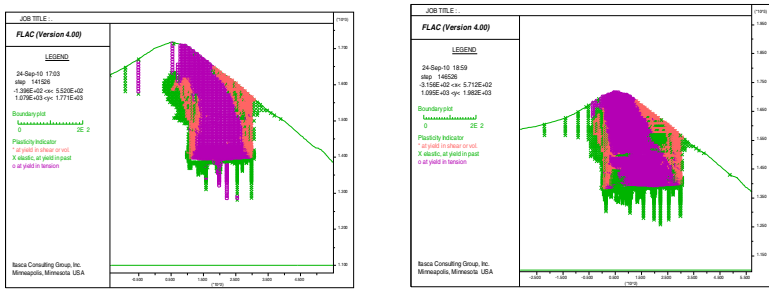
(c) The wall rock changes of the plastic section

(d) The wall rock changes of the plastic zone at 192m excavation zone when the ore body is excavated out

Fig. 3. (*continued*)

4   Summaries

Through analysis of the changes of stress at different steps and the change of the plastic zone, the following conclusion can be drawn:

A. At the initial stage, the rocks can manage to be self-stable(at about 32m excavation) ;the peak value of the working face is ever increasing, one is the result of the stress concentration transfer and the other is the result of the ore body sloping to the direction of dip;

B. The plastic zone is ever increasing, with the top working face under significant shearing effect and the top goaf under tensile effect;

C. The wall rocks mainly suffer from the tensile fracture; when the rocks are close to the ground surface (above +1460m) the yielding state is different from the lower part, because the stress is not even due to the terrain of the slopes.

D. The yielding area expands to the wall rocks in trapezoid.

References

[1] Wu, H., Hu, X., Bao, T.: Analysis of FLAC software in calculating the face stability of surrounding rock. Ground Pressure and Strate Control (4), 96–97 (2002)

[2] Sun, G., Li, J., Hu, X.: FLAC—3D-Based Stability Analysis of Mined-out Area. Metal Miine (2), 29–30 (2007)

[3] Xiao, H., He, X., Feng, T., Wang, E., Zheng, B.: Application of FLAC in the research on the coupling laws between EME and stress fields during the deformation and fracture. Journal of China Coal Society 29(6), 649–695 (2004)

[4] He, X., Wang, E., Liu, Z.: Experimental study on the electromagnetic radiation (EMR) during the fracture of coal or rock. In: Proceedings of the 1999 International Symposium on Mining Science and Technology, pp. 133–136. Science Press, Beijing (1999)

Performance Analysis of Solid Towed Array

Zhang Xiang and Li Shu-Qiu

The Lab of Oceanic Information Technology, Institute of Acoustics,
Chinese Academy of Sciences, Beijing 1000190, China

Abstract. Towed line array sonar has attracted attention for its property of low frequency and big aperture. As an important part of the wet-end, the towed array has developed rapidly. The liquid towed array is more mature for its being used earlier. Meanwhile, more studies on solid towed array have been taken for its special feature. To analyze the performances on the phase consistency of hydrophones and the beam forming performance of the solid towed array, the lake trail results are presented. The analysis results show that the solid towed array has good performance on phase consistency among different hydrophones and on beam forming. Especially under the condition of being towed, the solid towed array has better performance for it is less sensitive to the vibration noise.

Keywords: Solid towed array, Phase consistency, Beam forming performance.

1   Introduction

In the field of long distance detection and marine oil exploration, towed line array sonar is widely used as its characteristics of low frequency and large acoustic aperture[1][2][3]. As the major part of the towed line array sonar, towed array is mainly responsible for the acquisition, amplification and quantification of the acoustic signal, and the reliable transformation of the processed digital signal to the dry end. So the performance of towed array will directly determine the stability and reliability of the whole sonar system. Currently, there are two kinds of widely used hydrophones in towed array: piezoelectric hydrophone and fiber-optic hydrophone. The piezoelectric hydrophone is more mature, so the towed array composed by piezoelectric hydrophone has been widely used. In contrast, the fiber-optic hydrophone appeared later. Although much researches have been done on fiber-optic hydrophone at home and abroad and made some progress[4][5], the towed array composed by fiber-optic hydrophone is still far from perfect. As a result, the solid towed array we talk about in this article uses the traditional piezoelectric hydrophones. According to the different materials filled in the towed array, it can be divided into two categories: the liquid towed array and the solid towed array. Liquid towed array is applied earlier, so the technology of array formation is relatively mature; meanwhile, solid towed array has its own advantages, so researches on it have gradually increased in recent years. According to this situation, this article verifies the solid towed array by analyzing the lake trial data, and shows the performances of the solid towed array in detail.

2   Introduction of Piezoelectric Hydrophone and the Structure of Towed Array

2.1   Introduction of Piezoelectric Hydrophone

The piezoelectric hydrophone is usually made from piezoelectric ceramics. When pressure or tension is done to the cream, it will have the opposite-polarity charge at both ends, which will form the current through the loop. This effect is called the piezoelectric effect. If we put this kind of transducer in water, it will induce charge under the effect of sound waves, which constitutes a piezoelectric hydrophone.

In this lake trail, the hydrophones used in the solid towed array are of basically the same parameters: the sensitivity is -201 ± 1.5 dB V/1μPa @ 20 ℃ (8.9V/bar); the bandwidth is 1Hz ~ 5 kHz. The frequency band used in this trail is lower than 2 KHz, and the size of hydrophone is 38mm in length and 16.5mm in diameter.

2.2   Introduction of the Towed Array's Structure

As the primary functional unit of the towed line array sonar, the solid towed array mentioned in this article refers to digital array. The main functional modules in it are highly universal: the modules can be configured as needed flexibly without redesigning the circuit. In addition, the acoustic array meets the requirements of the slender array, and the diameter after the array formation is 39mm. The towed array used in this trail contains 16 piezoelectric hydrophones, and of course, we have fully considered the extensibility when designing. We can set the number of the hydrophone flexibly according to the actual needs. Since the output of the piezoelectric hydrophone is analog signal with small amplitude, amplification, sampling and quantification of the analog signal are required before transmission by data transmission system to the dry end to be processed [6]. Therefore, each towed array contains preamplifier modules, data conversion modules and a data transmission module. The connection of the modules mentioned above in towed array is shown in Figure 1(the distance between two adjacent hydrophones within towed array is 1m):
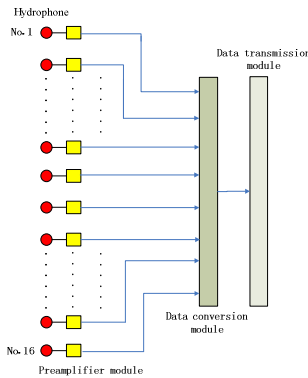
Fig. 1. Connections among modules in towed array

In order to meet the requirement that the towed array should accomplish zero buoyancy in the water, specific substance must be filled in it when the array is formatted. The difference between solid towed array and liquid towed array is the internal filling. The solid gel is used in solid towed array in the trail. Its main component is silicone gel and its density is about 0.94 g/cm3.

3   Performance Analysis of Solid Towed Array

For the modules in towed array (such as hydrophones, preamplifier modules, etc.) are surrounded by the filler, the nature of filler may have an impact on the working characteristics of modules, such as phase consistency of acoustic signals received by different hydrophones in the array and the beam forming properties. For the possible impact, we analyze the lake trial data and get the following results:

3.1   Analysis of the Phase Consistency

The phase consistency of array unit in line array is an important aspect of its performance, which will directly affect the beam bearing of the line array and the accuracy of target location results. The sound source was placed abeam of the solid towed array in the distance and transmitted single-frequency signals. So the phase of sound signals received by the adjacent array unit in towed array should be basically the same. In order to analyze the internal filling's influence on the phase of sound signals in solid towed array, we calculated the correlation coefficient of signals received by the adjacent hydrophones, and regard it as a measure of signal's phase consistency.

When testing phase consistency, the solid towed array kept static about 15 meters underwater. The sound source was placed about 1 km away from the towed array and about 10 meters underwater. It transmitted single-frequency signals of 400Hz and the sound source level was about 157 dB re 1μPa. Figure 2 shows the correlation coefficient fluctuation curve of the signals received by different pairs of hydrophones which are adjacent in solid towed array. The values of correlation coefficient vary from 0.6422 to 0.7428. In the x-axis, the coordinate 5, for example, denotes a pair of hydrophones with No.5 and No. 6 which are adjacent in the array.
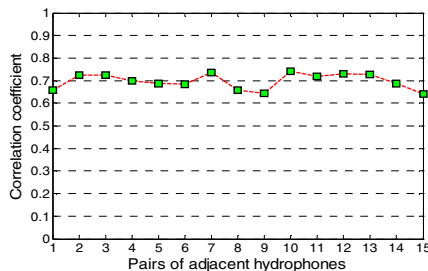
Fig. 2. Correlation coefficient of the signals received by different pairs of hydrophones

From the average correlation coefficient calculated from the acoustic signals, it can be seen that acoustic signals received by adjacent hydrophones in the solid towed array has high correlation and maintain a good degree of similarity.

3.2   Analysis of Beam Forming Performance

The beam forming performance is an important parameter of the line array, which directly reflects the accuracy of the target location of the line array. We analyze the beam forming performance of the solid towed array in static state and in towing state respectively. We also analyze the power spectral density of acoustic signal under above two test conditions. As the trail environment plays an important role in the signal quality, we use narrow-band signal processing method to analyze beam forming performance for better results [7].

3.2.1   Analysis of Beam Forming Performance in Static State

When testing the solid towed array in static state, it kept static about 15 meters underwater. The sound source is carried by a low-speed tugboat and kept 10 meters underwater. We did the test under tow different conditions: high SNR (signal to noise ratio) condition and low SNR condition.

3.2.1.1   Analysis of Trial Results under High SNR Condition

In the testing process, the tugboat moved at a low speed. When the tugboat moved to a location which was about 0.42 km away from the towed array(at that time, the azimuth was about 20° between the array and the sound source), the sound source began to transmit single-frequency signal of 700Hz and the source level was approximately 167 dB re 1μPa.
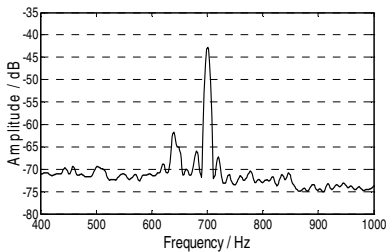
Fig. 3. The average power spectral density of 16 channel signals with high SNR in static state
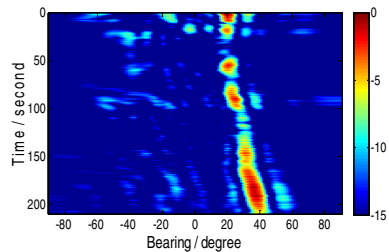
Fig. 4. Time-bearing tracks of 700Hz signal with high SNR in static state

Figure 3 is the average power spectral density of signals received by 16 hydrophones of solid towed array. From it we can see that signal is obvious at 700Hz point, and the ambient noise spectrum level is about-72.5dB at the signal receiving point.

Figure 4 illustrates the time-bearing tracks got by processing the signal received by solid towed array under the high SNR condition.

It can be seen clearly from the figure 4 that the sound source carried by the tugboat moved from 20° to 40° gradually. In static state with high SNR, the solid towed array's beam forming results are clear and it has good beam forming performance.

3.2.1.2   Analysis of Trial Results under Low SNR Condition

Similarly the sound source was carried by the tugboat and began transmitting single-frequency signal of 400Hz at about 0.96km away from the towed array with the azimuth about 60°, and the sound source level was about 157 dB re 1μPa. Unluckily, the trail environment at that time was poor, so the SNR was low. Figure 5 is the average power spectral density of signal received by 16 hydrophones of solid towed array, from which we can see that the amplitude of the 400Hz signal is small under low SNR condition, and the ambient noise spectrum level is about-62.2dB at the signal receiving point.
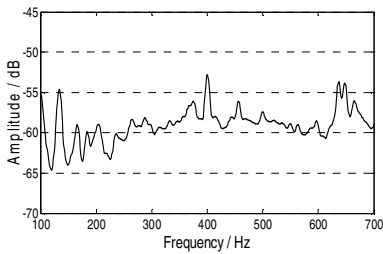
Fig. 5. The average power spectral density of 16 channel signals with low SNR in static state
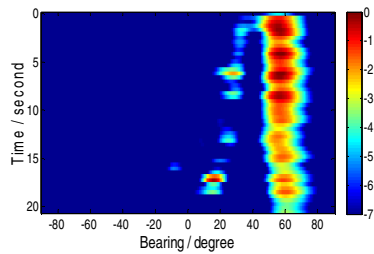
Fig. 6. Time-bearing tracks of 400Hz signal with low SNR in static state

Figure 6 illustrates the time-bearing tracks got by processing the signal under the low SNR condition.

In this trial process, the tugboat was far away from the towed array and its moving speed was low, so its orientation angle relative to the towed array remained almost unchanged in a short time. Figure 6 also reflects this fact. We can see that the solid towed array can also locate the target under low SNR condition.

3.2.2   Analysis of Beam Forming Performance in Towing State

When testing the solid towed array in towing state, it was towed by the tugboat. In order to maintain the towed array in a certain depth (setting to 12 meters underwater in this trail) during the towing procedure, the tugboat maintained its velocity at about 3 Navigation. The sound source in this test kept static about 10 meters underwater. The sound source began to transmit 800Hz signal when it was about 0.32km away from the towed array, and the sound source level was about 164 dB re 1μPa.

Figure 7 is the average power spectral density of 800Hz single-frequency signal received by 16 hydrophones in solid towed array, and the ambient noise spectrum level was about -67.6 dB at the signal receiving point.

Figure 8 illustrates the time-bearing tracks got by processing the signal received by the solid towed array.
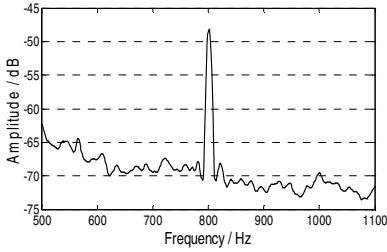
Fig. 7. The average power spectral density of 16 channel signals in towing state
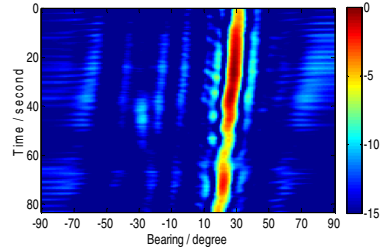
Fig. 8. Time-bearing tracks of 800Hz signal in towing state

It can be seen from Figure 8 that solid towed array can accurately locate the target in towing state, and the location result has narrower main lobe width and higher resolution ratio. Therefore, the solid towed array has better performance on beam forming in towing state. This is closely relative to that the solid towed array has low sensitivity to the noise generated by vibration when towing.

4   Summaries

This article introduces the solid towed array applied to towed line array sonar, and studies the performances of it by analyzing the data getting from the lake trail. From the results listed in this article, we can see that the acoustic signals received by adjacent hydrophones have high correlation and maintain a good degree of similarity. Meanwhile, the solid towed array has good performances on beam forming under both high and low SNR conditions, and the performance is better in towing state. All these advantages decide that the solid towed array has a very broad application prospect.

References

[1] Lemon, S.G.: Towed-Array History, 1917-2003. IEEE Journal of Oceanice Engineering 29(2), 365–373 (2004)
[2] Liu, M.-A.: Towed array sonar system development. Acoustics and Electronic Engineering (3), 1–5 (2006)
[3] Jiang, H.-Y., Zhao, W.-Z., Yan, C.-Z., et al.: Overview of the world's offshore oil and gas resources and the exploration models. Marine Petroleum Geology 13(3) (2008)
[4] Luo, H., Xiong, S.-D., Hu, Y.-M., et al.: Research on the fiber optic hydrophone applied in towed array. Applied Acoustics. Vol 25(2) (March 2006)
[5] Hill, D.J., Nash, P.J.: In-water acoustic response of a coated DFB fiber laser sensor. DERA WE2-9, 33–36 (2000)
[6] Yu, H.-B., Sun, C.-Y., Li, Q.-H.: Vanguard of submarine detection - towed array sonar. Physics 35(5), 420–423 (2006)
[7] Li, Q.-H.: Introduction to Sonar Signal Processing. Version 2. Ocean Press, Beijing (2000)

Based on Wavelet Transformation Mold Maximum Value Edge Examination Research

Huang Tiankai and Huang Shengzhong

Department of Mathematics and Computer Science
Liuzhou Teachers College
Liuzhou, Guangxi 545004, China
{gxlzhsz,htk3000}@163.com

Abstract. The image edge examination is digital image processing, the image analysis and the machine vision domain important research content. This article uses the wavelet transformation mold maximum value to carry on the image edge examination method. The wavelet transformation mold maximum value often uses for to examine the signal the singular value, may examine image all edges with this method the detail, but simultaneously also can examine some false edges and the noise spot. Through research on the optimization threshold to eliminate false edge, and provides the exact method to determine the threshold; gives the image edge detection based on wavelet transform modulus maximum algorithm and Matlab source code. The simulation experiment indicated that, the effect is extremely good, has certain fundamental research significance.

Keywords: Image signal processing, Edge examination, Mold maximum value, Threshold value.

1 Introduction

The image edge is the image basic characteristic, displays in the image for the image gradation information has the rapid change position, has reflected the image partial characteristic discontinuity. But this kind of discontinuity is because the goal displays in the scene the different depth, the different echoing characteristic and the different brightness cause. Therefore, the edge in some kind of situation on and image gradient related [1], the image edge examination is digital image processing, the image analysis and the machine vision domain important research content. In the image, gray values catastrophe point is located on the boundary of important targets in General, edge detection is to identify these mutations. Practical engineering application of image edge detection technology in many areas, such as pattern recognition, image matching, texture detection, in areas such as remote sensing, aviation, medicine also has a good application. In the recent several dozens years, the edge examination have had many classical operators, like Sobel operator, Prewitt operator, Robert operator, Laplac operator and so on. But they edge location imprecise on some image processing effect is not obvious. The wavelet and the multi-criterion analysis theory can reflect well the image gradation the change, is widely applied in aspects and so on filter, signal and imagery processing. The wavelet transformation mold maximum value algorithm is

opposite in Sobel and so on the classical operator, the edge examination effect is better. Some scholars describe signal irregularity [2] with the wavelet mold maximum value, but the signal singular point usually is in the signal has the important physics significance spot. To the image, the wavelet mold maximum value description is in the image the goal multi-criterion boundary. Edge detection method in the maximum detection operator can not only determine the location of mutations and graded, and can detect signal changes in the singularity. Because the wavelet transform pair strange characteristic especially is sensitive, causes its more suitable examination image the edge and the detail, to some kind of wavelet, the image edge corresponds to wavelet transformation partial mold maximum value [3]. How establishes the appropriate threshold value as well as filters out the partial mold maximum value spot which causes by the noise to become the wavelet edge examination a key question.

This paper presents a noise point of distinction between edge and edge detection algorithm to achieve the separation of noise and edge to solve the wavelet transform modulus maxima exist in the image edge detection problem of false edges. May know from the simulation experiment result, uses the wavelet transformation mold maximum value the algorithm to carry on the edge examination, all edge detail is examined all, meanwhile can examine some false edges, the effect is extremely good.

2   Based on Wavelet Transformation Partial Mold Maximum Value Algorithm

Wavelet transform has the advantage that in the time domain and frequency domain has good local properties, which can be used for image edge detection based on wavelet transform in spatial and temporal resolution in the domain level and the corresponding adjustment of the frequency: crude at low frequency and high frequency fine, it can focus on any details of the characteristics of the object being tested.

2.1   Algorithm Principle

Supposes is a suitable smooth dual function, satisfies the following condition:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\theta(x,y)\,dxdy=1 \quad \lim_{x^2+y^2\to\infty}\theta(x,y)\to0 \tag{1}$$

The introduction of mark

$$\theta_s(x,y)=\frac{1}{s^2}\theta\left(\frac{x}{s},\frac{y}{s}\right) \tag{2}$$

Where s is the scale, is under two dimensional Wavelet scale s has the following definition:

$$\phi^x(x,y)=\frac{\partial\phi(x,y)}{\partial x},\phi^y(x,y)=\frac{\partial\phi(x,y)}{\partial y} \tag{3}$$

The image after smooth function processing, under the criterion function, the two-dimensional wavelet transformation has two components:

$$\begin{vmatrix} W_j^x f(x,y) \\ W_j^y f(x,y) \end{vmatrix} = 2^j \begin{vmatrix} \dfrac{\partial (f*\phi_j)(x,y)}{\partial x} \\ \dfrac{\partial (f*\phi_j)(x,y)}{\partial y} \end{vmatrix} = 2^j (f*\phi_j)(x,y) \tag{4}$$

In which $*$ expresses the convolution, ∇f is gradient vector, $\xi_8(x,y) = \dfrac{1}{s^2}\xi\left[\dfrac{x}{s},\dfrac{y}{s}\right]$ Expresses the two-dimensional function $\xi(x,y)$ After scale factor s stretching transformation. This shows that the dyadic wavelet transform $W_j f(x,y)$ Is the signal $f(x,y)$ After smoothing by $\phi_j(x,y)$ -order partial derivatives of a, its gradient is proportional to the wavelet transform modulus of the mode value.

Modulus and the gradient direction can be expressed as:

$$M_{2^j} f(x,y) = \sqrt{\left|W_{2^j}^x f(x,y)\right|^2 + \left|W_{2^j}^y f(x,y)\right|^2} \tag{5}$$

$$A_{2^j}^x f(x,y) = \arctan \frac{W_{2^j}^x f(x,y)}{W_{2^j}^y f(x,y)} \tag{6}$$

The wavelet transformation partial mold maximum value defines as along the gradient direction partial maximum value, The mold maximum value is the edge intensity, with the argument vertical direction is the edge direction. Therefore, for any point in the digital image Therefore, for any point in the digital image (x,y), Gradient direction along the $M_{2^j} f(x,y)$ and the two neighboring pixels to compare the modulus values to determine whether the local maximum value of $M_{2^j} f(x,y)$, If the local maximum value is retained and marked singular point; if not local maxima, then be deleted, $M_{2^j} f(x,y)$ =0. At the same time choose a chain length threshold to filter out the noise of similar modulus maximum points and close to the maximum angle made by edge of the chain connect the dots, the edges of the image can be obtained.

Wavelet transform is actually a linear operator, said on a number of different scales linear combination of primitive components, and thus the flexibility to reflect the information on different scales. Suppose $\phi(x,y)$ is to meet the integral value of 1 and 0 at infinity tends to smooth two-dimensional differentiable function, then the two-dimensional wavelet can be defined as $\phi^x(x,y)$ and $\phi^y(x,y)$:

$$\phi^x(x, y) = \frac{\partial\phi(x, y)}{\partial x} \tag{7}$$

$$\phi^y(x, y) = \frac{\partial\phi(x, y)}{\partial y} \tag{8}$$

The two-dimensional image function in the scale 2^j and (x, y) position for the binary wavelet has two components:

$$\begin{vmatrix} W_j^x f(x, y) \\ W_j^y f(x, y) \end{vmatrix} = 2^j \begin{vmatrix} \dfrac{\partial(f * \phi_j)(x, y)}{\partial x} \\ \dfrac{\partial(f * \phi_j)(x, y)}{\partial y} \end{vmatrix} = 2^j (f * \phi_j)(x, y) \tag{9}$$

Where * denotes convolution, ∇f is the gradient vector, $\xi_8(x, y) = \dfrac{1}{s^2} \xi\left[\dfrac{x}{s}, \dfrac{y}{s}\right]$

After that two-dimensional function $\xi(x, y)$ of scale factor s stretching transformation. This shows that the dyadic wavelet transform $W_j f(x, y)$, $f(x, y)$ is the signal after smoothing by a $\phi_j(x, y)$-order partial derivatives, the gradient is proportional to the wavelet transform modulus values of the modulus value. Wavelet Transform is detected as extreme edge of the design is equivalent to the MALAT multi-scale edge detection method.

2.2   The Algorithm and Program Code

(1) For the original binary image, discrete stationary wavelet transformation, the wavelet transform coefficient matrix and the original image matrix of the same size.

(2) Transform coefficients, by the image horizontal and vertical wavelet coefficients

(3) Find the local maxima. For digital images at any point (m, n), if it is the gradient direction of the local wavelet transform modulus maxima coefficients point, for the 8 points in the neighborhood, 4 kinds of situations, The point (m, n) only for horizontal, vertical, $35°$, $135°$ in one direction, the direction corresponding to three pixels in the local wavelet transform modulus maxima coefficients. Therefore, any point (m, n) can be considered the midpoint in the direction of the 4 (m, n) where the gradient direction, points (m, n) are coefficients of the wavelet transform modulus on the direction of the local maxima Value point, If so, points (m, n) is the candidate edge points. Otherwise, the non-edge points. For any point (m, n), find local wavelet transform modulus maxima coefficients of the algorithm is:

① If $-\dfrac{\pi}{8}$ -8< $A^j f(x, y) < \dfrac{\pi}{8}$, Comparison points (m-1, n), (m, n), (m +1, n) 3 points corresponding to the modulus value.

② If $\dfrac{\pi}{8} < A^j f(x, y) < \dfrac{3\pi}{8}$, Comparison points (m-1, n+1), (m, n), (m +1, n+1)

3 points corresponding to the modulus value.

③ If $-\dfrac{3\pi}{8} < A^j f(x, y) < -\dfrac{\pi}{8}$, Comparison points (m-1, n), (m, n), (m +1, n-

1) 3 points corresponding to the modulus value.

④ If $A^j f(x, y) < -\dfrac{3\pi}{8}$ or $A^j f(x, y) > \dfrac{3\pi}{8}$, Comparison points (m, n-1),

(m, n), (m, n+1) 3 points corresponding to the modulus value.

If (m, n) is maximum, is the local maxima points. The generation of local maxima along the gradient direction is, in a range of values within the detection mode, the maximum value is retained, non-maxima were removed.

(4) To get the local maxima to determine its value is not greater than the given threshold, if so, determine the corresponding point in the image edge points. Judging by the results obtained above can be maintained the edge of the edge of the image weak image.

According to the algorithm, simulation using the original MATLAB code to achieve the following:

```
% Modulus maxima of wavelet edge detection for
clc;clear
RGB=imread('lena.jpg');
[X,map]=rgb2ind(RGB,128);
X=ind2gray(X,map);
SIZE=length(X);
m=1.0;
delta=2^m;
% Construct the partial derivative of Gaussian function
N=20;
A=-1/sqrt(2*pi);
for index_x=1:N;
   for index_y=1:N;
      x=index_x-(N+1)/2;
      y=index_y-(N+1)/2;
      phi_x(index_x,index_y)=A*(x/delta^2).*exp(-(x.*x+y.*y)/(2*delta^2));
      phi_y(index_x,index_y)=A*(y/delta^2).*exp(-(x.*x+y.*y)/(2*delta^2));
   end
end;
phi_x=phi_x/norm(phi_x);
phi_y=phi_y/norm(phi_y);

%   Do the ranks of the convolution of image
Gx=conv2(X,phi_x,'same');
Gy=conv2(X,phi_y,'same');
Grads=sqrt((Gx.*Gx)+(Gy.*Gy));
angle_array=zeros(SIZE,SIZE);
for i=1:SIZE
   for j=1:SIZE
      if (abs(Gx(i,j))>eps*100)
```

```
            p=atan(Gy(i,j)/Gx(i,j))*180/pi;
            if (p<0)
               p=p+360;
            end;
            if (Gx(i,j)<0 & p>180)
               p=p-180;
            elseif (Gx(i,j)<0 & p<180)
               p=p+180;
            end
          else
             p=90;
          end
          angle_array(i,j)=p;
       end
   end;
   edge_array=zeros(SIZE,SIZE);

   for i=2:SIZE-1
      for j=2:SIZE-1
         if  ((angle_array(i,j)>=(-22.5)  &  angle_array(i,j)<=22.5|    (angle_array(i,j)>=(180-22.5)  &
angle_array(i,j)<=(180+22.5)))
            if (Grads(i,j)>Grads(i+1,j) & Grads(i,j)>Grads(i-1,j))
               edge_array(i,j)=Grads(i,j);
            end
         elseif ((angle_array(i,j)>=(90-22.5) & angle_array(i,j)<=(90+22.5)) |
             (angle_array(i,j)>=(270-22.5) & angle_array(i,j)<=(270+22.5)))
            if (Grads(i,j)>Grads(i,j+1) & Grads(i,j)>Grads(i,j-1))
                 edge_array(i,j)=Grads(i,j);
            end
         elseif ((angle_array(i,j)>=(45-22.5) & angle_array(i,j)<=(45+22.5)) |
             (angle_array(i,j)>=(225-22.5) & angle_array(i,j)<=(225+22.5)))
            if (Grads(i,j)>Grads(i+1,j+1) & Grads(i,j)>Grads(i-1,j-1))
               edge_array(i,j)=Grads(i,j);
            end
         else
            if (Grads(i,j)>Grads(i+1,j-1) & Grads(i,j)>Grads(i-1,j+1))
               edge_array(i,j)=Grads(i,j);
            end
         end
      end
   end
   %   Removal of false edges
   MAX_E=max(max(edge_array).');
   edge_array=edge_array/MAX_E;
   threshold=0.15;
   % threshold=MAX_E;
   for m=1:SIZE
      for n=1:SIZE
         if (edge_array(m,n)>threshold)
            edge_array(m,n)=1;
         else
            edge_array(m,n)=0;
         end
      end
   end
   %   Display image and edge
   figure(1)
   subplot(2,2,1)
   imshow(X)
```

```
title('Original image')
subplot(2,2,4)
imshow(edge_array)
title('The method of image')
i=rgb2gray(RGB);
b1=edge(i,'canny');
b2=edge(i,'log');
subplot(2,2,2);
imshow(b1);title('canny image');
subplot(2,2,3);
imshow(b2);title('log image');
```

THE SIMULATION RESULTS

Fig. 1. Edge detection comparison chart

Is apparent from Fig.1: canny edge detection operator is the result of false edges detected more; LOG operator edge detection results, detected the edge of the more obscure, the phenomenon of simultaneous loss of the edge; this algorithm The best edge detection result, the affected small, marginal loss of situation does not exist, false edges less.

3   Summaries

The simulation results shows that a simple algorithm used in this paper, the image edges and details are detected by very accurate, and CANNY LOG operator and compare the edge detection operator, the conclusion is proposed based on wavelet transform modulus maxima The best edge detection. But less than the hard threshold is used, if the threshold value through the software automatically optimize, you can get a better edge image.

References

[1] Tin, Y., Qi, G.: Wavelet transform modulus maxima based approach to edge detection, vol. 2, pp. 102–105. Dalian Maritime University (2007)

[2] Chang, Y.-H., Chen, M.: Automatically set a threshold wavelet edge detection method. Computer 23, 286–308 (2007)

[3] Fang, Y.: An improved wavelet transform modulus maxima based approach to edge detection. Fujian Agriculture and Forestry University of High grams 7, 444–448 (2009)

[4] Yang, G., Liu, L.: Based on wavelet transform modulus maxima denoising adaptive threshold algorithm. Naval Engineering University 2, 43–47 (2007)

[5] Zhang, F., Zhang, Q.: Portugal, wavelet-based image edge detection algorithm. Sun Yatsen University: Natural Science 46(3), 39–42 (2007)

Numerical Simulation of Steel Solidification in Steel-Al20Sn Liquid to Liquid Bonding

Zhang Jun, Zhang Peng, Du Yun-Hui, and Yao Shasha

School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University

Abstract. Liquid to liquid bonding is a less procedure method for manufacturing composite plates. Based on this new bonding method the coupled temperature and flow fields of steel solidification in Steel-Al20Sn bonding process was simulated by Finite Element Method. The position of solidification line was analyzed under different pouring temperatures, casting-rolling speeds and cooling heat flux. The appropriate casting-rolling speed and cooling heat flux was obtained.

Keywords: Liquid to liquid bonding, Steel strip, Solidification, Coupled flow and heat transfer.

1 Introduction

Steel-Al20Sn composite plates are usually forming by steel back and Al20Sn alloy clad bonding together in some way. With steel back's high strength and Al20Sn cladding's excellent abrasion performance plus high heat conductivity, steel-Al20Sn composite plates have been a kind of ideal wear-resistant material for automobiles' bearing. It is now under a great demand of quantities every year.

At present the methods of Steel-Al20Sn composite plates are manufactured by solid to solid bonding and solid to liquid bonding. Solid to solid bonding method is that steel back and Al20Sn alloy clad are rolled and their bonding interface usually is considered as occlusal structure or a incomplete physical combination pattern, so the interface shear strength is unsatisfactory. Solid to liquid bonding method is that steel back and Al20Sn alloy molten is bonded directly immersed in flux which can partly prevent them from oxidation. Then the metallurgical bonding interface is obtained and its shear strength is higher. With aforementioned methods, the steel back's clean surface's oxidation and inclusions' existence will make the combination not perfectly. Moreover, pretreatment of steel back will lead to energy consumption.

Based on the solid to liquid bonding method, Zhang Peng presented a liquid to liquid bonding method [1-3], and steel liquid can form into boards in bonding nozzles. Bonding directly with steel molten and Al20Sn alloy molten in the condition of sealing, disturbance of oxygen is avoided and the effective combination ratio of interface will increase greatly. While liquid to liquid bonding needs complicated process, has many corresponding processing parameters. And metal flow behavior, temperature fluctuation and position of solidification line's control in the metal solidification process will significantly affect composite plates' quality. So the

numerical simulation method for research on processing parameters has a very important significance [4-6].

In this paper the flow field and the temperature field in the liquid to liquid bonding process of the composite plate are simulated numerically. And different pouring temperatures, casting-rolling speeds, cooling rates' influence on the position of solidification line, the velocity distribution and the temperature distribution is investigated.

2   Basic Hypothesis and Governing Equations

2.1   Basic Hypothesis

Assuming that the density of steel liquid in the solidification process does not vary with temperature, steel molten will flow with incompressibility, steadiness and variable viscosity. Ignoring heat transfer and flow in the direction of the plate width, a 2D model can be established.

2.2   Governing Equations

(1) Quality continuous equation

$$\frac{\partial \rho}{\partial t} + \frac{\partial (\rho u)}{\partial x} + \frac{\partial (\rho v)}{\partial y} = 0$$

Where u and v are the velocity vector components in the directions of x, y, ρ is the density and t is time.

(2) Momentum conservation equation

$$\frac{\partial (\rho v)}{\partial t} + \frac{\partial (\rho v u)}{\partial x} + \frac{\partial (\rho v v)}{\partial y} = \frac{\partial}{\partial x}(\mu \frac{\partial v}{\partial x}) + \frac{\partial}{\partial y}(\mu \frac{\partial v}{\partial y}) - \frac{\partial p}{\partial y} + S_v$$

Where μ is the effective viscosity, S_v is source item.

(3) Energy conservation equation

$$\frac{\partial (\rho T)}{\partial t} + \frac{\partial (\rho u T)}{\partial x} + \frac{\partial (\rho v T)}{\partial y} = \frac{\partial}{\partial x}(\frac{k}{c_p} \frac{\partial T}{\partial x}) + \frac{\partial}{\partial y}(\frac{k}{c_p} \frac{\partial T}{\partial y}) + S_T$$

Where c_p is the specific heat capacity, T is temperature and S_T is a viscosity dissipative term.

2.3   Latent Heat and Solidification Treatment

Crystallization latent heat will be released when the liquid metal solidificates, that is one of the solidification problem characteristics.

In FEM analysis, usual treatment methods for crystallization latent heat include temperature compensating method, equivalent heat capacity method, enthalpy method and so on. Adopting the equivalent heat capacity method here, the heat transfer process in the solid-liquid phase region with heat sources is equivalent to a heat transfer process without heat sources by revising specific heat C. And revised C is referred to as equivalent specific heat C^* according to the formula below [7].

$$\mathrm{C}^* = \begin{cases} C_s(T); & T \le T_s \\ \dfrac{C_s + C_l}{2} + \dfrac{L}{\Delta T}; & T_s \le T \le T_l \\ C_l(T); & T \ge T_l \end{cases}$$

Where C_s is solid-state heat capacity, C_l is liquid-state heat capacity, T_l is the solidification starting temperature, T_s is the solidification ending temperature, L is the crystallization latent heat and ΔT is the solidification cooling interval.

3   Boundary Conditions

(1) Pouring nozzle. The pouring temperature is given; the pouring rate is determined according to the casting-rolling speed, the volume conservation and nozzle line length.

(2) Horizontal liquid surface. Ignoring convective, it is set to an adiabatic boundary. Ignoring surface fluctuation, the vertical component of the given velocity is zero and the pressure is zero.

(3) Separate plug wall. It is set to an adiabatic boundary and components of velocity are zeros.

(4) Water-cooled boundary. Heat flux is given, and components of velocity are zeros.

(5) Outlet. It is set to casting-rolling speed, and the direction is vertically downward.

4   Simulating Results

Giving the Casting-rolling conditions that the strip pouring temperature T is 1525-1550℃, the cooling heat flux Q is 50000-80000J/(kg·℃), and the bonding nozzle height H is 0.6m. On the side of steel molten the pouring method of Immersing nozzle is adopted to simulate the temperature and velocity fields when the composite plates stabilized and molded.

4.1   Influence of Pouring Temperature on Temperature Fields

The temperature fields under different pouring temperatures are showed as figure 1.

Figure 1 shows that the pouring temperature has an effect on the solidification temperature fields. When the pouring temperature rises, the position of solidification line moves up. On the contrary, it will move down. But the pouring temperature's influence on the velocity field is little enough that can be ignored.

4.2   Casting-Rolling Velocity's Influence on the Temperature Fields

The temperature fields under different Casting-rolling speeds are showed as figure 2.

It shows that the casting-rolling velocity has a significant influence on the temperature fields and the position of solidification line. When the speed is low enough, steel molten in the flow channel will stay a long time, and then the cooling time becomes long and the steel molten cools well. Therefore the position of solidification line approaches to the flow channel entry. When the speed is faster,

before steel liquid cools sufficiently it has been pulled to the place near the flow channel entry, and then the position of solidification line is on the contrary moves down. The casting-rolling speed also affects the velocity field greatly. When the speed is high, steel molten stays in a laminar flow. As casting-rolling speeds up, turbulence will occurred.

| (a) 1525℃ | (b) 1550℃ |
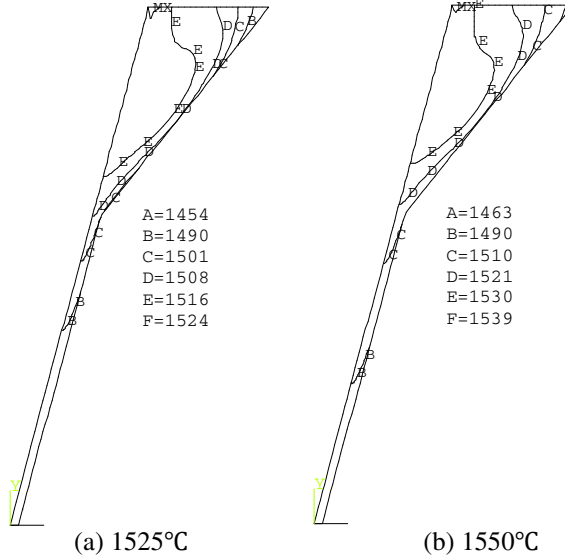
Fig. 1. Temperature field distribution under different pouring temperatures

| (a) 5mm/s | (b) 20mm/s |
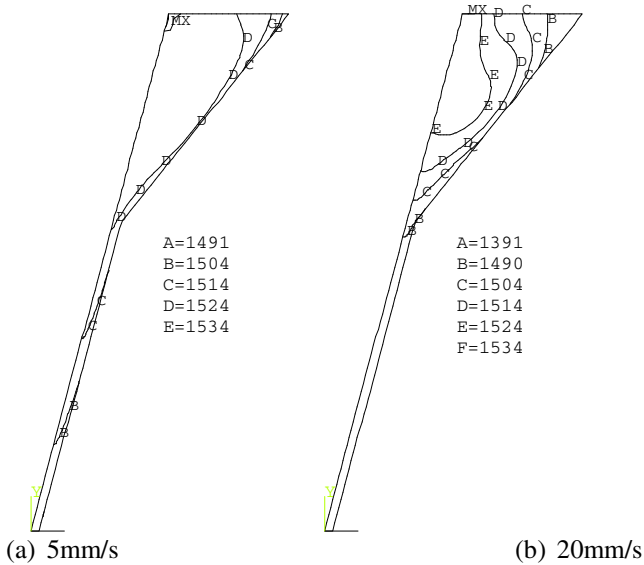
Fig. 2. Temperature field distribution under different casting-rolling speeds

4.3   Cooling Speed's Influence on the Temperature Fields

The temperature fields under different cooling heat flux are showed as figure 3.

In the figures it can be showed that the cooling heat flux has a large influence on the temperature fields. When steel liquid cools slowly enough, solidification becomes difficult and the position of solidification line is relatively low. When the cooling heat flux is power-up, steel liquid solidified fast and the position of solidification line approaches to the flow channel outlet.

A=1454
B=1490
C=1504
D=1514
E=1524
F=1534

A=1472
B=1490
C=1504
D=1514
E=1524
F=1534

(a) 50000 J/(kg·℃)   (b) 80000 J/(kg·℃)
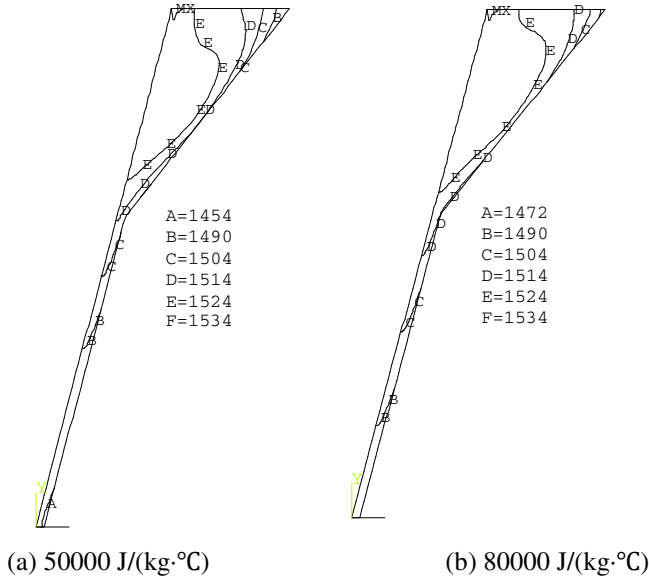
Fig. 3. Temperature field distribution under different cooling heat flux

5   Summaries

The simulation results show that liquid to liquid bonding method to manufacture Steel-Al20Sn composite plates is reasible. And pouring temperature affects temperature fields to a certain extent. The higher the pouring temperature is, the closer the position of solidification line is to liquid flow channel outlet. Casting-rolling velocity is the key factor for steel plates to form into shapes in the liquid flow channel. Flow channel cooling has greater impact on the steel solidification. The higher the cooling heat flux is, the closer the position of solidification line is to the entrance of the liquid flow channel.

References

[1] Kenneth: Metal-base Composite. National Defence Industry Press (1982) (in Chinese)
[2] Zhang, P.: Study on Steel-Aluminum Solid to Liquid Bonding Process and Theory. Northeastern University (1998) (in Chinese)

[3] Zhang, P., Du, Y.-H.: Study on Steel-Al20Sn Nonhomo-geneous Diffusion Bonding. ACTA Material Composite Sinica 1, 46–49 (2002) (in Chinese)

[4] Miao, Y.-C.: Numerical Simulation of Solidification of Twin-roll Strip Casting Process. Northeastern University (2001) (in Chinese)

[5] Jin, Z.-M., Hao, J.-C.: Mathematic Model of Solidification Point Position Control in Twin-roll Steel Strip Casting Process. The Chinese Journal of Nonferrous Metals 2, 694–695 (1998)

[6] Gupta, M., Sahai, Y.: Mathematical Modeling of Fluid Flow, Heat Transfer, and Solidification in Twin-roll Melt Drag Thin Strip Casting. ISIJ International 2, 144–152 (2000)

[7] Jin, Z.-M.: Simulation of Flow and Heat Transfer in Twin-roll Casting Process. Northeastern University (1999) (in Chinese)

An Urban Traffic Prediction Model Based on Temporal Data Mining in Shanghai City

Chen Hong

Nanjing Institute of Industry Technology

Abstract. A temporal data mining method was used to develop a short-term urban traffic flow volume prediction model in a signalized controlling situation, where the next green light phase traffic flow volume of a crossing entrance is predicted by historical green light phase traffic flow volume data of itself and upstream crossing entrances. An experiment is given, and the result is satisfying.

Keywords: Urban traffic management, Short-term prediction, Data mining.

1 Introduction

Traffic flow volume is a kind of important information in traffic management system. In the past years, many traffic flow prediction models have been proposed for predicting traffic flow volume. ARIMA(Autoregressive Integrated Moving Average) model and Kalman filtering model were chosen at forepart [1~2]. The method of recursive least-square estimation for the online prediction of traffic flow was applied in the SCATS system [3~4]. In the SCOOT system, the cycle flow profiles were applied to modeling the cycle-to-cycle variations in traffic flow model [5]. Recently, due to the stochastic and its nonlinear nature of the traffic flow, the neural network technique was employed [6~8]. A fuzzy-neural approach was used for urban traffic flow [9]. However, these prediction models are fit for predicting the traffic flow volume in a fixed time length situation, and most of them are under the situation of highway only. In urban traffic situation that has traffic signal controlling, a short-term prediction must consider the effect of the traffic signal controlling.

Han et al. proposed the methods to mine periodic segments and partial periodic patterns in temporal databases [10~11]. Algorithms for mining intertransaction association rules were presented in [12~13].

In this paper, the temporal data mining method is used to construct a short-term urban traffic flow volume prediction model, which employs a time delayed neural network (TDNN).

2 Traffic Flow Volume Data

The urban traffic flow volume data are a kind of temporal data, which are collected by detectors. The form of the traffic flow data series of a crossing entrance is as follow:

$$\cdots, CE(t), CE(t+1), CE(t+2), CE(t+3), CE(t+4), CE(t+5), \cdots$$

Where: CE(t) is a traffic flow data record, and it is the serial number. The traffic flow data record contains five fields.

Date: the date of making the record, marked as CE(t).date.

Phase Start Time: the start time of a phase of a crossing, and it is a certain second of the record date, marked as CE(t).phase_start_time. $0 \leq$ CE(t).phase_start_time< 86400

Phase: the phase, marked as CE(t).phase.

Phase Length: the time length of the phase holding , marked as CE(t).phase_length.

Volume: the value of traffic flow in a phase, marked as CE(t).volume. For example, Table1 records the traffic flow data at a crossing entrance.

Assuming that a phase period of the crossing consists of m phases, $\{Phase_1, Phase_2, \cdots, Phase_m\}$, the records with the ith phase can form a series, which is shown as follow.

$$\cdots, CE(t+i), CE(t+m+i), CE(t+2m+i), CE(t+3m+i), CE(t+4m+i), CE(t+5m+i), \cdots$$

Since the traffic flow volume of a crossing entrance can be ignored when it is in a red light phase, the traffic flow data series need to be considered in the green light phase. For simplicity, the green light phase traffic flow data series is denoted as follow.

$$\cdots, CE_g(t), CE_g(t+1), CE_g(t+2), CE_g(t+3), CE_g(t+4), CE_g(t+5), \cdots$$

3   Temporal Data Mining Method

The process of using the temporal data mining consists of three steps. Firstly, data preprocessing cleans the primitive data and transforms the traffic flow volume data into the velocity of flow. Secondly, set up traffic prediction model. Here, a TDNN is trained to predict the velocity of flow, and the predicted velocity of flow is used to calculate the traffic flow volume. Lastly, a standard is employed to compare the predictions given by the TDNN model with those observed, i.e., measure the capability of the proposed model in obtaining desirable traffic flow predictions.

3.1   Data Preprocessing

3.1.1   Data Cleaning

Default data process is a common problem in data mining. The data record is often lost by unknown reasons. In order to solve the problem, an average method is used.

In above records series, if CE(t+2m+i) is lost, the fill rules are,

$$CE(t+2m+i).date = CE(t+2m+i-1).date \tag{1}$$

$$CE(t+2m+i).phase\_start\_time$$
$$= CE(t+2m+i-1).phase\_start\_time + CE(t+2m+i-1).phase\_length \tag{2}$$

$$CE(t+2m+i).phase = CE(t+m+i).phase \tag{3}$$

$$CE(t+2m+i).phase\_length$$
$$= CE(t+2m+i+1).phase\_start\_time - CE(t+2m+i).phase\_start\_time \tag{4}$$

$$CE(t+2m+i).volume = CE(t+m+i)/2 + CE(t+3m+i)/2 \tag{5}$$

3.1.2   Data Transformation

Since the phase length is a variable, the traffic flow volume in a phase can be taken as a processed object. In order to get a standard processed object, the green light phase velocity of flow was selected as the processed object, which was denoted by v.

Therefore,

$$CE(t).v = CE(t).volume/ \ CE(t).phase\_length \qquad (6)$$

$$CE(t).volume = CE(t).v \ \times \ CE(t).phase\_length \qquad (7)$$

3.2   Traffic Prediction Model

A TDNN was employed to set up the velocity of flow in green light phase prediction model of a crossing entrance. The inputs of the TDNN are the historical velocity of flow in green light phase of itself and upstream crossing entrances.
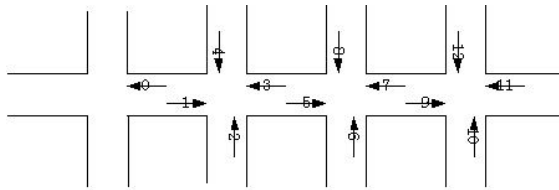
Fig. 1. An upstream and downstream relation map

The Fig.1 shows an upstream and downstream relation map. There are 13 marked crossing entrances in Fig.1, and they are marked as CE0, CE1, CE2, CE3, CE4, CE5, CE6, CE7, CE8, CE9, CE10, CE11, CE12. Assume that CE0 is the object crossing entrance. The traffic flow data series of CE0 are predicted by historical traffic flow data series of 12 crossing entrances. The next green light phase traffic flow data of CE0 is defined as $CE0_g(t)$. Several green light phase velocity of flow are selected for each upstream crossing entrance that are just before the green light phase that will be predicted. In the while, the object crossing entrance is also considered. Assuming that n green light phase traffic flow data records of the kth crossing entrance are selected, the data series is as follow,

$$CEk_g(t-1), CEk_g(t-2), CEk_g(t-3), \ \cdots, CEk_g(t-i), \ \cdots, CEk_g(t-n)$$

Where: $CEk_g(t-i)$ is the ith delayed record of the kth crossing entrance just before $CE0_g(t)$. The smaller value of i is, the newer the data is. A traffic flow data record in the series is called a datum point. A datum point selected for prediction is called as a delayed datum point. The velocity of flow of the selected delayed data points are the inputs of the TDNN.

3.3   Model Evaluation Measure

After setting up and training the TDNN, the root mean square error (RMSE) and equal coefficient (EC) between the predicted and the observed values of the predicted crossing entrance are used to measure the prediction capability of TDNN. RMSE is a measure for the error of the predicted. The smaller the RMSE is, the smaller the error is. EC denotes the comparability between the predicted series and the observed series. Here, the comparability is satisfiable if EC is bigger than 0.9 [14].

$$RMSE_{velocity} = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(CE0_g(t).v\_observed - CE0_g(t).v\_predicted)^2} \qquad (8)$$

$$EC_{velocity} = 1 - \frac{\sqrt{\sum_{t=1}^{T}(CE0_g(t).v\_observed - CE0_g(t).v\_predicted)^2}}{\sqrt{\sum_{t=1}^{T}CE0_g(t).v\_observed^2} + \sqrt{\sum_{t=1}^{T}CE0_g(t).v\_predicted^2}} \qquad (9)$$

$$RMSE_{volume} =$$
$$\sqrt{\frac{1}{T}\sum_{t=1}^{T}(CE0_g(t).volume\_observed - CE0_g(t).volume\_predicted)^2} \qquad (10)$$

$$EC_{volume} =$$
$$1 - \frac{\sqrt{\sum_{t=1}^{T}(CE0_g(t).volume\_observed - CE0_g(t).volume\_predicted)^2}}{\sqrt{\sum_{t=1}^{T}CE0_g(t).volume\_obesrved^2} + \sqrt{\sum_{t=1}^{T}CE0_g(t).volume\_predicted^2}} \qquad (11)$$

Where: T is the number of the data point in the predicted traffic flow series.

4   Experiments and Conclusions

A detail map and a sketch road map of XuJiahui District in Shanghai city are shown in the Fig.2 and the Fig.3 respectively. There are four crossings corresponding to A, B, C, D in Fig.2 and Fig.3.

   Crossing entrance 0 is the object entrance that will be predicted. The entrances of upstream crossing are 1,2,3,4,5,6,7,8,9., marked as CE0, CE1, CE2, CE3, CE4, CE5, CE6, CE7, CE8, CE9. For CE0, 2 delayed data points are selected. For CE1, CE2 and CE3, 3 delayed data points are selected respectively. For CE4, CE5 and CE6, 4 delayed data points are selected respectively. For CE7, CE8 and CE9, 5 delayed data points are selected respectively. In all, there are 38 delayed data points to be selected for prediction.

Fig. 2. The detail road map of XuJiaHui district in ShangHai
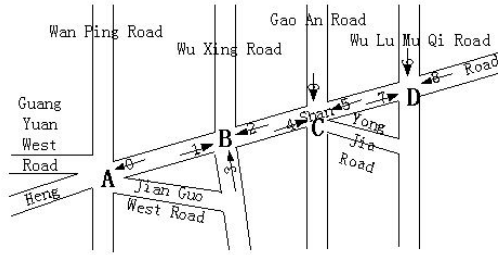
Fig. 3. The sketch road map of XuJiaHui district in ShangHai

After the prediction model is set up, two test sample sets, TS1 and TS2, are taken to assess the performance of the prediction model. Each test sample set is a data series with 100 data points. The predicted velocity of flow data series of test samples sets is shown in Fig.4, and their RMSE and EC are shown in Table1. The predicted traffic flow volume data series of the test sample sets are shown in Fig.5, and their RMSE and EC are shown in Table 2.
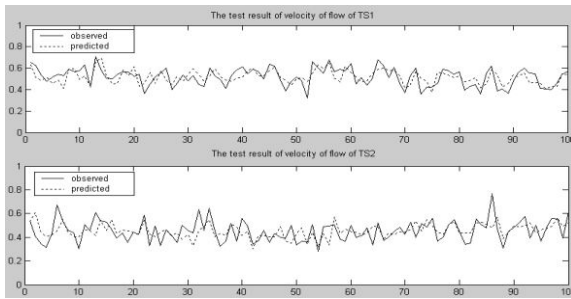
Fig. 4. The predicted and the observed velocity of flow data series of test samples sets
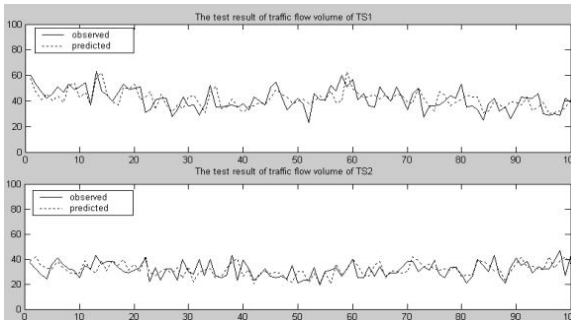
Fig. 5. The predicted traffic flow volume data series of test samples sets

Table 1. The RMSE and EC of the predicted velocity of flow data series of TS1, TS2

	TS1	TS2
RMSE	0.067519	0.084188
EC	0.93557	0.90846

Table 2. The RMSE and EC of the predicted traffic flow volume data series of TS1, TS2

	TS1	TS2
RMSE	4.2195	4.2449
EC	0.91438	0.90286

Test results in Fig.4 and Fig.5 show that the predicted curves track the observed curves well though the real traffic flow data contain big randomicity. Test data in the Table1 and the Table3 prove the RMSE of the predicted velocity of flow data series of TS1 and TS2 are less than 0.1 and the RMSE of the predicted traffic flow volume data series of TS1 and TS2 are less than 5. Each EC is bigger than 0.9, which shows that the comparability between the prediction and the observation is high. Considering the randomicity of the traffic flow, the TDNN model is satisfiable.

The temporal data mining method is used to set up a short-term urban traffic flow volume prediction model. The time length of prediction is alterable by using this model. Also, it is shorter than the time length comparing with other prediction models. As we known, the green light phase time is less than 120 seconds normally, therefore, this model will get a real time prediction value of traffic flow volume. The experiment result is satisfying. And it is feasible to set up urban traffic model based on temporal data mining.

References

[1] Shimizu, H., Yamagami, K., Watansbe, E.: Application of State Estimation Algorithms To Hourly Traffic Volume System. In: Proceedings of Second World Congress on ITS, Yokohama, pp. 72–77 (1995)

[2] Stephanedes, Y.J., Kwon, E.: Adaptive deman-deversion prediction for integrated control of freeway corridors. Transportation Research 1C, 23–42 (1993)

[3] Cremer, M., Keller, H.: A new class of dynamic methods for the identification of origin-destination flows. Transportation Research 21B, 17–32 (1987)

[4] Luk, J.Y.K.: Modelling and monitoring the performance of urban traffic control network. Australian Road Research Board SR43, Victoria, Australia (1989)

[5] Hunt, P.G., Robertson, D.I., Brethton, R.D., Winton, R.I.: SCOOT – a traffic responsive method of coordinating signals. TRRL LR1014, Transport and Road Research Laboratory, Crowthorne, UK (1981)

[6] Dougherty, M.S., Kirby, H.R., Boyee, D.: Using neural networks to recognize, Predict and model traffic. Artificial Intelligence Application to Traffic Engineering, 235–250 (1994)

[7] Ledoux, C.: An urban traffic flow model intergrating neural network. Transportation Research 5C, 287–300 (1997)

[8] Yun, S.Y., Namkoong, S.R., Shin, J.H., Choi, S.W.: A Performance Evaluation of Neural Network Models in Traffic Volume Forecasting. Mathematical and Computer Modelling 27(9-11), 293–310 (1998)

[9] Yin, H., Wong, S.C., Xu, J., Wong, C.K.: Urban traffic flow prediction using a fuzzy – neural approach. Transportation Research 10C, 85–98 (2002)

[10] Han, J., Dong, G., Yin, Y.: Efficient mining of partial periodic patterns in time-related databases. In: Proceedings 1999 International Conference Data Engineering, Sydney, Australia (1999)

[11] Han, J., Gong, W., Yin, Y.: Mining segment-wise periodic patterns in time series databases. In: Proceedings 1998 International Conference Knowledge Discovery and Data Mining, New York, pp. 214–218 (1998)

[12] Lu, H., Han, J., Feng, L.: Stock movement prediction and N-dimensional inter-transaction association rules. In: Proceedings 1998 SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, Seattle, WA, vol. 12, pp. 1–7 (1998)

[13] Tung, A.K.H., Lu, H., Han, J., Feng, L.: Breaking the barrier of transactions: mining inter-transaction association rules. In: KDD 1999, pp. 297–301. ACM (1999)

[14] Yang, Z.: Theory and model of urban traffic flow induction system. Renming Traffic Publisher, Beijing (2000) (in Chinese)

Numerical Simulation Study
of Military Shelter Defenses Fragment

Wang Chao and An Zhen-Tao

Ammunition Engineering Department, Ordnance Engineering College,
Shijiazhuang, Hebei 050003, China
junxiewangchao@sina.com

Abstract. It built the model of military shelter in the AUTODYN software, and analyzed the defense ability of shelter from fragment for the girder structure Ⅰ and the sandwich structure Ⅱ.Through analysis, the simulation result shows that the shelter is weak in defense the fragment, so we must take corresponding measure to prevent the shelter from destroying in order to improve its viability.

Keywords: Military shelter, Fragment defense, AUTODYN, Numerical simulation.

1   Introduction

Military shelter, a moveable box which is composed of metallic cover and the cored sandwich structure, can offer the needed work condition and environment defense, and is used to loading equipment and persons. [1]

As the most of military shelter is composed of metallic covers and cored sandwich structure, so the smooth aluminum cover is a fine reflector of radar wave and laser, meanwhile the radiator of air conditioner is a fine target-seeking heat source of infrared guidance, these characters make the military shelter a fine object being found out. Then, aluminum-covered sandwich plate is so weak that it is very vulnerable and easy to be destroyed by a fragment and becomes a heap of scrap iron. So, we must research on the fragment defense capability of shelter in order to take corresponding measure to improve its viability.

As the structure of the military shelter is very complicated, including the cover; polyurethane foam core; strengthened girder, etc. [2] The best research method is through experiment analyzing the bulletproof ability of shelter, but it costs a lot of money; then the traditional method of theory calculation is very hard meanwhile the precision is difficult to ensure.

This paper carries on the bulletproof ability experiment of military shelter making use of ANSYS AUTODYN[3] software, which makes the basic evaluation of the fragment defense ability of shelter.

2   Modeling and Simulation

2.1   Structure Simplify of Shelter

Generally speaking, the common shelter is a box structure, its basic dimension is 4012mm×2240mm×2100mm.The shelter is a sandwich, the inner and outer cover of the shelter is a 1.5mm aluminum-alloy plate, with the 50mm polyurethane foam core; along the border parallel to the width, it is strengthened by the square pipe with the profile dimension 40mm×40mm and 2mm in thickness, the inner cover of shelter and square pipe is strengthened by a hard wood of 10mm in thickness[4]. The girder and sandwich structure of the shelter is shown below in Fig1.
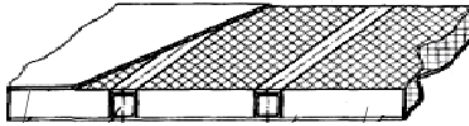
Fig. 1. Sketch map of the shelter wainscot

The study of shelter defenses fragment is mainly towards the structure of Fig1, because of symmetry of the shelter structure, we use 1/4 model for simulation.

2.2   Modeling

1） Fragment
The fragments select the IRON-ARMCO as its material, the main material parameters are listed as follows, Density: 7.89g/cm3; Bulk modulus: 164GPa; Shear modulus: 80GPa; Yield stress: 175MPa; Harding constant: 380MPa; Harding exponent: 0.32; Strain rate constant: 0.06; Thermal softening exponent: 0.55.

2） Aluminum Alloy
The shelter covers and the strengthen girder select AL2024T 351 as their materials, the main parameters are listed as follows, Density: 2.785g/cm3, Parameter C1: 5328m/s; Parameter S1: 1.338; Shear modulus: 27.6GPa; Yield stress: 265MPa; Harding constant: 426MPa; Harding exponent: 0.34; Strain rate constant: 0.015; Principal tensile failure stress: 400MPa; Erosion strain: 4.0.

3） Polyurethane Foam
The shelter selects polyurethane foam as its material, the main material parameters are listed as follows, Density: 0.12g/cm3, Bulk modulus: 6.0MPa; Shear modulus: 5.0MPa; Principal tensile failure stress: 1.06MPa; Erosion strain: 4.0.

4） Wood
As the material model of wood is hard to got, this paper select POLYCARB[5] as its substitute, this material is used in our life to substitute wood in many areas like construction and life article for use. The main material of the wood are listed as follows, Density: 1.2 g/cm3; Gruneisen coefficient: 0.61; Shear modulus: 1.0GPa; Yield stress: 80.6MPa; Plastic strain: 2.0; Erosion Strain: 4.0.

The military shelter model built in AUTODYN is shown in Fig2, the dimension is 120mm×80mm×53mm.
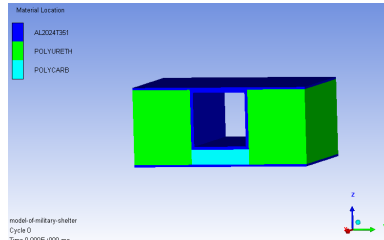
Fig. 2. Model of the military shelter

As the shelter is symmetry in X and Y axis, so we simulate the 1/4 model only. The whole dimension of the fragment is 8mm×16mm×40mm, its mass is 40g, the initial velocity of the fragment is 806.5 m/s along the minus Z axis and vertical to the surface of the military shelter. Suppose the girder as the structure Ⅰ, the sandwich as the structure Ⅱ, the initial state of the 1/4 model is shown below Fig3-0 and Fig4-0:
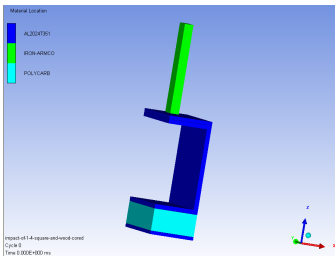
Fig. 3-0. Initial state of model Ⅰ   **Fig. 4-0.** Initial state of model Ⅱ
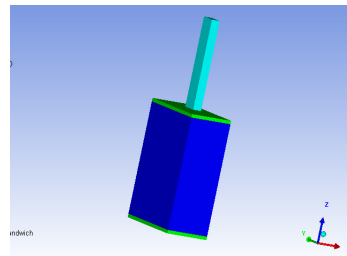
2.3   Simulation of the Models

As the models are built, we simulate the models in the AUTODYN software, the simulation is shown as follows:
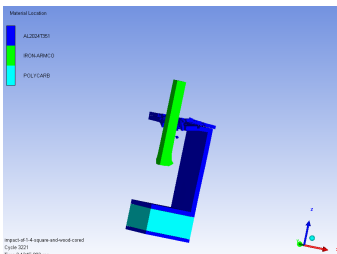
Fig. 3-1. Simulation process of the model Ⅰ   **Fig. 4-1.** Simulation process of model Ⅱ
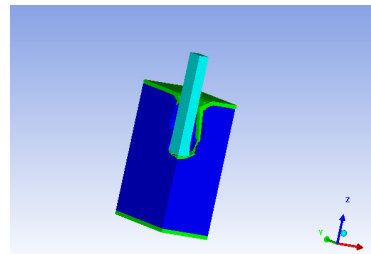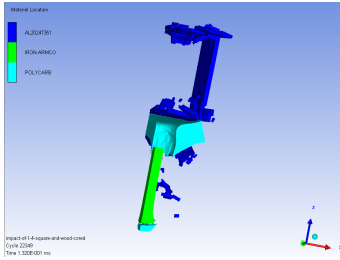
Fig. 3-2. Running results of simulation Ⅰ
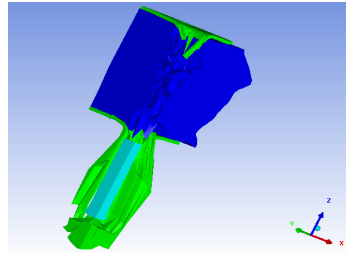
Fig. 4-2. Simulation result of simulation Ⅱ
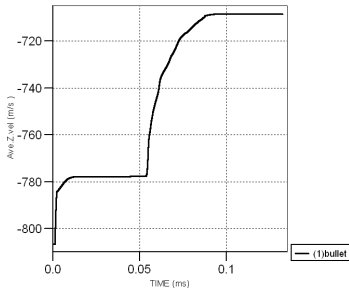
Fig. 3-3. Velocity-time relationship in Z axis of the bullet for model Ⅰ
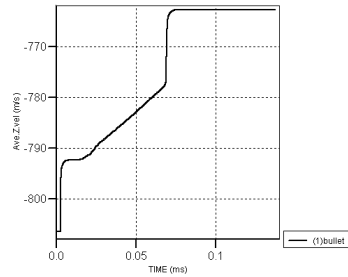
Fig. 4-3. Velocity-time relationship in Z axis of the bullet for model Ⅱ

3   Analysis of the Simulation Results

After simulation, for structure Ⅰ (Fig3-1~3):

1) As the fragment impacts on the structure Ⅰ, the residual velocity of the bullet is gradually decreased from 806.5 to 708.4m/s.

2) Between the time of 0~10.0μs, the fragment interacts with the outer cover of the shelter and square box, so the velocity decreases very quickly; between the time of 10.0~53.7μs, the fragment has penetrated the outer cover, the velocity of fragment holds the line as there is no obstacle; between the time of 53.7μs~61.6μs, the fragment interacts with the inner cover of square box, so the velocity decreases very quickly; between the time of 61.6μs~92.0μs, the fragment interacts with the wood and inner cover of the shelter, as the wood is relatively weak in intensity, so the velocity decreases some slowly; after 92.0μs, the shelter has been penetrated through, the velocity of the fragment maintains 708.4m/s.

For structure Ⅱ (Fig4-1~3):

1) As the fragment impacts on the structure Ⅱ, the residual velocity of the fragment is gradually decreased from 806.5m/s to 762.8m/s.

2) Between the time of 0~6.0μs, the fragment interacts with the outer cover of the shelter, so the velocity decreases very quickly; between the time of 6.0~68.0μs, the

fragment has penetrated through the outer cover of the shelter and interacts with the polyurethane foam, but the velocity decreases slowly as the foam is softer than the cover; between the time of 68.0~74μs, the fragment interacts with the inner cover of the shelter and the velocity decreases quickly; after 74.0μs, the shelter has been penetrated through and the velocity of the fragment maintains 762.8m/s.

4   Summaries

1) The fragment, whose dimension is 8mm×16mm×40mm and 40g in mass, could penetrate through both strengthened girder structure Ⅰ and the sandwich structure Ⅱ when its initial velocity is 806.5m/s。

2) The residual velocity of fragment after it penetrates through the shelter is 708.4m/s for structure Ⅰ and 762.8m/s for structure Ⅱ, the residual kinetic energy of fragment are 10728J for structure Ⅰ and 12000J for structure Ⅱ, which are far greater than the damage threshold value for person or equipment, so we should take necessary defense measure.

3) When penetrating different medium, the velocity decrease of fragment is different, it decreases most quickly in the aluminum alloy, and the wood takes the second place, then it decreases most slowly in the polyurethane foam.

References

[1] Dong, C.Y.: Technical progress of USA military shelter. Shelter and Ground Equipment (1) (2008)
[2] Dai, Y., Zhou, Z., Zhang, S., He, H., Guo, F.: Ultimate bearing capacity calculation and reinforcement analysis of square cabin under explosive impact load. Journal of Vibration and Shock 25(3) (2006)
[3] Autodyn theory manual Revision 6.1. Century Dynamics (2005)
[4] Ren, X.X.: The Structural Design and Performance Analysis of Universal Shelter with Composite Laminated Plates. Graduate School of National University of Defense Technology Changsha, Hunan (2008)
[5] Autodyn user manual Revision 6.1. Century Dynamics (2005)

Design on Simulation System
for Road Transport Service Training

Wang Chang Yong[1], Wang Wen Zheng[2], Liu Shan Shan[1], and Dai Lian Di[1]

[1] Military Transportation University
[2] Logistic Engineering University

Abstract. Taking the characters of Road Transport Service in information-based war into consideration, analyzing the necessity of road transport service training simulating system, the paper puts forward total and particular plan. At last, the simulating system of training based on road transport service validates the possibility with a concrete training item.

Keywords: Road Transport Service, Training, Simulation, Framework.

1 Introduction

Military training simulation which based on computer technology, is always applied as software, constituting a model to depict and compute the troop's moving, campaign process, damage and injury, organization and command, simulating the training item, in order to evaluate battle plan, aid in decision making, commander training[1]. Comparing with the traditional training, stimulation saves a mount of time and money, being effective and able to reduce the risk in live equipment training, therefore, it was being in military use extensively [2-3]. Road transport service training is seen as a function module in integrated training simulating system abroad, and from theory study to practical application inland [4]. But, some problems have arisen, for instance, function of training simulating system is limited, which can not meet all the demands; the simulated level has limits as well; there is gap between training and fact.

Therefore, the paper is on the basis of existing research, according to the content and demand of 《PLA land force training and exam outline (auto transport unit volume)》, taking character and rule of road military transport service in IT based warfare into mind, design and develop road transport service training stimulated system.

2 Developing Rule of System

When designing the system, the practical characters in military was taken in consideration, in order to keep the authority and quality of training, here are the rules [5]:

(1) Training content accords with the basic demand of 《PLA land force training and exam outline (auto transport unit volume)》, which has functions of cultivating commanders the ability to command troops and fit for the warfare;

(2) Training simulated circumstance must be vivid, which can serve circumstance like real warfare.

(3) Consistent information standard and data form should be used in system, and it also can format different kinds of data into standard form;

(4) System should has scientific structure and mechanism, which fit for different hardware, being reliable and durable. Interface is convenient. System is simple to manipulate;

(5) System can evaluate the students' integrated ability level of practical manipulation, and make the students know themselves;

(6) System should have compatibility, expansibility and flexibility to meet the development of training module and training itself.

3   Overall Framework

3.1   Function Design

According to 《PLA land force training and exam outline (auto transport unit volume)》 issued by headquarters recently and the purpose, road transport service training simulating system should have functions like theory and knowledge, training simulation, practical simulation and training evaluation and other assistant functions.

(1) Theory and knowledge function. It includes learning basic knowledge; professional technology; professional service knowledge; combat readiness knowledge of unit; information of railroad, waterway, aviation transport by unit; technology of manipulating correspondence equipment; transport command knowledge of unit in wartime; defense tactics knowledge of unit; transport service command knowledge of unit in non-war military operations.

(2) Training simulation function. It includes analyzing and estimating information, orders, reports and so on; taking march and transport of auto transport unit in control; service process is intervened by system at random, to train the commander to deal with paroxysmal affair; depicting commander's training decision project of all the items of topic, and the landform, weather, enemy's situation and so forth the elements which affects the service operations.

(3) Operation simulation function. It includes whole process simulation of motorized march by auto transport unit; simulation of all the items according to outline by auto transport unit; simulation of landform, weather, social circumstance and so on.

(4) Training estimation function. It mainly evaluates solutions of disposition operation and disposition project of all the items of topics; decision making estimation of dealing with paroxysmal affair is estimating possibility and rationality of solution; estimation of training effect is estimating simulation result of training items, checking whether it meets the demand of transport service or not.

3.2   System Structure

It uses structure methods to develop road transport service training simulating system. Structure means the relation among composing elements in system. Structure method emphasizes the rationality of software structure, and put forwards some rules to increase rationality, such as resolvability and abstract, independence of module,

information concealment and so on. Software structure of road transport service training simulating system depicts the structure of system by arrangement chart in design period, it is the structure module and transfer relation of compositive system.
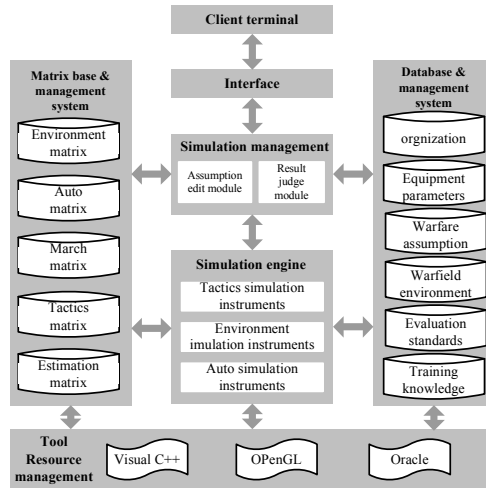
Fig. 1. Structure of road transport service training simulating system

As shown in Fig.1, road transport service training simulating system is consisted of 4 modules: simulation engine, simulation management, brace circumstance and human computer interaction interface. Besides, brace circumstance divided data brace circumstance, module brace circumstance and tool brace circumstance; human computer interaction offers a interface for teachers, students and computer. What should be cautious is there are interfaces for data communication among different modules.

4 Particular Design of System

Chapter 3 introduces simulation engine module, simulation management module and brace circumstance module in road service simulating system in detail.

4.1 Simulation Engine

Simulation engine adopts discrete event drive simulation mode, process of simulation flow is like Fig.2.

4.2 Simulation Management

Simulation management includes assumption edit module and outcome judge module.

Assumption edit module defines basic matrix, equipment parameters and so on in simulation. For instance, in an assumption, the type, number, command relation among autos and time which vehicle abates and being maintained. At the same time, assumption edit module configures road transport service training module.
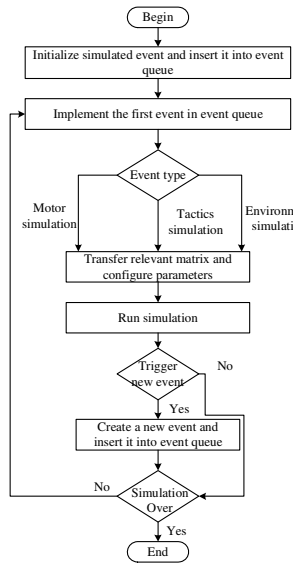
Fig. 2. System simulation engine flow

Outcome judge module offers a method to judge training outcome, for analyzing training outcome by students. There are two ways to judge the outcome: self judge and third judge. The third judge method must own director department, and the director judges the training outcome of training troops (students); self judge method is executed when students train freely and the training process comes to the end, system compares and analyzes the training outcome and the answer.

4.3   Brace Circumstance

Brace circumstance includes data brace circumstance, matrix brace circumstance and tool brace circumstance.

Data brace circumstance consists of all kinds of basic database and management system. Database is an necessary module for training simulation, includes organization database, equipment parameter database, campaign assumption database, battlefield circumstance database, estimation standard database and training knowledge base. Organization database stores military organization data, including soldiers' organization of all levels of troops, equipment configuration data and so on; equipment parameter database includes types of all equipment, performance parameters and so forth, such as type of transport truck, carrying capacity, high speed, oil cost, last maintaining time, trouble probability and so on. Campaign assumption database stores campaign assumption composed in advance, applied to training simulation in different situations. Battlefield circumstance database depicts different types of battlefield circumstance data, such as upland, mountainous region and plain and so forth. Estimation standard database shows the answer of training items, being used for director and students to estimate the training outcome. Training knowledge database stores all kinds of information computing knowledge, matrix data knowledge, battlefield situation knowledge, transport structure knowledge and so on, for students to use.

Matrix brace circumstance consists of matrix and management system. Matrix base stores all the matrixes needed in system, includes different weather, landform, march, autos, tactics and project estimation matrix and so on.

During the matrix constructing process, we consider dependability and verity the most. The process of constructing military training matrix is just like Fig.3.
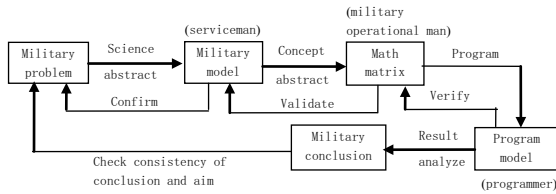
Fig. 3. Constructing matrix process figure

In order to increase the dependability and verity of the matrix, there are three transfers during the process of constructing matrix, each time it transfers, abstract, accepting and rejection, compression and predigesting process run once, all of them are impacted by students, the task of constructing system is simulating service training as similar as possible, processing the relation between idea condition and reality, at last, analyzing the result which the system computes, checking the consistency of aim and conclusion, then computing again until the conclusion is usable.

At first, according to the items of exam demands in 《PLA land force training and exam outline (auto transport unit volume)》, make sure task and aim of matrix, and participant in for service training, list the training service project and so on; Protract march line map, mark enemy & our own troops' position and road clearly; drafting combo box figure of matrix. The combo box contains training items and estimating rules, which are the basic of matrix. Constitute math matrix, includes reasoning and solving problem (analyzing method, experiment method and mix method are being used in system); make sure efficiency targets and rules; choose the best rule and condition. At last, constitute simulating logic figure, ascertain the output of combo box which can compute independently; and make certain general logic combo box and so on. Moreover, arrange the math relations among those combo boxes, ascertain part arithmetic of combo box and general arithmetic of matrix. Finally, develop the system in software engineer method with senior program language.

Tool brace circumstance can control and manage the tool software needed in road transport service training simulated software. The system is based on Visual C++ and 3D landform modeling function from OpenGL. Make use of Oracle to manage data. Thereby, tool brace circumstance is essential for system to run efficiently.

5   Example

In order to test the system, chapter 4 simulates an experiment with road transport service simulating system.

5.1   Training Items

Item：Auto unit transport organization
    War time：13:00 May 28, 2×××
    Astronomic time：13:00 May 28, 2×××
    Map：1:50000
    Background：
    Your company takes charge of transport, arrives at TaoLinzhuang warehouse (××,××), loads ××tons of ammunition, ××tons of material, ××tons of medical treatment equipment. All the material mentioned above should set off at ××: ××, along QingShan (××, ××), ZhuYuzi (××, ××), to warehouse in ZhaoCheng (××, ××), achieve there before ××: ××.
    Aim:

(1) Pretend to be the company commander, according to the order and the fact, make transport service plan, command the action.
    (2) Pretend to be the company commander, according to the paroxysmal affair (supplementary assumption,) judge the condition and process it.

5.2   Training Process

(1) Students log in the system, choose the transport task, it is always sent by order, inform, command, telephone call and so forth to auto unit.
    (2)   After receiving the task, the students analyze the unit condition, task, traffic, environment, enemy action firstly.
    (3) Drawing transport service project is the core content, which consisting of "drawing project" and "figure task" two function module, including elements such as task, organization, auto quantity and type, speed, time, load and unload position, distance and calculation and so on. In order to make sure the goods being transproted accurately in time , "auto computation" function computes time to somewhere, auto team's location at a time and march speed in limited time with math matrix, all the three methods take environment and firepower harassment in consideration.
    (4) According to transport service project, system shows the service process which logistic unit transport goods on determination plan dynamically , transfers the assumptions in database at random, sends it to the student to deal with,   gives the judgement until the mission is over. After training time and time again, students will enhance their transport organizing and command ability.

5.3   Training Effect

Road transport service training simulating system simulates the whole process of auto military combat readiness action, unit configuration, transport in wartime and other items stated in general outline, system involves elements as much as possible, the interface is understanding and easy to manipulate. At the same time, training with existing resources as much as possible makes much military and economical benefit.

6   Summaries

Road transport service training simulating system is complicated and valuable for auto military. The paper is primary production of road transport service training

simulating system, there are many problems needed to investigate and work in the system, for auto military transport service training simulation applied in troops.

The next step is extending functions based on the existing system, applying it in large scale, distributed military transport service simulation training.

References

[1] Zhang, Y.: Warfare Simulation Basic. High Education Publishing Company, Beijing (2004)

[2] Zhu, Y., Di, Y., Quan, H., Fan, S., Sun, S.: General Radar Training Simulating system and Key Technology. System Simulate Transaction 17(11) (2005)

[3] Lin, D., Zheng, X., Bao, C.: Design and Realization of Training View Simulating system in Unpiloted plane. Computer Application 27(B06), 107–108 (2007)

[4] Chen, J., Wang, X., Ming, Y.: Primary Explore of Road Transport Service Simulation Rehearsal System Based on DIS (1), 40–43 (2010)

[5] Zhu, X., Jiang, N.: Analysis and Design of Equipment Command Training Simulating system. In: System Simulation Technology and Application Communication Conference. Chinese Science and Technology University Publishing University, Hefei Anhui (2006)

$[r,s,t]$ – Colouring of One Kind of Join Graphs

Mo Ming-Zhong and Pan Yu-Mei

Department of Mathematics and Computer Science, Liuzhou Teachers College, Liuzhou, Guangxi, 545004, P.R. China
{momingzh,pym2003168}@163.com

Abstract. The concept of [r,s,t]-colourings was introduced by A. Kemnitz and M.Marangio in 2007 as follows: Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and $E(G)$. Given non-negative integers r, s and t, an [r,s,t]-colouring of a graph $G = (V(G), E(G))$ is a mapping C from $V(G) \cup E(G)$ to the colour set $\{0, 1, 2, \cdots, k-1\}$ such that $|c(v_i) - c(v_j)| \geq r$ for every two adjacent vertices v_i, v_j, $|c(e_i) - c(e_j)| \geq s$ for every two adjacent edges e_i, e_j, and $|c(v_i) - c(e_j)| \geq t$ for all pairs of incident vertices and edges, respectively. The [r,s,t]-chromatic number $\chi_{r,s,t}(G)$ of G is defined to be the minimum k such that G admits an [r,s,t]-colouring. In this paper, we determine the [r,s,t]-chromatic number for join graphs $O_m + C_n$.

Keywords: Empty graph, Cycle, Join graphs, [r, s, t]-colouring, [r,s,t]-chromatic number.

1 Introduction

The graphs we shall consider are finite, simple and undirected, unless stated otherwise, we follow the notations and terminologies in [1,2]. Let G be a graph. We denote its vertex set, edge set, minimum degree, maximum degree, order, size by $V(G), E(G)$, $\delta(G)$, $\Delta(G)$, $p(G)$ and $q(G)$, respectively, and vertex chromatic number, edge chromatic number and total chromatic number by $\chi(G)$, $\chi'(G)$ and $\chi_T(G)$ respectively. A (p,q) graph has order p and size q.

The join $G + H$ [3] of two disjoint graphs G and H is the graph having vertex set $V(G) \cup V(H)$ and edge set $E(G) \cup E(H) \cup \{xy| x \in V(G), y \in V(H)\}$.

Following the conception of vertex coloring, edge coloring and total coloring, in 2007, A. Kemnitz and M. Marangio[4] put forward the concept of [r,s,t]-coloring of graphs.

Let $G = (V(G), E(G))$ be a graph with vertex set $V(G)$ and $E(G)$. Given non-negative integers r, s and t, an [r,s,t]-colouring of a graph $G = (V(G), E(G))$ is a mapping C from $V(G) \cup E(G)$ to the colour set $\{0, 1, 2, \cdots, k-1\}$ such that $|c(v_i) - c(v_j)| \geq r$ for every two adjacent vertices v_i, v_j, $|c(e_i) - c(e_j)| \geq s$ for every two adjacent edges e_i, e_j, and $|c(v_i) - c(e_j)| \geq t$ for all pairs of incident vertices and

edges, respectively. The [r,s,t]-chromatic number $\chi_{r,s,t}(G)$ of G is defined to be the minimum k such that G admits an [r, s, t]-colouring.

It is a generalization of classical vertex coloring, edge coloring and total coloring of a graph, which has broad significant applications.

In [4], A. Kemnitz and M. Marangio gave the [r,s,t]-chromatic number of general graphs of the boundary and some related properties, also discussed the [r, s, t]-chromatic number of the complete graph of order n. Other results on [r,s,t]-colouring are presented in [5, 6]. In our paper, we determine the [r,s,t]-chromatic number for join graphs $O_m + C_n$.

2   Useful Results

We begin by stating some useful results from the literature. A. Kemnitz and M. Marangio [4] have proved the following properties for $[r, s, t]-$colourings.

Lemma 1. If $H \subseteq G$ then

$$\chi_{r,\,s,\,t}(H) \leq \chi_{r,s,t}(G)$$

Lemma 2. If $r' \leq r, s' \leq s, t' \leq t$ then

$$\chi_{r',s',\,t'}(G) \leq \chi_{r,s,t}(G)$$

Lemma 3.   For the $[r,\,s,\,t]-$chromatic number of a graph G, there holds

$$\max\{\, r(\chi(G)-1)+1,\, s(\chi'(G)-1)+1,\, t+1 \} \leq \chi_{r,s,t}(G) \leq r(\chi(G)-1)+s(\chi'(G)-1)+t+1$$

According to the definition of chromatic number, chromatic index and join graphs, we can get the conclusion easily:

Lemma 4.   For the join graphs $O_m + C_n$, there holds

(1) $\chi(O_m + C_n) = \begin{cases} 3 & n \text{ is even}; \\ 4 & n \text{ is odd}. \end{cases}$

(2) $\chi'(O_m + C_n) = \begin{cases} \Delta+1, & n = m+1; \\ \Delta, & \text{others}. \end{cases}$

3   The Main Results

Corollary 1. For join graphs $O_m + C_n$, it holds

(1)   If n is even and $n \neq m+1$, then

$\max\{2r+1,\, s(\Delta-1)+1,\, t+1\} \leq \chi_{r,s,t}(O_m + C_n) \leq 2r + s(\Delta-1) + t + 1$;

(2)   If n is even and $n = m+1$, then

$\max\{\, 2r+1,\, s\Delta+1,\, t+1\} \leq \chi_{r,s,t}(O_m + C_n) \leq 2r + s\Delta + t + 1$;

(3)   If n is odd and $n \neq m+1$, then

$\max\{\, 3r+1,\, s(\Delta-1)+1,\, t+1\} \leq \chi_{r,s,t}(O_m + C_n) \leq 3r + s(\Delta-1) + t + 1$;

(4)   If n is odd and $n = m+1$, then

$\max\{\, 3r+1,\, s\Delta+1,\, t+1\} \leq \chi_{r,s,t}(O_m + C_n) \leq 3r + s\Delta + t + 1$.

Proof. By Lemma 3 and Lemma 4, the conclusion is obviously.

Theorem 1. Let $O_m + C_n$ be a join graphs, n is even, if one of the following conditions holds, then, $\chi_{r,s,t}(O_m + C_n) = 2r + 1$.

   (1) $n \geq m + 2$, and $r \geq (n-1)s + 2t$;

   (2) $n = m + 1$, and $r \geq (n+1)s + 2t$;

   (3) $n \leq m$, and $r \geq (m+1)s + 2t$.

Proof. Let O_m is an empty graph and its vertex set $V(O_m) = \{v_1, v_2, \cdots, v_m\}$, cycle C_n vertex set $V(C_n) = \{u_1, u_2, \cdots, u_n\}$, edge set $E(C_n) = \{u_i u_{i+1} \mid i = 1, 2, \cdots, n; \ if \ i+1 > n \ then \ (i+1) \bmod n \ take \ place \ it\}$·

   (1)   If $n \geq m + 2$, then $\Delta = \Delta(O_m + C_n) = \max\{n, m+2\} = n$, it holds $\chi'(O_m + C_n) = \Delta = n$ by Lemma 4. Let c be a coloring of $O_m + C_n$ as follows:

  Colour the vertices of join $O_m + C_n$ with the $\chi(G) = 3$ colours 0, r, $2r$ as follows,

$$c(v_i) = 2r, \quad i = 1, 2, \cdots, m;$$
$$c(u_i) = \begin{cases} 0, & i \equiv 1 \bmod 2 \\ r, & i \equiv 0 \bmod 2 \end{cases}, \quad i = 1, 2, \cdots, n$$

and colour the edges of join $O_m + C_n$ with the $\chi'(O_m + C_n) = \Delta = n$ colours to obtain a proper $[r,s,t]$−colouring of $O_m + C_n$, hence $\chi_{r,s,t}(O_m + C_n) \leq 2r + 1$, by Corollary 1 (1) we have $\chi_{r,s,t}(O_m + C_n) \geq 2r + 1$, then $\chi_{r,s,t}(O_m + C_n) = 2r + 1$·

   (2)   If $n = m + 1$, then $\Delta = \Delta(O_m + C_n) = \max\{n, m+2\} = m + 2 = n + 1$, it holds $\chi'(O_m + C_n) = \Delta + 1 = n + 2$ by Lemma 4. Let c be a coloring of $O_m + C_n$ as follows:

    Firstly, colour the vertices of $O_m + C_n$ with three colours 0, r, $2r$ as follows:

$$c(v_i) = 2r, \quad i = 1, 2, \cdots, m;$$
$$c(u_i) = \begin{cases} 0, & i \equiv 1 \bmod 2 \\ r, & i \equiv 0 \bmod 2 \end{cases}, \quad i = 1, 2, \cdots, n$$

Secondly, colour the edges of $O_m + C_n$ with the $n + 2$ colours $t, t+s, t+2s, \cdots, t+ns, t+(n+1)s$ as follows,

$$c(v_i u_j) = t + [(i + j - 2) \bmod n]s, \quad i = 1, 2, \cdots, m; \quad j = 1, 2, \cdots, n.$$
$$c(u_i u_{i+1}) = \begin{cases} t + ns & i \equiv 1 \bmod 2 \\ t + (n+1)s & i \equiv 0 \bmod 2 \end{cases}, \quad i = 1, 2, \cdots, n \quad , \quad \text{if suffix}\quad i + 1 > n \quad \text{then}$$

$(i+1) \bmod n$ takes place $i + 1$.

   Since $\chi'(O_m + C_n) = \Delta = n$, hence the edge colouring c is proper, so we abtain a proper $[r,s,t]$−colouring of $O_m + C_n$, hence $\chi_{r,s,t}(O_m + C_n) \leq 2r + 1$. At the same time, it holds $\chi_{r,s,t}(O_m + C_n) \geq 2r + 1$ by Corollary 1 (2),   therefore $\chi_{r,s,t}(O_m + C_n) = 2r + 1$.

(3) If n is a even and $n \leq m$, then $\Delta = \Delta(O_m + C_n) = \max\{n, m+2\} = m+2$, it holds $\chi'(O_m + C_n) = \Delta = m+2$ by Lemma 4. Let c be a coloring of $O_m + C_n$ as follows:

Firstly, colour the vertices of $O_m + C_n$ with three colours $0, r, 2r$ as follows:

$$c(v_i) = 2r, \quad i = 1, 2, \cdots, m;$$
$$c(u_i) = \begin{cases} 0, & i \equiv 1 \bmod 2 \\ r, & i \equiv 0 \bmod 2 \end{cases}, \quad i = 1, 2, \cdots, n \cdot$$

Secondly, colour the edges of $O_m + C_n$ with the $m+2$ colours $t, t+s, t+2s, \cdots, t+(m+1)s$, since $\chi'(O_m + C_n) = \Delta = m+2$, hence the edge colouring c is proper, then obtain a proper $[r, s, t]$–colouring of $O_m + C_n$, hence $\chi_{r,s,t}(O_m + C_n) \leq 2r+1 \cdot$ At the same time, it holds $\chi_{r,s,t}(O_m + C_n) \geq 2r+1$ by Corollary 1 (1), therefore $\chi_{r,s,t}(O_m + C_n) = 2r+1 \cdot$

Theorem 2. Let $O_m + C_n$ be a join graphs, n is odd, if one of the following conditions holds, then $\chi_{r,s,t}(O_m + C_n) = 3r+1$.

(1) $n \geq m+2$ and $r \geq (n-1)s + 2t$;

(2) $n = m+1$ and $r \geq (n+1)s + 2t$;

(3) $n \leq m$ and $r \geq (m+1)s + 2t$.

Proof. Let empty graph O_m vertex set $V(O_m) = \{v_1, v_2, \cdots, v_m\}$, cycle C_n vertex set $V(C_n) = \{u_1, u_2, \cdots, u_n\}$, edge set $E(C_n) = \{u_i u_{i+1} \mid i = 1, 2, \cdots, n; if\ i+1 > n\ then\ (i+1) \bmod n\ take\ place\ it\}$. If n is odd, by Lemma 4 we have $\chi(O_m + C_n) = 4$.

(1) If $n \geq m+2$, then $\Delta = \Delta(O_m + C_n) = \max\{n, m+2\} = n$, by Lemma 4 we have $\chi'(O_m + C_n) = \Delta = n$. Let c be a coloring of $O_m + C_n$ as follows:

Firstly, colour the vertices of $O_m + C_n$ with four colours $0, r, 2r, 3r$ as follows:

$$c(v_i) = 3r, \quad i = 1, 2, \cdots, m;$$
$$c(u_i) = \begin{cases} 0, & i \equiv 1 \bmod 2, \quad i = 1, 2, \cdots, n-1, \\ r, & i \equiv 0 \bmod 2, \quad i = 1, 2, \cdots, n-1, \\ 2r, & i = n \end{cases}$$

Secondly, colour the edges of $O_m + C_n$ with the n colours $t, t+s, t+2s, \cdots, t+(n-1)s$, since $\chi'(O_m + C_n) = \Delta = n$ to obtain a proper $[r, s, t]$– colouring of $O_m + C_n$, hence $\chi_{r,s,t}(O_m + C_n) \leq 2r+1$ .And it holds $\chi_{r,s,t}(O_m + C_n) \geq 3r+1$ by Corollary 1 (3), then $\chi_{r,s,t}(O_m + C_n) = 3r+1$.

(2) If $n=m+1$ and n is odd, then $\Delta=\Delta(O_m+C_n)=\max\{n, m+2\}=m+2=n+1$, by Lemma 4 we have $\chi'(O_m+C_n)=\Delta+1=n+2$. Let c be a coloring of O_m+C_n as follows:

Firstly, colour the vertices of O_m+C_n with four colours $0, r, 2r, 3r$ as follows:

$$c(v_i)=3r, \quad i=1,2,\cdots,m;$$

$$c(u_i)=\begin{cases} 0, & i\equiv 1\bmod 2, \quad i=1,2,\cdots,n-1, \\ r, & i\equiv 0\bmod 2, \quad i=1,2,\cdots,n-1, \\ 2r, & i=n \end{cases}$$

Secondly, colour the edges of O_m+C_n with the $n+2$ colours $t, t+s, t+2s,\cdots,t+ns, t+(n+1)s$ to obtain a proper $[r,s,t]-$ colouring of O_m+C_n , hence $\chi_{r,s,t}(O_m+C_n)\le 3r+1$, by Corollary 1 (4) we have $\chi_{r,s,t}(O_m+C_n)\ge 3r+1$, then $\chi_{r,s,t}(O_m+C_n)=2r+1\cdot$

(3) If n is odd and $n\le m$, then $\Delta=\Delta(O_m+C_n)=\max\{n, m+2\}=m+2$, by Lemma 4 we have $\chi'(O_m+C_n)=\Delta=m+2$, Let c be a coloring of O_m+C_n as follows:

Firstly, to colour the vertices of O_m+C_n with four colours $0, r, 2r, 3r$ as follows:

$$c(v_i)=3r, \quad i=1,2,\cdots,m;$$

$$c(u_i)=\begin{cases} 0, & i\equiv 1\bmod 2, \quad i=1,2,\cdots,n-1, \\ r, & i\equiv 0\bmod 2, \quad i=1,2,\cdots,n-1, \\ 2r, & i=n \end{cases}$$

Secondly, to colour the edges of O_m+C_n with the $m+2$ colours $t, t+s, t+2s,\cdots,t+(m+1)s$, since $\chi'(O_m+C_n)=\Delta=m+2$, hence the edge colouring c is proper, and $r-(t+(m+1)s)\ge t$, so c is a proper $[r,s,t]-$ colouring, hence $\chi_{r,s,t}(O_m+C_n)\le 3r+1$, by Corollary 1 (3) we have $\chi_{r,s,t}(O_m+C_n)\ge 3r+1$, then $\chi_{r,s,t}(O_m+C_n)=3r+1$.

Theorem 3. If $n(\ge 3)$ is even and $s\ge\max\{r, 2t\}$, then

$$\chi_{r,s,t}(O_m+C_n)=(\chi'(O_m+C_n)-1)s+1\cdot$$

Proof. If $n(\ge 3)$ is even, by Lemma 4 we have $\chi(O_m+C_n)=3$, and C_n is a even cycle, then $n\ge 4$, hence

$$\chi'(O_m+C_n)-1\ge\Delta-1=\max\{n, m+2\}-1\ge 4-1=3$$

Let c be a coloring of O_m+C_n as follows:

Firstly, colour the edges of O_m+C_n with the $\chi'(O_m+C_n)$ colours $0, s, 2s,\cdots,(\chi'(O_m+C_n)-1)s$, we can make c is proper edge colouring.

Secondly, to colour the vertices of $O_m + C_n$ with three colours $t, t+s, t+2s$ as follows:

$$c(v_i) = t + 2s, \quad i = 1, 2, \cdots, m;$$

$$c(u_i) = \begin{cases} t, & i \equiv 1 \bmod 2 \\ t+s, & i \equiv 0 \bmod 2 \end{cases}, \quad i = 1, 2, \cdots, n.$$

Since $s \geq \max\{r, 2t\}$ we have $(t+s) - t = s \geq r$, $(t+2s) - t = 2s \geq r$, $(t+2s) - (t+s) = s \geq r$, $|ks - t| \geq t$, $|ks - (t+s)| \geq t$, $|ks - (t+ss)| \geq t$, which $k = 0, 1, 2, \cdots, \chi'(O_m + C_n)$, so c is a proper $[r, s, t]$– colouring, hence $\chi_{r,s,t}(O_m + C_n) \leq (\chi'(O_m + C_n) - 1)s + 1$, by Lemma 3 we have $\chi_{r,s,t}(O_m + C_n) \geq (\chi'(O_m + C_n) - 1)s + 1$, then $\chi_{r,s,t}(O_m + C_n) = (\chi'(O_m + C_n) - 1)s + 1$.

Theorem 4. Let $O_m + C_n$ be a join graphs, if $s \geq r + 2t$, then

$$\chi_{r,s,t}(O_m + C_n) = (\chi'(O_m + C_n) - 1)s + 1 \cdot$$

Proof. Let c be a coloring of $O_m + C_n$ as follows:

Firstly, to colour the edges of $O_m + C_n$ with the $\chi'(O_m + C_n)$ colours $0, s, 2s, \cdots, (\chi'(O_m + C_n) - 1)s$, we can make c is proper edge colouring.

Secondly, to colour the vertices of $O_m + C_n$ with the following 4 colours selected as follows.

In the interval $(0, s)$ is to select $t, r+t$, and in the interval $(s, 2s)$ is to select $\max\{s+t, 2r+t\}$, $\max\{s+r+t, 3r+t\}$ these 4 colours, by Lemma 4 we have $\chi(O_m + C_n) \leq 4$, so we can make c is proper vertices colouring, c is a proper $[r, s, t]$–colouring, hence $\chi_{r,s,t}(O_m + C_n) \leq (\chi'(O_m + C_n) - 1)s + 1$, by Lemma 3 we have $\chi_{r,s,t}(O_m + C_n) \geq (\chi'(O_m + C_n) - 1)s + 1$, hence

$$\chi_{r,s,t}(O_m + C_n) = (\chi'(O_m + C_n) - 1)s + 1 \cdot$$

4   Summaries

In this paper, we give the exact values of the [r, s, t]- chromatic number of the joint graph $O_m + C_n$ which depends on the parities of m and n, if r, s, t meet certain conditions.

References

[1] Bondy, J.A., Murty, U.S.R.: Graph theory. Springer, Berlin (2008)

[2] Chartrand, G., Lesniak, L.: Graphs and digraphs. Greg Hubit Bookworks, California (1979)

[3] Yap, H.P.: Total coloring of graphs. Lecture Notes in Mathematics, vol. 1623, pp. 1–11. Springer, Heidlberg (1996)

[4] Kemnitz, A., Marangio, M.: [r, s, t]- colorings of graphs. Discrete Math. 307, 199–207 (2007)

[5] Dekar, L., et al.: [r, s, t]-colorings of trees and bipartite graphs. Discrete Math. 310, 260–269 (2010)

[6] Xu, C., Ma, X., Hua, S.: [r, s, t]-coloring of $K_{n,n}$. J. Appl. Math. Comput. 31, 45–50 (2009)

Establish 3D Digital Elevation Model Based on SVG

Lin Wei-Yong[1] and Li Yan[2,*]

[1] School of Computer, South China Normal University, Guangzhou, Guangdong 510631
[2] Spatial Information Research Center, South China Normal University, Guangzhou,
Guangdong 510631
yanli@scnu.edu.cn

Abstract. This paper mainly discusses how to use Scalable Vector Graphics
(SVG) to create a Digital Elevation Model (DEM) and to animate it online; It
introduces an extended SVG element <replicate>, principle, the improved
interpolation algorithm with a triangulation network and the method to establish
a 3D DEM. Finally, two approaches of 3D DEM creation are compared, and the
experimental results showed that both of them have good visualizing effects but
different time cost for online displaying. Therefore, it is necessary to expand the
<replicate> as a basic SVG element to create 3D models and to adapt it for
creating 3D elements instead of using JS to realize 3D effect.

Keywords: SVG, <replicate>, 3D, DEM.

1   Introduction

SVG, developed by the W3C (World Wide Web Consortium) organization, is a new
generation of two-dimensional Vector Graphics. It is based on XML and used to
describe the hybrid Graphics Vector/bitmap. It is a kind of brand-new Vector
Graphics norms, and a kind of completely open two-dimensional vector data format.
Now it has been got strong supports from international well-known software vendors
especially Microsoft, Google and Netscape. It provides many basic graphic objects
and the precise control of the coordinate system, each graphic element can have their
own parameters and attributes, these characteristics make it effective to store image
data and convenient to express and display information.

However, SVG cannot support 3D model, SVG objects once define will fix on a
certain position, this goes against the establishment of 3D model. With the constantly
expand and thorough of 3D modeling research and application in all fields, and the
unceasing development of SVG technology, using SVG for the 3D modeling of
geographical entity will become important up. The current solutions are mostly by
writing JS to operate SVG objects, and then display simple 3D models. Lutz
Tautenhahn a German wrote a free JAVASCRIPT script library SVG-VML-3D [1],
this script can be according to the browser to select SVG or VML to display 3D
objects automatically. In IE VML will be used while in other browsers SVG. This
script supports 3D objects displaying, scaling, rotating, and simple illumination
processing and mouse event interaction. Jun Fujisawa and Anthony Grasso realize
2.5D effect in SVG tiny, mainly for the design of mobile GUI. They reach to the

---

effect by changing the values of "scale" and "translation" attributes of the SVG objects[2]. And David Dailey put forward a SVG extended element <replicate>, through this element objects can reproduce a number of new objects to reach 3D effect, however, this method can only handle objects that have the same number of points and proportional spacing [3]. Predecessors have done research on 3D modeling theory and technology based on SVG and have laid a foundation for the establishment of 3D model based on SVG, but less of these research achievements have applied to the establishment of 3D DEM. In order to solve this problem, this paper proposed a improved method of the <replicate> element for creating 3D DEM. It can be used to establish 3D DEM with practical dataset of contours which met inconsistent number of points and irregular spacing. And the DEM can also be animated the rotation in the multiple view angles.

2   Theory and Methods

SVG standard[4] defines 80 various elements of 17 classess. They involved basic graphics, text, images, and graphic animation elements, incident handling, hyperlinks, color gradient, transparency, filter effects and editing processes, synthesis and filling, etc. And SVG also support scripting language. Through script programming, SVG DOM elements and attributes can be visited and responded to a particular event, this can improve the dynamic and interactive performance on SVG. So, we can use SVG to design the various elegant, interactive vector graphic.

2.1   Definition of DEM

DEM is a orderly set used to describe the spatial location and terrain attribute distribution of surface unit, it is based on terrain data in a geographical scope, with mathematics transform and 3D forms such as the terrai, geophysics to display in 2D space. It is a data set within the scope of certain areas on the plane coordinates rules lattice outlets (x, y) and elevation (z) or latitude, longitude and altitude (h)[5]. Precision DEM can not only very vividly reveals a district topography and landforms, but also provides basic data for quantitative analysis of various terrain features and automatic drawing of different types of special charts.

2.2   Principal of SVG <Replicate> in Creating 3D

SVG <replicate> element is an extended SVG element put forward by David Dailey, which has a similar grammar rules with SMIL. The child node of <replicate> element named <replicateAttribute> is always with "from " and "to" attributes, and the JS can support to analyze the values of "from" and "to" attribute and calculate intermediate objects' attribute value using interpolation algorithm, after that the intermediate objects are copied to reach 3D effect.

The source code of a SVG demo using <replicate> element is showed below, and the result is show in Fig.1, among which (A) (B) (C) is corresponding to the "repeatCount" attribute value is 5, 10 and 20.

```
<svg xmlns="http://www.w3.org/2000/svg" xmlns:xlink="http://www.w3.org/
1999/xlink" xmlns:svgr="http://granite.sru.edu/svgr" onload="init(evt)"
width="100%">
<script type="text/javascript" xlink:href="repDD0.js" />
```

```
<path d="M -10 -10 -20 -20 z" fill-opacity=".4" stroke="black"stroke-width=".3"
stroke-opacity=".3">
<svgr:replicate repeatCount="20" >
<svgr:replicateAttribute attributeName="d"from="M 411 73, 487 201, 420 376,
217 240 z" to="M 511 73, 587 201, 520 376, 317 240 z"      />
<svgr:replicateAttribute attributeName="fill"   from="rgb(45,45,128)"to="rgb
(0,233,128)" />
 </replicate></path></svg>
```
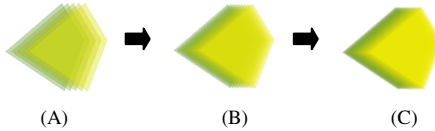

Fig. 1. Creating 3D model with <replicate>

2.3   Approach to Improve <Replicate> in Creating DEM

From the clue of the result mentioned in Fig.1, it means that <replicate> element can well realize 3D DEM effect. The result of 3D DEM is showed that 6 or 15 intermediate objects are inserted into the original object,   and the effect is good enough for creating DEM in Fig.2. Using another word, it is a possible approach to create 3D DEM and will be effectively and especially in establishing DEM.

However, there are some problems in which the "from" and "to" attributes of <replicateAttribute> element should have the same number of items after split into arrays, but the objects needed to be deal with in real most of the time have different number of items after split into arrays. In order to solve this problem, we try to use a triangulation network interpolation algorithm [6][7] to improve the interpolation algorithm of <replicate> element, and apply the method to the establishment of a practical 3D DEM.

Fig. 2. Example of path copy

2.3.1   Triangulation Network Interpolation Algorithm

The already by known data of inserting paths between two paths in SVG are the coordinates of these two paths, if apply to the establishment of DEM, then the elevation values of the contours will be also included. To realize the interpolation between two paths, the key lies in the establishment of the auxiliary line segment used to interpolate other paths, it request the auxiliary lines respectively in two vertices of the known paths, what's more, auxiliary lines can neither intersect with each other, nor intersect with the known path. When establish auxiliary lines, the algorithm using a method similar with the triangulation constructing between the known paths. Take the following chart as an example, the processes are described below:
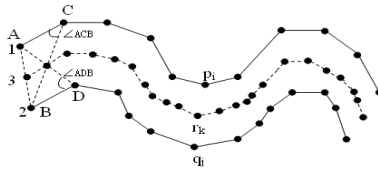
Fig. 3. Establishment of interpolated auxiliary lines

Suppose the known paths are path1 and path2, path1 is connected by p_1, p_2... p_m, and path2 is connected by q_1, q_2... q_n, and H paths will be inserted evenly. Then: ① Assign point p_1 to A, initialize variable i as 1; Assign point q_1 to B, initialize variable j as 1; Insert the first equivalent point in the line AB and initialize variable k as 1. The coordinate (x_{r1}, y_{r1}) of the first equivalent point r_1 of hth path can be calculated as followed:

$$
\begin{aligned}
&\alpha = h / (H+1) \\
&x_{r1} = x_A + \alpha \times (x_B - x_A) \\
&y_{r1} = y_A + \alpha \times (y_B - y_A) \\
&k = k+1
\end{aligned}
\tag{1}
$$

②if i<m, set point C as p_{i+1} ; else if j<n, set point D as q_{j+1} ;
      if($\angle$ACB>$\angle$ADB){
      Point C assigned to pointA(A=C);i=i+1;}else{
      Point D assigned to point B(B=D);j=j+1;}

Insert the kth equivalent point in the line AB，The coordinate (x_{rk}, y_{rk}) of the kth equivalent point r_k of hth path can be calculated as followed:

$$
\begin{aligned}
&x_{rk} = x_A + \alpha \times (x_B - x_A) \\
&y_{rk} = y_A + \alpha \times (y_B - y_A) \\
&k = k+1
\end{aligned}
\tag{2}
$$

③if i=m or j=n, then turn to ④, or return to ②
    ④if i<m, assign point p_{i+1} to A, use formula (2) to insert equivalent point in line AB and set i=i+1 until i=m; if j<n, assign point q_{j+1} to B, use formula (2) to insert equivalent point in line AB and set j= j+1 until j=n.

The algorithm shown above can insert H (here H is 1) paths between two paths which have different number of points (path1 contains 11points while path2 contains 13 points), and the dotted line shown in Fig.3 is the result.

2.3.2   Method for 3D Visualization

In order to check the 3D models from different angles, it needed to perform some rotation transformations. In 3D coordinate system, the rotations of an object can be decomposed to rotation around x-axis, y-axis, and z-axis. Suppose the coordinates of a 3D point before translate is (x', y', z'), and after the translation the coordinate is

(x'', y'', z''). Now let's turn the model around the z-axis by an angle of Φ. What will be the new position of after that? As it is easy to see, the z-component doesn't change, so we get z''=z'. But what about x' and y' ? When we look from the top of the coordinate system, we can see the model in view angle of Fig.4:
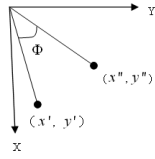
Fig. 4. (x',y',z') around the z-axis rotate Φ degrees

It seems to be difficult to find a formula which calculates x'' and y'' from x', y' and Φ, but when we do a separate calculation for x' and y', it gets much easier such as in Fig.5.

$$x'' = x' \times \cos(\Phi)$$
$$y'' = x' \times \sin(\Phi)$$

$$x'' = -y' \times \sin(\Phi)$$
$$y'' = y' \times \cos(\Phi)$$
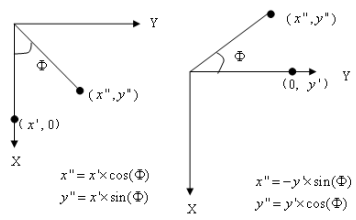
Fig. 5. Rotate separation

Now we can add the components for x, y and z and write it down properly:

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} \cos(\Phi) & -\sin(\Phi) & 0 \\ \sin(\Phi) & \cos(\Phi) & 0 \\ 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{3}$$

Similarly, we can get the formula (3) turning all the stuff additionally around the x-axis, y-axis by an angle of Φ:

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos(\Phi) & -\sin(\Phi) \\ 0 & \sin(\Phi) & \cos(\Phi) \end{bmatrix} * \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} x'' \\ y'' \\ z'' \end{bmatrix} = \begin{bmatrix} \cos(\Phi) & 0 & -\sin(\Phi) \\ 0 & 1 & 0 \\ \sin(\Phi) & 0 & \cos(\Phi) \end{bmatrix} \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} \tag{5}$$

3   Implementation and Results

3.1   Implementing Procedure

From the source code of 2.2, we know that there is only one object, a <path>, but the result has a number of objects. The principle to realize this effect is through the "Clone" function to reproduce a new object, then through the "Append" or "Insert" function to add the new object to the SVG file. And the realization process is shown in Fig.6 as follows:
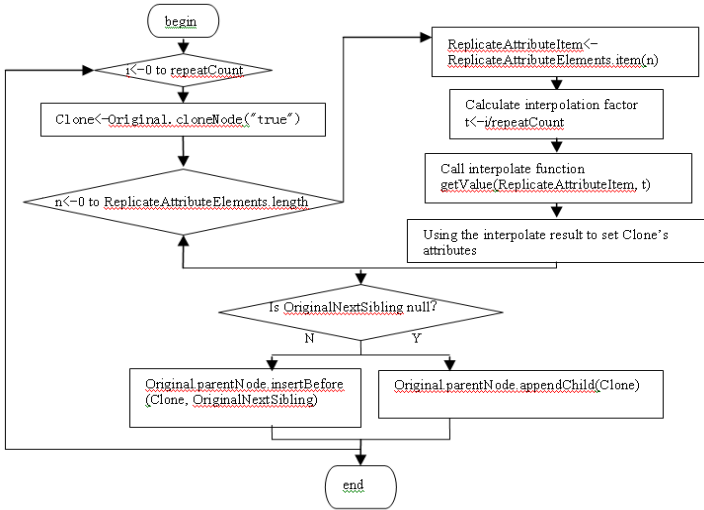
Fig. 6. Copy process of improved <replicate> element

As SVG is based on XML, it is easy to use the API of DOM to operate the SVG file. First get all the <replicate> elements, then traversal all these elements. Every element calls "replicate(ReplicateItem)" function to reproduce copies, and the process of the function is as below:

1) Initialization process
    ① Get the parent node of "ReplicateItem" and assign to the variable named "Original";
    ② Get the sibling nodes of "Original " and assign to the variable named "OriginalNextSibling";
③ Remove <replicate> element;
    ④ Get all the <replicateAttribute> elements and assign to the variable named "ReplicateAttributeElements";
    ⑤ Get the value of the "repeatCount" attribute of <replicate> element and assign to variable named " repeatCount ".
2) Copy process

As showed in Fig.6, new objects can be reproduced through the process, but how can it to reproduce new objects different from the original? The answer is by the interpolate function. From the example of 2.2, we know that every <replicateAttribute> element has "from" and "to" attributes. The interpolate function first get these two values and split them into arrays, then get the array items that have the same subscript, if the item is a num, then call the interpolate algorithm to calculate the interpolation, if not just append the item to the result. Suppose after split the two arrays are "fromArray" and "toArray", "subN" is the subscript, "repeatCount " is the total interpolation number, and "t" is the interpolation factor which is the result of the value of "i" divided by the value of "repeatCount", and the result is "interpolateResult", then the formula to calculate the ith interpolation is as followed:

interpolateResult = fromArray[subN] + (toArray[subN] - fromArray[subN])*t;

Every <replicateAttribute> element also has "attributeName" attribute, the value of which is set as one attribute of "Clone" object, and the interpolate result is set as the attribute value. If the attribute value of "attributeName" is related to shape or color, then the copied objects will have new shapes and colors, and this is why the new objects different from the original.

3.2   The Results

3.2.1   Online Creating 3D Model

Utilizing the core algorithm showed above and processing some abnormalities in practical applications and then good results of the 3D DEM have been achieved and showed in Fig.7.
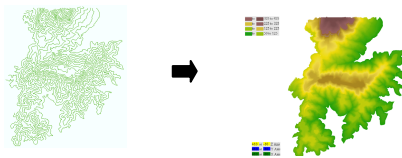
Fig. 7. 3D DEM result created from contour lines in SVG

3.2.2   3D Rotation Transformation

After that adding some buttons to control the view angels of the model in SVG, and through the click event of the buttons to control the rotation of the model. After rotate translation, the result of the model established in 3.2.1 is showed in Fig.8.
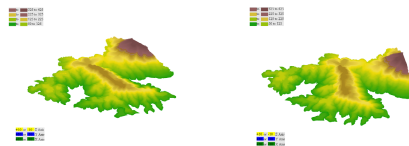
Fig. 8. DEM at two view angels

4   Discussion

Now <replicate> element can realize 3D effect through embedded JS in SVG file, after the loading of the SVG, the JS function begin to reproduce the copies of the original objects. Here approaches of the generation of 3D DEM by JS and JAVA are compared in order to find out a more sutiable way to generate 3D DEM on SVG. Tab.1 shows the time consumption of these two approaches.

Experimental environment: 1.CPU: Intel(R) Pentium(R) 4 Dicaryon; 2. CPU Clock Speed: 2.80GHz and 2.81GHz; 3.Memory: 1.5GB

Table 1. Time consumption of the two diffierent approaches

Approach	Number of contours	Number of points	Number of inserted contours	Number of inserted points	time consumption (unit: S)
JS	62	16816	298	178239	135.93
	33	13825	213	135792	126.21
	18	9256	115	83274	92.28
JAVA	62	16816	298	178239	43.63
	33	13825	213	135792	37.41
	18	9256	115	83274	27.87

5   Summaries

From the result of the 3D DEM on SVG, we can see that the effect of 3D DEM is good enough for practical use, so it is feasible to display 3D DEM on SVG. And from Tab.1 we know that when the number of the inserted points are the same, it is far more fast to generate 3D DEM on SVG by JAVA than by JS, but now SVG can only through embeding JS to generate 3D DEM real-time, in order to display 3D DEM on SVG more effective and real-time, it is important to   expand SVG to adapt to the establishment of 3D DEM.

References

[1] Lutz, T.: SVG-VML-3D 1.3 (November 2006),
    http://www.lutanho.net/svgvml3d/index.html
[2] Jun, F., Grasso, A.: Achieving 3D Effects with SVG. In: SVG Open 2008, Nuremberg, Germany, August 26-28 (2008)
[3] Dailey, D., Elder, E.J.: A proposal for adding declarative drawing to SVG. In: SVG Open 2009, October 2-4. Mountain View, CA (2009)
[4] W3C. Scalable Vector Graphics(SVG)1.1 Specification (EB/OL) (January 14, 2003 /July 5, 2005), http://www.w3.org/TR/SVG

[5] Li, Z.L., Zhu, Q.: Digital elevation model. Wuhan University Press, Wuhan (2001)
[6] Hu, W.-M., Wu, B., Ling, H.-B.: An Automatic Method for Contour Interpolation in Map Design. Chinese J. Computers (08), 847–851 (2000)
[7] Zeng, T.X.: Partial Contour Interpolation Algorithm Based on Field Analysis. Chengdu University (2009)

The Asymptotic Equipartition Property
for Nonhomogeneous Markov Chains Indexed by Trees

Peng Weicai and Chen Peishu

Department of Mathematics
Chaohu University, Chaohu, Anhui
238000, P.R. China

Abstract. In this paper, the tree T is a general tree. We prove the strong law of large numbers and the asymptotic equipartition property (AEP) for finite nonhomogeneous Markov chains indexed by trees. The results generalize some known results.

Keywords: Nonhomogeneous Markov chain, Tree, Strong law of large numbers.

1   Introduction

By a tree T we mean an infinite, locally finite, connected graph with a distinguished vertex o called the root and without loops or cycles. We only consider trees without leaves. That is, the degree of each vertex (except o) is required to be at least 2. Let σ, τ be vertices of a tree. Write $\tau \leq \sigma$ if τ is on the unique path connecting o to σ, and $|\sigma|$ for the number of edges on this path. For any two vertices σ, τ, denote by $\sigma \wedge \tau$ the vertex farthest from o satisfying

$$\sigma \wedge \tau \leq \sigma, \sigma \wedge \tau \leq \tau$$

The set of all vertices with distance n from the root o is called the n-th generation of T, which is denoted by L_n, $L_0 = \{o\}$. We denote by $T^{(n)}$ the subtree of a tree T containing the vertices from level 0 to level n, $T^{(n)}_{(m)}$ the subtree of a tree T containing the vertices from level m to level n.

Let t be a vertex of T, predecessor of the vertex t is another vertex which is nearest from t on the unique path from root o to t. We denote the predecessor of t by 1_t, the predecessor of 1_t by 2_t, the predecessor of N_t by $(N+1)_t$ and $0_t = t$, where N=0,1,2,.... We also say that N_t is the N-th predecessor of t.

Definition 1 (Tree-indexed nonhomogeneous Markov chains). Let T be a tree, S be a states space (finite or countable), $\{X_\sigma, \sigma \in T\}$ be a collection of S-valued random variables defined on the probability space $(\Omega, \mathbb{F}, P)$. Let

$$p = \{p(x), x \in S\} \tag{1}$$

be a distribution on S, and

$$(P_t(y \mid x)), x, y \in S, t \in T \tag{2}$$

be   stochastic matrices on S^2. If for any vertices t,

$$P\left(X_t = y \big| X_t = x \text{ and } X_s \text{ for } t \wedge s \leq 1_t\right)$$
$$= P(X_t = y \big| X_t = x) = P_t(y \big| x), x, y \in S \tag{3}$$

and
$$P(X_o = x) = p(x) \tag{4}$$

$\{X_t, t \in T\}$ will becalled S-value nonhomogeneous Markov chains indexed by   tree with the initial distribution (1) and transition matrix (2). If for all t

$$(P_t(y \big| x)) = (P(y \big| x)), x, y \in S \tag{5}$$

$\{X_t, t \in T\}$ will be called S-value homogeneous Markov chains indexed by   tree T.

It is easy to see that if $\{X_t, t \in T\}$ is a S-valued Markov chains indexed by a tree defined as above, then

$$P(x^{T^{(n)}}) = P(X^{T^{(n)}} = x^{T^{(n)}}) = p(x_0) \prod_{t \in T_{(N+1)}^{(n)}} P_t(x_t \big| y_t), \tag{6}$$

$$P(x^{T^{(n-N)}}) = p(x_0) \prod_{t \in T_{(N+1)}^{(n)}} P_t(x_{N_t} \big| x_{(N=1)_t}) \tag{7}$$

2   Strong Limit Theorems

Lemma 1 (see [4])   Let T be an infinite   tree. Let $\{X_t, t \in T\}$ be a T -indexed nonhomogeneous Markov chain with countable states space S defined as before, $\{g_t(x, y), t \in T\}$ be functions defined on S^2. For any given nonnegative integer N,

$$t_{n-N}(\lambda, \omega) = \frac{e^{\lambda \sum_{t \in T_{(N+1)}^{(n)}} g_t(X_{(N+1)t}, X_{N_t})}}{\prod_{t \in T_{(N+1)}^{(n)}} E[e^{\lambda g_t(X_{(N+1)t}, X_{N_t})} \big| X_{(N+1)_t}]} \tag{8}$$

where λ is   a   real   number.   Then $\{t_{n-N}(\lambda, \omega), F_{n-N}, n \geq N+1\}$ is   a   nonnegative martingale.

Theorem 1. Let T be an infinite tree, $\{X_t, t \in T\}$ be a T-indexed nonhomogeneous Markov chain with countable states space S, $\{g_t(x, y), t \in T\}$ be functions defined on S^2. For any given nonnegative integer N, let

$$H_n(\omega) = \sum_{t \in T_{(N+1)}^{(n)}} g_t(X_{(N+1)_t}, X_{N_t}),$$

(9)

$$G_n(\omega) = \sum_{t \in T_{(N+1)}^{(n)}} E[g_t(X_{(N+1)_t}, X_{N_t}) \mid X_{(N+1)_t}]$$

(10)

Let $\alpha > 0$, then

$$\lim_{n \to \infty} \frac{H_n(\omega) - G_n(\omega)}{a_n} = 0 \qquad \text{a.e. on } D(\alpha).$$

(11)

where

$$D_{(\alpha)} = \{\limsup_{n \to \infty} \frac{1}{a_n} \sum_{t \in T_{(N+1)}^{(n)}} E[g_t^2(X_{(N+1)_t}, X_{N_t}) e^{a|g_t(X_{(N+1)_t}, X_{N_t})|} \mid X_{(N+1)_t}] = M(\omega) < \infty\} \cap B$$

(12)

and

$$B = \left\{ \lim_{n \to \infty} a_n = \infty \right\}$$

(13)

Proof: By Lemma 1, we have known that $\{t_{n-N}(\lambda, \omega), F_{n-N}, n \geq N+1\}$ is a nonnegative martingale. According to Doob martingale convergence theorem we have

$$\lim_{n \to \infty} t_{n-N}(\lambda, \omega) = t(\lambda, \omega) < \infty \qquad \text{a.e.}$$

(14)

We have by (13) and (14)

$$\limsup_{n \to \infty} \frac{\ln t_{n-N}(\lambda, \omega)}{a_n} \leq 0 \qquad \text{a.e.}$$

(15)

By (8),(9) and (15), we get

$$\limsup_{n \to \infty} \frac{1}{a_n} \{\lambda H_n(\omega) - \sum_{t \in T_{(N+1)}^{(n)}} \ln[E[e^{\lambda g(X_{(N+1)_t}, X_{N_t})} \mid X_{(N-1)_t}]]\} \leq 0 \qquad \text{a.e.}$$

(16)

Let $\lambda > 0$, dividing two sides of (16) by λ, we have

$$\limsup_{n \to \infty} \frac{1}{a_n} \{H_n(\omega) - \sum_{t \in T_{(N+1)}^{(n)}} \frac{\ln[E[e^{\lambda g(X_{(N+1)_t}, X_{N_t})} \mid X_{(N-1)_t}]]}{\lambda}\} \leq 0 \qquad \text{a.e.}$$

(17)

Taking $0 \leq \lambda < -\alpha$, we arrive at

$$\limsup_{n \to \infty} \frac{1}{a_n} \{ H_n(\omega) - \sum_{t \in T^{(n)}_{(N+1)}} E[g_t(X_{(N+1)_t}, X_{N_t}) \big| X_{(N+1)_t}] \}$$

$$\leq \limsup_{n \to \infty} \frac{1}{a_n} \sum_{t \in T^{(n)}_{(N+1)}} \frac{\ln[E[e^{\lambda g(X_{(N+1)_t}, X_{N_t})} \big| X_{(N-1)_t}]]}{\lambda} - E[g_t(X_{(N+1)_t}, X_{N_t}) \big| X_{(N+1)_t}] \}$$

$$\leq \limsup_{n \to \infty} \frac{1}{a_n} \sum_{t \in T^{(n)}_{(N+1)}} \frac{E[e^{\lambda g(X_{(N+1)_t}, X_{N_t})} \big| X_{(N-1)_t}] - 1}{\lambda} - E[g_t(X_{(N+1)_t}, X_{N_t}) \big| X_{(N+1)_t}] \}$$

$$\leq \frac{\lambda}{2} \limsup_{n \to \infty} \frac{1}{a_n} E[g_t^2(X_{(N+1)_t}, X_{N_t}) e^{\lambda |(X_{(N+1)_t}, X_{N_t})|} \big| X_{(N+1)_t}]$$

$$= \frac{\lambda}{2} M(\omega) \qquad\qquad a.e. \quad \omega \in D(a) \tag{18}$$

where the first inequality follows by (17) and the fact that

$$\limsup_{n \to \infty}(c_n + b_n) \leq \limsup_{n \to \infty} c_n + \limsup_{n \to \infty} b_n$$

the second follows by the inequality $\ln x \leq x - 1 (x > 0)$

the third follows by the inequality $0 \leq e^x - x - 1 \leq \frac{1}{2} x^2 e^{|x|}$

Letting $\lambda \to 0^+$ $\lambda \to 0^-$ in (18), by (10) we have

$$\limsup_{n \to \infty} \frac{H_n(\omega) - G_n(\omega)}{a_n} \leq 0 \qquad a.e. \quad \omega \in D(a) \tag{19}$$

Let $-\alpha \leq \lambda < 0$. By (16), we similarly get

$$\liminf_{n \to \infty} \frac{H_n(\omega) - G_n(\omega)}{a_n} \geq \frac{\lambda}{2} M(\omega) \qquad a.e. \quad \omega \in D(a) \tag{20}$$

Letting $\lambda \to 0^-$, we can arrive at

$$\liminf_{n \to \infty} \frac{H_n(\omega) - G_n(\omega)}{a_n} \geq 0 \qquad a.e. \quad \omega \in D(a) \tag{21}$$

Combing (19) and (21), we obtain (11) directly. This completes the proof of Theorem 1.

Let $k \in S$, $S_n(k)$ be the number of k in the set of random variables $X^{T^{(n)}}$, and $S_n^N(k)$ be the number of k's N-th descendants in the set of random variables $X^{T^{(n)}}$, that is

$$S_n(k) = \sum_{t \in T^{(n)}} \delta(X_t), \tag{22}$$

$$S_n^N(k) = \sum_{t \in T^{(n)}} \delta_k(X_{N_t}) \tag{23}$$

Corollary 1. For any nonnegative integer N, $\quad S^N_{\ n}(k) \quad$ be the number of k's N-th descendants. If moreover

$$\lim_{|t| \to \infty} P_t(k \mid l) = P_t(k \mid l) \tag{24}$$

then

$$\lim_{n \to \infty} \frac{1}{\left|T^{(n)}\right|} \{S_n^N(k) - \sum_{l \in S} P_t(k \mid l) S_n^N(l)\} = 0 \tag{25}$$

Further more, for any positive integer m $<$n-N, we have

$$\lim_{n \to \infty} \frac{1}{\left|T^{(n)}\right|} \{S_n^N(k) - \sum_{l \in S} P^m(k \mid l) S_n^{N+m}(l)\} = 0 \tag{26}$$

Where $P^m(k \mid l)$ is the m-step transition probability determined by the transition matrix $(P_t(y \mid x))$.

Proof. Let $g_t(X_{(N+1)_t}, X_{N_t}) = \delta_k(X_{N_t}) \quad a_n = \left|T^{(n)}\right| \quad$ in Theorem 1, then (25) and (26) follow obviously.

Theorem 2. If (24) holds, then

$$\lim_{n \to \infty} \frac{S_n^N(k)}{\left|T^{(n)}\right|} = \pi(k) \tag{27}$$

$$\lim_{n \to \infty} \frac{S_n(k)}{\left|T^{(n)}\right|} = \pi(k) \tag{28}$$

Where $\pi = (\pi(0), \dots, \pi(b-1))$ is the uniquestationary distribution determined by transition matrix $(P(i \mid j)_{i,j \in S})$.

Proof. Obviously, (28) follows by setting N=0 in (32). Hence, we only need to proof (27). By Corollary 1,

$$\lim_{n \to \infty} \sup \frac{S_n^N(k)}{\left|T^{(n)}\right|} - \pi(k)$$

$$= \lim_{n \to \infty} \sup \frac{1}{\left|T^{(n)}\right|} \left| S_n^N(k) - \sum_{l \in S} P^m(k \mid l) S_n^{N+m}(l) + \sum_{l \in S} P^m(k \mid l) S_n^{N+m}(l) - \pi(k) \right|$$

$$\leq \sum_{l \in S} \left| P^m(k \mid l) - \pi(k) \right| \tag{29}$$

Since

$$P^m(k|l) \to \pi(k), m \to \infty, \tag{30}$$

by (30), the right hand side of (29) is arbitrary small for enough large m, which implies the equation (27). We complete the proof.

Let   T be a tree,

$$f_n(\omega) = -\frac{1}{|T^{(n)}|} \ln P(X^{T^{(n)}}), \tag{31}$$

will be called the entropy density. If $\{X_t, t \in T\}$ is a T-indexed nonhomogeneous Markov chain with states space   S , we have by (6),

$$f_n(\omega) = -\frac{1}{|T^{(n)}|} [\ln P(X_0) + \sum_{t \in T^{(n)}_{(N+1)}} \ln P(X_t|X_{1t})] \tag{32}$$

The convergence of $f_n(\omega)$ to a constant in a sense (L_1 convergence , convergence in probability, a.e. convergence) is called the Shannon-McMillan theorem or the entropy theorem or the AEP in information theory.

Theorem 3. Let T be an infinite tree, then

$$\lim_{n \to \infty} f_n(\omega) = \sum_{j \in S} \pi(j) H[P(0|j), ... P(b-1|j)] \tag{33}$$

.

Proof. By (9), (10) and (32),

$$\frac{H_n(\omega)}{|T^{(n)}|} = -\frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)}_{(N+1)}} \ln P_t(X_t|X_{1t}) = f_n(\omega) + \frac{\ln p(X_0)}{|T^{(n)}|} \tag{34}$$

$$\frac{G_n(\omega)}{|T^{(n)}|} = -\frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)}_{(N+1)}} E[\ln P_t(X_t|X_{1t})|X_{1t}]$$

$$= -\frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)}_{(N+1)}} H[P_t(0|X_{1t}), ... P_t(b-1)|X_{1t}] \tag{35}$$

If (18) holds, we can prove easily

$$\lim_{n \to \infty} \frac{1}{|T^{(n)}|} \sum_{t \in T^{(n)}_{(N+1)}} \{H[P_t(0|X_{1t}), ..., P(b-1|X_{1t}) - H[P(0|X_{1t}), ..., P(b-1|X_{1t})]\} = 0 \tag{36}$$

By Theorem 2, we get

$$\lim_{n \to \infty} \frac{1}{\left| T^{(n)} \right|} \times \sum_{t \in T_{(N+1)}^{(n)}} H[P_t(0 \mid X_{1_t}), \dots, P(b-1 \mid X_{1_t})]$$

$$= \lim_{n \to \infty} \frac{1}{\left| T^{(n)} \right|} \sum_{t \in T_{(1)}^{(n)}} \sum_{j \in S} H[P(0 \mid j), \dots, P(b-1 \mid j)] \delta_j(X_{1_t})$$

$$= \sum_{j \in S} \pi(j) H[P(0 \mid j), \dots, P(b-1 \mid j)] \tag{37}$$

by (34), (35), (36) and (37), (33) holds. This is the end of the proof.

References

[1] Benjamini, I., Peres, Y.: Markov chains indexed by trees. Ann. Probab. 22, 219–243 (1994)

[2] Berger, T., Ye, Z.-X.: Entropic aspects of random fields on trees. IEEE Trans. Inform. Theory 36(5), 1006–1018 (1990)

[3] Yang, W.-G., Ye, Z.-X.: The Asympotoic equipartition property for nonhomogeneous Markov chains indexed by a Homogeneous tree. IEEE Trans. Inform. Theory 53(9), 3275–3280 (2007)

[4] Huang, H.-L., Yang, W.-G.: Strong law of large numbers for Markov chains indexed by an infinitetree with uniformly bounded degeree. Science in China Series A: Mathematics 51(2), 195–202 (2008)

Design and Implementation
of Manchester Codec Based on FPGA

Chun-Ying Wang and Yan-Ling Zheng

Electronic and Electrical Engineering Department, Baoji University of Arts and Sciences, Baoji, Shaanxi 721016, China

Abstract. In this paper, the principle of Manchester II codec was introduced and analyzed, 4Mb/s Manchester II codec was designed and implemented in 1553B bus interface based on FPGA. Using the Verilog HDL language and Quartus II 8.0 of Altera company to program, realizing the synthesis, optimization and simulation of Manchester II codec and validating the function of encoding and decoding. The results suggest that the encoding and decoding modular can be implemented on FPGA and the transmission rate can be up to 4Mb/s, so make Manchester codec which is applied to 1553B bus interface, has ascendant anti-jamming characteristic and improve transmission rate.

Keywords: Manchester Codec, Verilog HDL, QuartusII, Simulation.

1 Introduction

Modern aircraft aviation and automotive systems were equiped with a lot of electronic equipment in order to meet the requirements such as emissions, safety and online diagnosis and so on, but it would be bound to cause the increase of wires. To further reduce harness and meet the need of the main electronic unit or real-time data information exchange between system and made the system tends to be best, it adopts the way based on the system bus. It can effectively connect the various control systemt to a comprehensive control system and greatly improve the system performance[1].

MIL-STD-1553, Digital Internal Time Division Command/Response Multiplex Data Bus, is a military standard for data bus. The data bus system has the advantages such as the operation detectability, the higher comprehensive performance and reliability. It is widely used in many military fields, gradually uses in non-military application fields and complements with CAN bus. At present, the domestic and international aerospace control systems have been widely applied in 1553B bus, so research 1553B bus interface circuit has significant value and become the domestic various IC design unit of great interest.

Manchester codec is one indispensable field of the1553B bus interface circuit. the performance of 1553B bus interface circuit is affected directly by the design of Manchester codec. In this paper a Manchester codec whose highest can be rate up to 4Mbps as a part of new type 1553B bus interface circuit is designed.

2   1553B Bus and Interface Function

The full name of MIL-STD-1553B is: Digital Internal Time Division Command/ Response Multiplex Data Bus. It transmits data by the twisted-shielded pair cable and achieves the signal transmission with the serial digital pulse, whose data transmission uses Manchester code form and transmission rate is up to 1Mb/s. The 1553B data bus has three kinds system:(1) the system bus controller (BC) : control bus data transmission; (2) the remote terminal (RT): response to commands, execute BC data transmission; (3) bus monitor (BM): selectively receiving data bus and kept the information. Each subsystems can connect to any a subsystem with 1553B bus and data is transmitted in the bus.

The standard allows for only three types of words as discussing in below: command word, status word and data word. Each word is consist of 16 bit of data plus, a sync pattern (3 bit), and a one bit parity providing a 20 bit word format. The effective information (16 bits) and parity is transmitted with Manchester code in the bus, each bit time interval is 1us. The width of synchronization shall be three bit times, with the sync waveform being positive for the first one and one-half bit times, and then negative for the following one and one-half bit times(command and status words) or with the sync waveform being negative   for the first one and one-half bit times, and then positive for the following one and one-half bit times(data word). The command word and the status word are different system type, so command and status words can always be distinguished from system type such as command word come from bus controller and status word come from the remote terminal[1].

MIL-STD-1553B bus interface shall complete the following functions: (1) the serial information change into the parallel information in bus or conversely, the parallel information change into the serial information. (2)when the message is received and sent, standard 1553B information words and messages can be recognized or generated.(3) complete information exchange between bus and processor, including the distribution of 1553B information address, the decoding of command word (or status word) or the return status word, sent data words, etc. Manchester codec complete the function that is the encoded and decoded of Manchester code, and to detect errors. It receives the Manchester code which has the valid sync, and decode the data, and identify the types and change serial into parallel, parity, etc.Or the parallel binary data from processor changes to Manchester code,and adds sync and parity bit, which has become the accord with the 1553B standard and output.
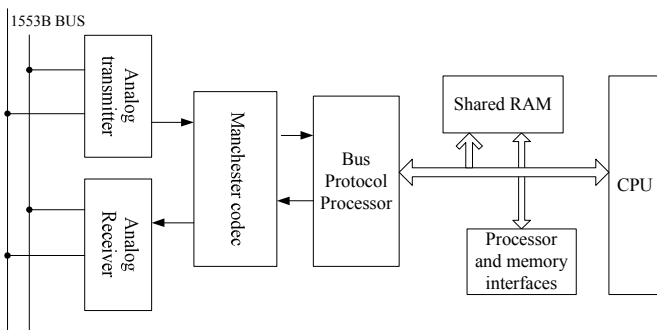
Fig. 1. 1553B bus interface block diagram

Codec is a part of the bus interface,which is consist of the level convert, Manchester codec, command/state word decoded logic, CPU and related circuit, as shown in Fig.1.

Besides codec can be realized by FPGA, the relevant circuit between the CPU, share RAM, decoded command/state word, memory processor and memory interface logic can also be realized by FPGA, in this paper the design of codec is introduced.

3   Manchester Code Principle

Data is transmitted by Manchester code in 1553B bus. Manchester code is dual polarity code, which is widely used in avionics integrated system. Manchester Ⅱ code is used to express binary value by jumping rather than level, which exists a phase jumping in every midpoint. Logic one represents a negative jumping from zero to one, which is dual phase signal "10". Logic zero represents a positive jumping from one to zero, which is dual phase signal "01". It can directly isolate timing clock from data, which contains the timing and does not require independent channel to transmit a timing information. When the code information is transmited, clock synchronization is also transmited to each other. Every code has jumping, not dc component, therefore it has the self synchronous ability and strong anti-jamming performance[2]. Encoding method is shown in Fig.2.
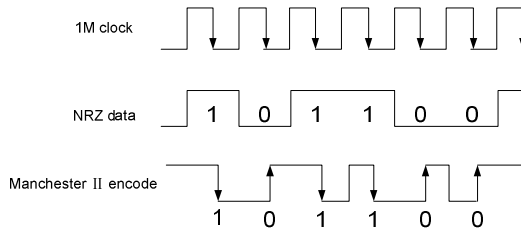
Fig. 2. Manchester Ⅱ encoding method

4   The Design of Manchester Encoder and Decoder

(1) The Design of Manchester Encoder
Manchester encoding module receives 16 parallel data through sending control module, then serial Manchester code data with 4Mbps is sent to analog transmitter/receiver and further sent to 1553B bus. It mainly realizes synchronization, the generation of parity, the change of code type and the conversion of parallel to serial.

The time benchmark of send unit is 48MHz. By Manchester II characteristics, the data "0" and "1" are encoded in "01" and "10", so the minimal clock of encoding units are 8MHz which is provided by 48MHz clock through divide frequency six and two.

Comnd is the type signal of synchronization. When comnd is the hign level, encoder prodeucs command word or status word; when comnd is the low level, encoder produce data word. Datain is the input signal(sixteen bits). The signal which

is generated from synchronization to shift   is the enable signal of shift and parity module. The output tx signal is the positive phase code of Manchester output, the ntx signal is negative phase code. The txready signal is the flag of encoded data[3].

Synchronization module generates different type synchronization head according to the demand of comnd signal, and start next module after completion of this module. The input parallel data carry out parallel-to-serial shift and generate parity, meanwhile encoding module realizes Manchester encoding of serial data(seventeen bits) which holds sixteen bits input data from shift module plus one bit for parity, then outputs it. Encoder logic block diagram is shown in Fig. 3.
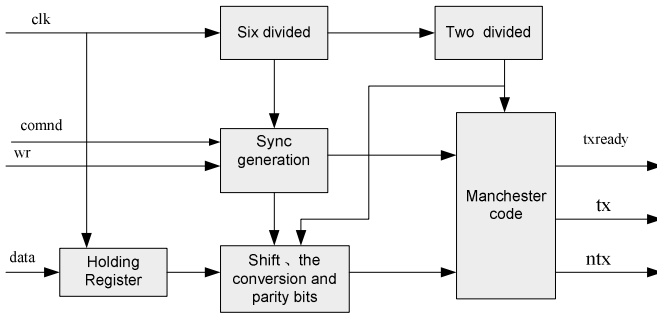
Fig. 3. Encoding logic block diagram

(2) The Design of Manchester Decoder

Manchester decoding process means decoder receives the Manchester encoding serial data from transceiver, achives below functions such as detecting synchronization head, detecting data, detecting Manchester II error , detecting parity and so on. Processed data is streamed into subsequent module to analysis by sixteen data line[4].

Decoder completes the decoding work, which uses the way of sampling counts to determine signal pulse width and jumping. First thing is the recognition synchronization head which has only two kinds of circumstances"10" and "01". Its waveform should have following four cases as shown in Fig. 4.
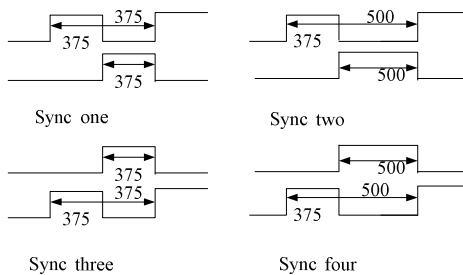
Fig. 4. Synchronization head type

If detected synchronization head is not the synchronization of the above four cases, then it is illegal synchronization head. If using different sampling clock frequency, it has different counting standards. The first half width of synchronization head is 375ns, according to the standard, the error range is 40ns, so only if the pulse is received whose width is (375+40), that is to say, the first half of synchronization head is received. If the sampling frequency is 8MHz, then the count value during three to four. That is to say, if the counter counts in this interval, the first half of synchronization head can be considered to receive.

After detecting synchronization head, code type conversion begins to work. In this module, there are two ways to realize its function. The one kind is to use sampling to determine the with and the level jumping of data bits, thus the decoding is realized. When the sampling frequency is 8MHz, one data bit can sample two bits, if the two bits is "01" or "10", which indicates data one or zero, otherwise it is wrong. The other kind is to use inverse that is XOR operation between input data and 4MHz clock in encoder to realize decoding[5].

After data completes serial/parallel and parity in this module, data is sent out by 16 parallel form. The logic block diagram is shown in Fig.5.
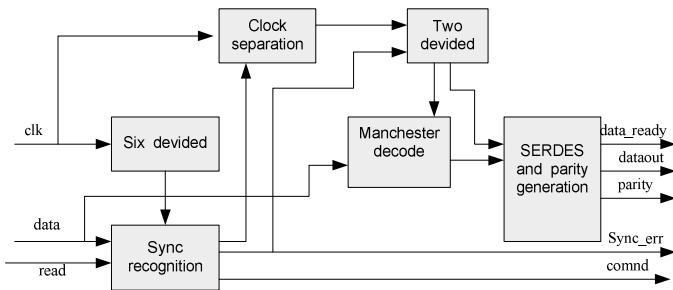
Fig. 5. Decoding logic block diagram

Decoding module input: 48MHz clock signal is clk; the enabling signal of decoder is read, only if the enabling signal is valid, the whole module starts to detect and decode bus data; data input signal is data. Decoding module output: output parallel data (sixteen bits) is dataout; the flag of output finish is data_ready; synchronization head output end is comnd; synchronization head error end is sync_err; the parity end is parity.

5   Simulation and Tested Results

In order to ensure the feasibility, the design is simulated and optimum Synthesis to improve the performance and the utilization rate of resources. Quartus II which has integrated simulation software also links to currently popular integrated simulation tools such as Synplify/Synplify Pro, Synplify/Synplify Pro, Leonardo Spectrum, CompilerII little Synopsys FPGA ModelSim, Active HDL, etc.Generally the third-party comprehensive tools hve the strong functions and the optimization effect, but integration tools in Quartus II software have its own advantages.

The design mainly adopts integrated simulation andcomprehensive tools in Quartus II software Manchester encoder and decoder simulation waveform are shown in Fig.6 and Fig.7.

Fig.6 illustrates the simulation waveform, when input data is 7521H. The clock source is 48MHz, the tx end is the sending data, parity bit is one.
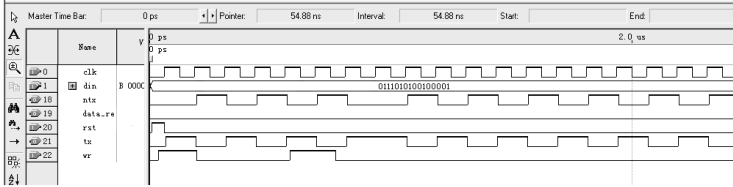
Fig. 6. Encoder simulation waveform

Fig.7 illustrates the simulation waveform, when input data is 01110101001000011B(i.e. input serial data is 7521H, parity bit is 1). The sampling clock is 8MHz, dataout end is output data.
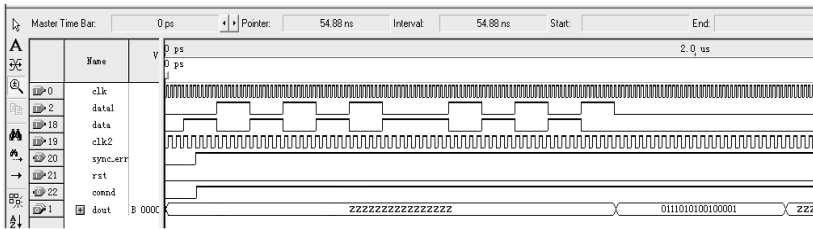
Fig. 7. Decoder simulation waveform

The design is implemented in Altera development board(CycloneII EP2C5Q208C8).

6   Summaries

This design uses Verilog HDL language as a design input, adopts the integrated simulation and synthesis tools in Quartus II software to complete the design of Manchester codec in 1553B bus interface. And the validation test of this design was implementated by CycloneII series high-speed FPGA chip development board. The The results suggest that the function of encoding and decoding modular were implemented by FPGA, and make this design has ascendant anti-jamming characteristic and improve transmission rate.

References

[1] DDC Corporation. ACE/Mini-ACE Series BC/RT/MT Advanced Communication Engine Integrated 1553 Terminal User' s Guide

[2] Li, Z.-G., Gai, Y.: The Design of the Manchester II Coders and Decoders for the 1553B Bus Based on FPGA. Metrology & Measurement Technology 4, 45–48 (2006)

[3] Chen, X., Luo, Z.-H., Chen, X.-G.: Manchester Encoding and Decoding Circuit Based on FPGA. J. of China Three Gorges Univ. 6, 85–87 (2008)

[4] Shi, H.-M., Ji, L., Xie, S.-Q.: The Implementation of 1553B Bus Interface Manchester Codec with FPGA. Microcontrollers & Embedded Systems 4, 42–46 (2004)

[5] Niu, Q., Jin, H.: The Design of the Manchester II Coders and Decoders for the 1553B Bus on FPGA. Metrology & Measurement Technique 37(9), 53–56 (2010)

Ishikawa's Iterative Algorithm of Nonexpansive Mappings and Averaged Mappings

Duan Qibin, Wang Erli, Wu Dingping, and Xie Xuping

Department of Mathematic, Chengdu University of Information Technology, 610225, P.R. China
wdp68@163.com

Abstract. In this paper we introduce and study Ishikawa's iterative algorithm of nonexpansive mapping and averaged mappings in the setting of real Hilbert spaces. The main results presented in this paper extend the classic result of convergence of Mann's iterative algorithm for nonexpansive mappings in the setting of Hilbert space and we prove the weak converge theorems.

Keywords: Ishikawa's iterative process, Averaged mapping, Nonexpansive mapping, Weak converge theorems.

1 Introduction and Preliminaries

Throughout this paper, we always assumes that H is a real Hilbert space with inner product $\langle \cdot , \rangle$ and norm $\|\cdot\|$, respectively.

Firstly, we recall the classic result of convergence of Mann's iterative method for nonexpansive mappings in the setting of Hilbert spaces. Let C is a closed convex subset of a Hilbert space, $T : C \rightarrow C$ is a nonexpansive mapping with a fixed point. Let $\{\alpha_n\}$ be a sequence in $[0,1]$ such that $\sum_{n=1}^{\infty} \alpha_n(1-\alpha_n) = \infty$. Then, initializing with $x_0 \in C$, the sequence $\{x_n\}$ generated by Mann's iteration:

$$x_{n+1} = (1 - \alpha_n)x_n + \alpha_n Tx_n , n \geq 0$$

converges weakly to a fixed point of T.

Definition 1.1. Assume H is a real Hilbert space. let $A, T : H \rightarrow H$ be nonlinear operators introduced below.

A mapping $T : H \rightarrow H$ is said to be nonexpansive if

$$\|Tx - Ty\| \leq \|x - y\|, \qquad x, y \in H .$$

A mapping $T : H \rightarrow H$ is said to be an averaged mapping if it can be written as the average of the identity I and a nonexpansive mapping S; that is

$$T = (1 - \alpha)I + \alpha S .$$

Where $\alpha \in (0,1)$ and $S : H \to H$ is nonexpansive .

A is monotone if $\langle Ax - Ay, x - y \rangle \geq 0$ for all $x, y \in H$.

A is β-strongly monotone, with $\beta > 0$, if

$$\langle Ax - Ay, x - y \rangle \geq \beta \|x - y\|^2 \text{ for all } x, y \in H.$$

A is υ-inverse strongly monotone(υ-ism),with $\upsilon > 0$,if

$$\langle Ax - Ay, x - y \rangle \geq \upsilon \|Ax - Ay\|^2 \text{ for all } x, y \in H.$$

Proposition 1.2. We have the following assertions:

i. T is nonexpansive if and only if the complement $I - T$ is $\frac{1}{2}$-ism.

ii.   If T is υ-ism and $\gamma > 0$,then γT is $\frac{\upsilon}{\gamma}$-ism.

iii. For $\alpha \in (0,1)$,T is α-averaged if and only if $I - T$ is $\frac{1}{2\alpha}$-ism.

iv.   If T_1 is α_1-averaged and T_2 is α_2-averaged, where $\alpha_1, \alpha_2 \in (0,1)$, then the composite $T_1 T_2$ is α-averaged ,where $\alpha = \alpha_1 + \alpha_2 - \alpha_1 \alpha_2$.

v.   If T_1 and T_2 are averaged and have a common fixed point, then $Fix(T_1 T_2) = Fix(T_1) \bigcap Fix(T_2)$.

Proposition 1.3. Let K be nonempty closed convex subset of a Hilbert space H . Let $\{x_n\}$ be bounded sequence which satisfies the properties:

- $\omega_\omega(x_n) \subset K$;
- $\lim\limits_{n \to \infty} \|x_n - x\|$ exists for every $x \in K$.

Then $\{x_n\}$ converges weakly to a point in K .

In 1974, Ishikawa [1] introduced an iterative sequence $\{x_n\}$ defined by

$$\begin{cases} x_0 \in D \\ x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T y_n \\ y_n = (1 - \beta_n)x_n + \beta_n T x_n. \end{cases} \tag{1.1}$$

Where $D \subset E$ is a nonempty convex subset, $T : D \to D$ is a mapping and both $\{\alpha_n\},\{\beta_n\}$ are parameter sequences in $[0,1]$ satisfying some conditions.

In this paper,we will apply the nonexpansive mappings and averaged mappings into the following Ishikawa's iterative process and study the convergence,

$$\begin{cases} x_1 \in C \\ x_{n+1} = (1 - \alpha_n)x_n + \alpha_n T y_n \\ y_n = (1 - \beta_n)x_n + \beta_n T x_n, \quad n = 0,1,2,\ldots \end{cases} \tag{1.2}$$

where $T : C \to C$ is a mapping in Hilbert space and where $\{\alpha_n\}$, $\{\beta_n\}$ are assumed to be a real sequence in $[0,1]$.

2   Main Results

Theorem 2.1. Suppose $T : C \to C$ is nonexpansive with a fixed point, where C is a closed convex subset of a Hilbert space. Assume that

$$0 < \liminf_{n \to \infty} \alpha_n < \limsup_{n \to \infty} \alpha_n < 1, 0 < \liminf_{n \to \infty} \beta_n < \limsup_{n \to \infty} \beta_n < 1 .$$

Then the sequence $\{x_n\}$ be defined by Ishikawa iterative algorithm (1.2) converges weakly to a fixed point of T.

Proof: Firstly, we prove that $\lim_{n \to +\infty} \|x_{n+1} - x_n\| = 0$.

For any $x^* \in Fix(T)$.

Noticing $Tx^* = x^*$, we have

$$
\begin{aligned}
\|Ty_n - x^*\| &= \|Ty_n - Tx^*\| \le \|y_n - x^*\| = \|(1 - \beta_n)x_n + \beta_n Tx_n - x^*\| \\
&= \|(1 - \beta_n)x_n + \beta_n Tx_n - (1 - \beta_n)x^* - \beta_n x^*\| \\
&\le (1 - \beta_n)\|x_n - x^*\| + \beta_n \|Tx_n - x^*\| \le \|x_n - x^*\|.
\end{aligned}
\tag{2.1}
$$

From (2.1), we get

$$
\begin{aligned}
\|x_{n+1} - x^*\| &= \|(1 - \alpha_n)x_n + \alpha_n Ty_n - (1 - \alpha_n)x^* - \alpha_n x^*\| \\
&\le (1 - \alpha_n)\|x_n - x^*\| + \alpha_n \|Ty_n - x^*\| \\
&\le (1 - \alpha_n)\|x_n - x^*\| + \alpha_n \|x_n - x^*\| \le \|x_n - x^*\|.
\end{aligned}
\tag{2.2}
$$

So the sequence $\{\|x_n - x^*\|\}$ is nonincreasing. Hence,

$$\lim_{n \to +\infty} \|x_n - x^*\| \text{ exists for all } x^* \in Fix(T).$$

By (1.2), we have

$$
\begin{aligned}
\|x_{n+1} - x^*\|^2 &= \|(1 - \alpha_n)x_n + \alpha_n Ty_n - x^*\|^2 \\
&= \|(1 - \alpha_n)(x_n - x^*) + \alpha_n (Ty_n - Tx^*)\|^2 \\
&= \langle (1 - \alpha_n)(x_n - x^*) + \alpha_n (Ty_n - Tx^*), (1 - \alpha_n)(x_n - x^*) + \alpha_n (Ty_n - Tx^*) \rangle \\
&= \langle (1 - \alpha_n)(x_n - x^*), x_n - x^* \rangle + \langle (1 - \alpha_n)(x_n - x^*), \alpha_n (Ty_n - x_n) \rangle \\
&\quad + \langle \alpha_n (Ty_n - Tx^*), Ty_n - Tx^* \rangle + \langle \alpha_n (Ty_n - Tx^*), (1 - \alpha_n)(x_n - Ty_n) \rangle \\
&= (1 - \alpha_n)\|x_n - x^*\|^2 + \alpha_n \|Ty_n - Tx^*\|^2 - \alpha_n (1 - \alpha_n)\|x_n - Ty_n\|^2 \\
&\le \|x_n - x^*\|^2 - \alpha_n (1 - \alpha_n)\|x_n - Ty_n\|^2 .
\end{aligned}
\tag{2.3}
$$

Simplify above

$$\left\| x_n - Ty_n \right\|^2 \le \frac{1}{\alpha_n (1-\alpha_n)} (\left\| x_n - x^* \right\|^2 - \left\| x_{n+1} - x^* \right\|^2) \tag{2.4}$$

So

$$\lim_{n\to\infty} \left\| x_{n+1} - x_n \right\|^2 = \lim_{n\to\infty} \alpha_n^2 \left\| x_n - Ty_n \right\|^2 = 0. \tag{2.5}$$

Then we get

$$\lim_{n\to\infty} \left\| x_{n+1} - x_n \right\| = 0. \tag{2.6}$$

Next we prove that $\omega_\omega(x_n) \subset Fix(T)$.

Suppose $p \in \omega_\omega(x_n)$ and $\{x_{n_j}\}$ is a subsequence of $\{x_n\}$ such that $x_{n_j} \to p$ weakly, thus, $x_{n_j+1} \to p$ weakly.

By (2.6), then we get

$$\begin{aligned}
\left\| x_{n_j} - Tx_{n_j} \right\| &\le \left\| x_{n_j+1} - x_{n_j} \right\| + \left\| x_{n_j+1} - Tx_{n_j} \right\| \\
&\le \left\| x_{n_j+1} - x_{n_j} \right\| + \left\| (1-\alpha_{n_j})x_{n_j} + \alpha_{n_j}Ty_{n_j} - (1-\alpha_{n_j})Tx_{n_j} - \alpha_{n_j}Tx_{n_j} \right\| \\
&\le \left\| x_{n_j+1} - x_{n_j} \right\| + (1-\alpha_{n_j})\left\| x_{n_j} - Tx_{n_j} \right\| + \alpha_{n_j}\left\| Ty_{n_j} - Tx_{n_j} \right\|.
\end{aligned} \tag{2.7}$$

Simplify (2.7)

$$\begin{aligned}
\alpha_{nj}\left\| x_{n_j} - Tx_{n_j} \right\| &\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}\left\| Ty_{n_j} - Tx_{n_j} \right\| \\
&= \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}\left\| (1-\alpha)(y_{n_j} - x_{n_j}) + \alpha(Sy_{n_j} - Sx_{n_j}) \right\| \\
&\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}(1-\alpha)\left\| y_{n_j} - x_{n_j} \right\| + \alpha_{n_j}\alpha\left\| Sy_{n_j} - Sx_{n_j} \right\| \\
&\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}\left\| y_{n_j} - x_{n_j} \right\| \\
&= \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}\left\| (1-\beta_{n_j})x_{n_j} + \beta_{n_j}Tx_{n_j} - x_{n_j} \right\| \\
&\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}\beta_{n_j}\left\| x_{n_j} - Tx_{n_j} \right\|.
\end{aligned} \tag{2.8}$$

By condition of Theorem 2.1 and (2.8), we obtain

$$\left\| x_{n_j} - Tx_{n_j} \right\| \le \frac{1}{\alpha_{n_j}(1-\beta_{n_j})}\left\| x_{n_j+1} - x_{n_j} \right\| \to 0. \tag{2.9}$$

So $p \in Fix(T)$. By proposition 1.2, we can see that $\{x_n\}$ converges weakly to a point of $Fix(T)$.

Theorem 2.2. Suppose $T : H \to H$ is α-averaged mapping with a fixed point, where C is a closed convex subset of a Hilbert space. Assume that

$$0 < \liminf_{n\to\infty} \alpha_n < \limsup_{n\to\infty} \alpha_n < 1, 0 < \liminf_{n\to\infty} \beta_n < \limsup_{n\to\infty} \beta_n < 1 \cdot$$

Then the sequence $\{x_n\}$ be defined by Ishikawa' iterative algorithm (1.2) converges weakly to a fixed point of T.

Proof. Since T is α-averaged mapping, $T = (1-\alpha)I + \alpha S$, where $\alpha \in (0,1)$ and $S : H \to H$ is nonexpansive.

Firstly, we prove that for any $x^* \in Fix(T)$, noticing $Tx^* = x^*$, we have

$$\left\| Ty_n - x^* \right\| = \left\| Ty_n - Tx^* \right\| = \left\| (1-\alpha)(y_n - x^*) + \alpha(Sy_n - Sx^*) \right\|$$
$$\leq (1-\alpha)\left\| y_n - x^* \right\| + \alpha \left\| Sy_n - Sx^* \right\| \leq \left\| y_n - x^* \right\| = \left\| (1-\beta_n)x_n + \beta_n Tx_n - x^* \right\|$$
$$= \left\| (1-\beta_n)x_n + \beta_n Tx_n - (1-\beta_n)x^* - \beta_n x^* \right\| \leq (1-\beta_n)\left\| x_n - x^* \right\| + \beta_n \left\| Tx_n - x^* \right\|$$
$$\leq (1-\beta_n)\left\| x_n - x^* \right\| + \beta_n \left\| x_n - x^* \right\| = \left\| x_n - x^* \right\|.$$
(2.10)

Simplify (2.10)

$$\left\| x_{n+1} - x^* \right\| = \left\| (1-\alpha_n)x_n + \alpha_n Ty_n - x^* \right\| = \left\| (1-\alpha_n)x_n + \alpha_n Ty_n - (1-\alpha_n)x^* - \alpha_n x^* \right\| \quad (2.11)$$
$$\leq (1-\alpha_n)\left\| x_n - x^* \right\| + \alpha_n \left\| Ty_n - x^* \right\| \leq (1-\alpha_n)\left\| x_n - x^* \right\| + \alpha_n \left\| x_n - x^* \right\| \leq \left\| x_n - x^* \right\|.$$

So the sequence $\left\{ \left\| x_n - x^* \right\| \right\}$ is nonincreasing. Hence,

$$\lim_{n\to+\infty} \left\| x_n - x^* \right\| \text{ exists for all } x^* \in Fix(T).$$

$$\left\| x_{n+1} - x^* \right\|^2 = \left\| (1-\alpha_n)x_n + \alpha_n Ty_n - x^* \right\|^2 = \left\| (1-\alpha_n)(x_n - x^*) + \alpha_n(Ty_n - Tx^*) \right\|^2 \quad (2.12)$$
$$= (1-\alpha_n)\left\| x_n - x^* \right\|^2 + \alpha_n \left\| Ty_n - Tx^* \right\|^2 - \alpha_n(1-\alpha_n)\left\| x_n - Ty_n \right\|^2.$$

$$\left\| x_n - Ty_n \right\|^2 = \frac{1}{\alpha_n}\left\| x_n - x^* \right\|^2 + \frac{1}{1-\alpha_n}\left\| Ty_n - Tx^* \right\|^2 - \frac{1}{\alpha_n(1-\alpha_n)}\left\| x_{n+1} - x^* \right\|^2. \quad (2.13)$$

by (2.10)

$$\left\| x_n - Ty_n \right\|^2 \leq (\frac{1}{\alpha_n}\left\| x_n - x^* \right\|^2 + \frac{1}{1-\alpha_n}\left\| x_n - x^* \right\|^2 - \frac{1}{\alpha_n(1-\alpha_n)}\left\| x_{n+1} - x^* \right\|^2) \quad (2.14)$$
$$= \frac{1}{\alpha_n(1-\alpha_n)}(\left\| x_n - x^* \right\|^2 - \left\| x_{n+1} - x^* \right\|^2).$$

By condition of Theorem 2.2 and (2.14), we obtain

$$\lim_{n\to\infty}\left\| x_{n+1} - x_n \right\|^2 = \lim_{n\to\infty}\alpha_n^2\left\| x_n - Ty_n \right\|^2 = 0.$$

So

$$\lim_{n\to\infty}\left\| x_{n+1} - x_n \right\| = 0. \quad (2.15)$$

Next we prove that $\omega_n(x) \subset Fix(T)$.

Suppose $p \in \omega_\omega(x_n)$ and $\{x_{n_j}\}$ is a subsequence of $\{x_{n_j}\}$ such that $x_{n_j} \to p$ weakly, thus, $x_{n_{j+1}} \to p$ weakly. By (2.6), then we get

$$
\begin{aligned}
\left\| x_{n_j} - Tx_{n_j} \right\| &\leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + \left\| x_{n_{j+1}} - Tx_{n_j} \right\| \\
&\leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + \left\| (1 - \alpha_{n_j}) x_{n_j} + \alpha_{n_j} Ty_{n_j} - (1 - \alpha_{n_j}) Tx_{n_j} - \alpha_{n_j} Tx_{n_j} \right\| \quad (2.16) \\
&\leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + (1 - \alpha_{n_j}) \left\| x_{n_j} - Tx_{n_j} \right\| + \alpha_{n_j} \left\| Ty_{n_j} - Tx_{n_j} \right\|.
\end{aligned}
$$

simplify the formula above, we have the follwing

$$
\begin{aligned}
\alpha_{nj} \left\| x_{n_j} - Tx_{n_j} \right\| &\leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + \alpha_{n_j} \left\| Ty_{n_j} - Tx_{n_j} \right\| \\
&= \left\| x_{n_{j+1}} - x_{n_j} \right\| + \alpha_{n_j} \left\| (1 - \alpha)(y_{n_j} - x_{n_j}) + \alpha(Sy_{n_j} - Sx_{n_j}) \right\| \\
&\leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + \alpha_{n_j}(1 - \alpha) \left\| y_{n_j} - x_{n_j} \right\| + \alpha_{n_j} \alpha \left\| Sy_{n_j} - Sx_{n_j} \right\| \\
&\qquad \leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + \alpha_{n_j} \left\| y_{n_j} - x_{n_j} \right\| \\
&\qquad = \left\| x_{n_{j+1}} - x_{n_j} \right\| + \alpha_{n_j} \left\| (1 - \beta_{n_j}) x_{n_j} + \beta_{n_j} Tx_{n_j} - x_{n_j} \right\| \qquad (2.17) \\
&\qquad \leq \left\| x_{n_{j+1}} - x_{n_j} \right\| + \alpha_{n_j} \beta_{n_j} \left\| x_{n_j} - Tx_{n_j} \right\|.
\end{aligned}
$$

Finally, we obtain that

$$
\left\| x_{n_j} - Tx_{n_j} \right\| \leq \frac{1}{\alpha_{n_j}(1 - \beta_{n_j})} \left\| x_{n_{j+1}} - x_{n_j} \right\| \to 0. \qquad (2.18)
$$

So $p \in Fix(T)$.

From the proposition 1.2, we can see that $\{x_n\}$ converges weakly to a point of $Fix(T)$.

This completes the proof.

References

[1] Ishikawa, S.: Fixed point and iteration of a nonexpansive mapping in a Banach spaces. Proc. Amer. Math. Soc. 73, 65–71 (1976)
[2] Qin, X., Cho, Y.J., Kang, S.M.: Viscosity approximation methods for generalized equilibrium problems and fixed point with applications. Nonlinear Anal. 72, 99–12 (2010)

[3] Xu, H.-K.: Iterative methods for the split feasibility problem in infinite-dimensional Hilbert space. Inverse Problems 26, 105018 (2010)

[4] Xu, H.-K.: Averaged Mapping and the Gradient-Projection Algorithm (in press)

[5] Hundal, H.: An alternating projection that does not converge in norm. Nonlinear Anal. 57, 35–61 (2004)

[6] Reich, S.: Weak convergence theorems for nonexpansive mappings in Bananch spaces. J. Math Anal. Appl. 67, 274–276 (1997)

Approximation of Implicit Iteration Process for Common Fixed Point of Composite General Asymptotically Nonexpansive Mappings

Chen Xiaomin[1], Xie Xuping[2], and Duan Qibin[2]

[1] Department of Information and Computing Science, Chengdu Electromechanical College, Chengdu, 610031, China
[2] Department of Mathematic, Chengdu Univ. of Info. Technology, 610225, P.R. China
wdp68@163.com

Abstract. We introduce the implicit iteration process with errors for approximation of common fixed points of composite general asymptotically nonexpansive mappings in Banach spaces, strong convergence theorem are established under suitable some conditions.

Keywords: Banach spaces, General asymptotically nonexpansive mappings, Common fixed points, Implicit iteration process.

1 Introduction and Preliminaries

The weak and strong convergence of implicit iteration process to a fixed point of asymptotically nonexpansive mappings and nonexpansive mappings in Banach spaces has been studied by many authors see [3-6].we consider the following finitely composite general asymptotically iteration process with errors.

$$\begin{cases} x_n = \alpha_n x_{n-1} + \beta_n S_{i(n)}^{k(n)} y_n + \gamma_n u_n \\ y_n = l_n x_n + s_n x_n + t_n S_{i(n)}^{k(n)} x_n + w_n S_{i(n)}^{k(n)} x_{n-1} \end{cases} \tag{1}$$

while $S_{i(n)}^{k(n)} = (1-\delta)I + \delta T_{i(n)}^{k(n)}$, $\delta \in (0,1)$ is a constant.

$$n = (k(n)-1)N + i(n), \quad i = i(n) = 1, 2 \cdots N, k = k(n), \quad T_{i(n)} = T_{n(\bmod N)}.$$

$$\alpha_n + \beta_n + \gamma_n = l_n + s_n + t_n + w_n = 1, \quad \alpha_n, \beta_n, \gamma_n, l_n, s_n, t_n, w_n \in [0,1].$$

Definition 1.1. T is said to be asymptotically nonexpansive, if there exists a sequence $\{k_n\} \subset [0, \infty)$ with $\lim_{n \to \infty} k_n = 0$

such that

$$\left\| T^n x - T^n y \right\| \le (1 + k_n) \left\| x - y \right\|, \quad \forall x, y \in D.$$

Definition 1. 2. T is said to be general asymptotically nonexpansive, if there exits a sequences

$\{k_n\},\{e_n\} \subset [0,+\infty)$ with $\lim_{n\to\infty} k_n = 0$, $\lim_{n\to\infty} e_n = 0$ such that

$$\|T^n x - T^n y\| \le (1+k_n)\|x - y\| + e_n \qquad \forall x, y \in D.$$

Let $T_i : D \to D, i = 1, 2 \cdots N$ be general asymptotically nonexpansive, there be two sequences

$\{k_{i(n)}\},\{e_{i(n)}\} \subset [0,+\infty)$ with $\lim_{n\to\infty} k_{i(n)} = \lim_{n\to\infty} e_{i(n)} = 0$, such that:

$$\|T_i^n x - T_i^n y\| \le (1+k_{i(n)})\|x - y\| + e_{i(n)}, \forall x, y \in D.$$

Lemma 1.1[1] . Let $\{a_n\},\{b_n\},\{c_n\}$ be three nonnegative sequences satisfying the following condition:

$$a_{n+1} \le (1+b_n)a_n + c_n, \forall\ n \ge n_0, \text{and} \sum_{n=1}^{\infty} b_n < \infty, \sum_{n=1}^{\infty} c_n < \infty.$$

Then $\lim_{n\to\infty} a_n$ exists.

Lemma 1.2[2]**.** Let E be a uniformly convex Banach space, a, b be two constants with $0 < a < b < 1$.Suppose that $\{t_n\}$ is a sequence in $[a,b]$ and $\{x_n\},\{y_n\}$ are two sequence in E. Then the conditions:

$$\limsup_{n\to\infty}\|x_n\| \le d,\ \limsup_{n\to\infty}\|y_n\| \le d,\ \lim_{n\to\infty}\|t_n x_n + (1-t_n)y_n\| = d,$$

imply that $\lim_{n\to\infty}\|y_n - x_n\| = 0$, where $d \ge 0$ is some constant.

2   Main Result

Theorem 2.1. Let E be a uniformly convex Banach space, D be a nonempty closed convex subset of E with $D + D \subset D$. Assuming $\{T_i\}_{i=1}^{N} : D \to D$ is a family of N general asymptotically nonexpansive mappings. Let $\{u_n\}$ be a bounded sequence in D, and $F = \bigcap_{i=1}^{N} F(T_i) \ne \phi$, $\{x_n\}$ defined by (1) with the conditions as follows:

(i) $\sum_{n=1}^{\infty} k_n < \infty$; $\sum_{n=1}^{\infty} e_n < \infty$; (ii) there exits a, b such that $0 < a < \beta_n < b < 1$, $0 < a < l_n$;

(iii) $\displaystyle\sum_{n=1}^{\infty}\gamma_n<\infty$;   (iv) the $\displaystyle\lim_{n\to\infty}\beta_n$, $\displaystyle\lim_{n\to\infty}l_n$, $\displaystyle\lim_{n\to\infty}t_n$, $\displaystyle\lim_{n\to\infty}s_n$, $\displaystyle\lim_{n\to\infty}w_n$ are exits.

Then, the following holds:

(I) The $\displaystyle\lim_{n\to\infty}d(x_n,F)$ exits, where $\displaystyle\lim_{n\to\infty}d(x_n,F)=\inf_{p\in F}\|x_n-p\|$;

(II) If T_l is semi-compact set, then the $\{x_n\}$ defined by (1) convergence strongly to a common fixed point of $\{T_i\}_{i=1}^{N}$.

Proof.
$$\| y_n - p \|\le l_n \| x_n - p \| + s_n \| x_{n-1} - p \| +$$
$$t_n \| S_{i(n)}^{k(n)} x_n - p \| + w_n \| S_{i(n)}^{k(n)} x_{n-1} - p \|$$
$$\le (l_n + t_n(1+\delta k_n)) \| x_n - p \| + (s_n + w_n(1+\delta k_n))$$
$$\| x_{n-1} - p \| + \delta(t_n + w_n)e_n \ . (2.1)$$

$$\| x_n - p \|\le \alpha_n \| x_{n-1} - p \| + \beta_n \| S_{i(n)}^{k(n)} y_n - p \| + \gamma_n \| u_n - p \|$$
$$\le \alpha_n \| x_{n-1} - p \| + \beta_n(1+\delta k_n) \| y_n - p \| + \beta_n\delta e_n + \gamma_n \| u_n - p \|$$
$$\le \alpha_{n-1} \| x_{n-1} - p \| + \beta_n(1+\delta k_n)\{(l_n + t_n(1+\delta k_n)) \| x_n - p \|$$
$$+ (s_n + w_n(1+\delta k_n)) \| x_{n-1} - p \| + (t_n + w_n)\delta e_n\} + \beta_n\delta e_n + \gamma_n \| u_n - p \|$$

$$\le [\alpha_n + \beta_n(1+\delta k_n)(s_n + w_n(1+\delta k_n))] \| x_{n-1} - p \|$$
$$+ \beta_n(1+\delta k_n)(l_n + t_n(1+\delta k_n)) \| x_n - p \| \qquad (2.1)$$
$$+[\beta_n(1+\delta k_n)(t_n + w_n) + \beta_n]\delta e_n + \gamma_n \| u_n - p \|$$

Let $\tau_n = [\beta_n(1+\delta k_n)(t_n + w_n) + \beta_n]\delta e_n + \gamma_n \| u_n - p \|$.

We have $\displaystyle\sum_{n=1}^{\infty}\tau_n<\infty$ from conditions (i)-(iv) and also:

$$[1 - \beta_n(1+\delta k_n)(l_n + t_n(1+\delta k_n))] \| x_n - p \|$$
$$\le [\alpha_n + \beta_n(1+\delta k_n)(s_n + w_n(1+\delta k_n))] \| x_{n-1} - p \| +\tau_n .$$

$$1 - \beta_n(1+\delta k_n)(l_n + t_n(1+\delta k_n)) \ge \frac{a}{3} > 0 .$$

where n is sufficient large. Hence

$$|| x_n - p || \le \frac{\alpha_n + \beta_n(1 + \delta k_n)(s_n + w_n(1 + \delta k_n))}{1 - \beta_n(1 + \delta k_n)(l_n + t_n(1 + \delta k_n))} || x_{n-1} - p ||$$

$$+ \frac{\tau_n}{1 - \beta_n(1 + \delta k_n)(l_n + t_n(1 + \delta k_n))}$$

$$\le \{1 + \frac{2(\beta_n k_n^2 - \gamma_n)}{a}\} || x_{n-1} - p || + \frac{2\tau_n}{a}$$

From conditions (i)-(iv) we have:

$$\sum_{n=1}^{\infty} k_n^2 < \infty, \sum_{n=1}^{\infty} \frac{2(\beta_n k_n^2 - \gamma_n)}{a} < \infty, \sum_{n=1}^{\infty} \frac{2\tau_n}{a} < \infty .$$

Hence   from   lemma   1.2   the   conclusion   (1)   holds:
$$\lim_{n \to \infty} d(x_n, F) = \inf_{p \in F} || x_n - p || \text{ exist.}$$

Let $$\lim_{n \to \infty} || x_n - p || = d, d \ge 0. \qquad (2.3)$$

From (2.1), (2.3) and (iv), we also have $\lim_{n \to \infty} || y_n - p || = d$.

Thus $\{x_n\}, \{y_n\}, \{u_n - S_{i(n)}^{k(n)} y_n\}$ is bounded sequences.

On one hand,

$$d = \lim_{n \to \infty} || x_n - p ||$$

$$= \lim_{n \to \infty} || \alpha_n[x_{n-1} - p + \gamma_n(u_n - S_{i(n)}^{k(n)}y_n)] + (1 - \alpha_n)$$

$$[S_{i(n)}^{k(n)}y_n - p + \gamma_n(u_n - S_{i(n)}^{k(n)}y_n)] || \qquad (2.4)$$

$$\limsup_{n \to \infty} || x_{n-1} - p + \gamma_n(u_n - S_{i(n)}^{k(n)}y_n) || \le \limsup_{n \to \infty} || x_{n-1} - p ||$$

$$+ \limsup_{n \to \infty} \gamma_n || u_n - S_{i(n)}^{k(n)}y_n || \le d.$$

$$\limsup_{n \to \infty} \left\| S_{i(n)}^{k(n)} y_n - p + \gamma_n(u_n - S_{i(n)}^{k(n)} y_n) \right\|$$

$$\le \limsup_{n \to \infty} [(1 + \delta k_n) || y_n - p || + \delta e_n] + \limsup_{n \to \infty} \gamma_n || u_n - S_{i(n)}^{k(n)} y_n ||$$

From lemma 1.2 we have:

$$\lim_{n \to \infty} || S_{i(n)}^{k(n)} y_n - x_{n-1} || = 0. \qquad (2.5)$$

Thus we derive at

$$\lim_{n\to\infty} \| x_n - x_{n-1} \| = \lim_{n\to\infty} \| \alpha_n x_{n-1} - x_{n-1} + (1 - \alpha_n - \gamma_n) S_{i(n)}^{k(n)} y_n + \gamma_n u_n \|$$

$$= \lim_{n\to\infty} [(1 - \alpha_n) \| S_{i(n)}^{k(n)} y_n - x_{n-1} \| + \gamma_n \| u_n - S_{i(n)}^{k(n)} y_n \|]$$

$$\leq \lim_{n\to\infty} (1 - \alpha_n) \| S_{i(n)}^{k(n)} y_n - x_{n-1} \| = 0.$$

So $\forall i = 1,2,...N, \quad \lim_{n\to\infty} \| x_n - x_{n+1} \| = 0$. \hfill (2.6)

On the other hand,

$$\| S_{i(n)}^{k(n)} x_n - x_n \| \leq \| x_n - x_{n-1} \| + \| S_{i(n)}^{k(n)} y_n - x_{n-1} \| + \| S_{i(n)}^{k(n)} y_n - S_{i(n)}^{k(n)} x_n \|$$

$$\leq \| x_n - x_{n-1} \| + \| S_{i(n)}^{k(n)} y_n - x_{n-1} \| + (1 + \delta k_n)[\| y_n - x_{n-1} \|$$

$$+ \| x_n - x_{n-1} \|] + \delta e_n \ . \hfill (2.7)$$

$$\| y_n - x_{n-1} \| = \| l_n x_n + s_n x_{n-1} + t_n S_{i(n)}^{k(n)} x_n + w_n S_{i(n)}^{k(n)} x_{n-1} - x_{n-1} \|$$

$$\leq t_n \| S_{i(n)}^{k(n)} x_n - x_n \| + w_n \| S_{i(n)}^{k(n)} x_{n-1} - S_{i(n)}^{k(n)} x_n \| +$$

$$w_n \| S_{i(n)}^{k(n)} x_n - x_n \| + (l_n + t_n + w_n) \| x_n - x_{n-1} \|$$

$$\leq (t_n + w_n) \| S_{i(n)}^{k(n)} x_n - x_n \| + [w_n(1 + \delta k_n) + (l_n + t_n + w_n)]$$

$$\| x_n - x_{n-1} \| + w_n \delta e_n \hfill (2.8)$$

Substituting (2.8) into (2.7)

$$\| S_{i(n)}^{k(n)} x_n - x_n \| \leq [w_n(1 + k_n)^2 + (1 + k_n)(2 - s_n) + 1] \| x_n - x_{n-1} \|$$

$$+ \| S_{i(n)}^{k(n)} y_n - x_{n-1} \| + (1 + \delta k_n)(t_n + w_n)$$

$$\| S_{i(n)}^{k(n)} x_n - x_n \| + (1 + \delta k_n) w_n \delta e_n + \delta e_n \ ,$$

$$[1 - (1 + \delta k_n)(t_n + w_n)] \| S_{i(n)}^{k(n)} x_n - x_n \|$$

$$\leq [w_n(1 + k_n)^2 + (1 + k_n)(2 - s_n) + 1] \| x_n - x_{n-1} \| +$$

$$\| S_{i(n)}^{k(n)} y_n - x_{n-1} \| + [(1 + \delta k_n) w_n + 1] \delta e_n \ ,$$

$$[1 - (1 + \delta k_n)(t_n + w_n)] \| T_{i(n)}^{k(n)} x_n - x_n \| \leq \frac{1}{\delta}[w_n(1 + k_n)^2 + (1 + k_n)(2 - s_n) + 1] \| x_n - x_{n-1} \|$$

where n is sufficient large and from condition (ii) and (iv):

$$1 - (1 + \delta k_n)(t_n + w_n) \geq \frac{a}{3} > 0.$$

From (2.5), (2.6) we can have

$$\lim_{n\to\infty} \| T_{i(n)}^{k(n)} x_n - x_n \| = 0. \tag{2.9}$$

We could see that T_l is full-continuous since it is continuous and compact, hence there must exist $\{T_l^{k(n_i)} x_{n_i}\} \subset \{T_l^{k(n)} x_n\}$ such that $\{T_l^{k(n_i)}\} x_{n_i} \to q$, for $\forall n > N$, $n = (k(n)-1)N + i(n)$, $i(n) \in \{1,2,...N\}$, thus

$$\| T_n x_n - x_n \| \leq \| x_n - T_{i(n)}^{k(n)} x_n \| + \| T_{i(n)}^{k(n)} x_n - T_n x_n \| \leq \| x_n - T_{i(n)}^{k(n)} x_n \|$$

$$+(1 + k_n) \| T_{i(n)}^{k(n)-1} x_n - x_n \| + e_n$$

$$\leq \| x_n - T_{i(n)}^{k(n)} x_n \| + (1 + k_n) \| T_{i(n)}^{k(n)-1} x_n - T_{i(n-N)}^{k(n-N)} x_n \|$$

$$+(1 + k_n)[(1 + k_n) \| x_n - x_{n-N} \| + e_n] + (1 + k_n)$$

$$\| T_{i(n-N)}^{k(n-N)} x_{n-N} - x_{n-N} \| + (1 + k_n) \| x_{n-N} - x_n \| + e_n$$

And from $n > N$, $n = (n - N)(\mod N)$, we have the following:

$$n - N = [(k(n) - 1) - 1]N + i(n) = [(k(n - N) - 1)]N + i(n - N)$$

that is, $k(n - N) = k(n) - 1, i(n - N) = i(n)$. From (2.6), (2.9) and above, we can have

$$\lim_{n\to\infty} \| T_n x_n - x_n \| = 0. \tag{2.10}$$

$\| x_n - T_{n+i} x_n \| \leq \| x_n - x_{n+i} \| + \| x_{n+i} - T_{n+i} x_{n+i} \| + (1 + k_n) \| x_{n+i} - x_n \| + e_n$.

From (2.6), (2.10) and $\lim_{n\to\infty} e_n = 0$.

We have $\lim_{n\to\infty} \| T_{n+i} x_n - x_n \| = 0$.

Therefore, $\lim_{n\to\infty} \| T_l x_n - x_n \| = 0, \forall l = 1,2,..., N$

By the fact that $T_n = T_{n(\mod N)}$.

Finally, because of $T_1, T_2, \cdots, T_N$ are uniformly continuous and $x_{n_i} \to q \in D$, thus $T_l q = q$, $q \in F(T_l)$. Since $\lim_{n\to\infty} \| x_n - q \|$ exist, therefore $\lim_{n\to\infty} \| x_n - q \| = 0$ that is, $\{x_n\}$ convergence strongly to a common fixed point to $\{T_i\}_{i=1}^N$.

References

[1] Tan, K.K., Xu, H.K.: Approximation Fixed Points of Nonexpansive Mappings by the Ishikawa teration Process. J. Math. Anal. Appl. 286, 351–358 (2003)

[2] Schu, J.: Weak and Strong Convergence of Fixed Points of Asymptotically Nonexpansive Mappings. Bull. Austral. Math. Soc. 43, 153–159 (1991)

[3] Chang, S.S., Tan, K.K., Lee, H.W.J., Chan, C.K.: On the Convergence of Implicit Iteration Process with Error for a Finite Family of Asymptotically Nonexpansive Mappings. J. Math. Anal. Appl. 313, 273–283 (2006)

[4] Thakur, B.S.: Weak and Strong Convergence of Composite Implicit Iteration Process. Applied Mathematics and Computation 190, 965–973 (2007)

[5] Nlsrakoow, Saejung, S.: A new three-step fixed point iteration scheme for asymptotically nonexpansive mappings. Appl. Math. Comput. 181, 1026–1034 (2006)

[6] Jeong, J.U., Kim, S.H.: Weak and strong convergence of the Ishikawa iteration process with errors for two asymptotically non-expansive mappings. Appl. Math. Comput. 181, 1394–1401 (2006)

Strong Convergence Theorems for Lipschitzian Demi-Contraction Semigroups in Banach Spaces

Wu Dingping

College of Mathematics, Chengdu University of Information Technology Chengdu, Sichuan 610225, China
`wdp68@163.com`

Abstract. The purpose of this paper is to introduce and study the strong convergence problem of the explicit iteration process for a Lipschitzian and demi-contraction semi-groups in arbitrary Banach spaces. The main results presented in this paper extend and improve some recent results announced by many authors.

Keywords: Lipschitzian, Demi-contraction semi-groups, The strong convergence.

1 Introduction and Preliminaries

Throughout this paper, we assume that E is a real Banach space, E^* is the dual space of E, C is a nonempty closed convex subset of E, R^+ is the set of nonnegative real numbers and $J : E \to 2^{E^*}$ is the normalized duality mapping defined by

$$J(x) = \left\{ f \in E^* : \langle x, f \rangle = \|x\| \cdot \|f\|, \|x\| = \|f\| \right\}$$

for all $x \in E$. Let $T : C \to C$ be a mapping. We use $F(T)$ to denote the set of fixed points of T and $" \to "$ to stand for strong convergence.

Definition 1.1. Let the one-parameter family $\Gamma := \{T(t) : t \geq 0\}$ of mappings from C into itself.

(1). Γ is called a Lipschitzian if the following conditions are satisfied: for each $x \in C$

(a) $T(0)x = x$;

(b) $T(s+t)x = T(t)T(s)x$ for any $t, s \in R^+$.

(c) the mapping $t \mapsto T(t)x$ is continuous;

(d) there exists a bounded measurable function $L : [0, \infty) \to (0, \infty)$, such that, for any $x, y \in C$,

$$\|T(t)x - T(t)y\| \leq L(t)\|x - y\|, \text{ for any } t \geq 0.$$

We denote $L = \sup_{t \geq 0} L(t) < \infty$, it is to see that every non-expansive semigroup is a Lipschitzian and pseudocontraction semigroup with $L(t) \equiv 1$.

(2). Γ is called a demi-contraction semigroup if $F(T(t)) \neq 0$ for any $t > 0$ and the conditions (a)-(c) and the following condition (e) are satisfied:

(e) there exists a bounded function $\lambda : [0, \infty) \rightarrow (0, \infty)$ such that , for any $t > 0, x \in C$ and $p \in F(T(t))$ there exists $j(x - y) \in J(x - y)$ such that

$$\langle T(t)x - p, \; j(x - p) \rangle \leq \|x - p\|^2 - \lambda(t) \|(I - T(t))x\|^2. \tag{1.1}$$

In this case, we also say Γ that is a $\lambda(t)$−demi-contractive semigroup..

Remark 1.2. It is easy to see that the condition (1.1) is equivalent to the following condition: for any $t > 0, x \in C$ and $p \in F(T(t))$,

$$\langle x - T(t)x, \; j(x - p) \rangle \geq \lambda(t) \|x - T(t)x\|^2. \tag{1.2}$$

The convergence problems of the implicit or explicit iterative sequences for non-expansive semigroup to a common fixed has been considered by some authors in the settings of Hilbert or Banach spaces (see, for example, [1-10]).

In 1998, Shioji and Takahashi [2] introduced the folloeing implicit iteration:

$$x_n = \alpha_n u + (1 - \alpha_n) \sigma_{t_n} x_n \tag{1.3}$$

for each $n \geq 1$ in a Hlbert space, where $\{\alpha_n\} \subset (0\ 1)$, $\{t_n\}$ is a sequence of positive real numbers divergent to ∞ and, for any $t > 0$ and $x \in C$, $\sigma_t(x)$ is the average given by

$$\sigma_t(x) = \frac{1}{t} \int_0^t T(t)x \, ds.$$

Under certain restrictions to the sequence of $\{\alpha_n\}$, they proved some strong convergence theorems of $\{x_n\}$ to a point $p \in \Psi := \cap_{t \geq 0} F(T(t))$.

In 2003, Suzuki [3] first introduced the following implicit iterative sequences process for nonexpansive semigroup in a Hilbert space:

$$x_n = \alpha_n u + (1 - \alpha_n) T(t_n) x_n \tag{1.4}$$

for each $n \geq 1$. Under appropriate assumptions imposed upon the sequences $\{\alpha_n\}$ and $\{t_n\}$, he proved that the sequence $\{x_n\}$ defined by (1.4) convergences strongly to a common fixed point of the nonexpansive semigroup.

In 2005, Xu [5] proved that Suzuki's result holds in a uniformly convex Banach space with a weakly continuous duality mapping. At the same time, he also raised the following open question:

Open Question([5]). We do not know whether or not the same result holds in a uniformly convex and uniformly smooth Banach space?

In 2005, Aleyner and Reich [1] first introduced the following explicit iteration sequence:

$$x_n = \alpha_n u + (1 - \alpha_n) T(t_n) x_n \tag{1.5}$$

for each $n \geq 1$ in a reflexive Banach space with a uniformly Gateaux differentiable norm such that each nonempty bounded closed and convex subset of E has the common fixed point property for nonexpansive mappings (Note that all these assumptions are fulfilled whenever E is uniformly smooth [3]) .

Also, under appropriate assumptions imposed upon the parameter sequences sequence $\{\alpha_n\}$ and $\{t_n\}$, they proved that the sequence $\{x_n\}$ defined by (1.5) converges strongly to some point in $\Psi := \bigcap_{t \geq 0} F(T(t))$.

Recently, in 2007, Chang, Yang and Liu [6] introduced the following composite iteration scheme in the framework of reflexive Banach with a uniformly Gateaux differentiable norm, uniformly convex Banach space and uniformly smooth Banach space with a weakly continuous duality mapping:

$$\begin{cases} x_{n+1} = \alpha_n u + (1 - \alpha_n) y_n, \\ y_n = \beta_n x_n + (1 - \beta_n) T(t_n) x_n \end{cases} \tag{1.6}$$

for each $n \geq 0$ for the nonexpansive semigroup $\Gamma := \{T(t) : t \geq 0\}$ of mappings from C into itself , where u is an arbitrary (but fixed) element in C and the sequences $\{\alpha_n\} \subset (0,1)$, $\{\beta_n\} \subset [0,1]$, $\{t_n\} \subset R^+$, and proved some strong convergence theorems for the iteration sequence $\{x_n\}$. In fact, the results presented in [4] not only extend and improve the corresponding results of Shioji-Takahashi [2], Suzuki [3], Xu [5] and Aleyner-Reich [1], but also give a partially affirmative answer for the open questions raised by Suzuki [3] and Xu [5].

In order to improve and develop of the results mentioned above, recently, Zhang [7, 8], by using the different methods, introduce and study the weak convergence problem of the implicit iteration processes for the Lipschitzian and pseudo-contraction semigroups in general Banach spaces. The results given in [7, 8] not only extended the above results, but also extended and improve the corresponding results in Li, Li and Su [9] and Ori [10] .

The purpose of this paper is to introduce and study the strong convergence problem of the following explicit iteration process:

$$\begin{cases} x_0 \in C \\ x_{n+1} = (1 - \alpha_n) x_n + \alpha_n T(t_n) y_n \\ y_n = (1 - \beta_n) x_n + \beta_n T(t_n) x_n , \end{cases} \tag{1.7}$$

for each $n \geq 1$ for the Lipschitzian and demi-contractive semigroup $\Gamma := \{T(t) : t \geq 0\}$ in general Banach spaces.

The results presented in this paper improve, extend and replenish the corresponding results given in [1-10] and many authors.

In the sequel, we make use the following lemmas for our main results.

Lemma 1.3. Let $J : E \to 2^{E^*}$ be the normalized duality mapping. Then for any $x, y \in E$,

$$\|x + y\|^2 \le \|x\|^2 + 2\langle y, j(x + y)\rangle$$

for all $j(x+y) \in J(x+y)$.

Lemma 1.4. ([6]) Let $\{\alpha_n\}$ and $\{\sigma_n\}$ be the sequences of nonnegative real numbers satisfying the following condition:

$$\alpha_{n+1} \le (1 + \sigma_n)\alpha_n$$

for all $n \ge n_0$, where n_0 is some nonnegative integer. If $\sum_{n=1}^{\infty}\sigma_n < \infty$, then the $\lim$ $\lim_{n\to\infty}\alpha_n$ exists. In addition, if there exists a sequence $\{\alpha_{n_i}\}$ of $\{\alpha_n\}$ such that $\alpha_{n_i} \to 0$, then $\lim_{n\to\infty}\alpha_n = 0$.

2   Main Results

Theorem 2.1. Let E is a real Banach space, C is a nonempty closed convex subset of E, Γ be a Lipschitzian and demi-contractive semigroup with a bounded measurable function $L(t)$ and a bounded function $\lambda(t)$ respectively, such that

$$L := \sup_{t \ge 0} L(t) < \infty, \quad \lambda := \inf_{t \ge 0} \lambda(t) > 0, \quad \Psi := \bigcap_{t \ge 0} F(T(t)) \ne \phi.$$

Let $\{x_n\}$ be the sequence defined by (1.3), where $\{\alpha_n\}, \{\beta_n\} \subset [0,1]$ and $\{t_n\}$ is a increasing sequence in $[0, \infty)$.

If there exists a compact subset K of E such that $\cup_{t \ge 0} T(t)(C) \subset K$ and the following conditions are satisfied :

(a) $\sum_{n=1}^{\infty}\alpha_n = \infty, \sum_{n=1}^{\infty}\alpha_n^2 < \infty, \sum_{n=1}^{\infty}\alpha_n\beta_n < \infty$;

(b) for any bounded subset $D \subset C$, $\lim_{n\to\infty}\sup_{x \in D, s \in R^+}\|T(s + t_n)x - T(t_n)x\| = 0$.

Then $\{x_n\}$ converges strongly to a common fixed point of the semigroup Γ.

Proof. (I) $\lim_{n\to\infty}\|x_n - p\|$ exists   for all $p \in \Psi$.

In fact, for any $p \in \Psi$, we have

$$\|y_n - p\| = \|(1 - \beta_n)x_n + \beta_n x_n - p\| \le (1 - \beta_n)\|x_n - p\| + \beta_n L\|x_n - p\|$$
$$\le (1 + L)\|x_n - p\|. \tag{2.1}$$

It follows from (2.1) that

$$\|x_{n+1} - p\| = \|(1 - \alpha_n)x_n + \alpha_n y_n - p\| \le (1 - \alpha_n)\|x_n - p\| + \alpha_n L\|y_n - p\|$$
$$\le (1 + L + L^2)\|x_n - p\|. \tag{2.2}$$

And from (1.7) ,we have

$$
\begin{aligned}
\left\| x_{n+1} - y_n \right\| &= \left\| (1-\alpha_n)x_n - \alpha_n T(t_n)y_n - (1-\beta_n)x_n - \beta_n T(t_n)x_n \right\| \\
&\leq \left| \alpha_n - \beta_n \right| \left\| x_n - T(t_n)x_n \right\| + \alpha_n L \left\| x_n - y_n \right\| \\
&\leq \left| \alpha_n - \beta_n \right| (\left\| x_n - p \right\| + \left\| T(t_n)x_n - p \right\|) + \alpha_n L \left\| x_n - (1-\beta_n)x_n - \beta_n T(t_n)x_n \right\| \\
&\leq (1+L)(\left| \alpha_n - \beta_n \right| + \alpha_n L) \left\| x_n - p \right\| .
\end{aligned}
$$

(2.3)

From (2.1)-(2.3), we have

$$
\begin{aligned}
\left\| x_{n+1} - x_n \right\| &= \left\| (1-\alpha_n)x_n + \alpha_n T(t_n)y_n - x_n \right\| \\
&= \alpha_n \left\| T(t_n)y_n - x_n \right\| \\
&\leq \alpha_n L \left\| y_n - p \right\| + \alpha_n \left\| x_n - p \right\| \\
&= \alpha_n L \left\| (1-\beta_n)x_n + \beta_n T(t_n)x_n - p \right\| + \alpha_n \left\| x_n - p \right\| \\
&\leq \alpha_n (1 + L + L^2) \left\| x_n - p \right\| .
\end{aligned}
$$

(2.4)

Since is Γ an demi-contractive semigroup with $\lambda > 0$, for the point x_{n+1} and p, there exists $j(x_{n+1} - p) \in J(x_{n+1} - p)$ such that

$$
\left\langle x_{n+1} - T(t_n)x_{n+1}, \, j(x_{n+1} - p) \right\rangle \geq \lambda \left\| x_{n+1} - T(t_n)x_{n+1} \right\|^2 .
$$

(2.5)

Thus, by Lemma 1.3 and (2.2)-(2.5), we have

$$
\begin{aligned}
\left\| x_{n+1} - p \right\|^2 &= \left\| (x_n - p) + \alpha_n (T(t_n)y_n - x_n) \right\|^2 \\
&\leq \left\| x_n - p \right\|^2 + 2\alpha_n \left\langle T(t_n)y_n - x_n, \, j(x_{n+1} - p) \right\rangle \\
&= \left\| x_n - p \right\|^2 + 2\alpha_n \left\langle T(t_n)y_n - T(t_n)x_{n+1}, \, j(x_{n+1} - p) \right\rangle \\
&\quad - 2\alpha_n \left\langle x_{n+1} - T(t_n)x_{n+1}, \, j(x_{n+1} - p) \right\rangle + 2\alpha_n \left\langle x_{n+1} - x_n, \, j(x_{n+1} - p) \right\rangle \\
&\leq \left\| x_n - p \right\|^2 + 2\alpha_n \left\| T(t_n)y_n - T(t_n)x_{n+1} \right\| \cdot \left\| x_{n+1} - p \right\| - 2\alpha_n \lambda \left\| x_{n+1} - T(t_n)x_{n+1} \right\|^2 \\
&\quad + 2\alpha_n \left\| x_{n+1} - x_n \right\| \cdot \left\| x_{n+1} - p \right\| \\
&\leq \left\| x_n - p \right\|^2 + 2\alpha_n L \left\| y_n - x_{n+1} \right\| \cdot \left\| x_{n+1} - p \right\| + 2\alpha_n \left\| x_{n+1} - x_n \right\| \cdot \left\| x_{n+1} - p \right\| \\
&\quad - 2\alpha_n \lambda \left\| x_{n+1} - T(t_n)x_{n+1} \right\|^2 \\
&\leq [1 + 2\alpha_n L(1+L)(\left| \alpha_n - \beta_n \right| + \alpha_n \beta_n L)(1 + L + L^2) \\
&\quad + 2\alpha_n^2 (1 + L + L^2)^2] \left\| x_n - p \right\|^2 - 2\alpha_n \lambda \left\| T(t_n)x_{n+1} - x_{n+1} \right\|^2 \\
&\leq \left[1 + 2(1 + L + L^2)^2((2 + L)\alpha_n^2 + \alpha_n \beta_n) \right] \left\| x_n - p \right\|^2 - \\
&\quad 2\alpha_n \lambda \left\| T(t_n)x_{n+1} - x_{n+1} \right\|^2 .
\end{aligned}
$$

(2.6)

This implies that $\left\| x_{n+1} - p \right\|^2 \leq [1 + 2(1 + L + L^2)^3 (\alpha_n^2 + \alpha_n \beta_n)] \left\| x_n - p \right\|^2$.

By the assumption (a), it follows from Lemma 1.4 that the limit $\lim_{n\to\infty}\|x_n - p\|$ exists for all $p \in \Psi$, so the sequence $\{x_n\}$ is bounded in C.

(II) $\liminf_{n\to\infty}\|x_n - T(t_n)x_n\| = 0$.

We first prove that

$$\liminf_{n\to\infty}\|x_{n+1} - T(t_n)x_{n+1}\| = 0. \tag{2.7}$$

Suppose $\liminf_{n\to\infty}\|x_{n+1} - T(t_n)x_{n+1}\| = \delta > 0$. There exists a positive integer n_0 such that

$$\|x_{n+1} - T(t_n)x_{n+1}\| \geq \delta\big/2, \text{ for each } n \geq n_0.$$

Suppose $M = \sup_{n\geq 1}\|x_n - p\|$. Thus it follows from (2.6) that

$$\|x_{n+1} - p\|^2 \leq \|x_n - p\|^2 - 2\alpha_n\lambda\|T(t_n)x_{n+1} - x_{n+1}\|^2 + 2(1 + L + L^2)^2((2 + L)\alpha_n^2 + \alpha_n\beta_n)\|x_n - p\|$$

$$\leq \|x_n - p\|^2 - \alpha_n\lambda\delta^2/2 + 2(1 + L + L^2)^3(\alpha_n^2 + \alpha_n\beta_n)M^2$$

for each $n \geq n_0$. This implies that

$$\frac{\alpha_n\lambda\delta^2}{2} \leq \|x_n - p\|^2 - \|x_{n+1} - p\|^2 + 2(1 + L + L^2)^3(\alpha_n^2 + \alpha_n\beta_n)M^2,$$

for each $n \geq n_0$. Hence, for each $m \geq n_0$, we have

$$\frac{\lambda\delta^2}{2}\sum_{n=n_0}^{m}\alpha_n \leq \sum_{n=n_0}^{m}(\|x_n - p\|^2 - \|x_{n+1} - p\|^2) + 2(1 + L + L^2)^3 M^2\sum_{n=n_0}^{m}(\alpha_n^2 + \alpha_n\beta_n)$$

$$\leq \|x_{n_0} - p\|^2 + 2(1 + L + L^2)^3 M^2\sum_{n=n_0}^{m}(\alpha_n^2 + \alpha_n\beta_n). \tag{2.8}$$

Let $m \to \infty$ in (2.8), we have

$$\frac{\lambda\delta^2}{2}\sum_{n=n_0}^{\infty}\alpha_n \leq \|x_{n_0} - p\|^2 + 2(1 + L + L^2)^3 M^2\sum_{n=n_0}^{\infty}(\alpha_n^2 + \alpha_n\beta_n),$$

which is a contradiction by the condition (a). Hence the conclusion (2.7) is proved.

Since $\{x_n\}$ is bounded and $\{t_n\}$ is increasing, it follows from (2.7) and the condition (b) that

$$\liminf_{n\to\infty}\|x_{n+1} - T(t_{n+1})x_{n+1}\| \leq \liminf_{n\to\infty}\{\|x_{n+1} - T(t_n)x_{n+1}\| + \|T(t_{n+1})x_{n+1} - T(t_n)x_{n+1}\|\}$$

$$= \liminf_{n\to\infty}\{\|x_{n+1} - T(t_n)x_{n+1}\| + \|T((t_{n+1} - t_n) + t_n)x_{n+1} - T(t_n)x_{n+1}\|\}$$

$$\leq \liminf_{n\to\infty}\left\{\|x_{n+1} - T(t_n)x_{n+1}\| + \sup_{z\in\{x_n\}, s\in R^+}\|T(s + t_n)z - T(t_n)z\|\right\} = 0.$$

By above (I), (II), and by the assumption, it follows that there exists a compact subset K of E such that $\cup_{t \geq 0} T(t)(C) \subset K$ and so there exists a subsequence $\{x_{n_k}\} \subset \{x_n\}$ such that

$$\lim_{n_k \to \infty} \|x_{n_k} - T(t_{n_k})x_{n_k}\| = 0, \quad \lim_{n_k \to \infty} T(t_{n_k})x_{n_k} = q. \tag{2.9}$$

for some point $q \in C$. Hence it follows from (2.9) that $x_{n_k} \to q$ as $n_k \to \infty$.

Next, we prove that

$$\lim_{n_k \to \infty} \|x_{n_k} - T(t)x_{n_k}\| = 0 \tag{2.10}$$

for all $t \geq 0$. In fact, it follows from the condition (b) and (2.10) that, for any $t > 0$,

$$\|x_{n_k} - T(t)x_{n_k}\| \leq \|T(t)x_{n_k} - T(t + t_{n_k})x_{n_k}\| +$$

$$\|T(t_{n_k})x_{n_k} - T(t + t_{n_k})x_{n_k}\| + \|x_{n_k} - T(t_{n_k})x_{n_k}\|$$

$$\leq (1 + L)\|x_{n_k} - T(t_{n_k})x_{n_k}\| + \sup_{z \in \{x_n\}, s \in R^+} \|T(s + t_{n_k})z - T(t_{n_k})z\| \to 0 \tag{2.11}$$

as $n_k \to \infty$. Since $x_{n_k} \to q$ as $n_k \to \infty$ and the semigroup Γ is Lipschitzian, it follows from (2.11) that $q = T(t)(q)$ for all $t \geq 0$, that is, $q \in \Psi$. Since $x_{n_k} \to \infty$ as $n_k \to \infty$ and the limit $\lim_{n \to \infty} \|x_n - q\|$ exists, which implies that $x_n \to q \in \Psi$ as $n \to \infty$. This completes the proof.

Remark 1.3. Theorem 2.1 not only extend and improve the corresponding results of Shioji and Takahashi [2], Suzuki[3], Xu [4] and Aleyner-Reich [1], but also give a partially affirmative answer for the open questions raised by Suzuki[3] and Xu [5].

References

[1] Aleyner, A., Reich, S.: An explicit construction of sunny nonexpansive retractions in Banach spaces. Fixed Point Theory and Appl. 3, 295–305 (2005)

[2] Shioji, N., Takahashi, W.: Strong convergence theorems for asymptotically nonexpansive mappings in Hilbert spaces. Nonliear Anal. 34, 87–99 (1998)

[3] Suzuki, T.: On strong convergence to a common fixed point of nonexpansive semigroup in Hilbert spaces. Proc. Amer. Math. Soc. 131, 2133–2136 (2003)

[4] Xu, H.K.: Astrong convergence theorem for contraction semigroups in Banach spaces. Bull. Austral. Math. Soc. 72, 371–379 (2005)

[5] Xu, H.K.: Inequonualities in Banach spaces with applications. Nonliear Anal. 16, 1127–1138 (1991)

 [6] Chang, S.S., Yang, L., Liu, J.A.: Strong convergence theorem for nonexpansive semi-groups in Banach spaces. Appl. Math. Mech. 28, 1287–1297 (2007)
 [7] Zhang, S.S.: Convergence theorem of common fixed points for Lipschitzian pseudo-contraction semi-groups in Banach spaces. Appl. Math. Mech. 30, 145–152 (2009)
 [8] Zhang, S.S.: Weak convergence theorem for Lischitzian pseudo-contract semi-groups in Banach spaces. Acta Math. Sinica, Chinese Series 26, 337–344 (2010)
 [9] Li, S., Li, L.H., Su, F.: General iterative methods for a one-parameter nonexpansive semi-group in Hilbert space. Nonlinear Anal. (2008), doi:10.1016/j.na.2008.04.007
[10] Osilike, M.O.: Implicit iteration process for common fixed points of a finite family of strictly pseudocontractive maps. J. Math. Anal. Appl. 294, 73–81 (2004)

Viscosity Approximation Methods for a Finite Family of Nonexpansive Non-Self-Mappings

Wu Dingping and Xie Xuping

College of Mathematics, Chengdu Univ. of Information Technology, Chengdu, 610225,
P.R. China
wdp68@163.com

Abstract. We introduce a viscosity approximation method for a finite family of nonexpansive non-selfmappings. Both strong convergence theorem and application of the main result are given. The result presented in this paper generalizes the result in [8,10].

Keywords: Nonexpansive non-self-mapping, Contraction, Common fixed points, Variational inequality, Weakly asymptotically regular.

1 Introduction and Preliminaries

The convergence about viscosity approximation methods for a finite family of nonexpansive non-self-mappings has been studied by many authors see [1-7]. Let E be a real Banach space and E^* be its dual. Gauge function that is continuous strictly increasing function φ defined on $R^+ := [0,\infty)$ such that $\varphi(0) = 0$ and $\lim_{r\to\infty} \varphi(r) = \infty$.

The mapping $J_\varphi : E \to 2^{E^*}$ defined by

$$J_\varphi(x) = \{ f \in E^* : (x, f) = \| x \| \| f \| \}, \| f \| = \varphi(\| x \|),$$

for all $x \in E$ is called duality mapping with gauge φ .

We know that a Banach Space E has a weakly sequential continuous duality mapping if there exists a gauge function φ such that the duality mapping J_φ is single-valued and continuous from the weak topology to the weak* topology.

Set $\phi(t) = \int_0^1 \varphi(t)dt$ for all $t \in R^+$, then it is well known that $J_\varphi(x)$ is the subdifferential of the convex function $\phi(\|\cdot\|)$ at x.Thus is easy to see that the normalized mapping $J(x)$ can also be defined as the subdifferential of the convex functional $\phi(\| x \|) = \| x \|^2 / 2$,that is,for all $x \in E$

$$J(x) = \partial\phi(\| x \|) = \{ f \in E^* : \phi(\| x \|) - \phi(\| y \|) \geq \langle y - x, f \rangle \text{ for all } y \in E \}.$$

and also know if E is smooth,then the normalized duality mapping is single-valued and norm to weak* continuous.

Recall a mapping $Q : C \to D$ is said to be retraction from C onto D if $Qx = x$ for all $x \in D$. A retraction $Q : C \to D$ is said to be sunny if $Q(Qx + t(x - Qx)) = Qx$ for all $t \geq 0$ and $x + t(x - Qx) \in C$.

A sunny nonexpansive retraction is a sunny retraction which is also nonexpansive, if E is a smooth Banach space, then $Q : C \to D$ is a sunny nonexpansive retraction if and only if the following holds:

$$\langle x - Qx, J(z - Qx) \rangle \leq 0, \text{ for all } x \in C, z \in D.$$

μ is Banach limit if $\| \mu \| = \mu_n(1) = 1$, $\mu_n(a_{n+1}) = \mu_n(a_n)$ for all $(a_0, a_1 \cdots) \in l^\infty$, and the following are well known:

(i) $\mu(a_{n+N}) = \mu(a_n)$ for any fixed positive integer N.

(ii) $\liminf\limits_{n \to \infty} a_n \leq \mu_n(a_n) \leq \limsup\limits_{n \to \infty} a_n$ for all $a_n(a_0, a_1, \ldots a_n) \in l^\infty$.

Let $\{x_n\}$ defined by the following scheme:

$$\begin{cases} x_{n+1} = \alpha_n f(x_n) + (1 - \alpha_n) Q T_{n+1} y_n \\ y_n = \beta_n x_n + (1 - \beta_n) Q T_{n+1} x_n \end{cases} \tag{1.1}$$

Our purpose is to applying viscosity approximation methods with Banach limit to prove $\{x_n\}$ defined by (1.1) converge strongly to a fixed point $q \in F$ which solves the variational inequality:

$$\langle (I - f)(q), J_\varphi(q - p) \rangle \leq 0, \quad f \in \Sigma_c, p \in F \tag{1.2}$$

The result presented in this paper generalizes the result in [8,10].

We need the following lemmas for the proof of our main results.

Lemma 1.1 [3]. Let E be a real Banach Space and φ a continuous strictly function on R^+ such that $\varphi(0) = 0$ and $\lim\limits_{r \to \infty} \varphi(r) = \infty$. Define $\phi(t) = \int\limits_0^1 \varphi(t) dt$ for all $t \in R^+$.

Then the following inequality holds:

$\phi(\| x + y \|) \leq \phi(\| x \|) + \langle y, j_\varphi(x + y) \rangle \forall x, y \in E$, where $j_\varphi(x + y) \in J_\varphi(x + y)$.

In particular, if E is smooth. Then one has $\| x + y \|^2 \leq \| x \|^2 + 2 \langle y, J(x + y) \rangle$ for $\forall x, y \in E$.

Lemma 1.2 [3]. Let $\{s_n\}$ be a sequence of non-negative real numbers satisfying

$$s_{n+1} \leq (1 - \lambda_n) s_n + \lambda_n \beta_n + \gamma_n, n \geq 0.$$

where $\{\lambda_n\}, \{\beta_n\}, \{\gamma_n\}$ satisfying:

(1) $\{\lambda_n\} \subset [0,1]$ and $\sum_{n=0}^{\infty} \lambda_n = \infty$.

(2) $\limsup_{n \to \infty} \beta_n \leq 0$.

(3) $\gamma_n \geq 0 (n \geq 0)$ and $\sum_{n=0}^{\infty} \gamma_n < \infty$.

Then $\lim_{n \to \infty} s_n = 0$.

Lemma 1.3 [3]. Let $a \in R$ be a real number and asequence $\{a_n\} \in l^{\infty}$ satisify the condition $\mu_n(a_n) \leq a$ for all Banach limit μ if $\limsup_{n \to \infty}(a_{n+N} - a_n) \leq 0$ for $N \geq 1$, then $\limsup_{n \to \infty} a_n \leq a$.

Lemma 1.4 [8]. Let C be a closed convex subset of a smooth Banach space E and T be a mapping from C into E. Suppose that C is a sunny nonexpansive retract of E. If T satisfies the nowhere-normal outward condition:

$$Tx \in S_x^c, \text{ for all } x \in C, \qquad (1.3)$$

where $S_x = \{y \in E : y \neq x, Qy = x\}$ and S_x^c is the complement of S_x.

Then $Fix(T) = Fix(QT)$.

Lemma 1.5 [8]. Let C be a closed convex subset of a strictly convex Banach space E and T be a nonexpansive mapping from C into E. Suppose that C is a sunny nonexpansive retract of E. If $Fix(T) \neq \phi$, then T satisfies the nowhere-normal outward condition(1.3).

2   Main Result

Theorem 2.1. Let E be a strictly convex and reflexive Banach space having a weakly sequentially continuous duality mapping J_{φ} with gauge function φ. Let C be a nonempty closed convex subset of E. Assume that C is a sunny nonexpansive retract of E with Q as the sunny nonexpansive retraction. $\sum_C = \{f : C \to C\}$ f be a contraction with coefficient $\xi \in (0,1)$. Let $T_1 \cdots T_N$ be nonexpansive mappings from C into E with $F := \bigcap_{i=1}^{N} Fix(T_i) \neq \varnothing$ satisfying the following conditions:

$$S := QT_N QT_{N-1} \cdots QT_1 = QT_1 QT_N \cdots QT_2 = \cdots = QT_{N-1} QT_{N-2} \cdots \ QT_1 QT_N,$$

$$\bigcap_{i=1}^{N} Fix(QT_i) = Fix(S).$$

Let $\{\alpha_n\},\{\beta_n\}$ be two sequence in $(0,1)$ which satisfies the following conditions:

$$\text{(i) } \lim_{n\to\infty}\alpha_n = 0 \text{ and } \alpha_n + \beta_n = 1; \text{ (ii) } \sum_{n=1}^{\infty}\alpha_n = \infty.$$

Let $f \in \Sigma_c$ and $x_0 \in C$ may be chosen arbitrarily. If $\{x_n\}$ is weakly asymptotically regular and defined by (1.1).Then $\{x_n\}$ converges strongly to $q \in F$, where q is the unique solution of the variational inequality (1.2).

Proof: Note that the definition of the weak sequential of duality mapping J_φ with gauge function φ implies that E is smooth.

Let $x_t = tf(x_t) + (1-t)Sx_t$ for $S = QT_N QT_{N-1}...QT_1$ and $t \in (0,1)$.

Then by theorem J[9], $\{x_t\}$ strongly converges to a point in $Fix(S) = \bigcap_{i=1}^{N} Fix(QT_i)$, which is denoted by $q = \lim_{t\to 0^+} x_t$, and by Lemma 1.4 and Lemma 1.5, $q \in F = \bigcap_{i=1}^{N} Fix(T_i)$.

First we prove:

$$\limsup_{n\to\infty}((I-f)q, J_\phi(q-x_n)) \le 0 \tag{2.1}$$

Since $\qquad\qquad x_t - x_{n+N} = (1-t)(Sx_t - x_{n+N}) + t(f(x_t) - x_{n+N}).$

By lemma 1.1, we have

$$\phi(\| x_t - x_{t+N} \|) \le \phi((1-t) \| Sx_t - x_{n+N} \|) + t(f(x_t) - x_{n+N}, J_\phi(x_t - x_{n+N})) \tag{2.2}$$

Let $p \in F$. Then

$$\| x_t - p \| \le t \| f(x_t) - p \| + (1-t) \| Sx_t - Sp \| \le \xi \| x_t - p \| + \| f(p) - p \|$$

So $\qquad\qquad \| x_t - p \| \le \dfrac{1}{1-\xi} \| f(p) - p \|.$

Then $\| x_{n+1} - p \| \le \alpha_n \| f(x_n) - p \| + (1-\alpha_n) \| QT_{n+1}y_n - p \| \le \alpha_n \xi \| x_n - p \| + \alpha_n \| f(p) - p \| + (1-\alpha_n) \| y_n - p \|$

For $\qquad\qquad \| y_n - p \| \le (1-\beta_n) \| QT_{n+1}x_n - p \| + \beta_n \| x_n - p \| \le \| x_n - p \|$

Thus

$$\| x_{n+1} - p \| \le (1-\alpha_n(1-\xi)) \| x_n - p \| + \alpha_n \| f(p) - p \| \tag{2.3}$$

Then by the nonexpansive of T_n and $f \in \Sigma_c$

$$\| x_1 - p \| \le (1-\alpha_1) \| T_1 y_0 - p \| + \alpha_1 \| f(x_0) - p \| \le (1-\alpha_1) \| y_0 - p \|$$
$$+ \alpha_1 (\| f(x_0) - f(p) \| + \| f(p) - p \|) \tag{2.4}$$

We can obtain $\| x_{n+1} - p \| < d$, d is a constant.

Hence $\{x_n\}$ is bounded, and so are $\{QT_{n+1}x_n\}$, $\{QT_{n+1}y_n\}$ and $\{f(x_n)\}$.

As a consequence with the control condition (i), we get

$$\| x_{n+1} - QT_{n+1}x_n \| \le \alpha_n \| QT_{n+1}x_n - f(x_n) \| + (1-\alpha_n) \| y_n - x_n \|$$
$$\le \alpha_n \| QT_{n+1}x_n - f(x_n) \| + (1-\alpha_n)(1-\beta_n) \| QT_{n+1}x_n - x_n \| \to 0 \, as(n \to \infty).$$

By using the same method, we have

$$\| x_{n+N} - QT_{n+N}QT_{n+N-1} \cdots QT_{n+1}x_n \| \to 0 \ (\text{as } n \to \infty).$$

And we can also have, $\forall n \ge 1$,

$$x_{n+N} - Sx_n = x_{n+N} - QT_NQT_{N-1} \cdots QT_1 x_n =$$
$$x_{n+N} - QT_{n+N}QT_{n+N-1} \cdots QT_{n+1}x_n \to 0 \ (n \to \infty). \tag{2.5}$$

From (2.5) we have

$$\| Sx_t - x_{n+N} \| \le \| x_t - x_n \| + \sigma_n,$$

where $\sigma_n = \| x_{n+N} - Sx_n \|$, and

$$\left\langle f(x_t) - x_{n+N}, J_\varphi(x_t - x_{n+N}) \right\rangle = \left\langle f(x_t) - x_t, J_\varphi(x_t - x_{n+N}) \right\rangle + \| x_t - x_{n+N} \| \varphi(\| x_t - x_{n+N} \|).$$

Thus it follows from (2.2) that

$$\phi(\| x_t - x_{n+N} \|) \le \phi((1-t)(\| x_t - x_{n+N} \| + \sigma_n))$$
$$+ t \left\langle f(x_t) - x_t, J_\varphi(x_t - x_{n+N}) \right\rangle + \| x_t - x_{n+N} \| \varphi(\| x_t - x_{n+N} \|) \tag{2.6}$$

Applying Banach limit μ to (2.6), we can obtain:

$$\mu_n(\phi(\| x_t - x_{n+N} \|)) \le \mu_n(\phi((1-t)(\| x_t - x_n \| + e_n))) + t\mu_n \left\langle f(x_t) - x_t, J_\varphi(x_t - x_{n+N}) \right\rangle$$
$$+ t\mu_n \| x_t - x_{n+N} \| \phi(\| x_t - x_{n+N} \|). \tag{2.7}$$

And it follows from (2.7) that:

$$\mu_n(\left\langle x_t - f(x_t), J_\varphi(x_t - x_n) \right\rangle) \le \frac{1}{t}\mu_n(\phi((1-t)\| x_t - x_n \|) - \phi(\| x_t - x_n \|))$$

$$\varphi(\| x_t - x_{n+N} \|) = -\frac{1}{t} \left\{ \int_{(1-t)\| x_t - x_n \|}^{\| x_t - x_n \|} \varphi(\tau)d\tau \right\} + \mu_n(\| x_t - x_{n+N} \| \varphi(\| x_t - x_{n+N} \|))$$

$$= \mu_n(\| x_t - x_{n+N} \| (\varphi(\| x_t - x_n \|) - \varphi(\tau_n)))$$

for some τ_n satisfying $(1-t)\| x_t - x_n \| \le \tau_n \le \| x_t - x_n \|$.

Since φ is uniformly continuous on compact intervals of R^+

$$\| x_t - x_n \| - \tau_n \le t \| x_t - x_n \| \le t(\frac{2}{1-\xi} \| f(p) - p \| + \| x_0 - p \|) \to 0 \text{ as } (t \to \infty).$$

And $q = \lim_{t \to 0} x_t$, we conclude from above that

$$\mu_n(\langle (I-f)(q), J_\varphi(q - x_n) \rangle) \le \limsup_{t \to 0} \mu_n(\langle x_t - f(x_n), J_\varphi(x_t - x_n) \rangle) \le 0. \quad (2.8)$$

Since $\{x_n\}$ is bounded, there exists a subsequence $\{x_{n_j}\}$ of $\{x_n\}$ such that

$$\limsup_{n \to \infty}(a_{n+N} - a_n) = \lim_{j \to \infty}(a_{n_j+N} - a_{n_j}) \text{ and } x_{n_j} \xrightarrow{\omega} z \in E.$$

This implies that $x_{n_j+N} \xrightarrow{\omega} z \in E$ since $\{x_n\}$ is weak asymptotically regular.

From the weak sequential continuity of duality mapping J_φ, we have:

$$\omega - \lim_{j \to \infty} J_\varphi(q - x_{n_j+N}) = \omega - \lim_{j \to \infty} J_\varphi(q - x_{n_j}) = J_\varphi(q - z)$$

And so,

$$\limsup_{n \to \infty}(a_{n+N} - a_n) = \lim_{j \to \infty}\langle (I-f)(q), J_\varphi(q - x_{n_j+N}) - J_\varphi(q - x_{n_j}) \rangle = 0.$$

Then Lemma 1.3 implies that $\limsup_{n \to \infty} a_n \le 0$, that is (2.1) holds:

$$\limsup_{n \to \infty}\langle (I-f)(q), J_\varphi(q - x_n) \rangle \le 0.$$

Next, we prove x_n converge strongly to q.

On the other hand, by Lemma 1.1, we derive at

$$\| x_{n+1} - q \|^2 \le (1-\alpha_n)^2 \| QT_{n+1}y_n - q \|^2 + 2\alpha_n(f(x_n) - q, J(x_{n+1} - q))$$
$$\le (1-\alpha_n)^2 \| T_{n+1}y_n - q \|^2 + 2\xi\alpha_n \| x_n - q \| \| x_{n+1} - q \| + 2\alpha_n(f(q) - q, J(x_{n+1} - q))$$
$$\le (1-\alpha_n)^2 \| x_n - q \|^2 + \xi\alpha_n(\| x_n - q \|^2 + \| x_{n+1} - q \|^2) + 2\alpha_n(f(q) - q, J(x_{n+1} - q)).$$

Now, by using gauge function φ, we define for every $n \ge 0$

$$\vartheta_n := \begin{cases} \dfrac{\|q - x_n\|}{\varphi(\|q - x_n\|)}, & \text{if } q \ne x_n \\ 0 & , \text{if } q = x_n \end{cases}$$

From $\sup\left\{\dfrac{\|q-x_n\|}{\varphi(\|q-x_n\|)}:q\neq x_n\right\}<\infty$, we obtain

$\limsup\limits_{n\to\infty}\vartheta_n<\infty$ and $J(q-x_{n+1})=\vartheta_{n+1}J_\varphi(q-x_{n+1})$, for all $n\geq 0$.

Then $\|x_{n+1}-q\|^2\leq\dfrac{1-(2-\xi)\alpha_n+\alpha_n^2}{1-\xi\alpha_n}\|x_n-q\|^2+\dfrac{2\alpha_n}{1-\xi\alpha_n}((I-f)q,J(q-x_{n+1}))$

$$\leq\dfrac{1-(2-\xi)\alpha_n}{1-\xi\alpha_n}\|x_n-q\|^2+\dfrac{\alpha_n^2}{1-\xi\alpha_n}M$$

$$+\dfrac{2\alpha_n}{1-\xi\alpha_n}\vartheta_{n+1}((I-f)q,J_\phi(q-x_{n+1})) \qquad (2.9)$$

where $M=\sup\limits_{n\geq 0}\|x_n-q\|^2$.

Putting $a_n=\dfrac{2(1-\xi)\alpha_n}{1-\xi\alpha_n}$ and $b_n=\dfrac{M\alpha_n}{2(1-\xi)}+\dfrac{1}{1-\xi}\theta_{n+1}((I-f)q,$

$J_\phi(q-x_{n+1}))$.

From (i), (ii) and (2.1), it follows that $a_n\to 0$, $\sum\limits_{n=1}^{\infty}a_n=\infty$ and $\limsup\limits_{n\to\infty}b_n\leq 0$.

Since (2.9) reduces to

$$\|x_{n+1}-q\|^2\leq(1-a_n)\|x_n-q\|^2+a_nb_n$$

From Lemma 1.2, we conclude that $x_n\to q$. This completes the proof.

References

[1] Browder, F.E.: Convergence of approximations to fixed point of nonexpansive mappings in Banach space. Archs Ration. Mech. Anal. 24, 82–90 (1967)

[2] Takahashi, W., Tamura, T., Toyoda, M.: Approximation of common fixed points of a Family of finite nonexpansive mappings in Banach space. Sci. Math. Jpn. 56, 475–480 (2002)

[3] Chang, S.-S.: Viscosity approximation methods for a finite family of non expansive mappings in Banach space. J. Math. Anal. Appl. 323, 1402–1416 (2006)

[4] Jung, J.S.: Convergence theorems of iterative algorithms for a family of finite non-expansive mappings in Banach Space

[5] Cioranescu, I.: Geometry of Banach Spaces, Duality Mappings and Nonlinear Problems. Kluwer Academic Publishers, Dordrecht (1990)

[6] Jung, J.S., Morales, C.: The mann process for perturbed m-accretive operators in Banach spaces. Nonlinear Anal. 46, 231–243 (2001)

[7] Liu, L.S.: Iterative processes with errors for nonlinear strongly accretive mappings in Banach spaces. J. Math. Anal. Appl. 194, 114–12 (1995)

[8] Matsushita, S., Takahashi, V.: Strong convergence theorems for nonexpansive nonself-mappings without boundary conditions. Nonlinear Anal. 68, 412–419 (2008)

[9] Jung, J.S.: Convergence theorems of iterative algorithms for a family of finite nonexpansive mappings in Banach spaces. Taiwanese J. Math 11(3), 883–902 (2007)

[10] Yoon, J.-H., Jung, J.S.: Convergence theorems on viscosity approximation methods for a finite family of nonexpansive non-self-mappings. Mathmatical and Computer Modeling 50, 1338–1347 (2009)

Formalizing Workflow Abstraction Modeling Process

Sun Shanwu[1] and Wang Nan[2]

[1] Network and Laboratorial Center,
Jilin University of Finance and Economics, Changchun, 130117, China
[2] Department of Management Science and Information Engineering,
Jilin University of Finance and Economics, Changchun, 130117, China

Abstract. The concept of the workflow fragment is proposed to formally describe the process of constructing the workflow abstraction model by tracking the workflow fragment class. Through perceiving specific workflows of the particular world, the components for constructing workflow abstraction models are generated, which gives the mechanism of knowledge learning, sharing and reuse to save the cost of redesign and remodeling of workflow models.

Keywords: Workflow fragment, Workflow object, Workflow object relationship, Workflow modeling.

1   Introduction

Workflow Management System(WfMS) is a system that completely defines, manages and executes "workflows" through the execution of software whose order of execution is driven by a computer representation of the workflow logic[1]. WfMS mainly provides support of workflow modeling and workflow executing. During the phase of workflow modeling, WfMS provides a Build-time environment for users to define, analyze and manage workflow models. And workflow executing provides a Run-time control to create, execute and manage a workflow instance, while the interactions with human users or application tools may happen. Workflow is the automation of a business process, in whole or part, during which documents, information or tasks are passed from one participant to another for action, according to a set of procedural rules[1]. To achieve the functions of workflow management, the business must be abstracted from the real world and described by a kind of formal method, which results in workflow model. It's necessary for workflow executing to create a well-defined workflow model. During developing WfMS, many approaches have been proposed to explore to find a general formal method to describe the business process[2,3,4] and some researchers analyze workflow modeling from different perspectives [5,6,7] and construct workflow models based on various frameworks[8,9,5].

These ideas focus on the design of workflow model in a certain business field and try to find out a general formal method to describe the workflow model. In this paper, we introduce the concept of the workflow fragment to formalize the workflow modeling process. Different from the designing ideas, we perceive specific workflows of the particular world and capture the components for constructing workflow abstraction models, which provides the mechanism of knowledge learning, sharing and reuse to save the cost of redesign and remodeling of workflow models.

2   Correlative Concepts

In this section, we introduce some concepts to formally describe the workflow and give the relations between them.

Definition 1. Workflow Object: A workflow object WfObj is a task step of a workflow, denoted by a 5-tuples, i.e. WfObj=(WfObjType, ExecutorType, InType, OutType, CondSet), such that:

- WfObjType is the type of the workflow object, showing what the task step does.
- ExecutorType represents the type of the one executing the task, namely, who is in charge.
- InType denotes the type of input resource the task deals with.
- OutType represents the type of output resource that has been dealt with by the task.
- CondSet is a set of conditions which must be held if the executing the task.

We divide the Workflow Object into **Fundamental Workflow Object (FWfObj)** and **Abstract Workflow Object (AWfObj)**. FWfObj is derived from perceiving the specific workflow of a particular field, however, AWfObj is the result of operating on one/more FWfObj/FWfObjs with the same WfObjType. So a many-to-one mapping between the set of FWfObjs and the set of AWfObjs can be built from the operation.

Definition 2. Workflow Object Relation: Suppose that WfObj$_1$ and WfObj$_2$ are two workflow objects, if WfObj$_2$ is able to deal with the output resource of WfObj$_1$, then there is a procedural relation between WfObj$_1$ and WfObj$_2$, denoted by WfObjRel(WfObj$_1$, WfObj$_2$).

    The relationship between two workflow objects shows the execution sequence of them. The relationship between FWfObjs is represented by FWfObjRel, and the one between AWfObjs by AWfObjRel. According to the mapping between the sets of FWfObjs and AWfObjs, we can also construct the mappings with workflow object relations involved.

Definition 3. Workflow Fragment: A workflow fragment WfF is a 4-tuples, i.e. WfF=(WfFType, WfObjs, WfObjRels, WfFCondSet), such that:

- WfFType represents the type of the workflow fragment, namely the executing field of the workflow.
- WfObjs is the task steps composing the workflow fragment.
- WfObjRels is the relation between WfObjs showing the executing sequence of the workflow fragment.
- WfFCondSet denotes the pre-conditions of the workflow fragment running.

Fundamental Workflow Fragment (FWfF) is composed of the elements of FWfObjs and FWfObjRels, which can be perceived from a specific running workflow of a particular world. So FWfObjs and FWfObjRels are captured not only directly from perceiving the real world but also from operating on FWfF. **Abstract Workflow Fragment (AWfF)** is generated from the mapping between the sets of FWfObjs and AWfObjs and between the sets of FWfObjRels and AWfObjRels.

    The relationship among the concepts is shown as Fig.1. The dotted line represents the mapping relationship between the two connected parts, however the full line are used to show the operation and the result.
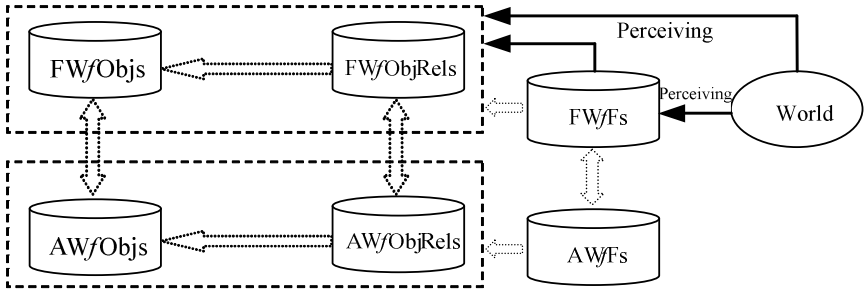
Fig. 1. The relationship among the concepts

3   Workflow Abstraction Modeling

The set of workflow fragments identifies some specific workflows with different workflow types. If we want to construct a workflow model of a particular field, the workflow fragments of the same type must be found out, which composes a **Workflow Fragment Class (WfC)**. The workflow fragments included in some WfC provide knowledge of workflow abstraction modeling of the particular workflow type. Through every workflow fragment, the relationships between workflow objects can be identified which together construct the corresponding model.

Procedure Workflow Abstraction Modeling

Input: set S of WfFs (including the mapping relation with set of WfObjs and WfObjRels), WfFType TYPE

Output: a workflow abstraction model of type TYPE M(Objs, ObjRels, Conds)

　//Objs and ObjRels respectively represent the set of workflow objects and the relations between them which compose the workflow abstraction model M. Conds is the set of constraints to assure M running correctly. Initially, Objs=ObjRels=Conds={}.

(1)　workflow fragment class with type TYPE WfC={};

(2)　For every WfF$_i$ ∈ S , if WfF$_i$. WfFType=TYPE, then WfC= WfC ∪ { WfF$_i$ };
　　//Generating the workflow fragment class of the given field TYPE

(3)　For every WfF$_j$ ∈ WfC, suppose that WfF$_j$=(TYPE, WfObjs$_j$, WfObjRels$_j$, WfFCondSet$_j$)

　　　For every WfObj$_k$ ∈ WfObjs$_j$, if WfObj$_k$ ∉ Objs,
　　　　　then Objs= Objs ∪ { WfObj$_k$ };
　　　For every WfObjRels$_k$ ∈ WfObjRels$_j$, if WfObjRels$_k$ ∉ ObjRels,
　　　　　then ObjRels= ObjRels ∪ { WfObjRels$_k$ };
　　　For every WfFCond$_k$ ∈ WfFCondSet$_j$, if WfFCondSet$_k$ ∉ Conds,
　　　　　then Conds = Conds ∪ { WfFCond$_k$ };

Note that the procedure of workflow abstraction modeling totally bases on the actual flows performed by specific participants under concrete constraints. So theoretically, as long as enough workflow fragments is perceived, an well-defined, accord-with-reality workflow abstraction model of a certain field must be constructed.

4   Summaries

In this paper, the process of workflow abstraction modeling begins with perceiving the real world to generate the specific workflow fragments, the set of workflow objects and the set of workflow object relationships. The workflow abstraction model is constructed by tracking the workflow fragments of a particular field which are captured from the real world, so it is a learning-abstraction process. The mechanism of sharing and reusing the components of perceived workflow fragments also save the cost of workflow redesign and remodeling. In the future, we will continue to explore indistinguishable workflow fragments to realize the hierarchical representation of workflow fragment class.

References

[1] WfMC. Workflow Management Coalition Terminology & Glossary, Document Number WFMC-TC-1011, Document Status- Issue 3.0. Technical report, Workflow Management Coalition, Brussels, Belgium (Februrary 1999)
[2] Sadiq, W., Orlowska, M.: Analyzing process models using graph reduction techniques. Information Systems 25(2), 117–134 (2000)
[3] Salimifard, K., Wright, M.: Petri net-based modeling of workflow systems: An overview. European Journal of Operational Research 134(3), 664–676 (2001)
[4] Bajaj, A., Ram, S.: SEAM. A state-entity-activity-model for a well-defined workflow development methodology. IEEE Transactions on Knowledge and Data Engineering 14(2), 415–431 (2002)
[5] Jablonski, S., Bussler, C.: Workflow Management: Modeling Concepts, Architecture, and Implementation. International Thomson Computer Press (1996)
[6] van der Aalst, W.M.P., van Hee, K.M.: Workflow Management – Models, Methods, and Systems. The MIT Press, Cambridge (2002)

[7] Tick, J.: Workflow Model Representation Concepts. In: Proceedings of 7th International Symposium of Hungarian Researchers on Computational Intelligence, HUCI 2006, Budapest, Hungary, November 24-25, pp. 329–337 (2006)

[8] Eshuis, R., Wieringa, R.: Verification Support for Workflow Design with UML Activity Graphs. In: the Proceedings of the 24th International Conference on Software Engineering Orlando, Florida, pp. 166–176 (2002)

[9] Tick, J.: Workflow Modeling Based on Process Graph. In: Proceedings of 5th Slovakian-Hungarian Joint Symposium on Applied Machine Intelligence and Informatics, Poprad, Slovakia, pp. 419–426 (2007)

Analysis and Comparison of Application of Reducing Mill by Pro / MECHANICA and ANSYS

Bailin Fan[1], Ganghan Huang[2], and Ma Quan[2]

[1] School of Mechanical Engineering University of Science and Technology Beijing, Beijing 100083, China
fanbailin@sohu.com
[2] University of Science & Technology Beijing
Hgh868@sina.com,
maquan0425@163.com

Abstract. Two kinds of finite element analysis were also carried out by the Pro / MECHANICA and the ANSYS on the reducing mill roll housing. The position of the maximum deformation，the maximum deformation, the maximum stress and the position of maximum stress were obtained .The results obtained by two analysis software were compared and very close, which verified the reliability of the results. The stress, deformation obtained by MECHANICA and ANSYS were different as algorithms were different for two software. The results obtained by MECHANICA were more reliable, closer to the actual value and the actual project.

Keywords: Reducing mill, Roll-housing, PRO/MECHANICA, Stress.

1 Introduction

Deformation of the hot-rolled steel after the uniform Seamless steel finishing is further processed in order to ensure the quality, expand the variety of unit sizes and improve productivity. [1]

Roll axis of the roll house of Reducing and sizing mill unit were designed to be the axis angle between 60 degrees each other, it form an equilateral triangle. Deformation of Seamless under the thermal state in the sizing process was crushed, deformed full pass, then the extension and deformation of the wall thickening in order to obtain the required dimensional accuracy and roundness of the finished steel. [2,3]

Roll-housing is the most important component of Reducing and Sizing Mill when monishing and regularizing the steel pipes. In rolling process, the forces on Roller completely passed to bearing base and then to Roll-housing. Therefore，the intensity as well as rigidity have great influence on Reliability and Precision of product[4,5]. Finite element analysis of Roll-housing, to achieve structural optimization design and implementation process of the parameters of the roll-housing design，lay pipe design basis of quality control.

2   The Finite Element Analysis and Optimization Software

With widely range of computer applications and developed, the finite element method have become numerical method in the last 40 years. It has great versatility and flexibility and can be used to solve complex boundary issues. Finite element analysis is valid modern method for the structural mechanics. It is widely used in the structure of static and dynamic characteristics, and then applied to solve the heat conduction, electromagnetic field, fluid mechanics and other continuity problems. The commonly finite element analysis software: ANSYS, PRO / MECHANICA and so on.

2.1   Pro / MECHANICA

Pro / MECHANICA PTC is the finite element software created by Parametric Technology Corporation(PTC) in the United States, it can directly use Pro / ENGINEER to analysis the geometric model, implementation, and Pro / ENGINEER, completely seamless integration. Pro / MECHANICA finite element software had the most significant features: to provide powerful parametric function of the geometric model sensitivity analysis and optimal design. The main module can be divided into three parts: [6,7]

(1) Pro / MECHANICA STRUCTURE: structural analysis software can be part models and assembly models of structural analysis and design optimization. It can be able to complete the analysis of species static analysis, modal analysis, buckling analysis, contact analysis, preload analysis and vibration analysis.

(2) Pro / MECHANICA THERMAL: Temperature analysis software can be part model or assembly model of steady state and transient temperature analysis, the temperature issue can also be based on sensitivity analysis and optimal design.

(3) Pro / MECHANICA MOTION: Motion analysis software package, institutional analysis and optimal design of kinematics can be used in three-dimensional static analysis, kinematics analysis, dynamic analysis, inverse dynamics analysis, and collision detection analysis.

2.2   ANSYS Software

ANSYS software have powerful analysis capabilities, including structural analysis, nonlinear analysis, thermal analysis, electromagnetic analysis, field analysis, fluid flow analysis, coupled field analysis. ANSYS mainly includes three parts: pre-treatment module, analysis and calculation module and post-processing module.[8,9]

PREP7 processing module provides a powerful modeling and meshing tools, and the user can easily construct a finite element model, including structural analysis calculation module SOLUTION (can be linear analysis, nonlinear analysis, and highly nonlinear analysis), fluid dynamic analysis, sound field analysis, electromagnetic analysis, piezoelectric analysis and coupled multi-physics analysis, also can simulate the interaction of a variety of physical media. The post-processing module can be calculated to display and output in various ways.

3   The Finite Element Analysis of Reducing Mill Roll-Housing

Aiming to meet the requirement of service life and improve the intensity of housing, weight reduction was also considered to make full use of materials and the optimum design was executed on Roll-housing. Research objects was the roll–housing to withstand the maximum force for twelve Reducing mill unit[10], shown in Figure 1. According to the Provided process parameters, the stress of roll-housing was researched and analyzed.

The roll-housing material was ZG270-500. In the Pro / MECHANICA integration mode, the material properties were defined. Three-dimensional solid model of the roll-housing was created by Pro / E.
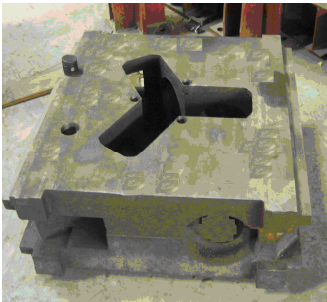
Fig. 1. Roll-housing
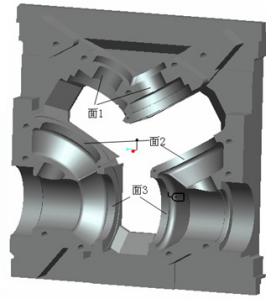
Fig. 2. The Force Plane
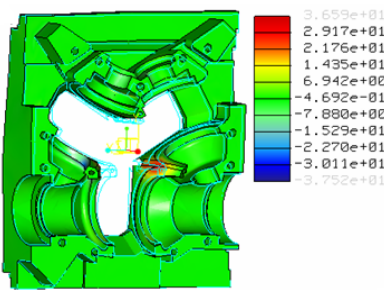
Fig. 3. Z-stress cloud
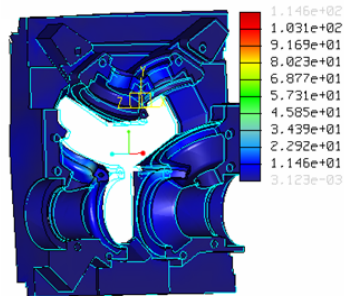
Fig. 4. Equivalent stress contour

3.1   The Force Analysis of the Roll-Housing

The stress analysis was studied for roll，roll axis and rolls consisting of bevel gear. The Position of Surface force of the Roll-housing and the size of the force in all directions were determined. 1,2,3 force plane were shown in Figure 2.

3.2   Applied Load

In the normal rolling process the roll-housing forces were very complex, there are several forces acting on the roll-housing:

1) The rolling force on the roll-housing through the roller bearing；
2) The friction torque of roll neck induced reaction through the roller bearing on the roll-housing;
3) Acceleration and deceleration cause inertial force on the roll-housing when Steel tube bite;
4) The sloping torques caused by all sorts of horizontal force gave rise to the reverse force in the roll-housing on below support.

The rolling force was the largest in the above all kinds of forces. The number of other force compared with rolling force was small. Therefore，other force influence on rolling housing was usually ignore when force was researched. Each force surface was set in figure 2. Table 1 showed the force of each surface.

Table 1. The force of each surface

force surface \ direction	1		2		3	
	left	right	left	right	left	right
X （N）	89110	−89110	89110	−89110	0	0
Y （N）	51450	51450	51450	51450	— 102900	— 102900
Z （N）	15210	15210	15210	15210	4380	4380

3.3   The Results and Stress Analysis

Through the finite element analysis stress for Z cloud，equivalent stress of convective and the biggest deformation image were obtained. Figure 3 demonstrated the Z-stress cloud. Figure 4 showed the Equivalent stress of convective. The maximum stress value all direction showed in table 2. The analysis results showed that bending cause stress concentration in the place of two gear tank intersection Angle. The value of the maximum equivalent stress was 114.6 MPa.Although the value was less than 150MPa which was the maximum allowable stress of the material, it was located at the sharp corners and it was easy to crack for using roll-housing in long-term. In addition，the stress in other parts of the roll-housing was very small and far less than the maximum stress value.

Table 2. The value of the maximum stress

Stress (MPa)	X	Y	Z	The max equivalent stress (MPa)	The vector of the max deformation mm
	69.76	20.86	36.59	114.6	0.07344
shear stress (MPa)	XY	YZ	XZ	The Maximum shear stress(MPa)	
	46.79	12.49	39.19	57.76	

4 Finite Element Analysis by ANSYS

To verify the correctness of the analysis results in Pro/MECHANICA, Three-dimensional model of Roll-housing by Pro / E was converted to the ANSYS though ANSYS-Pro / E interface. The stress，deformation and equivalent stress contour of roll-housing under the maximum rolling force were solved by loading，solving in ANSYS. The contour of equivalent stress was showed in Figure 5. The contour of deformation was showed in Figure 6. The results of finite element analysis by the ANSYS were illustrated in Table 3.

Table 3. The maximum stress, deformation of roll-housing by ANSYS

direction	X	Y	Z	XY	YZ	XZ
the maximum stress (MPa)	69. 304	20. 261	35. 976	44. 056	8. 222	40. 321
the maximum deformation (mm)	0. 0749	0. 0265	0. 0162			
the max Equivalent stress (MPa)	105. 868			The max Vector of deformation mm		0.0843
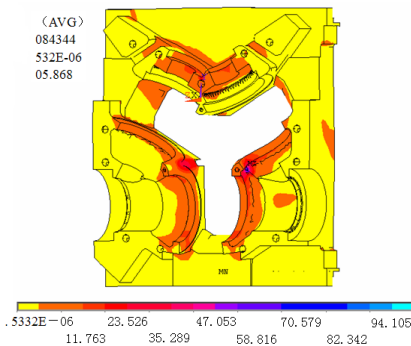
Fig. 5. Equivalent stress contour
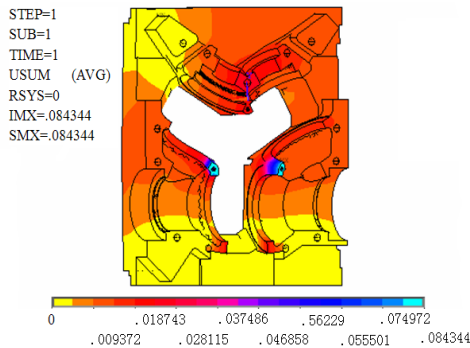
Fig. 6. Deformation cloud

We found that the Analysis value of maximum equivalent stress obtained by Pro / MECHANICA was greater than the maximum effective stress obtained by ANSYS in Table 2 and Table 3. The vector of the maximum deformation obtained by Pro / MECHANICA was less than the vector of the maximum deformation obtained by ANSYS.

5   Summaries

Two kinds of finite element analysis were also carried out by the Pro / MECHANICA and the ANSYS on the reducing mill roll housing. The position of the maximum deformation, the maximum deformation, the maximum stress and the position of maximum stress were obtained by defining materials, mesh, constraints imposed, loading and analysis. We drew the following conclusions:

1) The results obtained by two analysis software were compared and very close, which verified the reliability of the results.
2) The stress，deformation obtained by MECHANICA and ANSYS were different as algorithms were different for two software.
3) The maximum equivalent stress obtained by MECHANICA was larger than that by ANSYS. This is because the necessary simplification was made by ANSYS. The value of maximum equivalent stress was less than 150 which was maximum allowable stress of the material. Therefore, the results obtained by pro/MECHANICA were more reliable, closer to the actual value and the actual project.
4) The stress concentration, the maximum deformation position of roll-housing was obtained under Pro / MECHANICA.
5) We can fully test that the design is reasonable and lay the foundation for product development for the institute By using pro/MECHANICA and ANSYS to analysis stress and strain of reducing mill roll-housing. A research base is provided to optimize the design of similar products for the roll mill.

References

[1] Kang, Y.: Rolling Engineering, pp. 35–48. Metallurgical Industry Press, Beijing (2004)
[2] Ji, W.: SRM510-7 Rack Introduction to Micro-Tension Sizing Mill. Steel 34(6), 31–35 (2005)
[3] Zhang, H.: Desigh of Broadsiding Groove of 3 High Stretch Reducing Mill. Steel 34(4), 32–36 (2005)
[4] Zhang, F.: Research on Design of Passes of Rolls on Stretch-Reducing Mill. Steel Modern Manufacturing Engineering, 110–113 (October 2006)
[5] Zhang, J.: Analysis of Piston Finite Element Method Based on Pro/MECHANICA. Mechanical Engineer. (1), 55–57 (2006)
[6] Ye, Y.: Optimization Theory and Design of Mechanical, pp. 72–84. Metrology Publishing Press, Beijing (2001)
[7] Lu, J.P.P., Zeng, H.-F.: Optimized Design of Conduit Bracket under Pro/Mechanica Circumstance. Joumal of Kunming Metallurgy College 23(1), 35–39 (2007)

[8] Bathe, K.J.: Finite Element Procedures in Engineering Analysis, pp. 32–34. Prentice-Hall (1982)

[9] Iguchi, T., Owen, D.R.J., Liu, C.Q.: Analysis of rolling processes by dynamic explicit elastic-plastic finite element method. In: Processing of the 7th International Conference on Steel Rolling, Chiba, Japan, pp. 266–271 (1998)

[10] Yang, J.: Finite Element Analysis and Optimal Design of Roll-housing baseing. In: Pro/MECHANICA 2007. School of Mechanical Engineering University of Science and Technology, Beijing (March 2007)

Emotion Recognition from Physiological Signals Based on ASAGA

Lianzhe Zhou, Huanli Pang, and Hanmei Liu

School of Computer Science and Engineering
Changchun University of Technology
Changchun, China
zhoulianzhe@gmail.com,
{panghuanli,liuhanmei}@mail.ccut.edu.cn

Abstract. Research on emotion recognition from physiological signals has increased during last decade and is getting closer to achieve online recognition. In this paper, we discuss the Adaptive Simulated Annealing Genetic Algorithm, and use it to recognizing four emotional states from four physiological signals. The experiment shows that based on physiological signals, the ASAGA is feasible and effective.

Keywords: Emotion Recognition, Adaptive Simulated Annealing Genetic Algorithm, Physiological Signal.

1 Introduction

Emotion recognition is an important issue in advanced human-computer interaction. In everyday life people express their emotions through multiple modalities. Many efforts have been taken recently to recognize emotions using facial expressions, speech and physiological signals [1-5], current recognition systems are not yet advanced enough to be used in realistic applications.

Physiological signals are directly controlled by the human autonomous nervous system. The reactions are more spontaneous and robust compared to facial expression and voices that can be intentionally disguised. Therefore, using physiological signals as the research objects of emotion recognition will be more objective to recognize the internal emotion and affective changes. Picard et al were firstly engaged in emotion recognition with the features from physiological signals, other countries and scholars also joined in this field subsequently [6].

Four physiological signals such as skin temperature (SKT), Skin conductance (SC), electro-cardiogram (ECG) and respiration change (RSP) are often use to extract features for emotion recognition. Use some methods of feature selection such as FP, SFS, SBS, SFFS, SFBS, ANOVA to select features. And then use classification methods KNN、MLPN、LDF、CCA to classified physiological patterns for a set of four emotions (joy, anger, sadness, pleasure). A recognition accuracy of over 80% [3] on the average seems to be acceptable for realistic applications.

In this paper, Adaptive Simulated Annealing Genetic Algorithm (ASAGA) is introduced as a new attempt for emotion recognition. The algorithm embeds the simulated annealing algorithm in the genetic algorithm, should be able to escape from any local optima and thus has the potential to obtain a global optimal result.

2   Adaptive Simulated Annealing Genetic Algorithm

The Genetic algorithm (GA) [7] is randomized search and optimization techniques guided by the principles of evolution and natural genetics, having a large amount of implicit parallelism. GA is an adaptive heuristic search method based on natural selection and evolutional theory, and describes a realistic problem by gene and chromosome. The random initial solution would gradually approach the optimal solution by the genetic operation of reproduction, crossover and mutation. A solution is represented by chromosomes that are composed of genes. A set of solutions comprises a population that evolves over time through competition and controlled variation.

Generally, GA can not find the optimal solution of the problem while the space of solution is very large; but can easily seek for the best one when small. Since the standard genetic algorithm always made the individuals tend to the same structure, therefore, it is very difficult to catch the optimal solution in the global search. The Simulated Annealing (SA) [8] can fast find the local optimization, however, the ability of global searching of the algorithm is despondent. In some cases, The GA and the SA can be corporate to resolve problem of optimization.

SA has gained popularity in solving hard combinatory problems. SA belongs to a class of algorithms called probabilistic hill-climbing [9] which dynamically alter the probability of accepting inferior solutions. The SA methodology draws its analogy from the annealing process of solids. In the annealing process, a solid is heated to a high temperature and gradually cooled in order for it to crystallize. As the heating process allows the atoms to move randomly, if the cooling is done too rapidly it prevents the atoms from reaching thermal equilibrium. If the solid is cooled slowly, it gives the atoms enough time to align themselves in order to reach a minimum energy state. This analogy can be used in combinatorial optimizations with the states of the solid corresponding to the feasible solution, the energy at each state corresponding to the improvement in objective function and the minimum energy being the optimal solution.

This section describes a solution methodology based on flexibly and iteratively defined local solution search spaces and an integrated SA and GA solution search method.

2.1   Chromosome Representation

The chromosome is a string represented by a finite sequence of binary. In the application of feature selection, each chromosome C_k has n genes, Suppose $C_k = \{g_i | i=1,..., n\}$. Each gene, $g_i = 1$ means the i-th feature is been selected, whereas $g_i = 0$ denotes for exclusion.

2.2   Fitness Function

The Fitness Function is a particular type of objective function that prescribes the optimality of a chromosome in a genetic algorithm so that that particular chromosome may be ranked against all the other chromosomes. The fitness function is defined as:

$$f(x) = c - e(x)$$

Where $e(x)$ is the classification error rate of classification. And c is a positive integer such that is nonnegative and monotonic decreasing on $e(x)$.

2.3   Selection, Crossover, Mutation

The best selection strategy for picking the parents to be the base for new offspring chromosomes is often problem specific. We use Roulette wheel selection to randomly select chromosomes from the population. All candidate chromosomes occupy a section of an imaginary roulette wheel proportional to their fitness, and chromosomes are selected by generating a random number and pick the chromosome at that specific spot on the wheel.

Crossover operation can generate new individuals, and make new search in objective space. The new individuals inherit chromosome information from their parents, and are usually different from them. This operator selects two individuals within the generation and a crossover site and carries out a swapping operation of the string bits to the right hand side of the crossover site of both individuals. Crossover operations synthesize bits of knowledge gained from both parents exhibiting better than average performance. Thus, the probability of a better offspring is greatly enhanced.

Mutation is performed to introduce slight variations to allow for the exploration of states not generated through crossover. It may help the search escape local minima and find new areas of interest in the problem hyperspace. The mutation operation has been defined as following: From the string to be mutated, a random element is chosen and it is replaced by a different index of point in the range *{1, …, n}* such that no element is duplicated in the string.

2.4   Adaptive Mechanism

For multimodal functions, there is a need to be able to locate the region in which the global optimum exists, and then to converge to the optimum.. The goals with adaptive probabilities of crossover and mutation are to maintain the genetic diversity in the population and prevent the genetic algorithms to converge prematurely to local minima. Crossover probability P_c and mutation probability P_m are the main factors in affecting balanced search ability (global search ability and local search ability). This algorithm can better solve the problem of adjusting P_c and P_m dynamically and also fits to all kinds of optimal problem. Based on these facts, we adopt the adaptive probabilities to obtain the optimal solution as follows formulas:

$$P_c = \begin{cases} P_{c1} - \dfrac{(P_{c1} - P_{c2})(f' - f_{avg})}{f_{max} - f_{avg}} & f' \geq f_{avg} \\ P_{c1} & f' < f_{avg} \end{cases} \tag{1}$$

$$P_m = \begin{cases} P_{m1} - \dfrac{(P_{m1} - P_{m2})(f - f_{avg})}{f_{max} - f_{avg}} & f \geq f_{avg} \\ P_{m1} & f < f_{avg} \end{cases} \tag{2}$$

where f_{max} is the highest fitness value in the population; f_{avg} is the average fitness value in every population; f' is higher fitness value between two individuals; f is the mutation individual fitness; and we set $P_{c1}=0.9$, $P_{c2}=0.6$, and $P_{m1}=0.1$, $P_{m2}=0.001$.

2.5   Simulated Annealing

A simulated annealing search evaluates states in the search space one at a time, and new states are randomly selected from a set of neighboring states. At each distinct temperature, expeditions are made from the currently best state, producing a Markov chain M of state transitions. Not only states that decrease the objective function f are accepted, but also changes that increases it. The latter happens with a probability related to the temperature of the system, which is decreased at the end of each Markov chain. The probability defines as:

$$P = \begin{cases} 1 & if \quad f^* \leq f \\ \exp\left(\dfrac{f - f^*}{T}\right) & if \quad f^* > f \end{cases} \tag{3}$$

where f and f^* are the costs of the current and proposed configurations, respectively, and T is the temperature. The algorithm is run until a stopping condition is reached, typically a minimum temperature value, specified as part of the annealing schedule.

2.6   Algorithmic Description

The major steps of the proposed algorithm are:

Step1: Initialize the generation counter: $t \leftarrow 0$, Create randomly the original population $P(t)$, Initialize the temperature T_0;

Step2: Determining the fitness function of each chromosome in the population;

Step3: perform operator of selection: $P(t) = selection[P(t)]$;

Step4: Apply the crossover operator according to (1): $P'(t) \leftarrow crossover[P(t)]$;

Step5: Apply the mutation operator according to (2): $[P''(t)] \leftarrow mutation[P'(t)]$;

Step6: The child chromosomes are accepted according to (3), where f and f^* are the fitness values of parent and child chromosomes;

Step7: Perform operator of Simulated Annealing: $[P'''(t)] \leftarrow SA[P''(t)]$;

Step8: Let the current population be the new population: $P(t+1) \leftarrow reproduction[P(t) \cup P'''(t)]$;

Step9: If the convergence criterion is satisfied, stop, Otherwise: $t=t+1$, go to step 2.

3   Simulation Results

In this paper, the dataset used for simulation is obtained from Multimedia and Signal Processing Lab, Augsburg University, Germany [10]. The dataset contains physiological data of four sensors: SC, EMG, RSP and ECG. 25 data sets of one single subject consecutively expressing four emotional states: Joy, Anger, Sadness, Pleasure were collected. Statistic values (such as mean value, median value, standard deviation, first-order difference, second-order difference, minimum, maximum, minimum ratio, maximum ratio, etc.) as original features. Four emotion sates match along with four classes and each class has 25 samples, so there are 100 samples from this data.

In the experiment, the parameters are defined as follows: K-nearest K = 3, maximum iteration times T = 1000, scale of population is 50, general-crossover rate

$P_c = 0.8$, mutation rate $P_m = 0.1$, the most number of generations is 50, the temperature $T_0 = 70$, a=0.8. The feature set is selected by Adaptive Simulated Annealing Genetic Algorithm; the K-nearest neighbor is applied to classify the emotion classes.

The following tables represent the recognition rates by Adaptive Simulated Annealing Genetic Algorithm.

Table 1. Recognition rates of 4 emotions

Physiological signal	Feature number	Max recognition rate	Min recognition rate	Average recognition rate
SC	21	51%	40%	47.24%
EMG	21	68%	66%	67.35%
RSP	67	71%	65%	68. 62%
ECG	84	67%	59%	62.42%
All	193	85%	61%	69.83%

Tab.1 shows the recognition rates of four emotions base on Simulated Annealing Genetic Algorithm. From the last column we can conclude that the recognition results of EMG and RSP are better when using one physiological signal to recognition four emotions, and SC is the worst. Therefore it is difficult to recognize different emotional states by using signal physiological signal. While using the four physiological signals to identify the four emotions, it has been enhanced very apparently, with improved up to 85%.

Table 2. Recognition rates of single emotion

Physiological signal	Joy	Anger	Sadness	Pleasure	Average
SC	69%	59%	55%	43%	55%
EMG	71%	82%	88%	78%	76%
RSP	67%	100%	69%	68%	71%
ECG	89%	83%	76%	64%	86%
All	86%	100%	80%	82%	85%

Tab.2 shows the recognition rates of single emotion identified by single physiological signal. Form the table we can conclude that using RSP to recognize Anger can reach the highest recognition rate up to 100%.

4   Summaries

This paper applies ASAGA to emotion recognition from physiological signals, and use KNN for classification. The results of simulation shows that using a single physiological signal to recognition four emotions is feasible, and obtained good recognition results. Future work will focus on finding out some certain feature combinations, improve the recognition rate and represent relevant emotional states.

References

[1] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in humancomputer interaction. IEEE Signal Process. Mag. 18, 32–80 (2001)

[2] Healey, J.A.: Wearable and Automotive Systems for Affect Recognition from Physiology. PhD thesis, MIT, Cambridge, MA (May 2000)

[3] Picard, R.W., Vyzas, E., Healey, J.: Toward Machine Emotional Intelligence: Analysis of Affective Physiological State. IEEE Transactions Pattern Analysis and Machine Intelligence 23(10), 1175–1191 (2001)

[4] Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion Recognition Using Bio-Sensors. In: First Step Towards an Automatic System, Affective Dialogue Systems, Tutorial and Research Workshop, Kloster Irsee, Germany, June 14-16 (2004)

[5] Nasoz, F., Alvarez, K., Lisetti, C.L., Finkelstein, N.: Emotion Recognition from Physiological Signals for Presence Technologies. International Journal of Cognition, Technology and Work, Special Issue on Presence 6(1) (2003)

[6] Haag, A., Goronzy, S., Schaich, P., Williams, J.: Emotion Recognition Using Bio-sensors: First Steps towards an Automatic System. In: André, E., Dybkjær, L., Minker, W., Heisterkamp, P. (eds.) ADS 2004. LNCS (LNAI), vol. 3068, pp. 36–48. Springer, Heidelberg (2004)

[7] Holland, J.: Adaptation in Natural and Artificial Systems. University of Michigan Press, Ann Arbor (1975)

[8] Kirkpatrick, S., Gelatt, C., Vecchi, M.: Optimization by Simulated Annealing. Science 220, 671–680 (1983)

[9] Romeo, F., Sangiovanni-Vincentelli, A.: Probabilistic Hill-Climbing Algorithms: Properties and Applicatwm. Computer Science Press, Chapel Hill (1985)

[10] http://mmwerkstatt.informatik.uniaugsburg.de/ project_details.php?id=33

Information Management System of Skills Training Based on B/S Structure

Zheng Hua-An and Yu Dongwei

[1] Ningbo University
Ningbo, Zhejiang, China
`ZhengHuaan@nbu.edu.cn`
[2] Ningbo Vocational & Adult Education Teaching and Research Office
Ningbo, Zhejiang, China

Abstract. The article point out that the Skill Training is based on the use of the structure of B/S, and it uses the professional technology school as the target to analyze, it analyze the principle about the design of the management system. By analyzing the hardware of the present professional technology schools' systems and selecting the models of the software, we can get the guiding ideology about the development of the system.

Keywords: Browser/Server Structure, Skill Training, Technology, Information Management System.

1   Introduction

The training of the professional technology school becomes more and more important. Nowadays, Emphasis on the application of information management system. For example, in Ningbo Vocational Education Central School, the computer using training base and the numerical control technology training base are the country-class professional teaching training base which are supported by the center authorities of the finance. These conditions of the professional training will ensure the students' training strongly. Different professional technology schools mainly have their own bases like these.

How to synthetically bring the function of each department into play and finish the management about the information has become the directors' tasks. We come up with the idea about the B/S structure. It will be shown in the following fig.1:
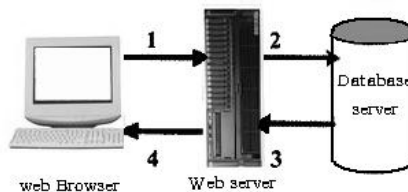
Fig. 1. Browser/Server Structure

We think deeply about the professional training information systems of the professional technology schools, we use the popular B/S structure as the technology route, visiting the servers through the users' ports. The essay begins with the network equipment of the technology schools, and then analyses the principle about the design of the systems, the environment of the hardware, software's structure. It's also important for us to consider the factor about the software platform. As a system project, the whole planning design has analyzed the school network systems' structure first, and will use the present software and hardware resources to design and develop the system as much as possible, and makes the expectant tasks and aims come true.

2   The Principle about the Design of the System

Professional training information system is consisted on several sub systems. When we divide the sub systems, we have considered the function of each sub system and the relationships among the sub systems. A technology training system about how to deal with the affairs in the professional technology schools' technology training information systems which usually are consisted on the technology training information offices, every subject research group, managers and out-school experts and so on, not only needs to consider the demand of the teachers, the managers and presidents and so on, but also should pay attention to the convenience of the training base and the work. The teachers can manage the things at home or any other office through the in-network or the Internet.

When we design the systems, we need to hold the key point according to the entirety. Analyze the function of each system around the target of the totality carefully; especially hold the process of the technology training information. Those can show the feature of the distributed. Management. At last it will achieve such following goals:

1. Achieve the management of the technology training information that is distributed and suitable to the professional technology schools.
2. Design the platform of the technology training service.
3. Build a set of management systems which are efficient, useful、 safe、 reliable, able to be transplanted.

To achieve those three targets, we must insist on following the principle of the use、 the principle that the early days are more important、 the principle that it can be developed, the principle that it is safe, the principle that it is systematic.

2.1   The Systematic Principle

At the prime time of the development of the system, analyzing and researching are the basal work to achieve the system. Consider the whole managers, decision makers, teacher system synthetically and plan the system by overall situation. For example, to manage the main affairs in the process of the technology training, and also the thesis, the patent and so on, we should continue on planning the system just by each working procedure that has been examined and approved and set the relevant limits of authority and tasks of the operation.

The system is achieved by the achievement of each function. And then, connect closely each link and make the technology training information activities be connected to

each other and also adjust to each other. Each module of the function can be independent when they are developing, but in the adjustment and integration, they need connecting to each other. For example, the statistics of the technology training work's quantity is the collection of the results of each technology training and it can also enter the personal and academy's collection chart. Thus, it develops the real sharing system of the technology training information data bank instead of the islet department using system.

2.2   The Principle That the Early Days Are More Important

Before the developing of the technology training system, we should fully listen to the opinions given by each college teacher and the leaders in the technology office, analyze carefully the present systems, and carry on the useful experience of the present systems. And to make the teachers to adapt to the new system, we should necessarily carry on and also improve the design of interface, the style of operation, the custom of operation. After analyzing the shortages of the present technology training information system, some new demands and targets need to consider the problem about the combination and we should make the new management system more advanced in technology, more steady in changing-over, more perfect in function.

2.3   The Development Principle

This principle require that the system should be easy to develop the function and without the changing of the structure.   Using programs or the new modules increase without a lot of changing about the past program's code. Every year, the professional technology schools may adjust the weight scale of the technology training data. The newest standard will also be changed. Thus, the developing of the system needs to have a good safeguard.

2.4   The Safety Principle

In order to ensure the normal work of the systems, the safety and reliability are necessary. Different personnel should have different limits of authority; otherwise, the data in the data bank may be changed at any time. Legal users also can not operate someone else's data. The work such as to prevent the illegal users from entering the systems is very important. To avoid the illegal incursions, the operations of the systems' software patches are required to be paid more attention to the fireproofing wall of the data bank's servers need operating in safety, the copies and safeguards of the data bank should be taken seriously. Coping the data bank at regular intervals can avoid the disasters which are brought by the results of the software safety accidents. The systems not only need to be clear with all kinds of the operations' limits of authority, but also should accredit the managers' new authentications again every school year.

3   The Analysis about the Network Structure

Professional technology schools' technology training system requires the present school network equipment and the Internet as its hardware basis. Out-network will

insure that this system works safely through the fireproofing wall. It is necessary to analyze and design the present school network hardware.

3.1   The Condition of the School Network's BUILDING

The schools this essay researches are divided into the fields such as every research office (group), training base, library, educational administration, electricity teaching center, scientific research office and laboratory building and so on. Every building uses the structure combined by three levels of the network. As the core of the school network, the network center will put the equipment including the management server of the technology training information, core exchange board and so on in itself. Different research offices (groups) achieve the connection of the in-network by using the wired and wireless network, establish the part network centers in each center laboratory building. Present network structure mainly adopts the connection of the light cable from the school network center (cadre) to each building. The whole system is the main communication network of the school network, and the center generator room is the core of the school network, and the system uses the CISCO(core exchange route equipment), nowadays, it can combine the china technological network and the China telegraphy at the same time and it can offer the network connection The main engine of the technology training information system will be put in the network center. Every network will combine every network center respectively to form the using of the networks.
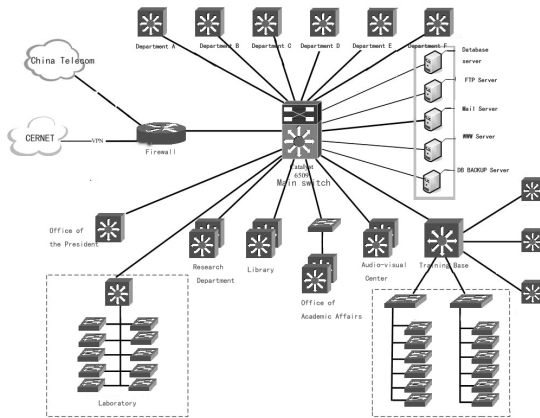
Fig. 2. The Technology School Topology

3.2   The Condition of the Server and the Operating Platform Hardware Equipment

The control center of the school's technology training system is established in the network center. That can make a full use of school network's fireproofing wall. And the preserve of the system can be done by the school network's technological staff.

Web server: IBM X3850, CPU Intel Xeon MP E7330 2.4GHZ, 4×1G DDR2 667, SCSI 2*73GB;

Data Base server: IBM X3850,   CPU Intel Xeon MP E7330 2.4GHz, 4×1G DDR2 667, SCSI 4*73GB;

Data reserve server: IBM X3850, CPU Intel Xeon MP E7330 2.4GHZ, 2×1G DDR2 667.

Because above all are mainly used in the school office OA system, so the main engine of the technology training information system needs to increase the purchase of some equipment on the base of the present school network.

4   The Choice of the System Software Environment

The developing environment and the developing means chosen before the technology training information system beginning the design and development are compared in many different conditions. And then, we will choose the relevant developing environment. Thus, it can bring us much convenience of the development. The essay concludes the following developing environment and means we have chosen by its own technology level and research object.

4.1   Developing Environment: Microsoft's Developing Environment Visual Studio 2008

Nowadays, it's the most popular Windows platform using program developing environment. And it has developed into the edition of 9.0 (Visual Studio 2008).Visual Studio can be used to establish the Windows platform's Windows using program and network using program, and also can be used to establish the network server, intelligent equipment using program and Office plug-in board.

4.2   The Choice of the Data Bank: SQL Server 2005

VS.NET can friendly support the present main data, such as SQL SERVER, ORACLE, MySQL and so on. According to recent years' data quantity, we choose the SQL Server2005 data bank that is suitable to the small and medium-sized system.

5   Summaries

We make a generic analysis for the professional technology school's technology training information system by analyzing above software、 hardware. We also have to move forward a single step to deeply analyze each school's feature, especially its different profession. Maybe their base trainings are different, and the manage processes also have differences. So it's important to finish the early analysis. And according to this essay's principles of the design, the professional technology school's technology training information manage system will play a more effective role.

References

[1] Huang, D.: Research on University Scientific Research Management Systems Based on B/S Pattern. Zhejiang University of Technology (2010)

[2] Crosier, S., Booth, B., Dalton, K., Mitchell, A., Clark, K.: ArcGIS basic tutorial. ESRI China (Beijing) Co., Ltd., Beijing (2004)

[3] Wang, F., Lin, R., Hu, Y.: Meteorological Data Services Based on Web Services. National Meteorological Information Center Communication platform, Beijing (2009) (in Chinese)

[4] Isotani, S., Mizoguchi, R.: Using Ontologies for an Effective Design of Collaborative Learning Activities. In: Proc. of International Conference on Artificial Intelligence in Education (AIED) (2007)

Research on Autobody Panels Developmental Technology Based on Reverse Engineering

Jun Ma and Qinghua Liu

Institute of Information Engineering of Jiaozuo University
2230600@sina.com

Abstract. For finding the shortest path in GIS (geographic information system), an algorithm of the approximate shortest path between two points on arbitrary triangular model is proposed in this paper. Its efficiency and better control accuracy are suitable to search for the shortest path. And three-dimensional visualization of the algorithm can be implemented by OpenGL with Visual C $^{++}$ 9.0.

Keywords: GIS(geographic information system), triangular grid model, the shortest path, subdivision iteration.

1   Introduction

In GIS, searching a shortest path is essential. In two-dimensional GIS environment, algorithms of choosing a shortest path are implemented based on Dijkstra algorithm. But with more and more attention paid to three-dimensional visualization technology, simple two-dimensional display is unable to meet requirements from majority people. For realistic and intuitive feeling, and providing assistance of three-dimensional spatial analysis and aided decision-making, that the shortest path is applied to displaying real three-dimensional terrain becomes of practical and immediate significance. Therefore, an algorithm of the approximate shortest path between two points on arbitrary triangular model is proposed in this paper.

2   Symbols Definition

To give an easy-understanding explanation of the algorithm, symbols definition and basic operation description are as follows:

(1) G represents a weighted graph of original triangular grid model
(2) G^i represents a subdivision graph created by the ith iteration, $G^0 = G$
(3) v^i represents set of vertices in G^i
(4) E^i represents set of edges in G^i
(5) v^i_j represents the jth vertex in G^i
(6) e^i_j (v_m , v_n) represents the jth edge and v_m, v_n are two endpoints of line e^i_j accordingly.
(7) $| e^i |$ represents the length of dege e^i

(8) InGraphG(v^i_j) determines whether v^i_j is a vertex of original triangular grid model in G^{iw}. If it is, the operation returns serial number of it; otherwise, returns null.

(9) InGraphG(v^i_j) determines whether vertices in V^i are vertices of graph G, if they are, the operation returns serial number of all the vertices of graph G; otherwise, returns null.

(10) Edge (v^i_j) returns serial number of the edge which vertex v^i_j is on, if InGraphG (v^i_j) returns null; otherwise returns serial number set of edges adjacent to InGraphG(v^i_j).

(11) OnSame Triangular (v^i_m, v^i_n) determines whether vertices vim , v^i_n in G^i are in the same triangular (condition of on the same edge is exclusive). If it is,return TURE; otherwise return FALSE.

(12) NeiborV (v^i_j) returns set of points adjacent to point v^i_j directly in subgraph G^i; if InGraphG(v^i_j) returns null, the adjacent points are adjacent to v^i_i directly and on the same Edge (v^i_j); otherwise are adjacent to v^i_j on Edge (InGraphG (v^i_j)).

(13) P^i indicates the shortest path between v_s (start point) and v_d (destination point) in G^i , expressed as P^i = (v^i_1, v^i_2, $\cdots$, vi$_n$; e^i_1,e^i_2, $\cdots$, e^i_m) , v^i_1= v_s , v^i_n = v_d, e^i_j= (v_j , v_{j+1}), j = 1, 2 , $\cdots$, n.

(14) L^i denotes length of a shortest path after the ith iteration, L^i = | e^i_1| + | e^i_2| + $\cdots$+ | e^i_m|

(15) SP^i (v^i_1, v^i_2, $\cdots$, v^i_n) indicates set of vertices from v^i_1 to v^i_n on a shortest path in G^i, and SP (v_1 , v_2 , $\cdots$, v_n) are vertices on the shortest path in initial graph G, equivalent to SP^0 (v^0_1, v^0_2, $\cdots$, v^0_n).

(16) Nodes (e^i_j) returns the serial number of two end points of edge e^i_j in graph G^i.

3   The Initial Shortest Path

Firstly, according to initial triangular grid model, weighted graph G is constructed. Then, a path tree of weighted graph G is set up point by point based on principle of Dijkstra algorithm. So the distance from starting point vs to destination v d in graph G can be calculated. The algorithm is as follows:

Step1.   for all v $\neq v_s$, L (v) $\leftarrow$w (v_s , v)
Step2.   L (v_s) $\leftarrow$0
Step3.   T $\leftarrow${ v_s}
Step4.   While (v_d is not part of T) do
            begin
Step5. search a point v′that is not part of T, and all points v fail to belong to T
            L (v′) $\leq$L (v)
Step6.   T $\leftarrow$T ∪{ v′}
Step7.   For (all points v are not part of T)
            if L (v) > L (v′) + w (v′, v)
            then L (v) $\leftarrow$L (v′) + w (v′, v)
    end

This is a strong polynomial time algorithm. T is set of points and there are the shortest path between them and destination point v_s. w(v_s , v) is the weight of edge from v to v_s (the starting point)。 if there is no directly connected edge between v and v_s,

w (v_s , v) = ∞.L (v) is the length of path from vertex v to v_s. When v_d is part of T, processing is over. Since points adjacent to each vertex of triangular grid model are stored, the operation of step5 is to search a point in union of adjacent points of each vertex in T. Fig.1 shows the initial shortest path from V_s to V_d according to original graph G with the above algorithm. The algorithm not only is fit for calculating the initial approximate shortest path, but also is suitable for each step of shortest path subdivision iteration.
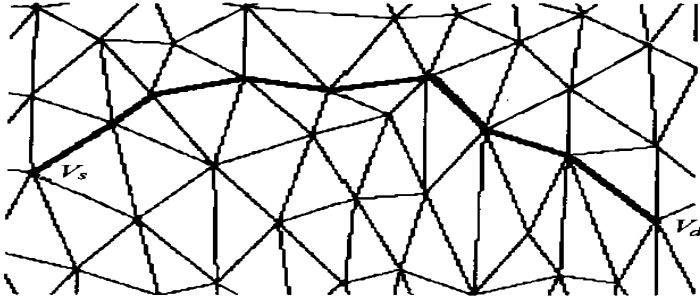
Fig. 1. The initial approximate shortest path

4   The Shortest Path Segment Iteration

According to the initial graph G, the shortest path is set of edges end to end of original triangular grid model, not the shortest path on grid surface. Therefore, the triangular around the vertices that are on the shortest path is subdivided, meanwhile vertices and edges after subdivision are added into G^0 , so as to form new subdivision graph G^1. Next, by calling the algorithm again described above, a new shortest path of subdivision graph is constructed, denoted as L^1. This iterative process continues repeatedly until the difference between L^{i+1} and L^i is less than a given error ε, which shows that a shortest path between two points on model is figured out by iteration. According to the ith iteration results, the （i+1） th iteration of the subdivision graph G^{i+1} is discussed as bellows.

The basic principle of constructing subdivision graph G^{i+1} based on creation of subdivision point, is inserting a new subdivision point on edge of graph G^i , then set up a new edge according to new inserted subdivision point. To keep balance of distance between new inserted subdivision points, the average number of insertion points η(the default parameter is set to 2) on each edge should be specified. And next, average length E^i of each edge in G^i is calculated, accordingly average distance ζ^i of inserted points will be figured out with the above results, described as $\zeta^i = E^{i/\eta}$. So, for every edge e (v^i_i, v^i_m) that needs subdivision furthermore, their number of subdivision points can be calculated, as the following equation shown:

$$m = | e^i_i - e^i_m | / mod\ \zeta^i - 1 \tag{1}$$

The creation of subdivision points employed in this article refers to determining whether S P(v1 , v2 , ⋯, v n) is part of SP^i (v^i1, v^i2, ⋯, v^in) firstly, and which is described as " existence condition of trapping sites".. If it is true, subdivision points are created according to "case 1" condition; otherwise in accordance with "Case 2".
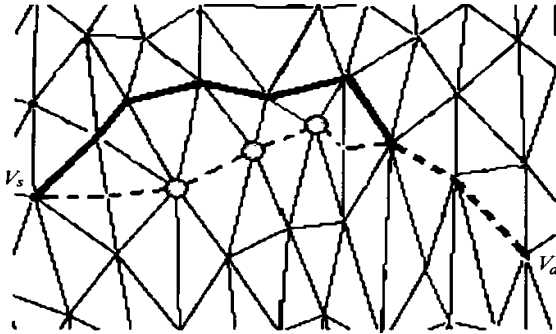
Fig. 2. The approximate shortest path after the 1st iteration
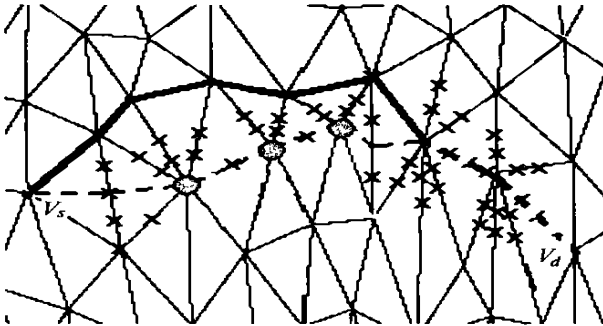
Fig. 3. Creation of subdivision points according to case 1
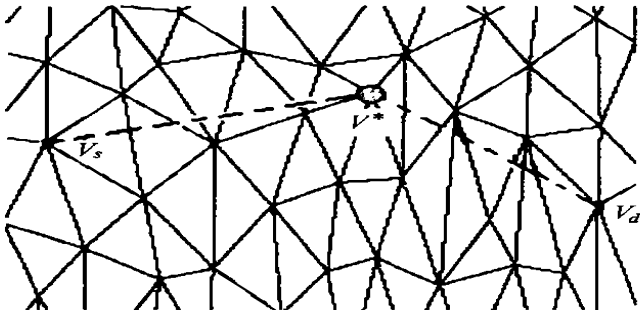
Fig. 4. Generation of "trap points" after iteration according to case 1
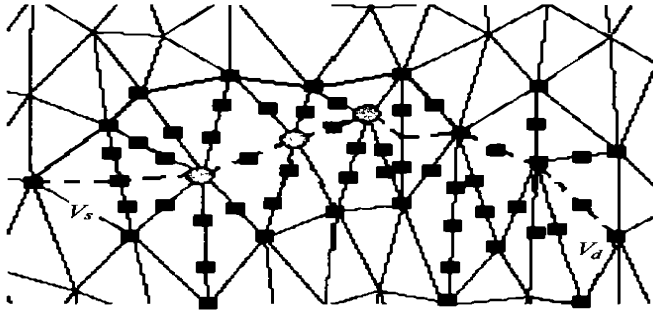
Fig. 5. Creation of subdivision points according to case 2
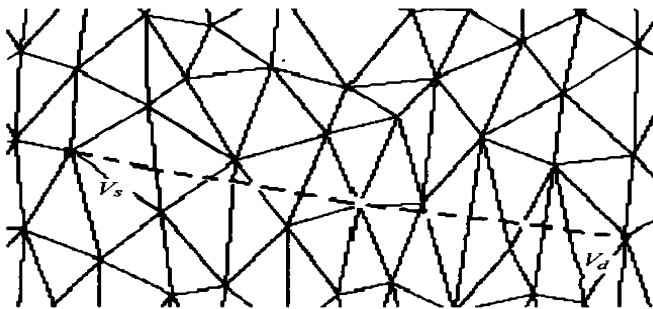
Fig. 6. Elimination of "trap points" after iteration according to case 2

Case 1:   V { $v_i^i|$ $v_i^i \in$ S P^i ($v_2^i, \cdots, v_{n-1}^i$) } is a set of points on the shortest path after the ith subdivision in graph G^i , and it is processed with operation of Nei2borV (v_i^i). For each associated edge e (v_i^i, v_m^i) of v_i^i , m additional points along edge e (v_i^i, v_m^i) are created according to equation (1). Newly generated subdivision points set and union of V { v_m^i | $v_m^i \in$ NeiborV (v_i^i) , $v_i^i \in$ V } and S P^i (v^i1, v^i2, $\cdots$, v^in) compose points in subdivision graph G^{i+1}.

Case 2: First, each point vii of set V { $v_i^i|$ $v_i^i \in$ InGraphG(S P^i (v_2^i, $\cdots$, v_{n-1}^i)) } will be coped with Edge (v_i^i) operation, so as to form set E{ $e_i^i|$ $e_i^i \in$ Edge (v_i^i) }. Next, for every edge of set E, subdivision points will be created according to equation (1). Union of All subdivision points set and V'{ v^i | $v^i \in$ Nodes ($e^{i)}$, $e^i \in$ E} compose points in subdivision graph G^{i+1}

According to "existence conditions of trapping points", creation of vertices of subdivision graph are divided into two cases, to avoid "trap points" generation in subdivision iteration of the shortest path. Fig.2 to Fig.6 are principle graphs of subdivision iteration based on the shortest path shown in fig.1.

Dotted line in fig.2 is the shortest path P1 after the 1st iteration, and 3 solid dots on the path are not part of P0, therefore, so trap existing condition is disestablished in the process of the 2nd iteration. If processed according to case 1, subdivision points will be created in the form of "×", as Fig.3 shown.

At 3 solid dots (possible trap points), the next iteration can be carried out only along the associated edges of the last iteration, which will lead to invariableness of V3 that is named as a "trap point". Consequently inflexion point is apparent in subdivision iteration result, as dotted line in Fig.4 shown. In this case, creation of subdivision points with method introduced in "case 2", will facilitate searching for a more reasonable moving position in the next iteration, as solid square points in Fig.5 shown, and as a result inflexion point is eliminated apparently, as the dotted line in Fig.6 shown. When "existing conditions of trap points" is established, subdivision points are created with method introduced in "case 1", with a faster subdivision.

In addition, the average number η of inserted points on each edge is equivalent to step length in the process of iteration and has an influence upon efficiency and accuracy of the algorithm. η is more large, and step length is more less consequently, then the number m of subdivision points on each edge is more large. Correspondingly, each iteration is inefficient, on the contrary, the accuracy of the shortest path is high. With a given η value, when discrepancy of length between shortest paths created by two consecutive iteration is less than the accuracy, it doesn't mean that the exact shortest path is found. When that occurs, η will be doubled, G^{i+1} is reconstructed again in accordance with G^i, meanwhile, new shortest path is searched out through iteration. If length discrepancy between shortest paths created by this time and the ith iteration is less than an specified accuracy ϵ, processing is over;otherwise, iteration will be continued with doubled η.

5   Summaries

In algorithm introduced in this paper, firstly the triangular grid model is represented as a weighted graph, and shortest path between two vertices of the weighted graph is constructed with Dijkstra algorithm. Then take it as initial approximate shortest path between two points, next, newly subdivided weight graph will be constructed by consecutive subdivision iteration of edges of related triangulars. In the end, Iterative approximation of shortest path between two vertices of grid model is implemented. And three-dimensional visualization of the algorithm is realized by OpenGL in Visual C^{++} 9.0.

References

[1] Yong, Z., Jie, C., Lei, L.: National Geomatics Center of data storage grid construction. Geographic Information World (04) (2007)
[2] Zhou, C.: A new era of geographic information system: GIS grid. Geographic Information World (04) (2007)
[3] Tiejun, C., Guo, L.: Grid-based geo-spatial information services key technologies. Surveying and Mapping (05) (2007)
[4] Wang, J.-Y., Zhu, Y.: On the grid and grid GIS. Surveying and Mapping (01) (2006)
[5] Wang, Y.: In intelligent transportation systems ijkstra algorithm efficient realize the D. Computer Engineering 33(6), 256–261 (2007)

[6] Company WangWenJing even method, ijkstra algorithm. J. the Realization of Surveying and Mapping Bulletin (8), 15–18 (2005)

[7] Wang, B.: Weeks, way of wearing bright. Robot path planning ij k the two-way D_Stra binary tree algorithm. Computer Engineering 33(10), 36–40 (2007)

[8] Yang, H., Xue, M.: Based on improved ant colony algorithm for railway network optimal Stroke path rules. Process and Project Work in Computer Application 46(3), 189–191 (2010)

Design and Implement of News Publishing System Based on MVC Design Pattern

Li Yong-Fei and Chen Zhen-Guo

Department of Computer
North China Institute of Science and Technology
Beijing, China
lyf518@ncist.edu.cn, czhenguo@gmail.com

Abstract. News Publishing System was a common web-based information release platform. A web-based news publishing system was developed with Java Web technology based on MVC design pattern, which included such functions as news displaying, news management and user management. And client-side programming technology, JavaBean and multi-tier design used for decoupling of code were analyzed.

Keywords: News publishing system, MVC, Client side programming, Multi-tier design.

1 Introduction

News Publishing System was a typical Web application, which had been widely used in information construction of various fields. For using the system as an information release platform, there were two types of user to involve: general news audience and news publisher, who was also the administrator of the system.

A News Publishing System with perfect functionality and fair well architecture was designed and implemented. Java Web technologies including JSP, Java Servlet, JDBC and Java Bean were used in developing. And MVC design pattern was employed for the system architecture.

2 Functions of News Publishing System

2.1 Function Modules of News Publishing System

Fig. 1 was function module diagram of News Publishing System. There were three modules, including news displaying, news management and user management.

2.2 Functions of News Publishing System

(1) News Displaying

The main function of news displaying was to display the news content, which achieved by two pages. One was the news listing page, which listing all the news title

Fig. 1. Function module diagram

in the order of publishing time; and the other was news viewing page, which showing the detail of a news article when the news title in news listing was clicked.

(2) News Management

The main function of news management was to add, edit, and delete news content. And in order to support more rich news content, a picture file of specify format and an attachment file of any format were allowed in the news content except for plain text.

To achiev the functions mentioned above, there were management page, adding page and editing page in news management module. In the management page, the links for adding, deleting and editing news were provided. In adding page, the function of adding news was implemented, and editing news in editing page. When adding, editing or deleting function had been executed, the management page was forward to.

(3) User management

The previous two functions were the main part of the news publishing system. In addition, as news displaying was to be used by all users, while news management could only accessed by administrator, there should be login with correct user name and password before using the news management function.

User management was used to manage the administrator's user name and password information, including adding, editing and deleting.

3   Design of News Publishing System

3.1   MVC Design Pattern

Pattern referred to a fixed solution for some problem abstracted from a recurring situation. Design pattern referred to the pattern used in software modeling and design process[1]. It would contribute to achieve a high quality and more efficient software when developing the system with appropriate use of design patterns.

MVC design pattern was currently governing architecture for building Web application. A Web application was divide into three parts, including Model, View and Controller, according to its different function for the purpose of minimizing the coupling between all objects constituting the system[2].

Model was the main part of an application, including the business data and business logic which implemented the system functions. When the model changed, the View would be notified and provided the ability to query the related state of model. Generally JavaBean was used as Model, which could be easily reused.

View was the user interface of an application, and responsible to obtain data from Model and show the data to user, and also accept user input data. There was not

business processing in the View. In web application based on request/response mode, View was in the browser side, and Model in the server side. Generally JSP was used as View.

Controller received request from the client and transferred the request into a certain behavior, which achieved by Model. After the behavior was completed, a appropriate View was selected by Controller to present the result data to the user, and fulfill the request.

The process of processing the request in MVC was as following: first, the user sent a request through the web browser; second, the Controller received the request, and passed it to the appropriate Model to do the business logic; third, the Model would address the result to the View in an appropriate way; fourth, the controller called the appropriate View JSP to show the result; fifth, the HTML document generated by JSP is returned to the browser as response, which presented the result to user. Fig. 2 was the structure of MVC Design Pattern[3].
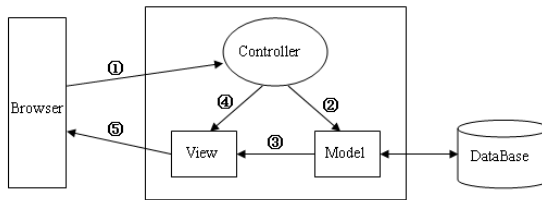
Fig. 2. The structure of MVC Design Pattern

3.2   Architecture of News Publishing System

Fig. 3 was the architecture of news publishing system based on the MVC design pattern. In the figure, C meant Controller, implemented with Java Servlet; M meant Model, implemented with JavaBean LogicBean; and V meant View, implemented with JSP. Additionally, the DBBean encapsulated the operations to access database, and the DataBase provided storage for news data and user data.

The solid lines indicated control flow, and dashed lines for data flow when processing the user request in Fig. 3.

The detailed control flow was: the HTTP request from client-side was sent directly to the Controller Servlet. The Servlet called the appropriate Model LogicBean
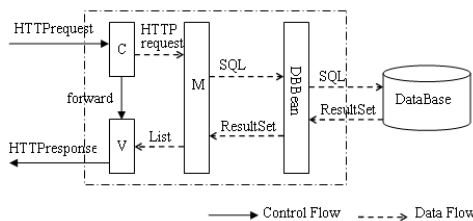
Fig. 3. Architecture of News Publishing System

according to the service type parameter in the request to do required business logic. And then the request was forward to the appropriate View JSP page by the Servlet. Finally, JSP page generated HTTP response and return back to the client to complete the processing of a request.

In this process, the detailed data flow was: the controller would pass the HTTP request containing all the requested data to the corresponding method of LogicBean. In LogicBean, if database operations is needed, the SQL statement would be generated and passed to the methods of DBBean. Then the DBBean sent the SQL statement to the database. After execution, the result set is returned from Database to the DBBean, and them from DBBean to the LogicBean. In LogicBean, the records in result set were to be processed into data object, and put into a list, then list was put into a public storage space, typically HttpServletRequest object. Finally, JSP acquired the data object from list and presented it to user.

4   Key Technology in Implementation of News Publishing System

There were several key issues to resolve in the implementation of news publishing system based on MVC architecture.

4.1   To Restrict the File Format of Upload Image in Client Side

When a news was published, an image could be uploaded to enrich the content of the news. But the file format of upload image could only be JPG or PNG considering the practical factors such as storage capacity. It is possible to restrict the file format by server-side program, but would increased the burden on network transport. Here JavaScript code in client side was used to achieve the goal, which reduced the communication between client and server.

The sample code was as follows:

```
<script language="javascript">
function checkNewsForm(theForm){
    var picValue=theForm.pic.value;
    var picForm=theForm.pic.value.substring((picValue.length-3),
                  picValue. length );
    if ((picForm!="jpg")&&(picForm!="png")){
        alert("Only jpg and png format supported!);
        return false;
    }
}
</script>
...
<form enctype="multipart/form-data" action="newsCenter?action=add"
    method="post" onSubmit="return checkNewsForm(this)">
...
    UPLOAD IMAGE：<input type="file" id="pic" name="pic"/>
...
</form>
```

4.2   To Encapsulate Database Record into Data Object for Data Transmission between Tiers

As the View of MVC design pattern, JSP was only responsible for present data, and should not import java.sql package to access the database result set. So, data object NewsBean was defined to encapsulate the record in the result set. Then each record in the result set corresponded to a data object, and all the objects were put into a list. By doing so, JSP was totally decoupled from database, and only treated with an ordinary object. It was illustrated in Fig. 4.

4.3   To Extract Code from JSP to Separate Processing and Display

In news listing page and news management page, it was necessary to display all the news title on the page. Apparently, it was too much to list all news title in one page because the total number of news would be hundreds or even more. Usually there would be a fixed number of news, for example 10, that be shown on one page. Then there should be processing code to decide which 10 news be shown on current page. As mentioned above, JSP was only responsible for present data, and there should not be too much code in the JSP to handle the problem. So a specialized JavaBean PageBean was developed, received the total number of news and the fixed number for one page and return the related values about news for current page. It was also illustrated in Fig. 4.
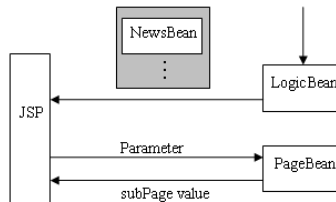
Fig. 4. Data object and PageBean

4.4   To Decouple the Components of System by Multi-tiers Design

When designing the system, multi-tiers design was adopted to divide the whole system into four tiers, including interface tier, business tier, data access tier and data tier. The system was decoupled with multi-tiers design. If the business logic of system was more complex, there could be more tiers to further decouple the system[4]. For example, DAO pattern could be used in data access tier, which separated the method definition and method implementation with DAO interface and DAO implement class. And O-R mapping could be introduced into data tier to add a persistence tier, which completely isolated the java code from the interaction with the specific database, and enabled to use the database through pure object-oriented way.

5   Summaries

News publishing system achieved such functions as news displaying, news management, and user management based on MVC design pattern and multi-tiers design, which would be helpful for information release.

References

[1] Liang, L.: Java Web Application Developing, Beijing: Publishing House of Electronics Industry (2007) (in Chinese)
[2] Liu, Y., Gao, L.-S., Wang, B.: Study and implement of distribution system based on J2EE and MVC design pattern. Computer Engineering and Design 28, 1655–1658 (2007) (in Chinese)
[3] JSP Architecture, JSP Model 1 architecture, JSP Model 2 architecture,
    `http://www.roseindia.net/tutorial/jsp/jsparchitecture.html`
[4] Qin, J.: Enterprise Java Development and Architecture. Publishing House of Electronics Industry, Beijing (2008) (in Chinese)

A Performance Measurement System for Supply Chain Based on AI and J2EE

Yongchun Wu and Zhenjian Jiang

School of Business, Shandong Jianzhu University, Jinan, China
clarence-wyc@126.com

Abstract. Performance measurement systems have become very popular and largely applied in organizations, as they provide not only a tool for measuring and controlling performance activities, but also a powerful management tool at the most strategic business level. The necessity for researching on a performance measurement system for supply chain based on AI and J2EE is discussed. Furthermore an integrated performance measurement system for supply chain based on AI and J2EE is proposed. Finally, the plan to design and implement the system is expounded.

Keywords: Performance measurement system, Supply chain, AI, J2EE.

1   Introduction

In the modern world the main focus of competition is not only between different companies but also between supply chains. The supply chain is a network of suppliers, factories, warehouses, distribution centers and retailers through which raw materials are acquired, transformed and delivered to the customer. Supply chains are growing increasingly complex, from linear arrangements to synchronized, multi-echelon, outward-facing networks of distributed servers. There is much more information that needs to be monitored than there was just a few years ago.

The overall performance of the supply chain significantly affects the financial health of all member companies. Therefore, an effective supply chain performance measurement process should be able to directly address performance areas that create sustainable profitability and financial strength. Performance measurement systems have become very popular and largely applied in supply chains, as they provide not only a tool for measuring and controlling performance activities, but also a powerful management tool at the most strategic business level. The performance measurement system will need to provide a reliable indication of the contribution of supply chain operations to the areas like growth, cost minimization, working capital efficiency and fixed asset utilization.

A robust, scalable and intelligent performance management system is the platform for improvement. It must be exception-based and allow users to automatically prevent problems, resolve issues, capture knowledge, and sustain improvements. The system must be able to handle an increasing number of users and amounts of information.

While it must be personalized and easy to use, it must also ensure high levels of security and privacy.

2   The Necessity for Researching on a Performance Measurement System for Supply Chain Based on AI and J2Ee

Many performance measurement tools are presented, for example the balanced socorecard, the supply chain operations reference model, the logistics scorecard, economic value added, and activity-based costing. Companies in a supply chain using above tools, face many problems, which are presented as follows.

1) Lack of specialized departments and professionals for the overall supply chain

In China, few supply chains have established specialized department for evaluating the performance. And capable professionals to do the job are rare. The current practice is that one company in the supply chain calls the relevant technological departments together to evaluate the performance when it is needed. Lacking the knowledge of industrial engineering and performance measurement, so-called experts from these departments have difficulties and limitations to fulfill the job.

2) Lack of advanced and standardized performance measurement procedure

The traditional performance measurement procedure is a series-connection one. Its total period is very long. This often results in doing repetitive work and wastes a lot of cost and time. For example, establishing the index system is the necessary work when evaluating performance.

3) Lack of advanced performance measurement tools

A large number of companies that now accomplish performance measurement in China still perform it manually without the help of a computerized system, which leads to low efficiency and affects their response speed seriously to the changing market and environment. Therefore, an advanced performance measurement tool is needed to change the current situation, increase efficiency, and boost the development of the companies.

4) Lack of intelligentized activies

There are many performance measurement tools, which are lack of intelligence activities. Therefore it is so difficult to automatically generate new evaluating reports according to the demands of users. A system based on AI is needed.

5) Lack of cooperative work

The performance measurement work often requires the team at different times and locations work together. Web-based performance measurement system meets this requirement. Built an integrated performance measurement system based on J2EE standards, performance measurement work will be successfully implemented in the supply chain.

To solve the above problems, it is necessary to standardize performeance measurement procedure, and to explore and research on an integrated performance measurement system based on AI and J2EE.

3   The Architecture of Performance Measurement System for Supply Chain Based on AI and J2EE

According to the J2EE typical framework, studied extensively and absorbed the international advanced performance measurement system thoughts, researched on performance measurement system practices in China, an integrated performance measurement system for supply chain based on AI and J2EE was framed. The architecture is shown in Figure 1.
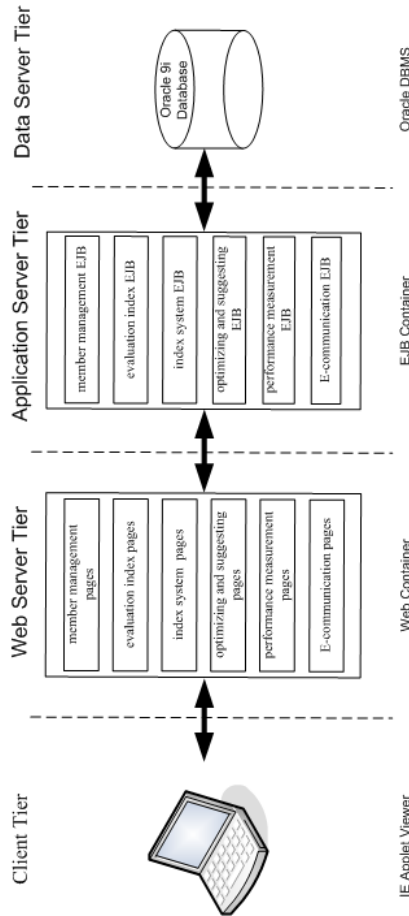
Fig. 1. The architecture of performance measurement system for supply chain based on AI and J2EE

4   The Function Designing of Performance Measurement System for Supply Chain Based on AI and J2EE

The performance measurement system has the main functions shown in Figure 2. The entire system is divided into member management, evaluation index, index system, optimizing and suggesting, performance measurement, as well as E-communication. The system integrates the well-known open source framework: Struts, Spring and Hibernate, designing MVC mode by using J2EE technology.
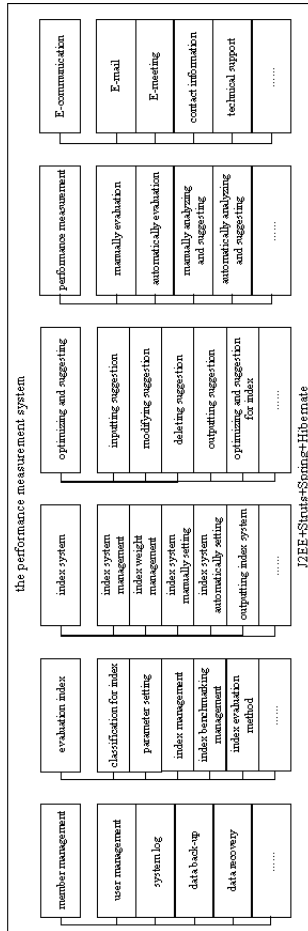
Fig. 2. The function architecture of performance measurement system for supply chain based on AI and J2EE

5 The Implementing of Performance Measurement System for Supply Chain Based on AI and J2EE

Most non-trivial web applications can be divided into at least four layers of responsibility. These layers are the presentation, persistence, business, and domain model layers. Each layer has a distinct responsibility in the application and should not mix functionality with other layers. Each application layer should be isolated from other layers but allow an interface for communication between them. For the presentation layer Struts will be used; for the business layer Spring will be used; and for the persistence layer Hibernate will be used.

1)The Presentation Layer

Here is what Struts is responsible for: (1)Managing requests and responses for a user. (2)Providing a controller to delegate calls to business logic and other upstream processes. (3)Handling exceptions from other tiers that throw exceptions to a Struts Action. (4)Assembling a model that can be presented in a view.(5)Performing UI validation.

For example, the system administrator will login the system. The key codes are provided.

The system administrator logining page-administratorLogin.jsp: The system administrator with the ID and password sends the requests.

```
<html: form action="administratorLogin.do" method="POST">
Struts-config.xml
<action path="/administatorLogin"
name="administratorLoginForm"
input="index"
scope="request"
type="org…web.struts.DelegatingActionProxy">
<forward path="/AdministratorContent.jsp"/name="success">
…
```

2)The Business Layer

Spring allows a sophisticated form of constructor injection as an alternative to setter injection as well. The objects are wired together by a simple XML file that contains references to objects such as the transaction management handler, object factories, service objects that contain business logic, and data access objects (DAO).

The way Spring uses these concepts will be made clearer with examples later in this article. The business layer should be responsible for the following: (1)Handling application business logic and business validation. (2)Managing transactions. (3)Allowing interfaces for interaction with other layers. (4)Managing dependencies between business level objects. (5)Adding flexibility between the presentation and the persistence layer so they do not directly communicate with each other. (6)Exposing a context to the business layer from the presentation layer to obtain business services. (7)Managing implementations from the business logic to the persistence layer.

The key codes of applicationContext.xml for Spring are provided.

```
    <bean name="/administratorLogin"
class="com.job.action.AdministratorLoginAction" singleton="false">
    <property name="administratorService">
    <ref local="administratorServiceProxy"></ref>
    </property>
    </bean>
    …
```

3)The Persistence Layer

Hibernate framework allows object-to-relational persistence and query service for Java. Hibernate has a medium learning curve for Java developers who are already familiar with SQL and the JDBC API. Hibernate persistent objects are based on plain-old Java objects and Java collections. Furthermore, using Hibernate does not interfere with the IDE. The following list contains the type of code that the developer would write inside a persistence framework: (1)Querying relational information into objects. Hibernate does this through an OO query language called HQL, or by using an expressive criteria API. HQL is very similar to SQL except the developer use objects instead of tables and fields instead of columns. There are some new specific HQL language elements to learn; however, they are easy to understand and well documented. HQL is a natural language to use for querying objects that require a small learning curve. (2)Saving, updating, and deleting information stored in a database. Advanced object-to-relational mapping frameworks like Hibernate have support for most major SQL databases, and they support parent/child relationships, transactions, inheritance, and polymorphism.

The key codes are provided, which can seek administrator with ID.

```
    <!--Access Data Interface IAdministratorDAO-->
    public interface IAdministratorDAO {
    public void saveOrUpdate(com.job.bean.Administratoradministrator);
    public void delete(com.job.bean. Administratoradministrator);
    public Administrator findAdministrator(String administratorID);}
    …
    <!--Display Data AdministratorDAO-->
    public class AdministratorDAO extends BaseAdministratorDAO implements
IAdministratorDAO{
    public AdministratorDAO() }
    public Administrator findAdministrator(String administratorID)
    …
    String hql="from Administrator as s where s.AdminID=: administratorID";
    …
```

6   Summaries

There are many deficiencies in implementing performance measurement in China. The JSP, Servlet, EJB components were based on J2EE platform, combining with the thinking of the three-tier structure and the design mode, which can simplify the development process, rapidly build the performance measurement system of a good scalability, maintainability, reliability and high availability, aiming to provide a

reference model for the development of computerized performance measurement systems, and to strongly push forward the implementation of performance measurement in China.

References

[1] Wu, Y., Han, N., Wang, H.: Implementation of the Computer Aided Business Process Reengineering System Based on UML and J2EE Technology, pp. 173–176. Computer Society Press (January 2008)

[2] Wu, Y., Jiang, Z., Han, N.: The Design and Implementation of the Integrated Supply Chain Management System Based on UML and J2EE Technology, pp. 1361–1365. Harbin Institute of Technology Press (March 2010)

[3] Xu, Y., Chen, J.-L., Li, Z.-B., Wang, Z.-X.: Study and Realization on Framework of Universal Performance Evaluation System for Supply Chain. Application Research of Computers, 152–154 (2004)

[4] Wang, M., Wu, Z.: Research and Design of Multi-tiered Fuzzy Comprehensive Performance Evaluation System for Suppliers Based on J2EE Technique. Computer Application and Software 25, 124–126 (2008)

[5] Zhao, X.-Y., Huang, X.-Y., Sun, F.-Q.: An Important Supplier and Customer Oriented Approach to Supply Chain Design. System Engineering 23, 34–38 (2005)

[6] Du, Y.-L., Zhu, X.-W.: Supplier Evaluation and Selection in Supply Chain Design. Journal of Tianjin University (Social Sciences) 9, 312–317 (2007)

Research on Tender Evaluation Intelligent Decision Support System of Engineering Project

Sun Xiujie, Lu Yanxia, Meng Yin, and Wu Yanan

Dalian Neusoft Institute of Information, Liaoning, Dalian, 116023

Abstract. As the rule of market competition, bidding embodies the principles of openness, fairness and impartiality. Tender evaluation is the first important stage, and which is also the key factors of bidding. However, the evaluation problem has more subjectivity and qualitative factors. It will make the fairness and impartiality in doubt. In order to much more fairness and impartiality, this paper adopts the theory of decision support system to improve the scientific and intelligent. And bidding is evaluated by the two common methods, which are "the lowest evaluated bid" and "percentile scoring system". At the end, this paper designs and implements the intelligent decision support system of tender evaluation.

Keywords: IDSS, tender evaluation, knowledge engineering.

1   Introduction

Construction project bidding is a method of choosing implementation department by the way of tender and bidding. How to determine the most satisfied successful bidder is a very critical problem [1]. However, because of the preference and capacity of experts, tender evaluation problems are always getting the surprised result. Therefore, it can't embody the fairness and impartiality [2].

Recently, more and more scholars are devoted to research the tender evaluation decision support system. However, most of the present evaluation systems belong to information system. Knowledge of evaluation expert is often represented by database. But database can't represent experience knowledge of expert. In this case, how to combine the knowledge engineering and intelligent decision support system(IDSS) will be the core problem. IDSS of tender evaluation will embody the scientific and advancement. At the same time, it can reduce the black case work.

IDSS combines artificial intelligent and DSS, applies the technology of expert system, and uses mankind knowledge to help solving the complex decision problem. And it is an assistant decision system. IDSS was putting forward by Bonczek in 80 decade [3]. Recently years, IDSS is gradually applied to many aspects. Such as macro economy, case mining, data mining, accounting, financial, environment management and ecological management, and so on. Based on the blackboard, this paper designs and realizes the tender evaluation IDSS of engineering project.

2   Method of Tender Evaluation

Tender evaluation is a very important link in the system of tender and bidding. How to choose the evaluation method will relate with whether the project most value can

be realized or not. A scientific, effective, impartial, and reasonable method will play a vital role, which can guarantee engineering quality, compress building cost, and shorten building period.

However, actual evaluation problems and methods are variety. Most of international problems adopt the method of "quantities tender, the lowest bidder winning". At the same time, the "percentile indicator evaluation and review technique" and "double-envelope" are usually used. Some scholars adopt MCDA (multi-criteria decision aid) method and its especial software to evaluate. In Chain still have some other methods to be used. These methods are "two-stage", "A+B", AHP, Fuzzy evaluation, and so on.

The most common evaluation in practical engineering project has two kinds of methods [4]. One of the methods is "percentile scoring system". Evaluation team will evaluate according to all standards in the tender files, such as quotation, quality, period, organization design, technical schema, managing results, and so on. By score to the bidding files, and choose the highest score [5]. The other method is "the lowest evaluated bid", which requires satisfying substantive requirements of bidding files. At the same time, the bidding price should be the lowest, and not be lower than reasonable scope of costs.

3   Process of Engineering Project Evaluation

By analyzing the tender evaluation process of engineering project, we can find the common characteristics. Most of tender evaluation can be included two stages, qualification pretrial and detail examination. Firstly, the qualification of bidder will be examined. Then, when they are qualified, experts will evaluate all aspects of bidder files from quotation, period, and construction design. In the period of pretrial, mainly including the following content:

(1) The validity of bidding files: if the bidder files have not signed according to the provisions, it will be invalid.

(2) The integrity of bidding files: if the files have not quotation, legal power of attorney, certification files, construction design and project manager profile, it will be invalid.

(3) The consistency of tender and bidding files: if the bidding files have not substantial react to the tender files, it will be invalid.

(4) The performance and credit of bidder: if the performance of recent 3 years can't meet the provisions, and in recently 2 years, have punishment of city level, it will be invalid.

In the period of detail examination, mainly including the following 7 aspects:

(1) The qualification of enterprise
(2) The credit of enterprise
(3) The experience of enterprise
(4) The organization design of construction
      ① The general layout
      ② The schema and method
      ③ The arrangement of planning
      ④ The arrangement of human, machine, and equipment
      ⑤ The quality, technology and safety guarantee

(5) The experience and qualification of project manager
(6) The experience and qualification of main construction human
(7) The reasonable of quotation

In the first stage should be very strict, however, most of projects have no pay attention to it. So, this paper uses the computer to finish the preliminary. The main process of tender evaluation is as follows:
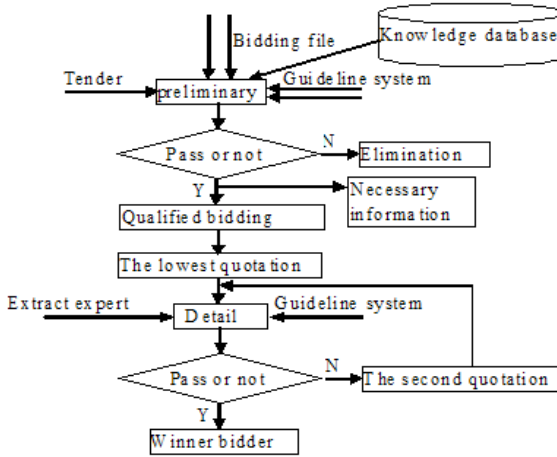
Fig. 1. The flow chart of the main process of tender evaluation

4   Construction of Engineering Project IDSS

Based on above evaluation methods and process, this paper designs an intelligent decision support system (IDSS) of tender evaluation. The system structure is as following:

The IDSS can be divided into four function modules: tender files management module, bidder files management module, evaluation decision support module, and reporting output management module.

Tender files management module: mainly including insert, update, delete, and select the tender files, at the same time, it can be searched of the bidder information in this module. In the DB, electronic tender file can be stored.

Bidder files management module: mainly including insert, update, delete, and select the bidder files, and electronic bidder file can be stored in the DB.

Evaluation decision support module: mainly be responsible for evaluation of bidder files. It is core part of whole system. In this section, including two kinds of evaluation modules, which is "the lowest evaluated bid" and "percentile scoring system". We can choose one of the methods for a special tender project. Of course, we can also adopt the two methods. Simultaneously, this module will give the comprehensive sequencing result, which can provide support to the decision maker.
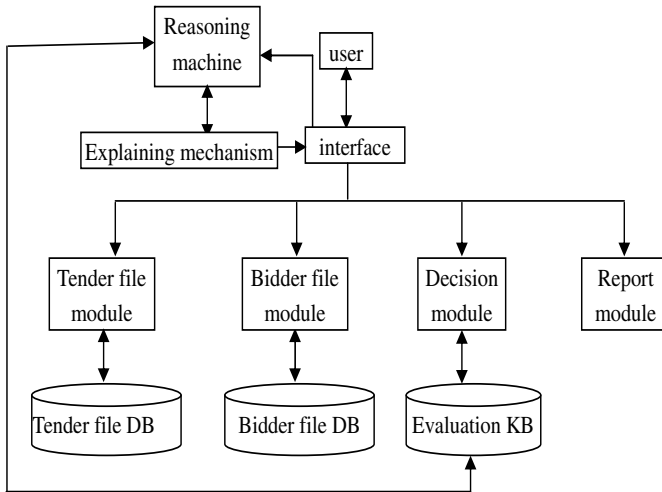
Fig. 2. The whole structure diagram of tender evaluation IDSS

Reporting output management module: mainly be responsible to output the bidder information, qualified bidder files, abandoned tender and comprehensive evaluation result.

5   Summaries

This paper researched common evaluation methods of engineering project, analyzed process of tender evaluation. And divided into two steps: qualification preliminary and detailed examination. This paper adopted related concept and theory of decision support system, designed and developed an IDSS of tender evaluation.

References

[1] Zou, T., Chen, H.: Project Bid Evaluation on the Basis of Extention AHP Method. Science Mosaic (8), 199–201 (2010)
[2] Zhang, J., Qu, Y., Wang, J.: Research on evaluation expert voice notification system technique. Shanxi Architecture 36(32), 279–281 (2010)
[3] Li, F.: New direction of DSS-Intelligent Decision Support System. Journal of Northeast China Institute of Electric Power Engineering 21(4), 86–89, 97 (2001)
[4] Wang, L., Yao, J., Liu, B.: Application of gray relative analysis method based on entropy weight in bid evaluation of power equipment. Journal of Shenyang Institute of Engineering: Natural Science (1), 32–34 (2010)
[5] Yang, X.: Application of Lowest Evaluated Bid Price in Project Bidding. Communications Standardization 22, 120–122 (2010)

Development and Realization of Product Configuration Management Function in Mass Customization

Li-Xia Wu

Computer Science Department, Baoji University of Arts and Sciences,
Baoji, Shaanxi 721016, China

Abstract. The product configuration in the Mass customization plays a very important function in the effective management of product data, in full use of existing resources, and in quickly realizing custom design products. This article discusses the modeling of product structure icon, and transformation model of multiple views of product structure is given. The use of acyclic graph version implements the version of parts and the management of current state. The author gives a detailed analysis of the development and role of a variety of configuration rules, and illustrates mode of action of configuration rules in the configuration search process. Data integration and exchange among Multi-systems and more strengthened configuration management can be achieved through the improvement of configuration management model and implementation, strengthening the endorsement of management and support of product data and information.

Keywords: Mass customization, Product configuration, BOM, Version management, Configuration rules.

1 Introduction

Product structure and configuration management is an important function of product data management. With electronic warehouse as its underlying support, and bill of material(BOM) as its core data organization, it links all the engineering data and documents that define the final products to realize data organization, management and control. It also provides different views of product structure and description for users or application system restricted by a certain goal or rule[1]. Thus, the product structure and configuration management are composed of product structure management and configuration management. Product configuration management, based on the product structure management, defines and implements a large number of data relationships, rules, limits to control the product data consistency, legitimacy and universality[2]. Product configuration is the process of reorganizing product information and data to reach the goal of configuration management.

The product structure as the center，the configuration management of product organizes product information in order to overcome the disorderly and unsystematic from the document management. Thus the serialization product management is effectively achieved. Its functions include multi-view management of material list, configuration rules, knowledge base management, version control of products and parts, as well as other systems of data integration and so on [3].

The main functions of product configuration management include: the creation and management of BOM, the multi-view modeling management and query of product structure, the version management of parts, the configuration rules management and the integration of MRPⅡ, ERP system, etc. The realization method of each function will be discussed as follows.

2   Creation and Management of BOM

BOM records the material information of parts and their assembly combination, which is the detailed presentation of product structure. The product structure constantly changes during design and production process, so that BOM will be also be updated accordingly. The creation and management of BOM refers to the automatic creation of BOM from the structure tree, keeping the consistency of both data, and also realizing the dynamic updates.

Product BOM is the nonlinear tree structure, each part of nonlinear structure can be considered as a node of the tree. The product is the root node, and each part the leaf node. According to the structural relationship, there will be different "layer" structure. The only relationship table can be recorded in the relational database is not tree structure, but the relation table. Therefore, one important problem of realizing BOM management is the splitting of relation table from tree structure and the reconfiguration of tree structure from the relation table.

According to the tree structure of database, BOM data can be divided into multi-layer BOM and single-layer BOM. Multi-layer BOM records the level number and the identification number, and the sequence number of parts in the tree structure in details. The same parts assembly in different levels of should be recorded repeatedly in the relation database table. For example, one layer structure of product contains A and B that are the set membership, which are contained in another layer. The multi-layer BOM does not distinguish among all records, which easily cause the data redundancy of database, and make the data maintenance become difficult; therefore, it is not good to the change of set membership of parts in BOM and the real-time updates of product structure (If not all the same set relationship in the database is updated, then parts assembly relation confusion would happen). But the single-layer BOM only record the parts material of adjacent levels, namely the relationship between parent and child. That is to say, each kind of parts assembly relation of product structure is only recorded once, and the assembly relation that has the same set relationship can be merged, unlike the multilayer BOM that repeatedly adds a record. The single-layer BOM usually adopts two tables (the table for parts and the table for product structure) to record the relation of product structure. To get all the parts information of the product structure, the above two tables performs joins operations [4].This structure can reduce data redundancy and improve data maintenance.

3   The Multiple Views Modeling and Inquires of Product Structure

The product is composed of parts, and parts are composed of components. The hierarchical relationships between products and parts usually adopt tree diagram. Figure 1 shows the assembly hierarchical structure of a product. The product is

represented by the boot node of the tree, the branch represents component relationship, the leaf nodes represent parts, and the nodes containing the branching structure represent components. It is obvious that tree diagram clearly expresses the hierarchy structure of products and the relationship between the parts.

Different department has different requirements for product structure. It means that parts of the same products are classified from different views. For example, the design department classifies the parts according to the product structure, the technology department according to the processing technology of parts, the manufacturing department according to the ability of technology and equipment, etc. Therefore, the product model in different views will have different forms. Different views need different data format, which is in accordance with multiple views of BOM.

While the multiple views of product structure are converted, the consistency and accuracy of the structure data in different views must be ensured.
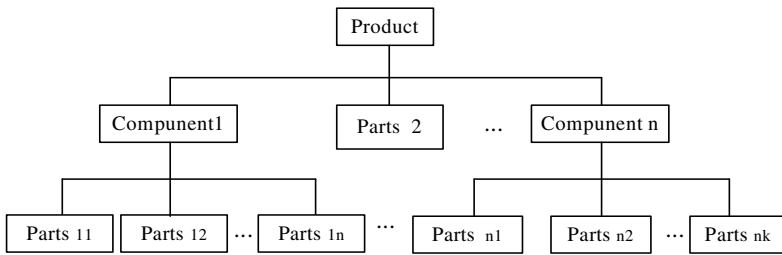
Fig. 1. Tree model of product structure

4   The Version Management of Product Components

The products may change in the process of design and manufacture along with the alteration of design scheme or other requirements. The changes from local correction to structural change will affect the change of parts. The changes of local alteration in the design process, without affecting main functions and interchangeability of the components, will form a new version rather than creating a new part. Conversely, when the structure of a component changes results in the influence of the function and interchangeability, a new part is created.

Multiple version of parts is formed in the process of design and change. The version management model of parts can be classified into: linear structure version model, tree structure version model, and directed acyclic graph version model[5]. Linear structure version model lists the components in chronological order according to version, each version at most has one parent- version, and only one son-version. Each version has only one unique identity, and the new version automatically is inserted into the chain tail, as shown in Figure 3 (a). Linear structure version model cannot distinguish from the redesigned version and the modified version and cannot present multiple paratactic versions. The tree structure version model represents the relationship between versions with tree structure. The version with only one root node does not have the parent-versions, and others all have parent-versions. A version with the branch means that it has several alternative new versions; if not, it means the son-version is the revised version, as shown in Figure 3 (b). Though the tree structure version model can distinguish from the redesigned version and the modified

version(in Figure 3 (b)),V1 , V2 and V0 belong to replacement relationship, V3,1, V3 belongs to the modified relationship), it can't correctly express the fusion and the derived relation of multiple versions. Acyclic graph structure model is shown in Figure 3 (c). It expresses the fusion and the derived relation between versions, and also can track version. But since it cannot express the logic gradation of versions (Figure 3 (c), does V4 belong to the third layer or the fourth layer?), the order and source produced by versions can only be represented by the node number.

（a）Linear structure version model

（b）Tree structure version model   （c）Acyclic graph structure model
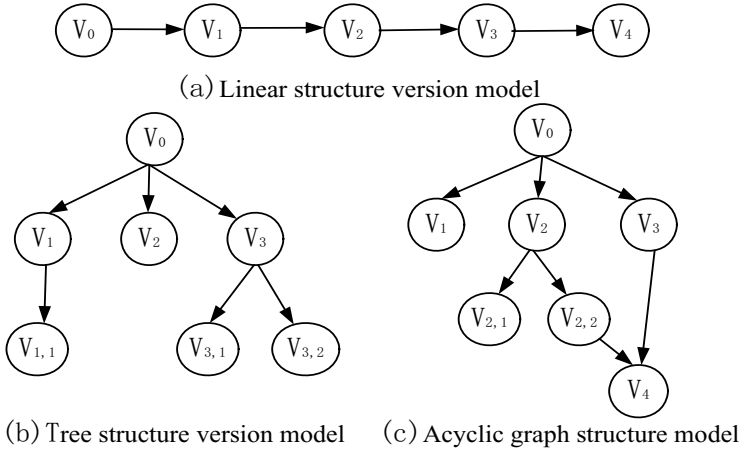
Fig. 2. Version model of parts

The changes of Version should be executed according to the following principles [6]:

The change of parts will be represented by the new version if it does not influence the interchangeability and by the new parts coding if it influences the interchangeability.

When the sub-components change, the components remain unchanged. When the sub-components of parts adopt new parts but does not affect the interchangeability, it is represented by the new version; if affecting interchangeability, the new parts coding should be adopted.

5   Configuration Rules Management

Configuration rules include the relations of parts assembly combination, the constraint relation, the version restrictions relations, attribute variables corresponding relation with structure, etc. According to different attributes, they are divided into mandatory rules and user-defined special rules [7]. Mandatory rules mainly record the effectiveness of the parts and the assembly constrains relations between them. The effectiveness is in connection with version, time and serial number while the constraint relations includes dependence, conflict, interchange, etc. The User-defined special rules are closely related to specific product object attribute, such as variables rules, which need to be made for specific product attributes.

The management of configuration rules is mainly the rational configuration rules and the effect of configuration search process. The draft and management of

mandatory rules is relatively easy. Assembly constraint has been partly formulated in the process of design, and it only needs determination and implementation in the use. The properties of configuration objects will be in a different state during different periods of time, which makes the selection of search process become easier. By contrast, customization and management of user defined rules will be more difficult. When it is used, the reasonable variable must be used to describe attributes of the object. Each product should have its own characteristics. For example, refrigerator variable of refrigeration power and working volume in air conditioning products are put together, but the refrigeration power might be considered in air-condition product. Once the variables are determined, the scope of value is determined. The value range of variable must include a series of characteristic value, which has better versatility. How to take specific value for variables in range is linked to the parts of configuration and the satisfaction of specific rules, which should be the different methods for specific products, and also need a lot of product design, use, and other aspects of knowledge and experience.

6   Summaries

This paper studies the basic function of product configuration management and its realization based on the configuration of realization. The creation and management BOM, the multi-view modeling management and query of product structure, the version management of parts and the configuration rules management can ensure the effectiveness of the process of product configuration. The product configuration under the large-scale customization needs to be built in a manufacturer with complete information management, which requests the integration and cooperation of multiple platforms. It also realizes the exchange, updates and reorganization of product information flow data between multiple systems, and provides information management of the whole life cycle of the products, and then reaches the goals of configuration management. Further improvement and optimization are needed for the configuration of complex product structure and custom management.

References

[1] Li, J.-M., Liu, N.-R., et al.: Product Data Standard and PDM, pp. 200–212. Qsinghua University Publishing Company, Beijing (2002)
[2] Tong, B.-S., Li, J.-M.: Product Data Management (PDM) Technology. Qsinghua University Publishing Company, Beijing (2000)
[3] Ye, X.-J.: Product Structure Management and Configuration Management Function of Product Daia Management. Computer Engineering and Applications 5, 29–32 (1995)
[4] Wang, X.-Z., Chen, X.-N., Pang, X.-M., et al.: A single- layer BOM and multi- views mapping. Mechanical Design and Manufacturing 2(4), 45–47 (2004)
[5] Yu, Y., Lu, J., et al.: Version Management Modle of PDM Based on Multicolor-Map. Journal of Computer-Aided Design & Computer Graphics 13(12), 1080–1087 (2001)
[6] Deng, L., Zan, X.-W., et al.: A Product BOM Management Method Based on Single Layer BOM. Journal of Chongqing University (Natural Science Edition) 25(6), 5–7 (2002)
[7] Yang, J.-H.: The Research of Product Structure and Product Configuration in Product Data Management System. Huazhong Normal University, Wuhan (2002)

The Analysis and Verification of Cooperative Workflow Based Interactive Abstract Graph

Liu Jian-Xin[1,2], Sun Lin-Lin[3], Du Yu-Yue[2], and Hong Yong-Fa[3]

[1] College of Information Science and Engineering,
Shandong University of Science and Technology,
[2] Modern Education, SUST
[3] College of Civil Engineering and Architecture,
Shandong University of Science and Technology

Abstract. To detect whether there are deadlocks in the analysis and verification of cooperative workflow, the method of Petri net accessibility analysis is often adopted. In order to alleviate possibility of state space explosion, this paper proposes the analysis approach based on interactive abstract graph. Interactive abstract graph only affects the elements of collaborative interaction, makes use of these elements of the partial order to discover whether there exists any deadlock situation by detecting inappropriate partial order of contradictory logic relations. Besides, it can reveal the hidden internal conflicts directly and its efficiency and accessibility are much higher. This method provides a new way for deadlock detection of cooperative workflow as well as concurrent system.

Keywords: Workflow, Deadlock, Interactive abstract graph, Petri net.

1   Introduction

The program of Client/Serve is not peer to peer, when the client asks for service, the server responds and offers the service, and the client accepts it, the task is completed. I t can be illustrated by Petri net model as follows [1].
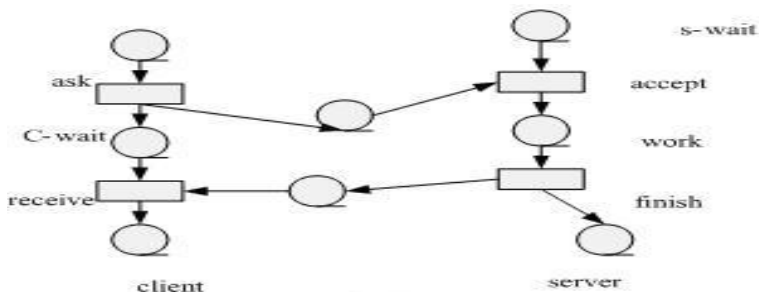
Fig. 1. The Petri net model of client/server program

As is shown in fig1, the occurrence of the event "accept" is restricted by the event "ask" and rather controlled by it. The event "receive", while running spontaneously, is

restricted by the event "finish" and rather controlled by it. Thus, the event whose occurrence is controlled by another is termed controlled event. While events like "finish" and "ask" that control the occurrence of other events are termed controlling event. The relationship between the controlled and the controlling is symbolized.

   ask→accept ; finish→receive;   pattern: controlling event→controlled event.

Interactive abstract graph deals with the very relationship between them. In the Fig.2, ● denote controlling event; ▭ denote controlled event. Controlled event is represented by the resistance-like symbol, for it is to occur after the corresponding controlling event occurred. While controlling event is not similarly restricted. The "line" in which the events occur represent a thread which corresponds with a working path [2] [3].

   The Petri net model of client/server can be transformed into Interactive abstract graph as is shown in Fig 2.
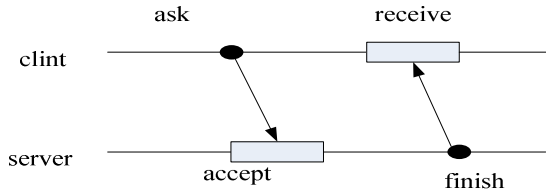
Fig. 2. The interactive abstract graph of client/server program

   Here we give the definition of interactive abstract graph as follows:

Definition 1: an interactive abstract graph is a 3-tuple $G = (T_s, \rightarrow, ---)$, where T_s is a set of events; →and-- is the set of relations on T_s as follows:

$$T_c = \{ t \mid t \text{ is the controlling event of } G \text{ graph}\}$$

$$T_r = \{ t \mid t \text{ is the controlled event of } G \text{ graph}\}$$

$$T_c \cup T_r \subseteq T_s$$

As to the interactive abstract graph having N threads, T_i is the event set of the number i thread , $i = 1,2...N$, $T_s = \bigcup\limits_{i=1}^{N} T_i$

Definition 2: the relations on the interactive abstract graph

   Relation $\rightarrow = \{ (t_a, t_b) \mid t_a \in T_i \wedge t_b \in T_j \wedge t_a \text{ controlling } t_b \}$

   Relation $-- = \{ (t_a, t_b) \mid t_a \in T_i \wedge t_b \in T_j \wedge t_b \text{ happened nest to } t_a \}$

   Preorder relation $\langle = \{ (t_a, t_b) \mid (t_a \rightarrow t_b) \vee (t_a -- t_b) \}$ means that t_a is bound to happen before t_b . Obviously, Preorder relation is anti-symmetric and transitive. Then the transitive chain forms the preorder route. There are some relations in Fig 2 as follows:

ask $\rightarrow$ accept   ; finish $\rightarrow$ receive ;

ask--receive   ; accept--finish ;

preorder route：ask $\langle$ accept $\langle$ finish $\langle$ receive

2   The Verification of Cooperative Workflow Based on Interactive Abstract Graph

An interactive abstract graph involves n threads. On each thread, the events are in total order. Only when all the controlled events meet the controlling events, can the thread moves forward. And its moving is affected by other threads. When it is interactive with another one, the existing total order may be affected indirectly via the controlling relationship. If in a preorder transfer chain there is an anti-sequence-order event pair, it suggests that the moving thread has logical contradiction, which hinders the thread from advancing, and results in a deadlock. The threads advance smoothly without deadlocks; the order relation is favorable and logical. Interactive abstract graph can help to judge whether the Petri model of cooperative workflow has deadlocks. It can be illustrated with the following instance [3] [4].
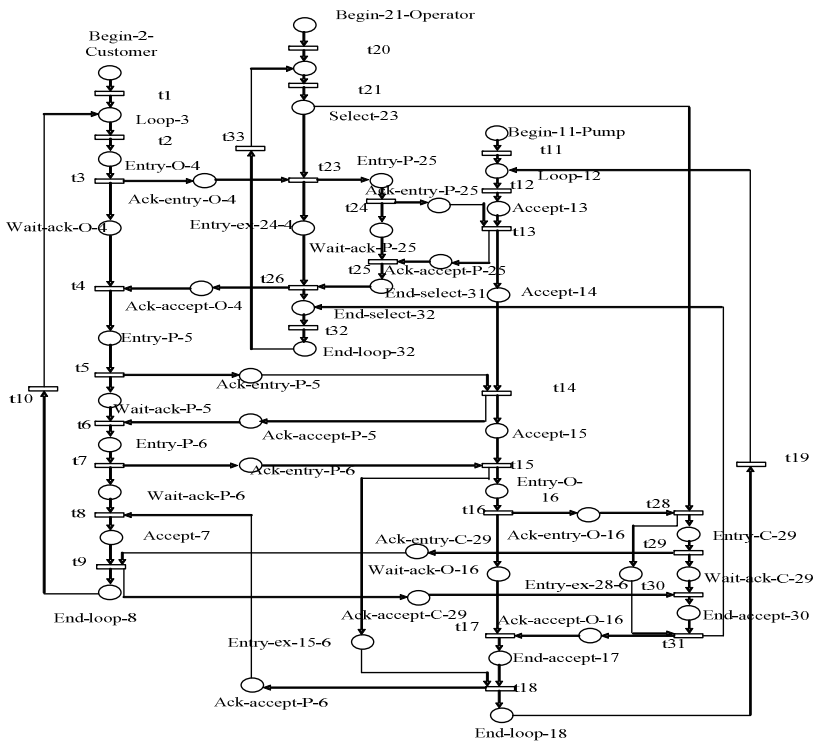
Fig. 3. The Petri net model of automatic stations

The auto gas station problem is typical of cooperative workflow. The customer comes to the station to fuel the automobile, involving the cooperative interactive process between the customer, the operator and the pump. The process goes like this. The customer anticipates a sum from the operator; the operator starts the pump to fuel the automobile and then shuts off the pump and charges the customer. The bill paid, the operator and the pump are restored to their position. The workflow cycle is completed. The cooperative workflow model of the auto gas station problem is as follows. Fig 3 is the Petri net model of automatic stations.

To make the reachable graph less complex, The Petri net model of automatic stations can be divided into 3 sub-models shown in Fig 4.
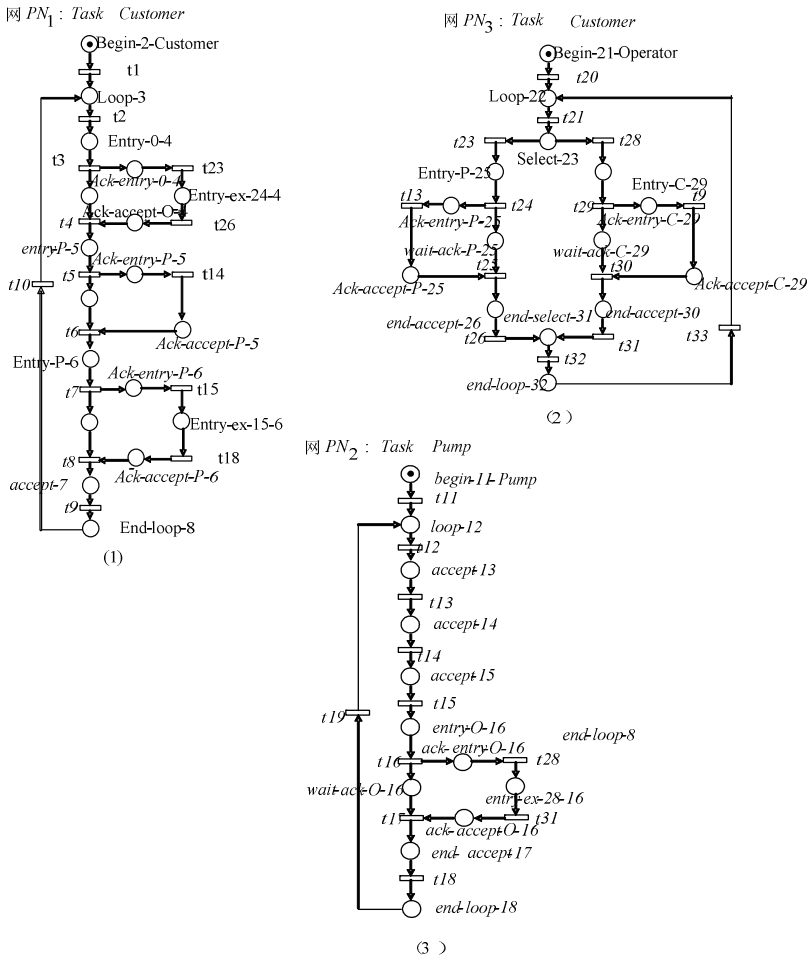
Fig. 4. The subnet model of graph 3

As Customer carries out the sentence Pump. Finish and waits for the response, and the task Pump, to accept the transfer by finish, needs to execute the sentence Operator. Charge (requiring the cabinet to offer the change). Similarly, the task operator, to undertake the transfer by charge, needs to execute Customer. Charge. (Requiring the customer to accept the change.) To undertake this transfer, in turn, it has to execute the sentence Customer Accept, which follows Pump. Finish. From the reachable graph, the Petri net gets into the state of a deadlock after executing the transition t_7, t_{16}, t_{29}. Therefore, the program ends up a deadlock and cannot advance. With the approach of the reachable graph, the execution of the program while the deadlock occurs is judged by the reachable indication. The indication is the vector set of all the tokens in places, and consequently the judgment is troublesome. And the following is the approach of interactive abstract graph to the auto-gas-station problem.

3   The Approach of Interactive Abstract Graph to the Auto-Gas-Station Problem

The same auto gas station problem is dealt with. The process goes like this. The customer anticipates a sum from the operator; the operator starts the pump to fuel the automobile and then shuts off the pump and charges the customer. Fig 5 is the previous interactive abstract graph [5].
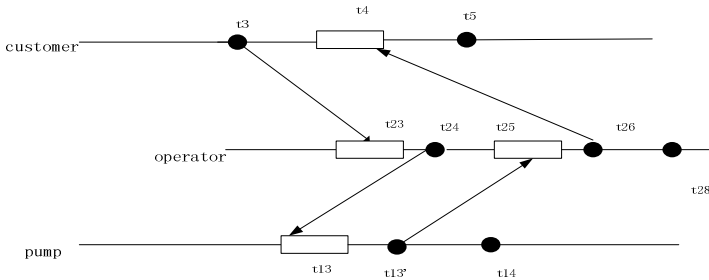
Fig. 5. The previous interactive abstract graph

The interactive abstract graph alone can also demonstrate concurrent running conditions. To explore the dynamic feature of the system, we can simulate the functioning.

t_c , t_o , t_p represent the events to occur in the three threads. At the beginning, $t_c = t_3$, $t_o = t_{23}, t_p = t_{13}$

(1) for c, $t_c = t_4$ when t_3 occurs; for o, t_3 controlling t_{23} , t_{23} occurs, hence t_{24} occurs. $t_o = t_{25}$;

For p, t_{24} controlling t_{13} , $t_p = t_{14}$

2) since t_{13} controlling t_{25} , for o : t_{25} occurs , then t_{26} occurs, $t_o = t_{28}$

(3) since t_{26} controlling t_4 , t_4 occurs, $t_c = t_5$

This phase is completed smoothly, $t_c = t_5, t_o = t_{28}, t_p = t_{14}$

With reference to Fig 5, in the sequence of the preorder path: $t_3 \langle t_{23} \langle t_{13} \langle t_{25} \langle t_{26} \langle t_4$, no anti-sequence order events occur. Like wisely, no anti-sequence order events occur in other paths. It reveals the system works smoothly.

Now Fig 6, the graph demonstration of the latter phase, is analyzed.
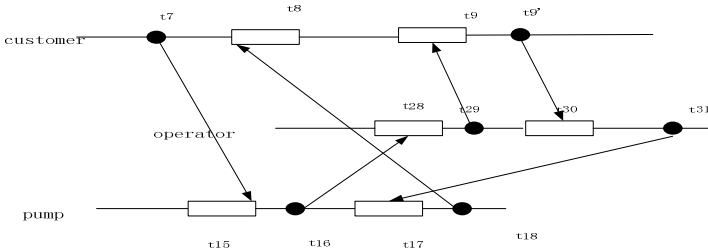
Fig. 6. The later interactive abstract graph

Running simulation:

Starting: $t_c = t_7$, $t_o = t_{28}, t_p = t_{15}$

$t_7 \rightarrow t_{15}$, $t_c = t_8, t_{15}$ occurs,   and then t_{16} occurs, $t_p = t_{17}$

$t_{16} \rightarrow t_{28}, t_{28}$ occurs,   and then t_{29} occurs, $t_o = t_{30}$

$t_{29} \rightarrow t_9$, as for customer, t_8 has not occurred, therefore t_9 cannot occur, customer is blocked. In t_8 operator and pump cannot advance as scheduled. The system gets into a deadlock. Meanwhile $t_c = t_8, t_o = t_{30}, t_p = t_{17}$

With reference to Fig 6, starting with t_7 , a preordered path of the interactive abstract graph is $t_7 \langle t_{15} \langle t_{16} \langle t_{28} \langle t_{29} \langle t_9 \langle t_{30} \langle t_{31} \langle t_{17} \langle t_{18} \langle t_8$ However, $t_8 \langle t_9$, this contradiction shows the system do not work, and a deadlock emerges.

The case is: the operator informs customer of paying the bill at t29 and then the operator waits for the payment (wait-ack-c-29) , the customer takes out money (accept-7) , and after taking out money (t_9 occurs)   he informs the operator (t_{30} occurs) . Therefore it is supposed to be $t_{29} \rightarrow t_8$ .It is figured out that logical contradiction inappropriate control arcs produced has resulted in the deadlock. The pre-order-path approach can reveal the logical contradiction.

It can be found that figuring out the preorder path is closely related to $\langle^+$, the transitive closure to compute the preorder relation. Therefore, whether cooperative workflow has deadlocks can be detected by computing the transitive closure.

4   The Further Research Work

The following problems should be tackled, handling the circulating arc, grasping the essence of cooperative and interactive attributes, compressing the original net, easing computation, avoiding state-space explosion.

References

[1] Murata, T., Shenker, B.: Detection of Ada Static Deadlocks Using Petri Net Invariants. IEEE Transactions on Software Engineering 15(3), 314–326 (1989)
[2] Godefroid, P., Wolper, P.: Using Partial Orders for the Efficient Verification of Deadlock Freedom and Safety Properties. In: Formal Methods in System Design, vol. 2(2), pp. 149–164. Kluwer. Academic Publisher (1993)
[3] Duri, S., Buy, U., Devwapalli, R., Shatz, S.M.: Using State Space Reduction Methods for Deadlock Analysis 34(3), 51–60 (1993)
[4] Corbett, J.C.: Evaluating Deadlock Detection Methods for Concurrency Software. IEEE Transactions on Software Engineering 22(3), 161–180 (1996)
[5] Du, Y., Jiang, C.: Formal Representation and Analysis of Batch Stock Trading Systems by Logical Petri Net Workflows. In: George, C.W., Miao, H. (eds.) ICFEM 2002. LNCS, vol. 2495, pp. 221–225. Springer, Heidelberg (2002)

A New Method for Measuring Firewall Throughput

Liu Guowei, Qian Xiubin, Yan Tengfei, and Huang Shaoqing

Beijing Information Security Test And Evaluation Center

Abstract. Problems of current method for measuring firewall throughput are analyzed, and a new method for measuring firewall throughput is given. Samples are measured by the two methods, the results are compared, and the advantages of the new method are reached.

Keywords: Throughput Measuring, Performance, Stability, Pass rate.

1 Introduction

With the deepening of information, the amount of data transmitted over the network increases explosively. This requires higher performance network equipments. Firewall as a common device to isolate intranet and extranet, is also to be required higher performance. And in test area it is required to give accurate performance information of firewall. Throughput as an important firewall performance indicator, is widely used. But the current dichotomy throughput measured method has some problems, and cannot show the true performance of firewall.

2 Problems of Current throughput Measured Method and Result Representation

In 2010, our laboratory organized an internal capacity comparison, the test engineers were divided into four groups to give the test results. When the firewall throughput was measured, the testing condition, tested sample and it's configuration, testing equipment, testing method, and so on, were exactly the same. The throughput measurement results were as show in Tab.1. It can be seen that the second group result was totally different. It was found by analyzing the measuring process, that for some reason on the dichotomy point of 25%, the throughput measurement was passed, so the throughput that the second group given was 25%. But on the point 25%, it was not passed, when the other three groups measured. So the other three results were very different with the second group.

Now we usually use dichotomy approaching to get the throughput value. In theory, this method is perfect. But actually, there are some problems. Dichotomy method requires that the sample must be very stable. That is the throughput measurement result should have good repeatability. But now the structure of information equipment is more and more complex, and the equipment has more and more functions. So there are more and more factors to affect the throughput measurement, which leads to an unstable interval of throughput. If the cut-off point of dichotomy is right on the unstable range, the result of throughput measurement will be very different. The experimental data later in this paper can also prove this.

Table 1. Four groups measuring results

Group NO.	1	2	3	4
Throughput Measured Result (%)	18.12	25.00	18.11	18.12

Note: Throughput is expressed by percentage of wire-speed, test data packet size is 64 bytes.

On the other hand, the representation of the current throughput measured method just gives a single value (or gives the average of multiple measurements), regardless of what kind of data processing, stability of the sample cannot be expressed, and the result cannot reflect the actual performance of the sample. The throughput value given by this method is only a reference value.

The problems of the current throughput measured method are summarized.   It cannot rule out the impact of sample instability, and the single value result cannot accurately reflect the performance of samples. The result can only give a reference value, and cannot give the stability indicator.

3   The New Throughput Measuring Method

Firstly a concept, pass rate, is given. On a throughput value point, throughput is measured repeatedly, the measured times that packet loss rate is 0 are the ratio of the total measured times. That is the pass rate on the throughput value.

According to statistical theory, in the case of infinite measurements, the pass rate of higher throughput value must be lower than the lower throughput value or be equal. Based on this theory, a new method for measuring throughput and result representation is given. Throughput is measured enough times on each throughput value, and the pass rate of each value is calculated. Then throughput is as the x-axis, the pass rate is as y-axis, the two-dimensional curve is drawn. The example curve is as show in Fig.1.
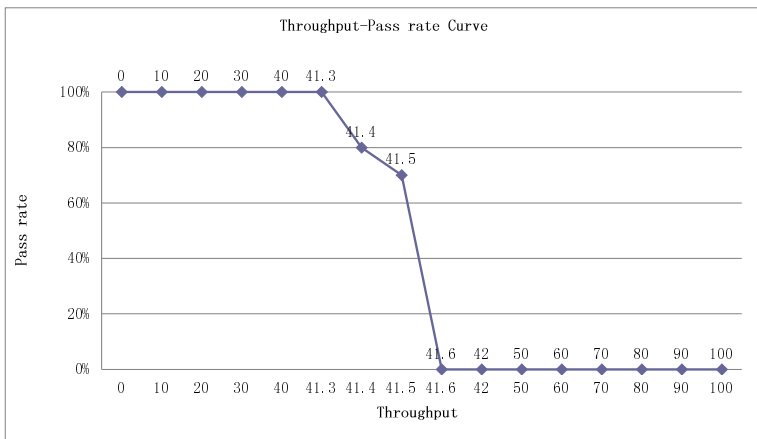
Note: Throughput is expressed by percentage of wire-speed, test data packet size is 64 bytes.

Fig. 1. Throughput-pass rate example curve

This curve can be divided into three parts according to the pass rate.

The range of throughput whose pass rate is 100% (as in the example curve 0 to 41.3 part of throughput), this is a stable range. The highest 100% pass rate throughput point (as 41.3 in the curve) can indicate the quality of performance, higher value of the point indicates higher performance.

The range of throughput whose pass rate is from the last 100% point to the first 0 point (as in the curve 41.3 to 41.6 part of throughput), this is an unstable range. The changes of this range in the size and slope can reflect the stability of the sample. The big range and great slope changing indicate bad stability. Vice versa indicates good stability.

The range of throughput whose pass rate is form the fist 0 pass rate point to the curve end (as in the example curve 41.6 to the end part of throughput). In this range, the measurement totally cannot be passed.

Through such measurement and representation, it is able to indicate the performance and the stability of the sample. That is the new method.

4   Experimental Data

This paper selects three firewalls (marked as Sample A, Sample B, Sample C) as the measured samples. Respectively, they are measured by the proposed method and the traditional method, and the results are compared.

The measured firewall is only open IP access policy between intranet and extranet. Test data packet size is 64 bytes. Throughput is expressed by percentage of the wire-speed. It is measured 10 times on each throughput value to facilitate the calculation, when the proposed method is used. It is measured 3 times and the average is given, when the traditional method is used.

The result of Sample A is as follow.
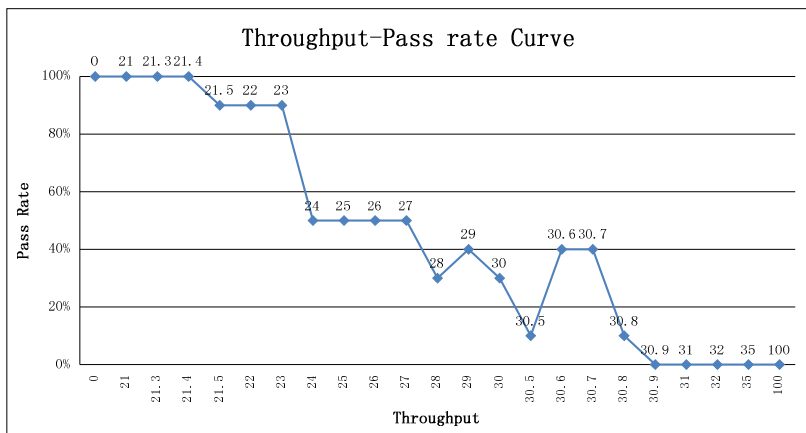
Fig. 2. The throughput measured result of Sample A by the proposed method

Table 2. The throughput measured result of Sample A by the traditional method

NO.	1	2	3	Average
Throughput Measured Result (%)	27.64	31.25	24.46	27.78

The result of Sample B is as follow.
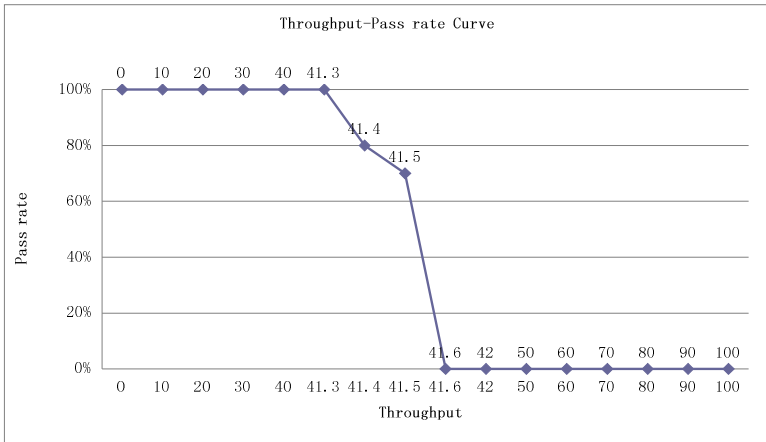
Fig. 3. The throughput measured result of Sample B by the proposed method

Table 3. The throughput measured result of Sample B by the traditional method

NO.	1	2	3	Average
Throughput Measured Result (%)	37.49	41.55	40.61	39.88

The result of Sample C is as follow.
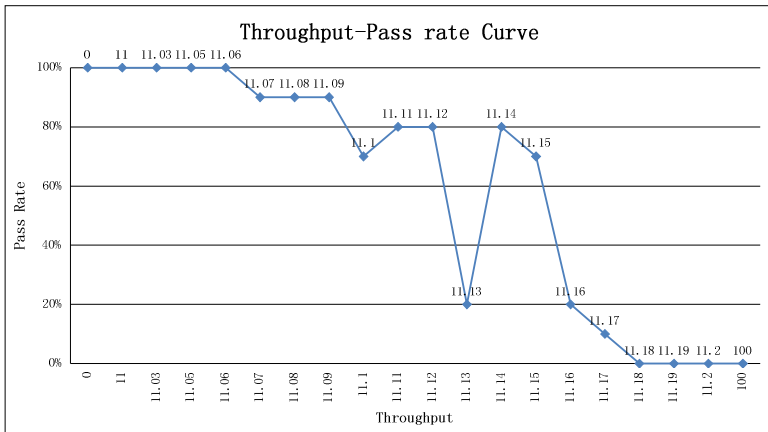
Fig. 4. The throughput measured result of Sample C by the proposed method

Table 4. The throughput measured result of Sample C by the traditional method

NO.	1	2	3	Average
Throughput Measured Result (%)	11.17	11.13	11.23	11.18

5   Data Analysis

It can be seen from the two-dimensional curve of the three samples (measured by the proposed method) that throughput value of the last 100% pass rate of the Sample A is 21.4%, Sample B is 41.3%, Sample C is 11.06%. Therefore, the performance is sort: Sample B> Sample A> Sample C.

The unstable range of the three samples can be seen that Samples A is 21.4% ~ 30.9%, the slope change is large; Sample B is 41.3% ~ 41.6%, the slope change is small; Sample C is 11.06% ~ 11.18%, the slope change is slight. Therefore, the stability is sort: Sample C> Sample B> Sample A.

On each sample, the 3 values measured by the traditional method are very different. It is calculated by a formula to indicate the stability of the samples. The calculations are as follow.

Sample A: (Maximum measured value - the minimum measured value) / average = 24.4%;
Sample B: (Maximum measured value - the minimum measured value) / average = 10.2%
Sample C: (Maximum measured value - the minimum measured value) / average = 0.9%

The calculated result shows that the Sample A is the most unstable, the Sample C is the most stable. The result of Sample A measured by the traditional method has the greatest deviation, which shows that the more unstable the sample is, the greater deviation the results have.

The average value of Sample A measured by traditional method is within the unstable range, the average value of Sample B falls on the left side of the unstable range, the average value of Sample C falls on the right of the unstable range. So the average value of the traditional method is not accurate, it is only a reference value, and cannot show the true performance of the sample.

Comparing the two measuring methods and result representations, it can be seen that the new method is more scientific, more realistic to show the actual performance of samples. And the new method can indicate the stability of the sample, however the traditional method cannot.

6   The Advantages of the New Method

According to statistical principles, measurements are repeated, and statistical analysis is done. So the impact of sample instability is avoided. Expressed by two-dimensional curve, the performance and stability of the sample can be clearly shown.

After properly modified, the method can be used for measuring other firewall performance indicators, such as delay, back to back. On the other hand it can be used for measuring other network equipments except firewall, such as routers, switches and so on.

7   Discussion

1. How many times should it be measured on each throughput value? In theory the more times the sample is measured, the more accurate the result is. But the actual needs to limit measuring times based on reliability requirements and confidence interval. For example, on 95% confidence interval how many times it needs to be measured?

2. Which throughput value should be chose to measure? That is how to split the throughput value. It should be done according to accuracy and convenience requirements.

3. The existing measuring equipments are designed according to traditional method. If the new method is used, there are too much manual interventions and statistics work.

References

[1] RFC 2544
[2] Freedman, D., et al.: Statistics. W.W.Norton & Company Inc. (1991)
[3] Lin, C., Shi, X.-Q., Hu, B., et al.: Network Performance Testing and Analysis. Higher Education Press (2009)

[4] Application Delivery Testing Methodology. Ixia Technologies Co. Ltd. (2009)
[5] Cheng, J.-X., Chen, W.-L., Song, J.: Compatability Relation and Compatability Kernel. Journal of Anhui University 4(28), 1–4 (2004)
[6] Yao, Y.Y.: Relational Interpretations of Neighbourhood Operators and Rough Set Approximation Operators. Information Science 111, 239–259 (1998)
[7] Galton, A.A.: A Generalized Topological View of Motion in Discrete Space. Theoretical Computer Science 305, 111–134 (2003)

A Design and Research of Image Detection System Based on Embedded

Hao Wang

School of Electronic and Information Engineering
Liaoning University of Technology
Jinzhou, China
wanghao6811@163.com

Abstract. Design and research an image detection system based on the embedded and involve cameras driver programming, detection methods for the campaign target and the module of getting or displaying image. This image detection system which is based on an embedded image detection technology has designed optimization methods of image detection and how to acquire image information. Image detection is used to complete for real-time video image detection and it is mainly adopted in video surveillance system, which can make people detect collected image, mark activities and acquire or show image information.

Keywords: image detection system, optimization algorithm, image acquisition.

1 Introduction

In recent years, with the rapid development of society and economy, the image detection was used by more and more people or places, especially in video surveillance systems. This has designed an image detection system built on the embedded platforms. Meanwhile, through the connection with USB cameras, we may detect images on the analysis of the detection algorithms research in order to obtain and display images. Along with the rapid economic development, image detection technology is expanding, improving the detection efficiency greatly and reducing the cost of labor with its simple, clear characteristics, and is becoming an important component of many systems.

2 The Overall Structure

This image detection system consists mainly of s3c2410 microprocessor and connects cameras through USB. System chart show as Fig.1.
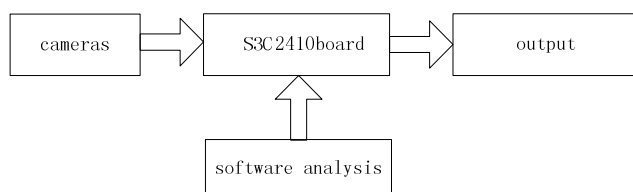
Fig. 1. System chart

For the design of this system, we select the company's samsung s3c2410 processor, which contains embedded chips s3c2410 and is based on ARM9 framework, owing basic frequency of 202 MHz, flash of 64MB and SDRAM of 64MB. In the design of hardware platforms, we select Linux as the operating system. Moreover, Linux operating system has a fine performance, public source and it is widely used for operating system with its small size, a higher execution speed and a good network performance advantages.

For building software platforms, it includes guiding bootloader, transplanting linux kernel and transplanting root filesystem. Here, we use u-boot because it can support multiple hardware infrastructure and a number of operating systems.

3   The Design Process of Image Detection Modules

3.1   Transplant of USB Driver Module

Driver Module of USB is shown as Fig.2.
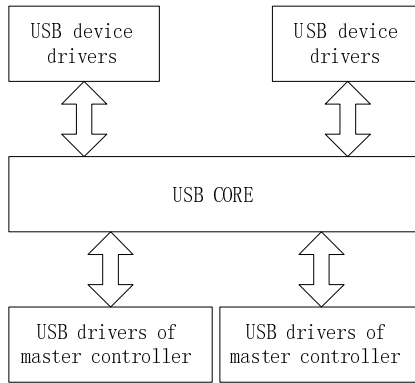
Fig. 2. Driver Module of USB

(1)New media directory in the directory linux-2.6.24.4/drivers/usb, copy gspcav1-20100611.tar.gz to media and decompress it. Modify Kconfig and Makefile in the directory linux-2.6.24.4/drivers/usb.

```
[root@vm-dev usb]$ pwd
/work/kernel/linux-2.6.24.4/drivers/usb
[root@vm-dev usb]$ vi Kconfig
Add source "drivers/usb/media/Kconfig"
[root@vm-dev usb]$ vi Makefile
Add obj-$(CONFIG_USB_SPCA5XX) += media/
(2) new Kconfig and Makefile in the media in order to add gspcav1-20100611
[root@vm-dev media]$ pwd
/work/kernel/linux-2.6.24.4/drivers/usb/media
[root@vm-dev media]$ vi Kconfig
```

```
#
# USB Multimedia device configuration
#
comment "USB Multimedia devices"
depends on USB
config USB_SPCA5XX
tristate "USB SPCA5XX Sunplus/Vimicro/Sonix jpeg Cameras"
depends on USB && VIDEO_DEV
```
(3) Modify Makefile of gspcav1-20100611
[root@vm-dev gspcav1-20100611]$ pwd
/work/kernel/linux-2.6.24.4/drivers/usb/media/gspcav1-20100611
[root@vm-dev gspcav1-20100611]$ vi Makefile
gspca-objs := gspca_core.o decoder/gspcadecoder.o
obj-$(CONFIG_USB_SPCA5XX) += gspca.o
clean:
rm -f *.[oas] .*.flags *.ko .*.cmd .*.d .*.tmp *.mod.c
rm -rf .tmp_versions
(4) Compile kernel
①Multimedia devices --->Video For Linux
②USB support ---> Support for Host-side USB
--- USB Host Controller Drivers
OHCI HCD support
③--- USB Multimedia devices
USB SPCA5XX Sunplus/Vimicro/Sonix jpeg Cameras

3.2   Algorithm Optimization of Image Detection

1) Detect movement target
This paper uses campaign target detection algorithms based on frame difference to judge whether is movement target in the monitoring area.

The process of detecting movement goal is shown as Fig.3.

Concrete steps of detecting movement goal are shown below:

a）the conversion of gray scale
Extract data of the current frame image $f_t(x,y)$ and forward a frame $f_{t-1}(x,y)$ from a series of video datas. First turn it into gray scale if collected images are the RGB color picture. Convert it according to the formula:

$Gray_t(x,y)= B(x,y)*0.11+ G(x,y)*0.59+R(x,y)*0.3$

Meanwhile, (x,y) is the coordinate of pixels in the image, R,G and B represent the red, green and blue stress components of the pixel. $Gray_t(x，y)$ represents gray-scale value after the conversion.

b）Differences management
Operate between $f_t(x,y)$ and $f_{t-1}(x,y)$ to obtain the absolute of temporal difference.   In order to get difference image $D_t(x，y)$:

$D_t(x,y)=| f_t(x,y)- f_{t-1}(x,y)|$

C）binaryzation
Deal with difference image $D_t(x,y)$ through binaryzation according to the formula:

$$P_t(x\ ,\ y) = \begin{cases} 1, D_t > T1(ForeGround) \\ 0, D_t <= T1(BackGround) \end{cases}$$

It is foreground when difference data is greater than threshold T1, else it is background. Here, choosing threshold is very important. This paper uses the way of the largest class of difference to choose threshold.
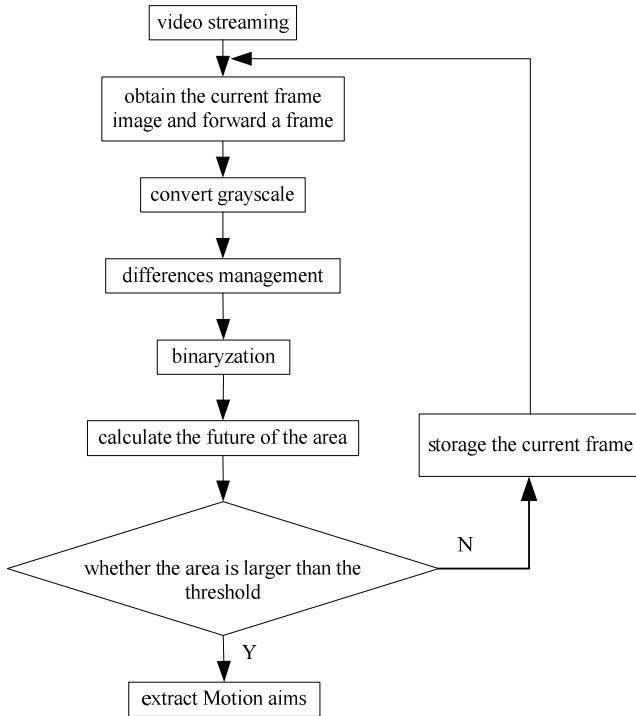
Fig. 3. Video collection process

Its concrete steps are:

Sum of all pixels of difference image is N and its grey level is L. Here, L is equal to 256 and N is equal to 320*240. Sum of all pixels which have i grey level is equal to n_i, so the probability of various degrees is:

$p_i = n_i/N$, i=0,1,..........,i-1

Image is separated into two areas and its threshold is t. Pixel and degrees whose gray degrees is from 0 to t constitute a region between a and b for pixels whose gray degrees is from t+1 to L-1. Their probability is:

$$P_A = \sum_{i=0}^{t} P_i$$

$$P_B = \sum_{i=t+1}^{L-1} P_i$$

Average gray degrees of regional a and b represent for:

$$u_A = (\sum_{i=0}^{t} P_i * i) / P_A$$

$$u_B = (\sum_{i=t+1}^{L-1} P_i * i) / P_B$$

The general average of image is

$$u = P_A * u_A + P_B * u_B$$

Variance between class between regional a and b is:

$$g(t) = P_A (u_A - u)^2 + P_B(u_B - u)^2$$

When t of g(t) is the largest, it is requirement of the threshold.

d）Decision-making part

There is a movement target if the value in the image of the edges of the area beyond a certain threshold. Threshold is fixed according to the size of the monitoring area. If the monitoring area is larger, the threshold will be little some. And if the monitoring area is smaller then the threshold will be great.

In these circumstances, this text proposed that we calculate the area of future regional using regional weighted. Binary image is divided into the n*n a small area and each region was given to a different value according to the concern and objectives of the different weight. Meanwhile, people may judge a motion after comparing the future of the area with the prescribed threshold.

When judging that there is a goal in the picture of the current frame, people can deal with the objective analysis about this module. When there is no objective judgement movement, it is stored in the current frame.

2) Establish Background model

The text proposed the method of background update based on common areas. Its basic idea is that people can deal with stored m frame of video image without movement in regular intervals. Separate each frame picture into n*n small areas, then calculate mathematical expectation and variance of each regional pixel. Here, M is equal to 5 and N is equal to 4.Then steps are:

①Time interval is T and read stored m frame image without movement video every T time;

②The resolution of image is 320*240 and separate each frame picture into 80*60 small areas. There are 16 small areas and calculate mathematical expectation and variance of each area. At last, write them into vector(E_{ji},F_{ji}).($1 = <i< = 5$, $1 = <j< = 16$);

③Calculate average vector(A_j ,B_j) of same area of m frame image:

$$A_j = (\sum E_{ji}) / M, \quad B_j = (\sum F_{ji}) / M \qquad i = 1,2,\ldots,M, j = 1,2,\ldots,16$$

④Calculate vector of same area and minimum distance of average vector in this area:

$$D_j = \min|(E_{ji}, F_{ji}) - (A_j, B_j)| \quad i = 1,2\ldots\ldots.M, j = 1,2\ldots\ldots.16$$

3) Extract Motion aims

(1) Differences management

To obtain the absolute between two frames:

$$D_t(x,y) = | f_t(x,y) - Gray_t(x,y)|$$

(2) Separate and extract the campaign target

Deal with difference image $D_t(x,y)$ through binaryzation:

$$P_t(x, y) = \begin{cases} 1, D_t(x, y) > T2(ForeGround) \\ 0, D_t <= T2(BackGround) \end{cases}$$

Image acquisition and display

(a)Get the source code and decompress
[root@vm-dev webcam]# tar –vxjf jpegshow.final.tar.bz2
[root@vm-dev webcam]# ls
jpegshow
[root@vm-dev webcam]#
(b)Code analysis
Modify Makefile and make include and lib point to their path。

4   Summaries

This image detection system has been based on embedded and has designed optimization methods of image detection and acquired image information in order to provide a good foundation for information transport and real-time detecting. Moreover, it has a high reliability and reduces cost of software and hardware greatly. It is becoming an important component of many systems.

References

[1] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
[2] Lu, Y., Zhang, C.: The video and image acquisition of embedded linux computer programming. Techniques and Maintenance (2006)
[3] Hui, S.C., Wang, F.: Remote Video Monitoring Over the WWW. Multimedia Tools and Applications 21(2) (2003)
[4] Ran, H.C.: The fire detection system on neural network. Fire Safety Science 9(1), 34–38 (2000)
[5] Piccinini, F.: Quality control of screen printing media. CFI-CE-RAMIC Forum International 82(12) (2005)

The Research of RSA-Based Undeniable Signature Method

Xin Li[1] and Chunxiao Liu[2]

[1] Computing Center
Liaoning University of Technology
Jinzhou, China
`lg_lx@163.com`
[2] College of Electronics and Information Engineering
Liaoning University of Technology
Jinzhou, China
`lg_lx@126.com`

Abstract. In order to resolve the problem of the non-repudiation of digital signatures, this paper analyzes the RSA signature scheme and the undeniable signature scheme, based on this an improved undeniable signature algorithm and the confirmation protocol using the legitimacy of the signature verification key validation is proposed. Experiments show that the method can ensure non-repudiation of digital signatures to achieve a secure exchange of documents promising security technologies.

Keywords: digital signature, RSA, undeniable signature.

1   Introduction

Extensible Markup Language (XML) has become the network data representation and information exchange standard [1]. With a wide range of XML applications in various fields, using XML to transmit information need more security requirements, and its security is also more influenced by people's attention. XML digital signature technology in the protection of transaction information confidentiality, integrity, authenticity and anti-denial is particularly one important part. The current general XML digital signature [2,3,4] can only be signed to ensure data integrity, not to ensure non-repudiation of the signature itself. However, there are very few XML undeniable signature on the research, the literature [5] although proposed a method with XML undeniable signature, but did not discuss the undeniable signature scheme, the confirmation protocol and repudiation protocol, and therefore it can't confirm the security. This paper studies RSA mechanism and undeniable signature scheme based on an improved undeniable signature algorithm, and applied to the structure of XML digital signature <SignatureMethod>, formate a new XML undeniable signature methods, including the generation of signature, protocol and deny recognition agreements. The method can fully guarantee the integrity of transaction data, verifiable and non-repudiation.

This paper is organized as follows: Section 1 describes principles of XML digital signature; Section 2 describes the traditional RSA undeniable signature algorithm and

signature scheme based on the study and improvement of this undeniable signature algorithm; Section 3 of undeniable signature algorithm completeness, reliability, and non-repudiation; Section 4 proposed undeniable signature algorithm based on improved the XML undeniable signature method; Section 5 concludes.

2   RSA-Based Undeniable Signature Algorithm

2.1   RSA Digital Signature Principle

An RSA key pair using two large prime numbers is the result after operation: one is the public key [7], known to many entities; the other is the private key, in order to ensure its confidentiality and integrity, must be strictly controlled and only its owner can use it. RSA encryption algorithm use the basic features of a key in the keys of the encrypted message can only be decrypted with the other, which also reflects the RSA system, the non-symmetry [8].

RSA digital signature is generated by an entity with its private key to encrypt plaintext. This encryption allows an entity to send messages to multiple entities without prior exchange secret key or encryption key, the receiver with the sender's public key can decrypt [9].

RSA digital signature process is as: $S = m^d$ mod n, where m is the message, S is the result of digital signature, d and n is the message the sender's private key. Message decryption process is as: $m = S^e$ mod n, where e and n is the message the sender's public key.

RSA-based undeniable signature has some important properties. First, the verification process to confirm the agreement confirmed to interact with to achieve. Then, using the denial protocol, confirmed that those who can prove a forgery. Confirmed forgery need to meet the following properties: for a given message and signature, the signature confirmation protocol output valid, then for the same input, denied that an agreement will not be output, then it is a forgery. Confirm and deny the combination of these two agreements, while protecting the signature of the recipient and the signer, and digital signatures to keep the traditional features of non-repudiation [10]. Therefore, there are three main undeniable signature components: signature generation algorithm, confirm the agreement and denied the agreement.

2.2   RSA-Based Undeniable Digital Signature Algorithm

Definition 1: For a positive integer k, let $[k] \overset{def}{=} \{1, ..., k\}$, Z_n^* representative of an integer model for the multiplicative group of n, $\phi(n) = (p-1)(q-1)$ for this group of the order.

For an element $w \in Z_n^*$ ord (w) represent w the order in multiplicative group Z_n, ord<w> representatives orde of w in multiplicative group(Z_n^*).<w> representatives subgroup generated by elements $w \in Z_n^*$.

Define a set $N=\{n \mid n = pq, p<q, p=2p'+1, q= 2q'+1,$ and p, q, p', q' are primes$\}$. System by the signer set as follows: Choose a random element n$\in N$, Choose elements e, d$\in \phi(n)$, makes ed $\equiv$ 1 mod ϕ (n); Choose a (w, S_w), Among them $w \in Z_n^*$, $w \neq 1$, $S_w = w^d$ mod n; Setting public-key parameter is a tuple (n, w, S_w), Set the private key is (e, d). For here pk representatives all the above generated tuple (n, w, S_w) collection. Specifically, w set as a fixed value $w = 2$.

2.2.1 Generate Signature

Ordinary RSA signature use the signature operation on the message m to sign, that is calculated $S_m = w^d$ mod n, output (m, S_m) pairs. More precise, in the exponentiation of the message m before message m first passes through the appropriate encoding format (such as one-way hash) for processing, so the results can not be forged to ensure the signature. Given a message m, with a deal on behalf of the m encoded output.Therefore, m the results of the signature will be $S_m \overset{def}{=} \overline{m}^d$ mod n. For (w, S_w) on for, with S_w representative w^d mod n, namely direct to w rather than $\overline{w}$ exponentiation.

2.2.2 Confirmed Agreement

Figure 1 shows the digital signature confirmation protocol, which consists of a confirmed and a validation to interact to perform.Agreement is a public key input parameters,namely, $(n, w, S_w) \in \Pi K$ and $(m, \hat{S}_m)$ pairs. If $\hat{S}_m$ is a valid signature message m, The prover to the verifier to confirm the validity of the signature;on the contrary, if the signature is illegal, No prover proved the validity of the signature. Confirmed agreement of the specific implementation is shown in Figure 1.

Input: Public key parameters $(n, w, S_w) \in \Pi K$, $m \in Z_n^*$ and Valid signature $\hat{S}_m$, Prover that the person's private key $(d, e) \in [\phi(n)]^2$

1) Verifier select i, j $\in _R[n]$，calculation Q $\overset{def}{=} \hat{S}_m^{2i} S_w^j$ mod n , then Verifier will send Q as the question to Prover;

2) Prover calculation A $\overset{def}{=}$ Q^e mod n, then Prover will send A as the answer to Verifier;

3) Verifier verify A = $\overline{m}^{2i} w^j$ mod n , if the equation was established, then Verifier accepted $\hat{S}_m$ as digital signature on message m, or that the signature is uncertain.

Fig. 1. Confirmed agreement

2.2.3 Deny Agreement

Figure 2 give of digital signature deny agreement. The input is public key parameters agreement, namely $(n, w, Sw) \in \Pi K$,and $(m, \hat{S}_m)$ pair. If $\hat{S}_m \notin \Sigma\Gamma(m)$,then Prover can Confirm the validity of signature for Verifier; On the contrary, if $\hat{S}_m \in \Sigma\Gamma(m)$, then no Prover can confirm that the signature is legitimate to Verifier. The concrete implementation process denying agreement as Figure 2.

Input：Prover's private key $(d, e) \in [\phi(n)]^2$, Public key parameters $(n, w, S_w) \in \Pi K$，$m \in Z_n^*$ and Non_valid signature $\hat{S}_m$

1) Verifier select i = 4b and j $\in$ R[n]，Among them b $\in$ R[k]，calculation Q1 = $\overline{m}^i w^j$ mod n and Q2 = $\hat{S}_m^i S_w^j$ mod n , then Verifier will send (Q1，Q2) as question to Prover;

2) Prover calculation $Q_1/Q_2 = \left(\dfrac{\overline{m}}{\hat{S}_m^e}\right)^i$, And through the test of all possible b，Among them b $\in$ [k]，calculation i= 4b。

   If such an i value existence，then Prover let A = i，Otherwise suspend。Then，Prover will send answer message A to Verifier;

3) Verifier verify A = i. if the equation was established, then Verifier refuse $\hat{S}_m$ as digital signature of message m, otherwise, the digital signature uncertain.

Fig. 2. Deny agreement

3   Undeniable Signature Method of Completeness, Reliability and Undeniable Analysis

3.1   Confirmed Agreement the Completeness and Reliable Properties

Theorem 1. Let $(n, w, S_w) \in \Pi K$, then Confirm agreement the completeness and reliability are respectively:

Completeness: preset $S_m \in \Sigma I\Gamma(m)$, if prover and verifier follow sign to confirmed agreement, then Verifier always accepts S_m as a legitimate signature.

Soundness: Even through a lot of calculation, Prover persuade Verifier accept the possibility of signature $\hat{S}_m \notin \Sigma I\Gamma(m)$ will not exceed the $\frac{O(1)}{P'}$.

Proof: Completeness. Observe confirmed agreement can see, if the existing order to 2 additional factors, $\hat{S}_m$ ability from the signature of the elimination of such additional factors (such factors in $\Sigma I\Gamma(m)$ the definition is allowed).

Soundness. This article draws on the literature [13] part of the proof of the reliability of proof to confirm the agreement. Prover can deceive (or convince) Verifier accepts the signature $\hat{s}_m \notin \Sigma I\Gamma(m)$ the best possible situation is that he can select the A maximum likelihood (which Verifier selected values of i and j related) testing by Verifier. When the prover saw the question Q that from   Verifier(And based on his knowledge of the $\hat{s}_m$, m, w, d, e and n) Select A, the reliability of the theorem proved in the i and j need to get on the Prover select A for some of the information available at the time.

In the confirmed agreement, Verifier from the set [n] in the value of i and j randomly selected. To facilitate the analysis, assume that these values from $[\phi(n)]$ in the selection and confirmed by i or j exceeds the scope of the possibility of deception, Namely, i or $j \notin [\phi(n)]$, the possibility of using π said, most for $2 \times \frac{n - \phi(n)}{n}$. So, the next assumptions $i, j \in_R [\phi(n)]$.

First we define $I(Q) = \{i \in [\phi(n)] : \exists j, Q = \hat{S}_m^{2i} S_w^j \bmod n\}$. Because $\hat{S}_m \notin (m)$, so for the $\alpha \in Z_n^*$, $\mathrm{ord}(\alpha) > 2$, $\hat{S}_m = \alpha \overline{m}^d$. Step 3 of the protocol, the verifier checks

$$A = \overline{m}^{2i} w^j = \alpha^{-2ei} \hat{S}_m^{2ei} S_w^{ej} = \alpha^{-2ei} Q^e \tag{1}$$

Whether established.

Because α is in the first set, then for any A, the number   of i satisfy equations (1), and the number of i make founded $\alpha^{2i} = A^{-d} Q$, two the number of i maximum likelihood does not exceed $\phi(n) / \mathrm{ord}(\alpha)$.Given Q, V for i selection evenly distributed in the i(Q). This is because for each $i \in I(Q)$, there are the same number of j value can satisfy the equation $Q = \hat{S}_m^{2i} S_w^j \bmod n$. Therefore, P can succeed for the possibility of cheating at most $\frac{\phi(n) / \mathrm{ord}(\alpha)}{|I(Q)|}$.

3.2   Deny Agreement the Completeness and Reliability

Theorem 2: Let $(n, w, S_w) \in \Pi K$, Then deny agreement the completeness and reliability are respectively:

Completeness: hypothesis $\hat{S}_m \notin \Sigma(m)$ if Prover and Verifier follow sign to deny agreement, Verifier always accepts $\hat{S}_m$ not a legitimate signature of m.

Soundness: hypothesis $\hat{S}_m \in \Sigma\Pi\Gamma(m)$, for a Prover, even through a lot of calculation, also cannot with more than $\dfrac{1}{k} + \dfrac{O(1)}{P'}$ the possibility of persuading Verifier refuse to sign $\hat{S}_m \in \Sigma\Pi\Gamma(m)$.

Proof: Completeness. hypothesis $\hat{S}_m = \alpha\overline{m}^d$ among them ord $(\alpha) \geq p'$,this equation can set up the reason is $\hat{S}_m$ and $\overline{m}^d$ in Z_n^* .So α existence and $\hat{S}_m \notin (m)$, ord $(\alpha) \geq$ p'. Only when $\text{ord}(\dfrac{\overline{m}}{\hat{S}_m^e}) < 4k$, Prover will cannot find a Prover i value.orde $\text{ord}(\dfrac{\overline{m}}{\hat{S}_m^e}) = \text{ord}(\dfrac{\overline{m}}{\alpha^e \overline{m}}) = \text{ord}(\alpha^e)$. Because $(e, \phi(n)) = 1$,so ord $(\alpha e) \geq p'$. When select k<<p',Proved the completeness of a statement.

Soundness: In order to make the Prover to persuade Verifier, make verifier believe $\hat{S}_m$ is not a legal signature. Prover must send A value to Verifier, make $A = i$. Because $\hat{S}_m \in \Sigma\Pi\Gamma(m)$，can think $\hat{S}_m = \alpha\overline{m}^d$ established, among them ord $(\alpha) \leq 2$.So $Q_2 = \hat{S}_m^i S_w^j = \alpha^i \overline{m}^{di} w^{dj} = (\overline{m}^i w^j)^d$. According to the definition 1,if $m \in <w>$,then ord (w) = $2p'q'$ and $m2 \in <w>$.When $\overline{m}^4 \in <w>$, $\exists$ 1 make $w^l = \overline{m}^4$, so $Q_1 = \overline{m}^i w^j = w^{lb+j}$, $Q_2 = (\overline{m}^i w^j)^d = w^{(lb+j)d}$. With a mass of computing power Prover, we can compute r value causes the $Q_1 = w^r = w^{lb+j}$. In order to calculate the i, Prover still need to find b, meaning he needs to solve equation. Hypothesis $j \in R$ $[\phi(n)]$, then for every possible values b, there will be $\dfrac{\phi(n)}{\text{ord}(w)}$ a possible value j says the best prover can do random speculation given a 1/k probability value. According to $j \in R[\phi(n)]$, we can get $\dfrac{1}{k} + \dfrac{O(1)}{P'}$.

4   Summaries

Digital Signature in E-commerce more widely recognized digital signature and nonrepudiation becomes necessary, analyzing and undeniable RSA signature scheme signature mechanism based on an improved non-repudiation signature algorithm, and gives the completeness, reliability, and undeniable proof. Application undeniable signature confirmation and denial protocol agreement to ensure the security of e-commerce transactions.

References

[1] World Wide Web consortium extensible markup language (XML) 1.0, 3rd edn. (EB/OL). W3C Recommendation (February 4, 2004)

[2] Dournaee, B.: XML security. McGraw-Hill, New York (2001)

[3] Selkirk, A.: XML and security. BT Technology Journal, 23–34 (2001)

[4] Liu, Y.-Z., Wang, H.-X.: Principle of XML digital signatures and the implementation in EC. Journal of Hefei University of Technology (Natural Science) 29(10), 1298–1301 (2006)

[5] Sun, L., Li, Y.: XML undeniable signatures. In: Proc. of the CIMCA-IAWTIC, pp. 231–236 (2005)

[6] XML-Signature syntax and processing. W3C Recommendation (February 12, 2002), `http://www.w3.org/TR/xmldsig-core/`

[7] Chaum, D., van Antwerpen, H.: Undeniable Signatures. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 212–216. Springer, Heidelberg (1990)

[8] Gennaro, R., Krawczyk, H., Rabin, T.: RSA-based undeniable signatures. Journal of Cryptology 13(2), 397–416 (2000)

[9] Miyazaki, T.: An improved scheme of the Gennaro-Krawczyk-Rabin undeniable signature system based on RSA. In: Proc. of the ICISC, pp. 135–149 (2000)

[10] Chaum, D.: Zero-Knowledge Undeniable Signatures. In: Damgård, I.B. (ed.) EUROCRYPT 1990. LNCS, vol. 473, pp. 458–464. Springer, Heidelberg (1991)

[11] Mathematical Markup Language (EB/OL), `http://www.w3.org/TR/2003/REC-MathML220031021/`

[12] Gennaro, R., Jarecki, S., Krawczyk, H., Rabin, T.: Robust and Efficient Sharing of RSA Functions. In: Koblitz, N. (ed.) CRYPTO 1996. LNCS, vol. 1109, pp. 157–172. Springer, Heidelberg (1996)

Development of Touch Screen Driver Based on S3C2410 under Linux

Chen Hao-Wei, Bai Feng-Shan, and Liu Hai-Jing

College of Electronic Information Engineering
Inner Mongolia University
Hohhot, PRC
howardchw@163.com, eefs@imu.edu.cn,
walhl520@126.com

Abstract. Touch screen as a kind of simple and convenient input device has been already used in consumer electronics and industrial control areas. S3C2410 is a embedded processor, which is with the ARM920T core produced by Samsung Corp. This paper introduces the architecture of touch screen and its driver under Linux based on ADS7843 connected with S3C2410 platform. Both hardware architecture and software design of the system are described.

Keywords: Touch Screen, Driver, Linux, S3C2410, ADS7843.

1   Introduction

Touch screen is a new kind of input device. The user only needs with the finger or other hard objects to touch the icon or the characters displayed on monitor, the computer will do the operation. Because of the simple operation, touch screen has been widely used in lots of areas. Such as: service inquiries in telecommunication bureau, tax bureau, and bank, the industrial control, the military command, the computer game and so on. S3C2410 is an ARM920T core embedded processor which is released by Samsung Corp. This Processor has high performance and low power consumption. Use its interface to be possible to connect with touch screen conveniently. In this paper, hardware interface circuits between S3C2410 and ADS7843 are designed, and embedded Linux driver of touch screen is developed.

2   Embedded Linux Driver

The device driver controlls the interaction between the operating system and the hardware. When the driving program is first accessed, it needs to use entry function init_module() firstly. Its prime task is to register this equipment to the Linux Kernel. For the byte device, calling the function register_chrdev is to finish the register. After registering successfully, this device has obtained the main device number and the inferior device number assigned by system, and establishes connection with the file system. When the byte device driver registered to Linux Kernel, it increases a data

structure named device_struct in the byte device structure chrdevs. This device's main identifier serves as the index of this structure. Each class of the structure namely device_struct includes two elements: a pointer of the registered device driver's name and a pointer of a group of file operation [1].

The registration is to fill the file_operations structure pointer which provides by the module into some item of the device_struct data structure array. Though calling the function mknod() and creating a node represent the device, the application program may visit it. Moreover, when the device unloading, it needs to replace the device registered value and logging out from the system.

The Linux operating system completes the transfer from the user space to the kernel space through the system calling and the hardware interrupion. Fig.1 is an entire process of a device driver module to load, unload and the system call dynamically [2].
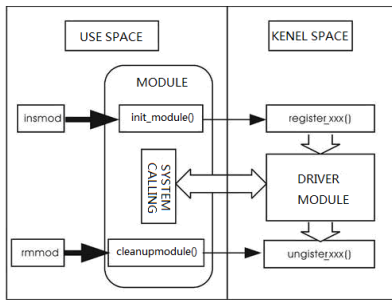
Fig. 1. Process of loading, unloading and the system calling

In the progress of embedded system development, we complete merely several interface functions: read, write, open, ioctl and get the application system need. One of the driver's function is to finish the function pointer in the file_operations structure.

3   The Hardware Component of Touch Screen

Touch screen is a kind of position device. It is composed of the touch examination part and the touch screen controller. The touch examination part has been installed in front of the monitor screen, used in examining the position of screen touched by user. And the main function of the touch screen controller is to receive the touch information from the touch detector, transform it to the electronic contact coordinate, then give CPU again. It can also receive the order from CPU and perform it at the same time.

According to its technical principle touch screen is possible to divide into 5 kinds: Vector pressure sensing type, resistance type, electric capacity type, infrared-like type and superficial sound wave type. What this design uses is four resistance type touch screen. This touch screen's outside surface is the glass plate with the resistances, the polyester surface layer is hanging on the glass face, separated from the glass face through the small transparent insulated pellet. The internal surface has a conduction layer. When the screen is touched, the conduction layer and the glass surface layer have the electronic contact, and the touch position is expressed by the produced voltage [3].

Hardware connection mode in this paper is that touch screen controller ADS7843 which has SPI interface connects with four resistance type touch screen. The most remarkable characteristic in this way is the quick speed of response, the high sensitivity, the great reduction of communication time between the microprocessor and the touch screen controller, thus promoted microprocessor's operating efficiency.

ADS7843 as the touch screen controller, the pattern of connection with S3C2410 is shown in Fig.2. Due to the advantage of the differential working mode, it has been used in the hardware circuit.
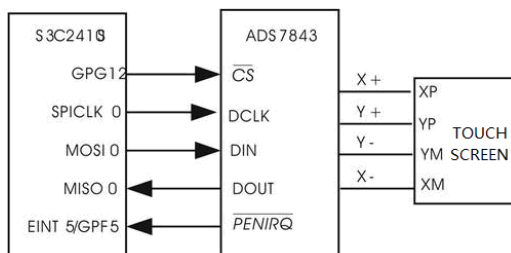
Fig. 2. Touch screen hardware interface schematic diagram

4   Design of Touch Screen Driver

As shown in Tab.1, the information in the TS_RET structure is the driver provided to the upper application, using to store the touch screen returning value. Through reading the information of interface, the upper application program gets the data from the driver, and does the operation according to the returning value.

Table 1. The main structures of driver

Structure Name	Function	Main Parameters
TS_RET	Information of pen touch	short pressure; short x; short y; short pad
TS_DEV	Information of touch screen controller	PenStatus; buf[MAX_TS_BUF]; wq; lock
s3c2410_fops	Interface of driver an-d application	open:s3c2410_ts_open; read:s3c2410_ts_read; release:s3c2410_ts_release;

In the program, the pen touch queue is a ring, until the pointer of the beginning and the end get the same position, the queue will read the pen touch information, and the process will be arranged to sleep. The wq is the queue for waiting, contains a lock variable and a sleeping process table. When several processes wait for some event, Linux will record these processes in this waiting queue. Its function is when there is no pen touch event, blocking the upper layer to read the information, until the pen touch event occured. The lock use the self locking pattern, self locking is based on the shared variable. The function may obtain the lock through setting a special value to the variable. But other function which needs lock will verify that the lock is available.

The s3c2410_fops is the interface for the kernel to call the driver, its function is mapping for the standard interface. Its expressed method is not the standard C grammar, but is one kind of special expansion in the GNU compiler using name to initialize the structure field. The benefit is that the structure is clear and easy to understand, many questions will be avoided after the structure has been changed [4].

5   Interrupts and Interrupt Handing

The data exchange between the device and the system is achieved by the interruption in this design. Because the entire processing of touch screen driver is quite complex and time-consuming, thus the touch screen driver process is impossible to complete in the interrupt service. Generally the interrupt processing is cut two parts in the Linux operating system. The interrupt handling is the upper portion. While receiving an interruption it starts to execute immediately, but only does the strict time limit works. For example, reply to the received interruption or reset the hardware. These works are finished in the situation that all interruptions are forbidden, the work which to be allowed to complete later will be able to postpone to the lower portion [5]. When touch the screen, the output value has a vibration process from the ADS7843, is that there is a unstable time from the ADS7846 output, this process is probably 10ms. In order to avoid this unstable time, the processing function for interrupt handling uses the mechanism of the kernel timer. Because the processing will happen after the lower portion interruption has occurred 50ms, and it has been possible to process the long time pen touch event.

The driver processes the interruption through using the request_irq function to register and activate an interrupt handling program. The function definition is as follows:

int request_irq(unsigned int irq, void(*handler)(int, void *, struct pt_regs *), unsigned long irq_flags, const char *dev_name, void *dev_id)

The parameter irq expresses the interrupion number needed to apply. The parameter handler is an interrupt processing subroutine registered to the system, and is called by system when the interruption happened. The parameter dev_name is the device name. The parameter dev_id is the device identifier. The parameter irq_flags is the application option, it decided the interrupt handling characteristics, and the most important part is that judges the interrupt handling response immediately or delay.

The interrupt output pin of the touch screen controller ADS7843 connected with the interrupt controller through the external interrupt 5 in this design. When the screen has been touched, the interrupt service s3c2410_isr_tc() which the interrupt number is IRQ_EINT5 will occur. The interrupt handling flow chart is shown in Fig.3.

The timer has been set 50ms in the function s3c2410_isr_tc(). Therefore, after the hardware interrupted 50ms, the timer interrupt will be triggered. The timer interrupt program is ts_timer_handler(), this program has implemented the function of the interrupion's lower portion. After vibration time, if the effective event does happen, then get the coordinate of the screen, reset and activate the timer at 100ms. The purpose is that if the pen move is dragging, get the coordinate figures per 100 ms, and

store them in the buffer. If the pen move is not dragging, after get a coordinate figure, inquire the value which represents the pin's condition when enter the function ts_timer_handler() in second time. If the value becomes high level, set the variable which represents touch screen's condition tsdev.PenStatus at PEN_UP, and release the timer, prepare for processing the effective event next time. Timer interrupt service flow chart is shown in Fig.4.
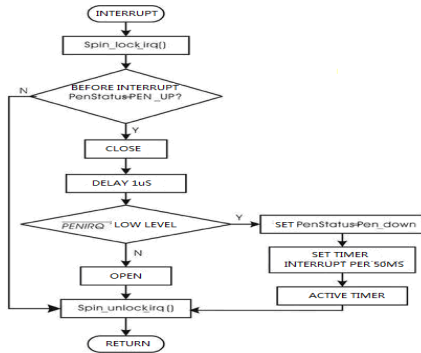
Fig. 3. The flow chart of hardware interrupt handling
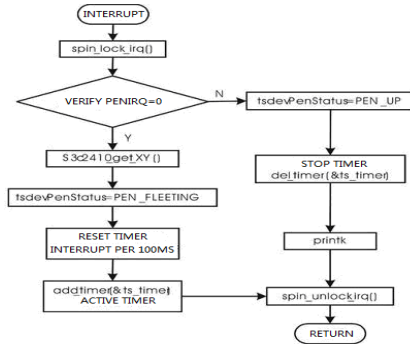
Fig. 4. The flow chart of timer interrupt service

6   Capture the Data from Touch Screen

In the program of the timer interrupt, if get the value from the pin EINT5/GPF5 is low level, it indicated that the effective event has occurred, then get the information through the function s3c2410_get_XY().

ADS7843 has many kinds of time sequence, and the time sequence adjusted how the device and CPU work together. The coordinate transformation will be active every 16 clock cycles in this design. Through the method of averaging the values many times the value of X coordinate will be got, every step in the cyclic program must be

completed in 8 clock cycles, and the data processing is strictly in accordance with the time sequence. The method of getting Y coordinate and X coordinate is similar.

7   Summaries

This paper mainly introduces the application of touch screen based on S3C2410, and the development of the driver under the embedded Linux OS. Provides one kind of solution for the embedded system using graphical interface. Based on these approaches, may extend many kinds of functions, and will be applied in various kinds of product under embedded system widely.

References

[1] Liu, M.: Design of embedded system interface and development of Linux driver, pp. 83–105. Press of Beihang University (2007)
[2] Wei, S.-D.: Development of embedded Linux application handbook, pp. 240-360. Posts & Telecom Press (August 2008)
[3] Wei, Y.-M., Geng, Y., Zhong, S.-Y. Linux device driver programming, pp. 55–60. China Electric Power Press (2006)
[4] Duan, P., Ding, C.-J., Han, Y.-H.: Research and Exploitation of Touch Screen Device Driver Based on Embedded System. Microcomputer Information, 59–60 (May 2008)
[5] Zhang, Y.-C., Zhuang, Y.: Design of Touch Screen Based on Linux Embedded System. Control Engineering of China, 321–323 (May 2009)

Building 3D Web Map Applications Based on Google Earth and ArcGIS Server

Haifeng Huang and Wei Liu

[1] College of Civil Engineering and Architecture, China Three Gorges University, Yichang, P.R. China
[2] Electrical Engineering and Renewable Energy School, China Three Gorges University, Yichang, P.R. China

Abstract. Map services have gradually become an important aspect of web services and applications. Especially with the popularity of some 3D virtual earth software such as *Google Earth* (*GE*)*, World Wind* and so on, the demand of building 3D web map applications (3D-WMA) is getting more and more. Among the many technology platforms and combinations, building 3D-WMA based on *GE* and *ArcGIS Server* is a good choice. This paper firstly introduces the advantages of *GE* geobrowser and *ArcGIS*, then the system architecture which comprises professional spatial data management, professional map publishing services, and *KML* services load in geobrowser is analyzed. Next, some system implementation skills and methods includes making maps, publishing maps as *KML* services, and using *KML* services in *GE* geobrowser are elaborated. At last, a case of 3D-WMA that is 3D digital tour presentation system is introduced, which demonstrates the good practicability of building 3D web map applications based on *GE* and *ArcGIS* Server. In fact, the system set up mode, architecture can be applied to all other 3D-WMA construction.

Keywords: Web map services, Geobrowser, KML, Google Earth, ArcGIS Server.

1 Introduction

In recent years, with the rapid development of web geographic information system (WebGIS), remote sensing (RS), Internet technology and so on, three-dimensional web map applications (3D-WMA) have been widespread concerned because of its better practicality. Moreover, some geobrowsers' popularity such as *Google Earth* (*GE*), *NASA World Wind*, *SkylineGlobe*, *ArcGlobe*, etc has also greatly promoted the applications [1].

  GE has been the most successful virtual globe geobrowser for a few key reasons [2]: it is free; it is fast; and it has its own markup language (*KML*), which allows anyone to display and easily share their own data. And now GE can be embedded into web pages with the release of *Plugin* and *API* that makes it an ideal 3D web map geobrowser. However, for a complete professional map web application, *GE* has not server-side functions include geospatial data management and map distribution. As the most mature professional GIS software platform, *ArcGIS* can manage geospatial data by *ArcSDE*, make maps by *ArcMap*, and publish maps by *ArcGIS Server*, what

can combine with *GE* web geobrowser to build powerful and dynamically professional 3D web map applications.

Although *GE* has provided free 3D terrain and some RS data, some professional GIS spatial data are still necessary. Consequently, the core of building 3D-WMA is that publishing GIS data as map services, then using these services to acquire, load, display and control all professional data in geobrowser, and *KML* is the link between professional GIS map server and *GE* geobrowser. Around the core, this paper firstly elaborates the system architecture of 3D-WMA based on *ArcGIS*, *GE* and *KML*, then system implementation is introduced include making maps, publishing map services by *ArcGIS Server*, using *KML* services in *GE* web geobrowser. At last, a case is introduced and the conclusion demonstrates the practicality of 3D-WMA based on *ArcGIS Server* and *GE* web geobrowser.

2   System Architecture

Now, there are more and more open web map services and *API* on Internet, which promote a new trend for development of web map applications, and that is to combine local professional GIS data and map services with open web map data and services [3]. The advantages of doing so include saving costs of basic geographic background data acquisition and update, reducing the difficulty of system development, shortening the development cycle, and so on.

The construction of 3D-WMA based on *ArcGIS Server* and *GE API* can follow above patterns, and the core points are as follows:

- Building centralized and unified professional GIS spatial database in RDBMS by *ArcSDE*.
- Making professional maps by using local GIS data, and publishing these maps to *KML* services through *ArcGIS Server*.
- Embedding *GE* geobrowser into web pages by using *Plugin*, while using free 3D terrain and RS images data provided by *GE*, 3D-WMA can also load, display and control professional GIS data by linking these *KML* services.

Overall, 3D-WMA architecture can be divided into 3 main tiers (or parts), which are professional GIS database, *KML* Services and *GE* geobrowser (Fig.1).
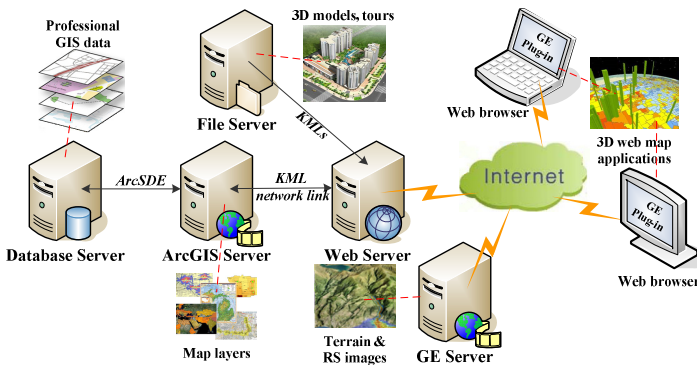
Fig. 1. System architecture of 3D-WMA based on *GE* and *ArcGIS*

(1) Professional GIS database. Using *ArcSDE*, all local professional GIS data can be saved into RDBMS such as *Oracle*, *SQL Server*, etc. According to the use, the GIS data can be divided into base maps data and operational maps data. Base maps data provide a framework onto which dynamic operational data can be displayed, and these data are relatively static and do not change frequently. Operational maps data are the ones that end users will view and work with to perform their daily work. In *ArcGIS* 10.0, different management methods are applied to base and operational maps data in order to improve display performance, which is very important for construction of web map applications. It should be noted that the terrain and RS images in *GE* are actually base maps data, however for the reasons of low resolution, non-distinct, or low accuracy, etc, it is necessary to use more suitable base maps data such as high-resolution RS images in 3D-WMA.

(2) *KML* services: *ArcGIS Server* can publish map and image services as a *KML* network link through Representational State Transfer (*REST*), and also can return results as *KML* when end users query map layers [4]. More important is that the *KML* is dynamically generated, meaning that users who view the *KML* will always see the latest maps and data from *ArcGIS Server*. Well, this is the basic and key function requirement of web map applications and is also the absence of *GE*.   At the same time, some special information, e.g. 3D models and tours can by authored in *KML*, however *ArcGIS* can't support and publish this kind of *KML* services. In view of the *KMLs*' invariability, it is suitable to save all these *KML* files that are prepared beforehand in a File Server, and then can be load into *GE* geobrowser at anytime.

(3) *GE* geobrowser: As the presentation platform, *GE* can be embedded into web pages [5], which on one hand terrain and RS images data provides by *GE* will be loaded automatically as base and framework, on the other hand professional map layers data can be loaded, displayed and controlled through *KML* network link published by *ArcGIS Server*, *KML* files include 3D models and tours as well. When end users open the web page which has embedded *GE* for the first time, the *GE Plug-in* need to be setup, after that, they can use 3D-WMA conveniently just like open common web pages.

Besides above 3 main parts, Web Server is indispensable for a complete 3D-WMA, and the function of which is publishing websites embedded with the applications.

3　System Implementation

The main task of building 3D web map applications include making maps in *ArcMap*, publishing maps as *KML* services by *ArcGIS Server*, and using map services in *GE* web geobrowser by *API*.

(1) Making maps. All professional GIS layers data can be made to different maps singly or in combination, and each map will be released as a separate *KML* services. Some map production strategies and skills will greatly affect the network data transmission speed and data load, display efficiency in web geobrowser, thus affecting the practicality of 3D-WMA. So, building effective maps combinations according to specific content and purpose of professional GIS layers, what is an important work to construct 3D-WMA successfully. There are some basic maps making skills as follows:

- Making a map by overlaying all GIS layers does not recommend, because the multi-layers map services will take a long time for data transmission & load in web client, which greatly damage the user experience, so making different maps according to base map layers and operational map layers is more suitable.
- Lower layer brightness and increase transparency should be set for base map, in order to ensure the operational map is clearly visible. And at the same time, attachment information include label, rendering and symbol system should be avoided or reduced. This can minimize the amount of data transmission, thereby improving system responsiveness.
- Grouping the layers with great amount of points, lines or polygons and setting different display scale can optimize display effect by limiting the number of points displayed at higher altitudes. Setting *HTML Popup* is useful for layers that will be shared with end users, and these web contents will be automatically added to *KML* services and loaded in *GE* geobrowser. And legends also can be set in layout view of maps, then these legends will be embedded in *KML* and displayed as a screen overlay in clients.
- Map to *KML* and Layer to *KML* tools in *ArcToolbox* can generate static *KML* documents quickly, by which maps can be made finer.

(2) Publishing maps as *KML* services. By using *ArcCatalog* or *ArcGIS Server Manager*, *mxd* map files can be directly published as *KML* services and creating *KML* network links, and *KML* network links are useful because they reference dynamic *KML* rather than static *KML* that never changes or updates. In order to improve system performance, some methods can be used as follows:

- Use cached map services in place of dynamic map services, especially for base map services. *ArcGIS Server* can precompute basemaps as cached map services, which are stored as a series of map image tiles on server for a range of map scales and levels of detail (*LOD*), this will significantly increases performance.
- If cached map services have supported coordinate system such as *WGS84*, or any other cylindrical projection, the ground overlay in *KML* can be displayed by using *Regions*, which is a hierarchical collection of images and are recommended for optimum performance.
- The *KML* services that have been published can be checked in web browser through *KML* network link, and the *URL* is just like *http://<web server name>/<instance name>/rest/services/<service_name>/<service_type>/kml/<service>.kmz*, the service_type is either *MapServer* or *ImageServer*.

(3) Using *KML* services in *GE* web geobrowser. By *GE API*, *KML* files or services that contain professional GIS maps, layers, 3D models or tours can be load into *GE Plugin*. Just like maps making and *KML* services publishing, to ensure 3D-WMA speed and practicality, the following items should be paid attention to:

- According to user operation procedure, loading suitable *KML* at the appropriate time rather than loading all *KML* services or files once the *GE Plugin* pages are opened. In normal conditions, when initial page is opening, the base maps *KML* should be loaded according to *LOD*, yet the operational

maps, 3D models or tours *KML* should be loaded separately when user concerned.

- For *KML* services published by *ArcGIS Server*, there is a *layerOptions* (layer drawing options) which contains *composite*, *separateImage* and *nonComposite* value need to be set in *KML* network link according to specific application requirement. *Composite* means all layers will be displayed as a single composite image, and the layers cannot be turned on and off in the client; *separateImage* means each layer will be displayed as a separate image; *nonComposite* means vector layers will be displayed as vectors and raster layers as images. So *composite KML* has less data quantity and poor graphic quality, and *nonComposite* is the exact opposite.

4   A Case of 3D-WMA

2008, "Ecological-cultural tourism circle in western Hubei Province" regional development policy and planning was put forward by Hubei Province, central of China. Building 3D digital tour presentation system (3D-DTPS) is an important and basic informatization construction task [6]. So, based on *ArcGIS* and *GE API*, the 3D_DTPS was constructed.

The professional GIS data include high resolution remote sensing images, administrative divisions, tourist attractions, tourist routes, 3D tourist spots and flight roaming. In order to increase performance, the first two data was set as base map layers, the tourist attractions and routes data as operational maps, then these maps were published as *KML* services by *ArcGIS Server*. As for 3D tourist spots and flight roaming, were generated and saved as *KML* files.

When the main page which embedded with *GE Plugin* was opened, the base maps and some most famous tourist attractions would be load firstly (Fig.2), and as users gradually zoom in the map, more and more tourist attractions would be displayed. "*Layers*" in control panel shows the hierarchical list of operational map layers and features, users can control the display or hide. "*Query*" allows user to search and fly to specific tourist spots, routes and so on.
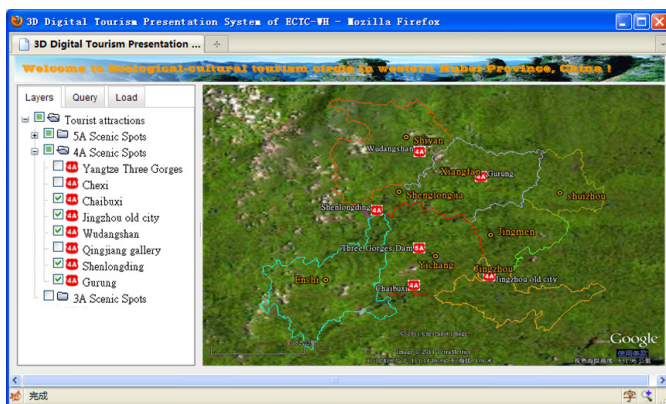
Fig. 2. The initial page embedded with *GE* geobrowser of 3D-DTPS

Fig. 3. 3D model and tour in *GE* geobrowser of 3D-DTPS

In "*Load*" panel, there is a list of all 3D tourist spots models and flight tours. To ensure 3D-DTPS has good efficiency and speed, these data would not be load by default. If user wants to view interested 3D models or roam scenes, they can select corresponding checkboxes, and the static *KML* files would be load into *GE* geobrowser automatically. Fig.3 shows the 3D model of Three Gorges Dam, a national 5A scenic spot, and there is also a tour at *GE Plug-in*.

Practical operation effect shows that 3D_DTPS has not only advantages of fast running speed, excellent user experience inherited from *GE* platform, but also can integrate professional geospatial data seamlessly by *ArcGIS KML* services, which have good practicability.

5   Summaries

As technology advances, the demand of building 3D web map applications is getting more and more. Among the many related software platforms, *GE* is an ideal 3D web map geobrowser, and *ArcGIS* can provide perfect professional geospatial management and dynamic map publishing services, so it is a good choice to build 3D-WMA based on *GE* and *ArcGIS*. The core of this method is using *ArcGIS Server* to publish professional maps as *KML* services and then calling these services in *GE* geobrowser, thereby the 3D-WMA that integrates professional spatial information with 3D base map can be established.

References

[1] Rakshit, R., Ogneva-Himmelberger, Y.: Application of Virtual Globes in Education. Geography Compass, 1995–2010 (2008)
[2] Grossner, K., Clarke, K.: Is Google Earth, 'Digital Earth' - Defining a Vision. In: Proceedings of the 5th International Symposium on Digital Earth, pp. 1–13 (2007)

[3] Studio, S., Hill, N., Cotton, W., Gibb, A.: Data visualization applications in virtual globe software. In: Proceedings of the ACM International Conference on Digital Libraries, p. 435 (2008)

[4] ArcGIS Resource Center (2011),
    `http://resources.arcgis.com/content/web-based-help`

[5] Google Earth Api. (2011), `http://code.google.com/intl/en/apis/earth/`

[6] Gao, T., Du, J., Sun, Z., Jia, Y.: Digital tourism integrated service system realization. In: ICNSC 2010, pp. 232–237 (2010)

The Design and Implementation of the Mobile Facility Communication Software Based on Plug-In

Zhi-Yan Xu, Yan-Pu Zhang, and Yu-Qiang Chen

LangFang Artillery Command College, 3rd Department, Langfang HeBei 065000
yuqiang.chen@ia.ac.cn, chenyuq2002@163.com

Abstract. Based on the encapsulation of the communication protocol, the plug-in technology was adopted to design the communication software for the mobile facilities. We achieved the separation between the communication manner and the communication content. Even the changes of the plug-in will not be spread to the application of upper layer. It denoted the "plug and play". The XML technology is used to provide a unified data structure for every layer of the software. The design can improve the opening, transplantable and generalization, and can shorten the development period.

Keywords: plug-in, communication, XML.

1   Introduction

With the development of the portable terminal based on embedded system, it is common to use electric products, such as handset and palm PDA. It's necessary to make use of the computer to manage their personal information by software. It will be practical significance for the management software, which is compatible with multiple communication facilities, and can communicate with PC in different communication manners. A software designed idea was presented based on plug-in. Multiple communication facilities can be   supported with the "plug and play" functional plug-ins. Firstly, the system structure design of the software was introduced. Secondly, several technologies were used during the development of the software. Finally, the functional plug-in and the application layer design were presented.

2   System Structure

A king of communication software must have some functions, such as identical user interface, and communicating with the PC according to different types of handset or diverse communication manners, and file transmission or personal information management. The software includes three layers. They were Data Transmission Layer, Communication Facility Service Layer and Application Layer as the following the fig.1.

As to the different communication manners, such as IrDa or Bluetooth, the bottom protocols were encapsulated in the data transmission layer based on the analysis of the protocol. The corresponding plug-in was designed to provide the unified data transmission interface, such as erecting connection, receiving and sending methods. With the interface, the plug-in can communicate with the upper layer.
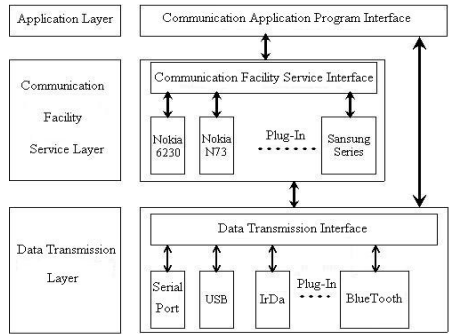
Fig. 1. Communication Software Structure Based on Plug-in

In the communication facility service layer, the practical communication facility service plug-in is designed, such as the Nokia serial handset plug-in engine or Sansung serial handset plug-in engine. With the uniform encapsulated interface, the information service was provided for the upper layer. With the data transmission interface, the plug-in can be related with the data transmission style. In the application layer, the process of setting communication way is the process of accessing data transmission layer. On the other hand, with the communication facility service interface, the user can be connected with the practical handset to get information and achieve the designed functions. The software can be a good handset partner, with which, the user can save the personal information efficiently.

The modularized design has better compatibility. To be connected with PC, the corresponding plug-in is loaded according to the practical communication facility. The content and the ways can be isolated. The software will be open and transplantable.

3   System Implemental Technology

Plug-in Technology

What is called plug-in[1-3] is a segment of executable program. The essence lies in extending and strengthening the function of the software without changing the platform. The user can customize their owner functional plug-in according to the demands. The plug-in can be compiled and tested separately, but cannot be executed independently. That is the "plug and play" software development. The plug-in is independent with the main program, and they were independent among the different plug-ins. With the unified interface, the inner changes will not affect each other.

Dynamic link libraries or ActiveX controls can be plug-in. Dynamic link libraries is a very common style. The encapsulated case program or resource can be shared by other application program. The emphases of the paper are the DLL file. Compared with the Static link libraries, the called function code will not be copied to the executed file of the application, but only some necessary information is added into the function. The windows system will not load the DLL file into the process space, until the application program is loaded into the memory. The linked oriental information

can be used to find the corresponding function. For example, during using a DLL, the Win32 system will assure that there is only one copy in the memory through memory mapping.

The Operation Mechanism of Plug-In

The plug-in DLL provided the fixed output functions which will be called by the application program. At the same time, the main program can get the interface pointer by loading the plug-in.

There were two steps during the loading of the plug-in. The application kernel will get the handle and the address of the plug-in. Firstly, the main program loaded the plug-in. the function CMainFrame::LoadPlugIn() was used to obtain the purpose. Secondly, the main program got the process address of the plug-in through the Windows API GetProcAddress(m_hMobileDevLib,"CreateEngine"), where m_hMobileDevLib was the handle, and "CreateEngine"was the output function. The application program could call the PASCAL EXPORT CloseEngine(PHEngine* pEngine) to unload the plug-in.

XML Technology

XML is a simple method to mark and distinguish the content.The user can custom the any complex structure within the rule of the XML syntax. The information is transmitted by common text, not binary format. That sounds better for exchanging data between different platforms. It is the reason why the HTML technology was so popular by dealing with text. XML provided a uniform method to describe and exchange data. XML technology has promoted the advantage to a higher level for its expansibility. The user can extend it at will. Further more, the mechanism is standardized.   As a marked language, XML has many characters:

1. Simple. XML is designed dedicatedly. It is composed of some rules, which is simple and can be used to create marked language. A parse program can be used to dispose the new marked language. XML is taken as a world language for reading and writing by anyone. Its function is unified. Such as the created marks are displayed in pairs, and are depended on the new code criterion.

2. Openness.

3. High efficiency and good expansibility. The document segment can be reused. The user can invent and make use of their own labels, and can share them with others. In XML, we can define an infinite labels.

4. Internationalization. The new coding criterion supports mix text, which is compiled by the main language.

Based on the merits and applied future, the data structure is designed by unified format in the paper.

Where the parameters can be chosen during the setting of the communication manners, and the content can be determined by setting the communication facility. The concrete setting was illustrated as the Fig.2.

Fig. 2. Communication Setting

4   Plug-In Engine Design of Mobile Facility

Design of Communication Mode in Data Transmission
As to the Nokia6230 handset, we assume that it was connected with PC by IrDa. Firstly, as to be connected with the concrete facility, the IrDa protocol should be encapsulated in the IrDa plug-in. It is the basis. And then, the plug-in can achieve connection with the PC, shut up and property setting and so on.

Secondly, the information was transmitted faultlessly between the handset and PC. The socket technology is applied to meet IrDa or Bluetooth communication, which is a simple point-to-point format. As a general interface, the socket was taken as a special pipe, which could support all kinds of network with interactive communication mechanism. At first, the port will be opened at the bottom facility. The facility will set up the connection after validating the request. In the data transmission layer, the plug-in engine process and the socket connection were created. Finally, the communication facility service layer will parse and respond on the applied request, then feedback. The process was illustrated as the Fig.3.

Several main interface functions were listed as following:

1. IrDAPlugThreadProc(void* lpParameter)
Description: static void WINAPIV IrDAPlugThreadProc(void* lpParameter)
Function: PC initiated the IrDa plug-in and created the process after being connected with the handset.
2. Send(char* buffer,int size)
Description: virtual int Send(char* buffer,int size)
Function: sending files or request.
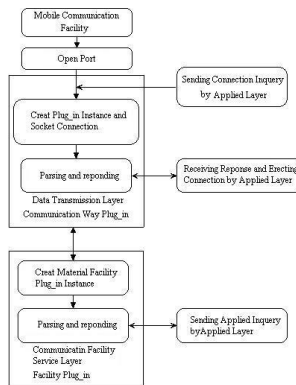
Fig. 3. File Transmission Flow Table

3. Recv(char* buffer,int size)
Description: virtual int Recv(char* buffer,int size)
Function: receiving data and reponse.

Within the frame, the functions and applied scope can be extended.

Communication Facility Plug-in Design in the Communication Facility Service Layer

Firstly, we will analyze the communication protocol as to the concrete facility. Here is an example of Nokia6230, whose protocol is based on OBEX IrDa[4-6]. We took file transmission as an example to describe the analysis process.

1. Setting up the connection

Request and response are the basic operations in OBEX. As to the Nokia6230 handset, byte 1 to 3 is the   constant initial data of the data packet. The 4th byte is the operational code. Byte 5 and 6 is the size of the packet. The following table is an example of the request on the model of the connected facility.

Table 1. Data Packet on Requesting for Model

Byte0, 1, 2	Byte3	Byte4,5	Byte6…
0x14 0x00 0x10	0x1b	0x00 0x06	0x00 0x01…

Where the byte0,1,2 denote the request, byte3 denotes the destination. Byte4,5 denote the size of the packet. The facility received the data and returned the response after parsing the data packet. The response data was listed as the following tab.2.

Table 2. Response Data Packet

Byte0, 1, 2	Byte3	Byte4,5	Byte6…
0x14 0x10 0x00	0x1b	0x00 0x19	0x01 0x44…

After identifying the facility, the connection was erected.

2. File transmission

Every mobile facility has its own file transmission protocol. We first initialized the file system when we wanted to manage the files. An initializing file system data packet would be sent, which was listed in the tab.3.

Table 3. Initializing File System Data Packet

Byte0, 1, 2	Byte3	Byte4,5	Byte6…
0x14 0x00 0x10	0xdb	0x00 0x0e	0x00 0x02…

Where byte3 denotes the request of initialization, and byte 6 denotes the content of the packet. After receiving the response as the tab.4, the normal file operation could be done, such as upload, download, update, delete and so on.

Table 4. Response Data Packet

Byte0, 1, 2	Byte3	Byte4,5	Byte6…
0x14 0x00 0x10	0xdb	0x00 0x05	0x02 0x00…

Based on the analysis of the transmission protocol, we can get to know the rule of the protocol.

3. Close the file system

After finishing the operation of the file system, the application layer should send a corresponding request data packet as tab.5.

Table 5. Finishing Operation Request

Byte0, 1, 2	Byte3	Byte4,5	Byte6…
0x14 0x00 0x10	0xd9	0x00 0x18	0x00 0x02…

where the byte3 was the operational end code. The application layer would go on processing after receiving the response.

Secondly, the main interface design was listed as following,

a. Asynchronous transmission interface: the interface should provide enough buffer and overtime limit.

b. Data packing interface: according to the protocol rule, the data is sent after packing a full data packet.

c. Data parsing interface: parsing the data packet and get the information.

With the design, we can achieve the separation between the mode and the content. At the same time the robust is ensured.

5   Design and Implement of Application Layer

As to the connection between the communication software and the PC, a unified application program interface was designed as the following Fig.4.
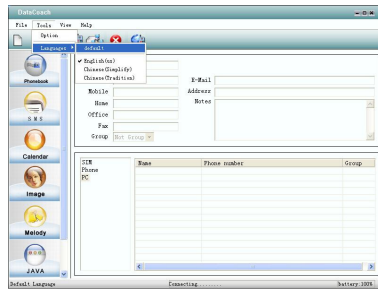
Fig. 4. Application Program Interface

With the XML technology, we can achieve the following functions.

- The user can choose the communication ways, parameters and communication facility. So that he can mange the personal information easily.
- According to the different facility or plug-in, the corresponding information will be displayed in the system functional zone.

There will be some hints in the state column zone. For example, the handset model will be displayed in the zone after being connected with the PC.

6   Summaries

As to the mobile facility communication software based on the plug-in, there are several merits. One is the task distribution according to the demands, the "plug and play" was achieved. It is easy for the system development and upgrade. Secondly, it's very convenient for the user. According to the different communication ways, the user can choose different plug-ins. Furthermore, the design can achieve the separation between the communication way and communication content efficiently.

References

[1] Zhu, R., Cai, W.-D., Liu, C.-G.: Remote Monitor System based on plugins. Microelectronics & Computer 21(11), 89–95 (2004)
[2] Chen, F.-M., Chen, Q.: Design and implementation of reusable software based on thought of plug-in. Computer Engineering & Design 26(1), 172–173 (2005)
[3] Zhang, J.-J., Zhang, W.-Y., Xue, L.-W.: A kind of plug-in based software architecture. Journal of HeFei University of Technology 28(4), 398–401 (2005)
[4] Zhu, X., Pang, H., Lin, H., Realizing of PDA Communication by User-defined Infrared Protocol Based on WinCE. Journal of Electronic Measurement and Instrument (supp. 1), 433–438 (2005)
[5] Xu, Y., Hu, C., Yao, G.-L.: Design and Implementation of Infrared Transfer Based on IrDA Object Exchange Protocol for Portable Terminals. Chinese Journal of Electron Devices 30(1), 215–218 (2007)
[6] Chen, F.-M., Chen, Q.: Design of Serial Communication Based on IrDA. MicroComputer Information 22(11), 169–171 (2006)

Formal Analysis of Authentication Protocol Based on Directed Graph Model

Changchun Li, Kangnian Wang, and Xingtao Zhu

ChongQing Communication Institute
ChongQing China
Sunnyspring_li@yahoo.com.cn

Abstract. The formal description of authentication protocol and its security analysis is one of the important issues in the protocol analysis. In this paper, a method based on directed graph is presented to formally depict the authentication protocol specification. Meanwhile, a converse-searching algorithm based on directed graph is also presented to construct the protocol messages. This method is used to analyze WOO-LAM authentication protocol, and then a new flow and its attacking trace have been founded.

Keywords: Authentication protocol, Formal protocol analysis, Directed graph.

1 Introduction

Security protocols are more and more used for transmitting data safely and protecting computer system. But the design and analysis of the security protocols are very complicated and error-prone. Since the end of the 80s, BAN logic [2] is successfully used for analyzing the security of the security protocols; many methods are designed to analyze them. All these methods need to formulize the protocol before analyzing. For example, BAN logic needs protocol-idealized step to translate the protocol specification into the symbols which particularly belong to the BAN logic. Strand space [1] uses 'bundle' to formally express the protocol specification, which reflects some relationship between protocol run and its principals. CSP [4] depicts protocol specification with the event, progress and trace, which is more close to the real run environment of the cryptographic protocols.

Among these methods, many details about the relationship among the protocol messages, steps and principals are omitted. However, these details are very important for analyzing the security properties of the security protocols. Losing these details does affect the efficiency of these methods [5] and [6].

In order to solve this problem, the paper presents a directed graph based method to describe the protocol run. This method defines a model to construct the protocol messages. Each protocol step is decomposed into the sequence actions. The construction rules of the directed graph of the protocol are defined, which can be used to uniquely describe the protocol specification. Based on this, we present a converse-searching algorithm (CSA) used for constructing the protocol messages. CSA algorithm can find the trace of attacking.

This paper is arranged as follows. In section 2, the message computing model and the building rules of the directed graph of the protocols are defined. Then, the

converse-searching algorithm is presented in section 3. it is proved that this method is effective through the analysis of Woo-Lam authentication protocol, and also find a new flow and its attacking trace in section 4. At last, we summarize this paper briefly in section 5.

2   Directed Graph Based Protocol Model

The implementation of a protocol run is completed by the exchanging of the protocol messages. According to Delov-Yao model [7], we have security protocol system model as Fig.1. In this model, intruder and communication media are unified to a special entity: environment, i.e. all messages sent by protocol principals can be owned also by environment, hence by intruder. We denote by PID the set of protocol principals, K the set of keys owned by principals, N the set of nonce generated by principal in the protocol run, D the set of information transmitted among principals. We denote by [m] the set of message that comes from the decomposing of message m. The decomposing rules are shown in Tab.1. We also denote by IP(i) the set of messages owned by principal P in protocol step i. We have IP(i) = IP(i-1) ∪ [mi], while m is the message that P received in step i. So we have message deductive system (table 2), while gen(m, i, P) means message m is generated by P in the step i, send(m, k) means message m is sent in the step k.
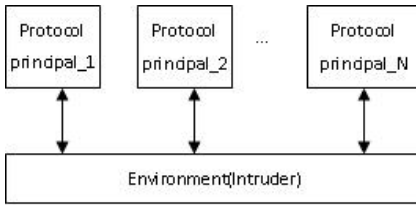
Fig. 1. Security protocol system model

Table 1. Decomposing rules

$$[M, N] = [M] \cup [N]$$

$$[M] = \{M\}, \text{ if } M \in (P_{ID} \cup K \cup N \cup D)$$

$$[\{M\}_k] = \{\{M\}_k\}, \text{ if } k \notin I_P(i)$$

$$= [M], \text{ if } k \in I_P(i)$$

Table 2. Message deductive system

$$(1)\frac{m \in I_p(i)}{I_p(i) \mapsto m}, (2)\frac{gen(m, i, P)}{I_p(i) \mapsto m}, (3)\frac{send(m, k), k \leq i}{I_p(i) \mapsto m},$$

$$(4)\frac{I_p(i) \mapsto \{m\}_k, I_p(i) \mapsto k}{I_p(i) \mapsto m}, (5)\frac{I_p(i) \mapsto m, I_p(i) \mapsto k}{I_p(i) \mapsto \{m\}_k},$$

$$(6)\frac{I_p(i) \mapsto < m, n >}{I_p(i) \mapsto m}, (7)\frac{I_p(i) \mapsto < m, n >}{I_p(i) \mapsto n},$$

$$(8)\frac{I_p(i) \mapsto m, I_p(i) \mapsto n}{I_p(i) \mapsto < m, n >}$$

Definition 2.1. Message similitude: if two protocol messages have same form, i.e. the components in the same positions in them have the same message types and different values, we say these messages are similar, which is denoted by ≈.

A protocol message M is built up from cleartest and chippertest. The former is generated in the current step or comes from some other steps prior to the current step, while the later is generated by encrypt or decrypt operation. So we denote by r_G, r_α, r_δ, and r_∂ the four message computing actions respectively.

In order to express the inner relationship among the messages, the steps and the principals in a protocol run, we present a directed graph based protocol model. The directed graph is consisting of the nodes and the directed side; the directed side indicates the flow direction of the message. Therefore, a protocol run can be expressed to Π(N, C), where N is the set of nodes and C is the set of directed sides.
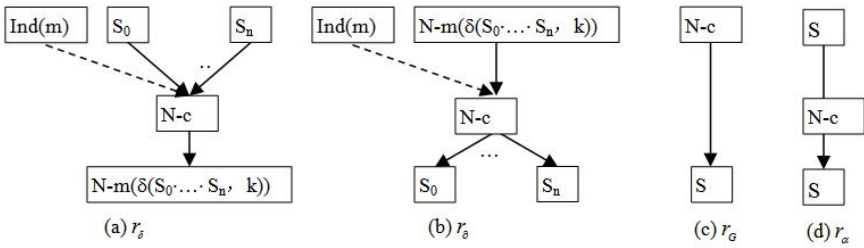
Fig. 2. Three kinds of atomic directed graphs

Definition 2.2. Computing node: when principal executes one of the four message computing actions, we call it a computing node which is denoted by N-c. if the computing node N-c belongs to the protocol step i, the content of this node is denoted by (Act, Own, $I_P(i)$, $R_\Pi(i)$), where Act denotes one of the four computing actions, Own is the operator of this action, $R_\Pi(i)$ is the protocol rules at the ith protocol step.

Table 3. Constructing algorithm of the directed graph

i=1
For not the end of the protocol
Decide on who is the operator of this step, e.g. P, and its $I_P(i-1)$.
Compute $[m_i]$.
For each $m \in [m_i]$
According to m, select one of the four message computing actions.
Set the computing node as N-c: r, P, $I_P(i)$, $R_\Pi(i)$), r∈ { r_G, r_α, r_δ, r_∂ }
Next
i=i+1
$I_P(i)=I_P(i-1)\cup[m_i]$
Next

Definition 2.3. Message nodes: if message m is the parameter of a computing action, no matter it is an input message or an output message, it is called a message node, which is denoted by N-m. Especially, if m is not a parameter of any computing action, then m is an isolated message node.

Definition 2.4. Indicating node: in the directed graph of the protocol run, if the key parameter or the computing operator is decided by the node when a computing node executes an action on message m, then this node is called Indicating node, denoted by Ind(m).

Definition 2.5. Directed side: if node N-m or Ind(m) is the input node of node N-c, then there exists a real or broken line which points from N-m or Ind(m) to N-c. If it is the output node of node N-c, then there exists a real line which points from N-c to N-m.

According to the Def.2.2-2.5 and four computing actions, there are four atomic directed graphs shown in Fig.2.

The constructing algorithm of the directed graph of the protocol is shown in Tab.3.

As an example, Tab.4 and Fig.3 show the Woo-Lam protocol specification and its directed graph respectively.

3   Formal Analysis Based on Directed Graph Model

According to the directed graph of the protocol and the message deductive system, we present a converse-searching algorithm. This algorithm is used to find an other directed graph different from the normal one to generate message m. Because it begins with m, and computes back to the start point of the protocol, so we call it Converse-Searching Algorithm (CSA). The details of this algorithm are shown in Tab.5. Let A and B be the legal principals, and CSA(m, A, P) means message m is generated by P and sent to A in the protocol run initialized by them. We denote by E the environment that can interact with the protocol run. While $m_j(i)$ means $m_j(i)=m_i$, but is generated in step j. if message m is generated by P through computing actions Act* added to message m', then we denote by Input(m) the input message m' of Act*, by Operator(m) the action operator P.

Theorem 3.1. It is decidable whether CSA(.) can be stopped.

Prof. Let $CSA(m_i, A, B)$, which means m_i is generated by B and sent to A. If A and B are the legal principals, so it is must be hold that the formula $\{IB(i)-\{m_i\}\} \mapsto m_i$ is true. Then $CSA(m_i, A, B)$ will be stopped in step (1). If A is a legal principal, and B is an illegal principal, i.e. environment. If m_i satisfies (2), then the algorithm will be stopped. Otherwise, (3) and (4) must be executed. In (3) and (4), only three principals can take part in the protocol, i.e. the legal A and B, the illegal E. so the protocol runs that can be initialized by these three principals are limited. According to (3) and the constructing algorithm of the directed graph of the protocol, this algorithm begins with message m_i, while $i \geq 1$ and searches back to the start point of a protocol run initialized by two of A, B, and E. The start point of a protocol run is decidable, so the

Table 4. Woo-Lam authentication protocol

Message 1	A -> B : A
Message 2	B -> A : N_b
Message 3	A -> B : $\{N_b\}K_{as}$
Message 4	B -> S : $\{A, \{N_b\}K_{as}\}K_{bs}$
Message 5	S -> B : $\{N_b\}K_{bs}$

Table 5. CSA algorithm

CSA(m_i, A, P)

(1) If $(I_P(i)-\{m_i\}) \mapsto m_i$

    Then return true and stop.

(2) If m_i is the start point of a protocol run, or i=1

    Then return false and stop.

(3) If m'_{i-1}= Input(m_i), P= Operator(m_i), P∈{A, B}

    Then CSA(m'_{i-1}, P, E). E∉{A, B}

(4) If find $m_j \approx m_i$, P= Operator($m_j(i)$), P∈{A, B}

    Then CSA($m_j(i)$, P, E)， E∉{A, B}.
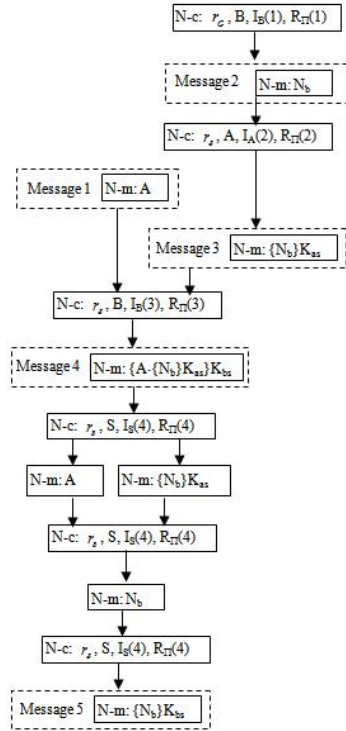
Fig. 3. Directed graph of Woo-Lam protocl

algorithm can be stopped. As for step (4), except for it will find a similar message mj and begin with it, it will actually fall into step (3). So the theorem 3.1 is hold.

Under the 'free hypothesis' [3], there exist two ways that can be used by the intruder to attack the protocol: one is that the intruder uses the keys and messages owned by him to construct a fake message which satisfies the protocol specification or to decrypt a chiphertext; the other is that the intruder initializes a protocol run with a legal principal, and makes it to generate messages that can be used to attack an other protocol run. Therefore, from the generalized point of view, when the intruder uses the protocol specification as a tool to attack the protocol run, the core of this attacking method is to construct the fake messages which satisfy the protocol specification and can be accepted by the legal protocol principals.

In other words, according to the directed graph model presented in this paper, if the number of the directed graphs of a message is greater then one, it means that this message can be constructed through an other way except for the legal one. As we have seen above, CSA algorithm can be used to find out all directed graphs of a message. So if the return of CSA(mi, A, E) is true, it means that mi can be generated by environment, i.e. there exists an intruder that can fake message mi. and if mi is for the goal of the authentication protocol, then there exists an attack, and CSA(mi, A, E) can find out the attacking trace. The next section gives an example.

4   The Analysis of Woo-Lam Authentication Protocol

In this section, Woo-Lam authentication protocol is used as an example to formally analyze the protocol based on directed graph model and CSA algorithm. Its specification and directed graph are respectively showed in Tab.5 and Fig.3.

According to the protocol, m5={Nb}Kbs is used by B as the goal to authenticate A. So we can compute CSA(m5, B, E) to find out whether there exist other directed graphs to construct m5.

For the sake of the limitation of the paper length, we briefly explain the computing progress of CSA(m5, B, E). as for step (3) in CSA, the input(m5) is {I, { Nb }Kis } Kbs, which can be build from CSA({I, { Nb }Kis } Kbs, E, B). and there exists a protocol run initialized by E and B, while B is the responder and E is the initiator. So the return of CSA({I, { Nb }Kis } Kbs, E, B) is true, and this attack is show in Tab.6. as for step (4), we can find { Nb }Kas≈{ Nb }Kbs, so CSA({ Nb }Kbs, B, E) is true in a protocol run initialized by E and B, while E is the responder and B is the initiator. This attack is a new attack and is shown in Tab.7.

Table 6. Attacking trace one

Run 1 message 1: I (A)-> B : A
Run 1 message 2: B ->I (A) : N_b
Run 2 message 1: I -> B : I
Run 2 message 2: B ->I : *
Run 2 message 3: I -> B : { N_b }K_{is}
Run 2 message 4: B ->I (S) : {I, { N_b }K_{is} } K_{bs}
Run 1 message 3: I (A) -> B : *
Run 1 message 4: I(B) -> S : {I, { N_b }K_{is} } K_{bs}
Run 1 message 5: S -> B : { N_b }K_{bs}

Table 7. Attacking trace two

Run 1 message 1: I (A)-> B : A
Run 1 message 2: B ->I (A) : N_b
Run 2 message 1: B -> I (C) : B
Run 2 message 2: I (C) -> B : N_b
Run2 message 3: B -> I (C) : { N_b }K_{bs}
Run 1 message 3: I (A) -> B : *
Run 1 message 4: B -> I (S) : *
Run 1 message 5: I (S) -> B : { N_b }K_{bs}

5   Summaries

In order to account for the limitation in the formal depiction of the protocol specification, this paper presents a new method to formally describe the protocol run: the directed graph based model. Based on this model, a converse-searching algorithm is presented. First of all, a message deductive system is used to build up message from IP(i), while decomposing rules of message is used to decompose message. Then the constructing rules of the directed graph of the protocol are defined to indicate how to build up a directed graph through protocol run. Furthermore, through the converse-searching algorithm one can find out one or more traces to build the directed graph of the protocol message. At last, the formal analysis of Woo-Lam authentication protocol is used as an example to prove the validity of this method. We find a new flow and its attacking trace.

However, if the intruder can fake the protocol message, does it mean that the intruder can successfully attack the protocol run? Or which message should be selected by the intruder to construct to attack the protocol run? This question comes down to the security properties of the protocol. This is the question that will be

researched next, i.e. how to formally express the security properties and the relationship between the protocol messages and the security of the protocol run.

References

[1] Javier Thayer Fabrega, F., Herzog, J.C., Guttman, J.D.: Strand Spaces: Proving Security Protocols Correct. Journal of Computer Security, 191–320 (July 1999)

[2] Burrows, M., Abadi, M., Needham, R.: A Logic of Authentication. ACM Transactions on Computer Systems, 18–36 (February 1990)

[3] Meadows, C.: Formal Methods for Cryptographic Protocol Analysis: Emerging Issues and Trends. IEEE Journal on Selected Areas in Communication, 44–54 (January 2003)

[4] Lowe, G.: Breaking and Fixing the Needham-Schroeder Public-key Protocol Using FDR. Software-Concepts and Tools, 93–102 (August 1996)

[5] Mao, W.: An Augmentation of BAN-like Logics. In: 8th IEEE Computer Security Foundations Workshop, pp. 44–55 (March 1995)

[6] Woo, T.Y.C., Lam, S.S.: Authentication for Distributed Systems. Practical Cryptography for Data Internetworks, 19–41 (January 1992)

[7] Dolev, Yao, D.: On the Security of Public Key Protocols. IEEE Transactions on Information Theory 29, 198–208 (1983)

Research of Customer Classification Based on Rough Set Using Rosetta Software

Liangzhong Shen and Shenkai Chen

[1] Information Management Department
City College of Wenzhou University
Wenzhou, China
[2] Department of Computer Science,
City College of Wenzhou University
Wenzhou, China

Abstract. Customer classification is an important aspect of customer relationship management for an enterprise. This paper uses Rosetta software which is based on rough set theory to process the data from a customer investigation. The rules generated not only show the factors influencing the customer's purchase more, but also clarify which kind of customer will own higher loyalty degree.

Keywords: data mining, rough set, customer classification, Rosetta software.

1   Introduction

The continuous economic development leads to continuous change of enterprise marketing concept, which has been transformed from product-centric to customer-centric. Therefore the customer has already become a decisive strength for the enterprise's future development. More and more enterprises have adopted this new management mechanism of customer relationship management to improve the relationship between the enterprise and the customer, namely enhancing the customer's satisfaction degree and loyalty degree. The enterprise knows that different customer has different contribution to him and different influence for his future development. That is why more and more valuable data such as the customer's basic information, purchasing record and product feedback, has been collected and analyzed to excavate the hidden and potentially useful information. Though some early methods like statistics for customer classification have returned satisfactory results, they can hardly meet the challenges of today's explosive data and complicated customer analysis. The new classification methods based on data mining technologies have opened up a new way.

In this paper, Rough Set theory is introduced for customer classification. Compared with other data mining technologies, rough set theory has many advantages, such as no information loss, flexible, extendable. It is a new mathematical tool to process indefinite and inconsistent data and imprecise questions. As a powerful tool of analyzing indefinite aspects, rough set theory has obtained widespread application in machine learning, data mining, policy-making analysis, process control, and pattern

recognition, and so on. For example, Hong [3] and Tian [5] discussed the application of Rough Set Theory in analyzing the evaluation of course design and crm intellectualization.

The following paper is organized as: firstly the principle of Rough Set theory is introduced, secondly an investigation is carried out and some related data about product selection are collected for analysis, and then Rosetta software is used to perform the attribute reduction and rule generation, the intermediate result is compared to that calculated according to rough set theory, and finally explanations are provided so that the enterprise can follow to make better marketing strategy in the near future.

2   Rough Set Theory

Rough set theory was first put forward by Professor Z.Pawlak in Poland on 1982. The core of the theory is upper and lower approximation set. Rough set theory applies the indiscernibility relation and data pattern comparison based on the concept of an information system with indiscernible data, where the data is uncertain or inconsistent.

2.1   Basic Concepts of Rough Set

1) Information system
An information system is a 4-tuple s=(U,A,V,f), where U is the universal object set, U = $\{x_1,x_2,...x_n\}$; A = C∪D represents the attribute set, A = $\{a_1,a_2,...a_n\}$,subsets C and D are respectively conditional attribute set and decisional attribute set; V is the corresponding value set of the attribute set A, V=$\{v_1,v_2,...v_n\}$;f:U*A->V is the information function $(X_i, a_j)\in v_j$.

In the classification problems, an information system is also regarded as a decision table which is shown in fig1:
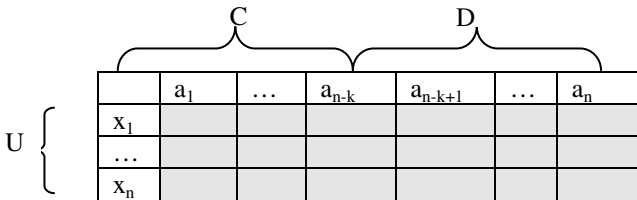
Fig. 1. A decision table

2) Lower and upper approximation
Let R be an equivalence relation on U, and the object subset X ⊆ R, the R-lower approximation set of X is denoted as follows:

$$R\_(X) = \cup \{x| \ x \in U, [x]_R \subset X\} \qquad (1)$$

And the R-upper approximation set of X is denoted as follows:

$$R^-(X) = \cup \{x| \ x \in U, [x]_R \cap X \neq \varphi\} \qquad (2)$$

The R-lower approximation set is the union of all equivalence classes contained in X, while the R-upper approximation set is the set of equivalence classes which have non-empty intersection with X. And this means that the lower approximation is the set of

objects that can be positively classified as members of X, the upper approximation is the set of objects possibly belonging to X.

3) Attribute dependability and attribute importance

Let $C \subseteq A$ and $D \subseteq A$, the positive region of classification U/D to the set of attribute C is as follows:

$$POS_C(D) \quad = \cup \; P\_(D) \tag{3}$$

$POS_C(D)$ contains all objects in U that can be classified to one class of the classification U/D by attribute C. The dependency of D on C is as follows:

$$k = r_C(D) = card(POS_C(D)) \;/\; card(U) \tag{4}$$

An attribute a is said to be dispensable in C with respect to D if $POS_C(D) = POS_{C-\{a\}}(D)$, otherwise a is an indispensable attribute in C with respect to D.

Suppose attribute a is joined in C, the importance degree of a on U/D is as follows:

$$SGF(a,C,D) = r_C(D) - r_{C-\{a\}}(D) \tag{5}$$

4) Attribute reduction set and the core

The minimum set of attributes that preserve the indiscernibly relation is called attribute reduction set. Let $C \subseteq A$ and $D \subseteq A$, C is the condition attribute; D is a set of decision attribute. We will say that $C' \subseteq C$ is a D-reduction of C. If C' is a minimal subset of C such that

$$r_C(D) = r_{C'}(D) \tag{6}$$

Thus attribute reduction enables us to make decisions employing minimal number of conditions. The intersection of all reductions is called core. The core stands for the constant information. It is the basis of the reductions and the knowledge attribute sets that cannot be eliminated. Similarly, we can reduce values of conditional attributes from attribute reduction to get a set of tidy. The maximum number of conditional attribute values of a rule is removed without decreasing the classification accuracy of the rule; this process is called as value reduction.

3   Example of Customer Classification

In order to have a deep understanding about the product and the customer, an investigation was made to collect the corresponding data. The main steps of the process are as follows: (1) define ten decisive variables about the product that the enterprise think the customer will care for and achieve the customer's feedback; (2) calculate the customer's potential value according to the current value, discretize the customer's potential value to the loyalty in a simple way (3) reduce the attribute set based on rough set theory by using Rosetta software and compare the intermediate result to that calculated by rough set theory to improve the consistency; (4) generate the decisive rules and explain the meaning of the rules.

3.1   Define Ten Decisive Variables

The analyzed product is a kind of some famous beer brand, so that the ten decisive variables related to the product chosen are shown in Tab.1:

Table 1. List of Ten variables

Code	Variable	Code	Variable	Code	Variable
C1	taste	C4	price	C7	convenient
C2	smell	C5	brand	C8	alcoholicity
C3	packing	C6	advertisement	C9	stuff
				C10	safety

The customer needs to select the key variables influencing his purchase decision. If the customer selects the variable, then the corresponding value is marked as 1, otherwise marked as 0. Other related investigative questions are omitted here.

3.2   Calculate the Customer's Potential Value

The current value of a customer is defined as the total profit value that the customer brings to the enterprise during the period from past to present, namely the customer's past profit contribution (PPC) and the calculation formula is as follows:

$$PPC = \sum_{i=0}^{N} \pi_p(i)(1+d)^{N-i} \tag{7}$$

Thereinto, i is the life cycle interval index, d is the discount rate, N is the length of life cycle, and $\pi_p(i)$ is the historical contribution function。

The potential value of a customer refers to the present amount of the total profit that the customer would bring to the enterprise in the future based on the condition of purchase increase. It depends on the incremental purchase, cross purchase and the possibility of new customer recommendation. It is represented as expect Future Profit Contribution (FPC) and the calculation formula is as follows:

$$FPC = \sum_{i=N+1}^{N+E+1} \frac{\pi_f(i) + P(i)}{(1+d)^{i-N}} \tag{8}$$

Thereinto，E is the expected length of life cycle, $\pi_f(i)$is the future contribution function representing the expected profit brought to the enterprise based on the unchanged purchase model, P(i) is the potential contribution function representing expected profit brought to the enterprise based on the changed purchase model. The value of $\pi_p(i)$ can be obtained from the customer's previous transaction and $\pi_f(i)$ can be obtained with linear regression of customer's previous transaction. Based on formula (7) and (8), the current value and potential value of the customer is computed and several representative records are shown in tab.2. The customer's loyalty degree is decided by the potential value in a simple way. If the potential value is less than 5000, then marked as 1; if the potential value is bigger than 10000, then marked as 3; if the potential value is between 5000 and 10000 then marked as 2.

Table 2. Customer's current and potential value and loyalty degree

Cno	Current Value	Potential Value	Loyalty degree
1	4133	6199	2
2	3774	4303	1
3	5307	7960	2
4	6986	10479	3

Combined with the investigation data that a customer provide and his corresponding loyalty transformed from his potential value which is calculated by way of formula (7) and (8), twenty customer's record is selected as shown in tab.3:

Table 3. Twenty customer investigation data

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	D1
1	1	0	0	1	0	0	0	1	1	0	2
2	1	0	1	1	1	0	1	0	1	0	1
3	0	1	1	1	1	1	0	0	0	1	2
....	....	....	....	....	....	....	....	....	....	....	....
....	....	....	....	....	....	....	....	....	....	....	....
19	1	1	0	0	1	1	1	0	1	0	1
20	1	0	1	1	0	1	0	1	0	0	3

3.3   Reduce the Attribute Set Using Rosetta

ROSETTA is a toolkit for analyzing tabular data within the framework of rough set theory. It is designed to support the overall data mining and knowledge discovery process: From initial browsing and preprocessing of the data, via computation of minimal attribute sets and generation of if-then rules or descriptive patterns, to validation and analysis of the induced rules or patterns.

The upper data of Tab.3 is imported into Rosetta software, and a genetic algorithm is implemented for computing minimal attribute sets. The attribute reduction set {C1, C2, C3, C4, C5} is shown is fig.2, so it is clear that the attribute set {C6, C7, C8, C9, C10} are redundant attributes.

Fig. 2. The attribute reduction set

From Section II we know that an attribute a is said to be dispensable in C with respect to D if $POS_C(D) = POS_{C-\{a\}}(D)$, otherwise a is an indispensable attribute in C with respect to D. and also $POS_C(D) = \cup P\_(D)$. At present, the value of C and D are as follows:

C = {C1, C2, C3, C4, C5, C6, C7, C8, C9, C10}
P = {C1, C2, C3, C4, C5}   P = C – {C6, C7, C8, C9, C10}
D = {D1}
U/D = {{2,5,14,15,18,19},{1,3,6,12,13},{4, 7, 8, 9, 10, 11, 16, 17, 20}};
U/C = {{1},{2},{3},{4},{5},{6},{7},{8},{9},{10},{11},
      {12},{13},{14},{15},{16},{17},{18},{19},{20}}
   $POS_C(D)$ = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20}
U/P =   {{1},{2},{3},{4},{5},{6},{7},{8},{9},{10,16},
      {11},{12},{13},{14,18},{15,19},{16},{17},{20}}
   $POS_P(D)$ = {1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20}

The result $POS_C(D) = POS_P(D)$ shows that the core attribute set {C1,C2,C3,C4,C5 } is correct.

3.4   Generate the Decisive Rules Using Rosetta

Through the above reduction, tab.3 has been changed into tab.4:

Table 4. The reduction data

	C1	C2	C3	C4	C5	D1		C1	C2	C3	C4	C5	D1
1	1	0	0	1	0	2	11	1	1	1	1	0	3
2	1	0	1	1	1	1	12	0	1	0	1	1	2
3	0	1	1	1	1	2	13	0	1	1	1	0	2
4	0	0	1	0	1	3	14	0	0	1	1	1	1
5	1	1	1	1	1	1	15	1	1	0	0	1	1
6	1	1	0	1	0	2	16	0	0	1	1	0	3
7	0	0	1	0	0	3	17	0	0	0	1	1	3
8	1	0	0	0	0	3	18	0	0	1	1	1	1
9	1	1	1	0	1	3	19	1	1	0	0	1	1
10	0	0	1	1	0	3	20	1	0	1	1	0	3

Afterwards we can obtain decision-making rules by using Rosetta software, the result is shown as fig.3:

Fig. 3. Generated rules

From above, we see 17 rules are generated. Take the first rule as an example: C1(taste) ^ C4(price) -- > D1(2), this means if the customer think C1 and .C4 are the key variables for him to purchase the product, then the customer's loyalty degree would be 2. As for the enterprise, they can realized that the customer cares for the product's taste, smell, packing, price, brand more than others, so they should place more attention to these product aspects; and also, from the subsequent customer feedback, they can classify correctly which customer is loyal and how to adjust the subsequent marketing strategy according to the customer's feedback.

4   Summaries

This paper uses Rosetta software which is based on rough set theory to process the data gathered from the customer feedback. The customer feedback is related directly to his loyalty degree which is calculated in accordance with his previous transaction. The rules generated can let the enterprise know better how to improve their products and how to classify their customer accurately.

References

[1] Pawlak, Z., Grzymała-Busse, J.: Rough Sets. Communication of the ACM 38, 89–95 (1995)
[2] Pawlak, Z.: Rough set theory and application to data analysis. Cybernetics and Systems 29, 661–688 (1998)
[3] Liu, H., Ma, Y.: Research of performance evaluation of course design based on rough set theory. In: Proceeding of 2009 4th International Conference on Computer Science & Education, pp. 1005–1008 (2009)
[4] Tang, B.: Customer relationship management. Beijing Higher Education Press (2003)
[5] Tian, J., Zhang, G., Lv, F.: The research of crm intellectualization based on rough set. In: 2009 International Forum on Information Technology and Applications, pp. 181–185.
[6] http://www.lcb.uu.se/tools/rosetta/index.php

A Protocol to Support Mobile Computing for Publish/Subscribe Middleware

Tao Xue[1] and Tao Guan[2]

[1] College of Computer Science
Xi'an Polytechnic University
Xi'an, Shaanxi, China
`xthappy@gmail.com`
[2] Department of Computer Science and Application,
Zhengzhou Institute of Aeronautic Industry Management,
Zhengzhou, Henan, China
`timm.guan@gmail.com`

Abstract. Existing publish/subscribe middleware are mostly optimized for static systems where users are fixed. In this paper, we present a novel protocol to support mobile clients for publish/subscribe middleware. We describe necessary steps of the protocol and how to solve the problem of message loss and duplication. The experiment shows that our protocol can effectively address the challenges raised by emerging mobile applications.

Keywords: Publish/subscribe, middleware, mobile computing.

1 Introduction

Due to the characteristics of wireless communication (for example, narrow bandwidth, unreliable, disconnected, etc.), it is not suitable to support mobile applications using traditional RPC middleware (CORBA, RMI, DCOM, etc.). In contract, publish/subscribe middleware has greater advantage than traditional middleware; it has characteristics of loosely coupled and asynchronous mode. But most publish/subscribe middleware are designed for static systems and not fully satisfied the additional demand of mobile computing. These extra requirements include moving among different assess points, unstable connection and guarantying reliability of message without missing and repeating.

In this research area, some literatures [1-5] have present solution to solve these problems. In this paper, we present a new protocol MobilePSM which allows that mobile clients can use publish/subscribe middleware to move among different access points and ensure messages' integrity.

2 The Problem

When distributed publish/subscribe middleware is used to support mobile system, mobile clients can transparently connect and receive message in different broker nodes. Transparency must be ensured by publish/subscribe middleware. And following two requirements should be meet.

Integrity: When a mobile client moves from one event broker to another event broker, it can still receive messages which are sent to it during the moving period. Even if it is disconnected passively with the old connection, the message can not be lost and duplication.

Ordering: Mobile clients should receive messages according to the sending order.

3   MobilePSM Protocol

In this section we propose a protocol called MobilePSM, which ensures all requirements of mobile application in the previous section. The implementation of protocol requires two prerequisites: First, to store the receiving messages of mobile clients temporarily in the moving period, each event broker of publish/ subscribe middleware should maintain a first-in-first-out (FIFO) buffer for each connected mobile client (subscribers); Second, each published message carries a unique identifies of publisher and message.

- Basic process

The basic process of the protocol is shown in Figure 1. Mobile client C moves from event broker B1 to event broker B2 (Step 1). The underlying runtime library of C establishes a connection with B2, it automatically resends subscription message to B2 (Step 2). B2 processes subscription by routing algorithm (Step 3) so the messages sent to C will be sent to B2. In addition, B1 will store the message sent to C in this period to its buffer temporarily.

Then B2 establishes a connection to B1, sends "get" command (Step 4). After B1 receiving the command, it knows that C has been moved to B2. Then it sends a unsubscribe message to the relevant event broker (Step 5). This will change the routing table, so the message sent to C will stop be sent to B1. B1 will send the message of C to B2 orderly (Step 6), and then B2 sends to C.
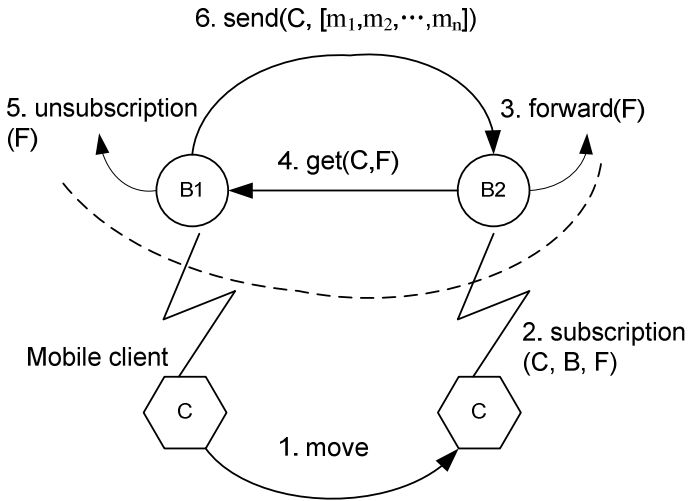
Fig. 1. The basic process of MobilePSM protocol

The basic protocol cannot guarantee integrity of message. The message may be lost or duplicated.

Loss of the message: In the time B2 invokes forward operation and the new routing has not been established, B1 receives get command, and then cancels C's subscription and deletes the old routing. In this time, the message sent to C will be abandoned and the message will be lost.

Duplication of the message: Like forward operation, unsubscribe operation is asynchronous too. In other words, when it is called, the routing may not be deleted. The message sent to C will be spread in the two routing paths.

- Improved algorithm

We use a synchronization algorithm to solve the problem of message loss. As shown in Figure 2, after B2 called forward operation, it can not immediately issue get command to B1, but sending a sync command. When B1 receives the command it begins to periodically send test message called MTEST. The content of MTEST is matched with the subscription of client C and its publisher identifier is a special test identifier called IDTEST. If the system has established a routing to B2, MTEST will reach B2. Clearly the intention of synchronization is testing whether the new routing has been established. When B2 receives the message of sending to C (whether it is MTEST), the new routing has been established. Then B2 sends get command to B1. When B1 receives the command it stops sending MTEST and cancels subscription of C. At last, it will send the message of C to B2.

8. send(C, [m_1,m_2,$\cdots$,m_n])

5. publish(m_{test})

7. unsubscription (F)

3. forward(F)

4. sync(C,F)

B1   B2

6. get(C,F)

Mobile client

2. subscription (C, B, F)

C   C

1. move
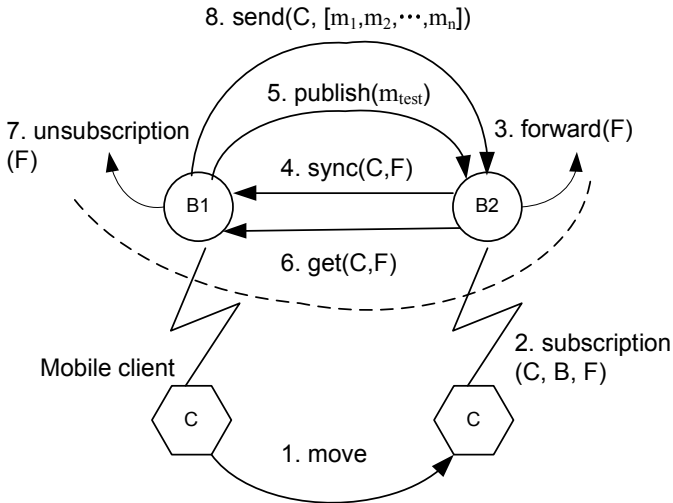
Fig. 2. MobilePSM protocol with synchronous algorithm

Since each message has a unique identity, B2 will remove the duplicated message through checking the history buffer and its temporary buffer. Finally, the combined message are orderly sent to C.

4   Test and Evaluation

In experiment we deploy 10 event broker nodes in 10 PCs and one message publisher program in a PC. The publisher sends some type of messages periodically. And the interval range of the cycle is from 100ms to 2000ms. In addition, we also deploy a mobile client which receives the message publisher send. In the same configuration, we compare with other two algorithms: "Subscribe-Unsubscribe-Subscribe (SubUnsub)" algorithm and movein-moveout algorithm. Figure 3 shows the loss of the message in different algorithms.
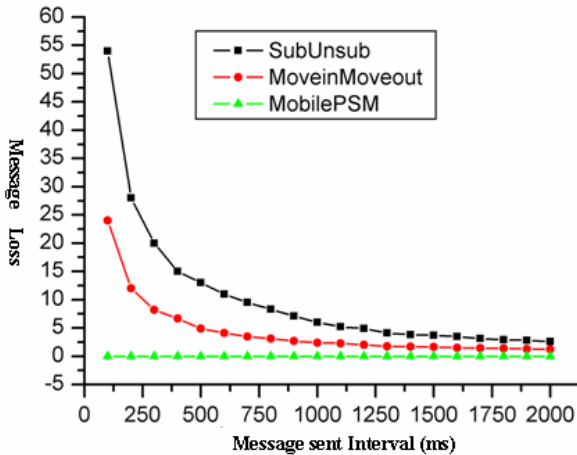
Fig. 3. The loss situation of the different algorithms

From the experimental results, we can see MobilePSM protocol can fully guarantee integrity of message. However message loss in different levels appears in the other two algorithms. In SubUnsub algorithm message loss is the largest.

Both MobilePSM and movein-moveout algorithms merge duplicate message so message duplication does not emerge. But SubUnsub algorithm has the problem.

5   Summaries

Publish/subscribe middleware has loosely coupled and multicast features. Compared with RPC middleware it has better advantage to support mobile computing. But nowadays most of the publish/subscribe middleware systems are designed for static and fixed network. And the needs of mobile computing do not be considered. The paper propose MibilePSM protocol that need not any central control or communication mechanism outside publish/subscribe middleware. This protocol ensures the message integrity and application transparency.   The experiment shows that it completely meets the requirements of mobile computing, and better than other major algorithms. The next step of the study is to further improve the protocol performance and enhance its scalability.

References

[1] Huang, Y., Garcia-Molina, H.: Publish/subscribe in a mobile environment. In: Proceedings of the 2nd ACM International Workshop on Data Engineering for Wireless and Mobile Access (MobiDE 2001), Santa Barbara, CA (May 2001)

[2] Sutton, P., Arkins, R., Segall, B.: Supporting disconnectedness – transparent information delivery for mobile and invisible computing. In: First International Symposium on Cluster Computing and the Grid, Brisbane, Australia, pp. 277–287. IEEE/ACM (2001)

[3] Cugola, G., Di Nitto, E., Fuggetta, A.: The JEDI Event-Based Infrastructure and Its Application to the Development of the OPSS WFMS. IEEE Trans. on Software Eng. 27(9), 827–850 (2001)

[4] Caporuscio, M., Carzaniga, A., Wolf, A.: An Experience in Evaluating Publish/Subscribe Services in a Wireless Network. In: Third International Workshop on Software and Performance, Rome, Italy (July 2002)

[5] Fiege, L., Gartner, F.C., Kasten, O., Zeidler, A.: Supporting Mobility in Content-Based Publish/Subscribe Middleware. In: Endler, M., Schmidt, D.C. (eds.) Middleware 2003. LNCS, vol. 2672, pp. 103–122. Springer, Heidelberg (2003)

System Design of Virtual Human Science Museum

Jianxia Ge

Center of Modern Education Technology, Xin Xiang Medical University,
Xinxiang, Henan, China, 453003

Abstract. In recent years, with the use of internet and computer graphics program develop quickly, virtual reality technology emerges in an endless stream. According to this article, research museum of human sciences by Xin Xiang medical university for target, to achieve virtual museum of human sciences roaming with computer interact for the purpose, using virtual reality project development and design thinking, detailed analysis and design on the Science Museum interior scene on the human body, indoor specimens, specimens knowledge points, roaming the scene specimen interaction, design a three-dimensional forms on the internet showing the body Museum, visitors in the Library Walking Browse and interactive learning with the body structure of the concept.

Keywords: virtual reality, virtual museum, body Museum.

1 Introduction

Medicine is a special subject. The object of their research is human. The beginning of the learning is from form of the human body. Medical tests is used by live animals, cases of real human specimens and real medical records, thorough, cost is very high. The best use of limited resources to make by the university, reduce the education cost, ease the tension in the experimental teaching resources. In this paper, the background of the real human Museum, virtual reality technology into medical education, the Science Museum on the virtual human design, to describe his ideas and methods, so that more people have the opportunity to understand the mysteries of the human body, understanding the various parts of the body the structure, to promote independent learning, teaching methods and teaching methods of reform, development of teaching ideas, which realization of educational innovation has important reference value.

Virtual Reality (Virtual Reality, referred to as VR) is a virtual world can create and experience of computer systems [1], which make full use of computer hardware and software technologies to provide a real-time, three-dimensional virtual environment (Virtual Environment), provide visual, auditory, tactile and other sensory simulation, the user just like being in general, in real time, directly observed within the three-dimensional objects [2].

2 Virtual Human Museum System Requirements Analysis

An area of the museum of human sciences by Xin Xiang medical university is about 600 square meters. the collection exhibits more than a thousand pieces, including

embryogenesis, organ systems, pipe mold, plastic body, and mummies, etc., exhibits traditional anatomical techniques and plastics, transparent, mold, fault and other special technical methods, from different angles to show the complex structure of human visitors, to help visitors interpret the mystery of the human body, self-understanding, inspires visitors to learn medical knowledge on the human interest. Requires the user to interact with the system to complete the mouse and keyboard, in the Science Museum visitors enter the virtual human, not only can walk here complexes, and click on any of the human body specimens, that pop up in this sample point of the interaction of learning and knowledge resolution window specimens were also able to push past, pulling away, rotation, translation and other operations, so that visitors fell throughout the body of its clinical , to achieve the body at the Science Museum online remote roaming and self-learning interact. Science Museum's layout and structure of the human body show in Fig.1.

```
                    ┌──────────────────────────┐
                    │ Museum of human sciences │
                    └──────────────────────────┘
         ┌───────────────────┬───────────────────┐
  ┌──────────────┐   ┌────────────────┐   ┌───────────────┐
  │ Left pavilion│   │ Middle pavilion│   │ Right pavilion│
  └──────────────┘   └────────────────┘   └───────────────┘
```

| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |

1---front wall cabinet Embryo, 2---plastic body, 3---human body system, 4---skeletons specimen
5---after the wall cabinet skeletons, 6---human skeletal structure, 7---male and female mummy
8---human museum exhibition panels, 9---TV display screen, 10---plastic body, 11---pipe type casting
12---human organ, 13---after the wall cabinet vascular, 14---bonsai in front

Fig. 1. The layout of the structure of human Museum

3   Virtual Human System Implementation Process Design Museum as Below

Fig.2 includes these steps:

(1) Get data which is housing structure and entity object, effective processing of data sorting.

(2) Modeling indoor scene and body structure specimens in 3Ds MAx

(3) After modeling, leads to nmo. Files from 3Ds MAX for Virtools further processing.

(4) Control by nom. File in Virtools

(5) Optimization of integrated leads to vmo. Files

(6) Embedded in web page, edit subject knowledge, web publishing

4   Virtual Human Design of the Science Museum

The design of virtual reality is the use of human Museum of software engineering thinking and progressive software development model for virtual reality human

Museum needs analysis, design and coding [3]. The modular, hierarchical design, development and design level of clear, rational structure of human Museum. The Virtual human science museum of all structure shows in Fig.2.

```
                    ┌─────────────────┐
                    │  Museum design  │
                    └─────────────────┘
```

| Indoor scene | Indoor | Knowledge point | Roaming | Interface |
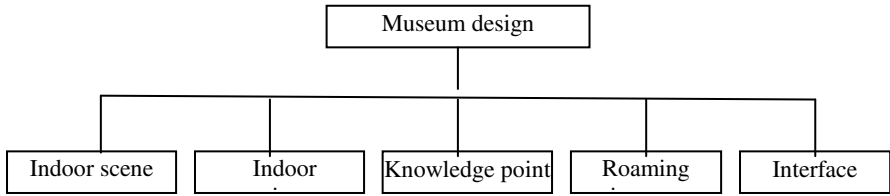
Fig. 2. Virtual Museum of the overall structure of human

4.1   The Design of Virtual Human Museum Interior Scenes

Human Science Museum virtual reality room designed by The layout of the interior scenes, indoor environment, housing decorative design composed of indoor environment including cabinets, wardrobe, glass containers, air conditioning, TV, curtains, panels and Bonsai. The interior design of the hierarchy of the scene shown in Fig. 3.

The design uses virtual reality programming language in the background node, view node, the node coordinates, inline nodes, group nodes, re-defined node, re-node, node geometry, face nodes, Texture node, the time the sensor node, dynamic interpolator node, the script node, routing the design and development, and implementation using inline subroutine call node to achieve hierarchical, modular and component design.

```
                    ┌──────────────────────┐
                    │ Interior scene design │
                    └──────────────────────┘
         ┌──────────────────┬──────────────────┐
┌─────────────────┐ ┌──────────────┐ ┌──────────────┐
│ Housing structure│ │ Environment  │ │  Decoration  │
└─────────────────┘ └──────────────┘ └──────────────┘
```

| Wall cabinet | Wardrobe | Glass container | Panels | Bonsai |
| Aircondition | Curtains | TV | Letters | Equipment |
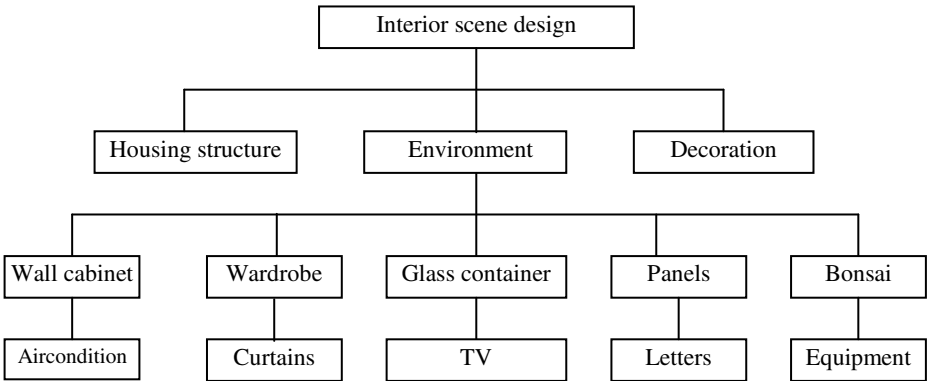
Fig. 3. Museum interior scene virtual human design hierarchy structure

4.2   The Design of Virtual Human Specimens of the Science Museum Interior

Museum Indoor virtual human specimens are specimens of human body structure. It includes human embryogenesis, organ systems, and local anatomy, cast, and plastics specimens. Specific developmental stages are all part of the embryo and abnormal specimens, specimens of fetal skeleton transparent; Bone from the motion system,

joints, skeletal muscle, the digestive system of the esophagus, stomach, liver, pancreas; from the respiratory system, nose, throat, trachea, lung, the genitourinary system, kidney, bladder, prostate, uterus; Power from the blood circulation organs - heart, to the whole body of the large and small blood vessels; mummies; various blood vessels, trachea, bile duct and other pipeline cast specimens; In addition there are "sprint", "Guardian," "longing," " face to face "," layup "and plastic body. Levels of the specimens according to the design of modular design, every level design down. Show in figure 4:
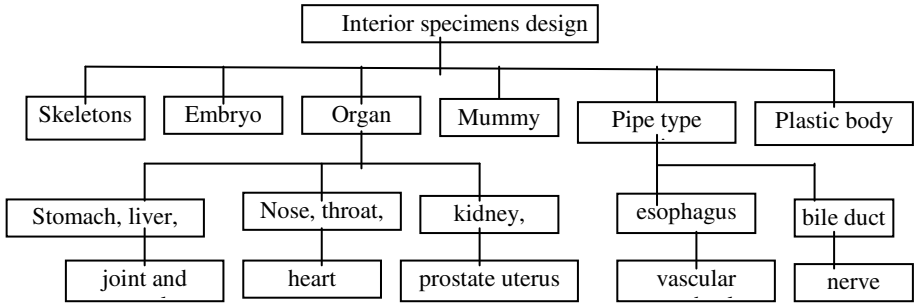
Fig. 4. Interior specimens design structure

Three-dimensional design of the human skeleton as shown in Figure 5:
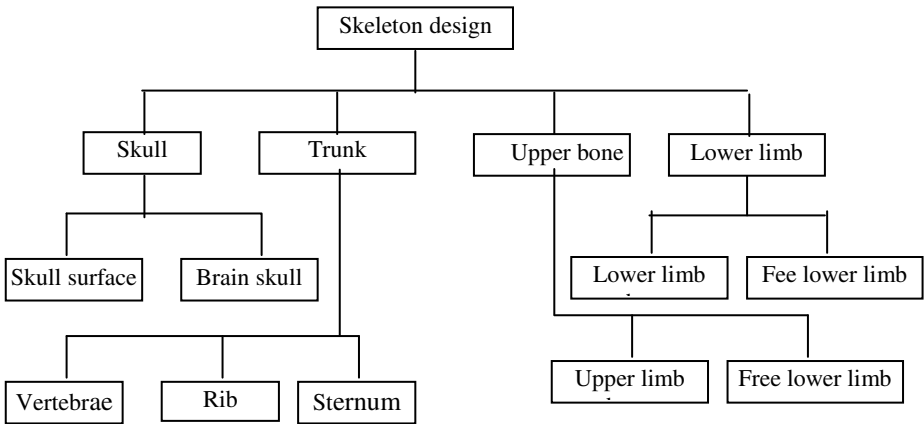
Fig. 5. Three-dimensional design structures

Reality human bone from the skull, trunk bones, on the limb and lower limb, of which the skull by the skull, brain, skull surface composition of bone from the trunk vertebrae, ribs and sternum and the upper limb from the shoulder girdle, the free form on the limb, lower limb bone from the lower limb, lower limb free form.

4.3   The Structure of Specimens of Virtual Human Design Knowledge Points

The body structure which is a straightforward, practical, strong morphological course, is the system anatomy, regional anatomy, histology, three basic courses of the combination of traditional medicine, mainly on the composition of the human body system, organ morphology, were embryogenesis Development Law and other related knowledge.

Knowledge of human body structure, design of virtual human is constructed on the basis of the Science Museum, follow the teaching of design, should include: learning objectives and learning for the content of the knowledge structure and characteristics of learners, and the selection of instructional media for three-dimensional structure model of the human body in different design point of their knowledge, the use of virtual reality technology of three-dimensional human body model learners the freedom to control, click on the trigger point of knowledge shows that this form is bound to show the three-dimensional visualization of the learning process, students are more likely to acceptable and can stimulate learner interest.

(1) Identify learning objectives
Learning goal is to learners in medical system which is in virtual teaching, should meet the learning outcomes or standards, and is a virtual human body structure to learn the starting point and foundation design. It controls the system determine the learning content, learning media choices. Virtual human anatomy learning goal is to make learners to master the body morphology of each organ system and their relationship, understanding embryonic development and human growth and development of various body parts and functional significance.

(2) To determine the teaching content
Medical virtual human is not the structure of learning content directly to the content of the book "moved" to the virtual human museum. It need to go through screening and treatment to enable the system to learn the contents of both to ensure the integrity of their knowledge but also the structure of the medical technology advantage of virtual humans.

(3) The analysis of knowledge structure
The so-called knowledge structure is the relationship between knowledge content and its associated form. The human body is composed of several independent knowledge unit compositions. Therefore, the analysis focuses on the knowledge structure of the contents of each unit of knowledge points and the relationship between knowledge and its associated sites.

(4) Characteristics analysis of learners
Characteristics of the learner is prepared to understand the learning activities and their characteristics, it is the basis for subsequent design work of teaching. The so-called learning readiness is a concept in educational psychology, refers to the learners engaged in new learning, the existing level of knowledge or level of psychological development of the original study on the adaptability of the new [4]. Characteristics of the learner include: analysis and knowledge levels of learners starting point for understanding the learning environment and learning ability and other aspects.

(5) instructional media selection

Teaching content according to the characteristics of different media means choosing the appropriate performance of a virtual teaching system in the medical teaching media selection key. Only full play the advantages of various media forms in order to achieve a multiplier effect of the teaching. Therefore, the virtual human body specimens during the instructional media selection and design, you first need to goal according to the teaching requirements and the characteristics of teaching content in the media choose the right form of expression, the special structure of the human body, if only presented as a picture is difficult to help learners on the human body structure to form an intuitive understanding of the content. And a three-dimensional virtual human body model, the overall structure of the human body, internal organs, blood vessels and nerves can be independently rendered, these independent models can be combined together to form different organic systems and three-dimensional virtual human body model, subject knowledge and the three-dimensional model associated with the same time, the use of virtual reality technology to the learner the freedom of three-dimensional human body model manipulation, so that students are free to choose viewing angle, viewing distance, to observe the structure of the human body. When the mouse to move the trigger of a three-dimensional model of the human body structure, the system will display the name of the model and knowledge points, is to help learners understand more about the human body structure learning and memory.

4.4   Virtual Human System Interaction Roaming Design Museum

Roaming in the virtual environment system, in order to facilitate the interaction between the user and the scene need to provide a variety of roaming control. Including the viewpoint of the forward, backward, left and right shift, reset observation point position, left rotation, and rotate the view. In the virtual scene, there is to create a simulation of the observer's point of view camera, when the viewpoint position and direction of change in sight, I saw part of the scene will change.System to obtain real-time view of the location information and change the parameters to re-draw the scenes, to achieve the user in the process of moving from different positions, different angles scene.

Interactive mode with the desires of the user roaming is carried out in a complex three-dimensional virtual scene roaming, mainly rely on the keyboard and mouse as input means. Principle is realized in advance for the mouse, keyboard, set the corresponding output device such as the operation content, when the user presses a function key, that is, the appropriate action to obtain the information, and information to the scene of the drawing module. Then adjust the parameters of plotting module, re-draw the scene. As users continue to operate, the scene will show a different attitude and perspective.

This article, the user presses the "↑" key to move forward, press the "↓" key back. Press "←" button to the left to go, "→" key right away. Users can move the mouse to change the view looking the direction of the virtual camera, surveying the scene.

4.5   Virtual Human System Interface Design Museum

One is to interface design beautiful, more importantly, to facilitate access by those who operate and use. Interface design is mainly creative design, creative design of multimedia applications and important source of active [6]. Good ideas to make

application system not only unique, but also greatly improves system availability and visibility. The system interface design including system design and the main interface sub-interface design, interactive samples.

4.5.1   Design of the Main Interface

In order to facilitate visitors to browse the virtual human Science Museum, Science Museum, the user create a virtual human body the main interface, the main interface design consists of three parts, the left is the body About the Science Museum, Science Museum, the body video, Virtools browser to download and use methods; the middle generation of human embryonic development, the arts body, the Science Museum system entrance virtual human right is the study of historical figures related to the human body. Framework as the main interface show in Fig.6.

This design can provide users with a good interface, click on the body Museum video, the explanation sounds appealing to watch videos Human Science Museum; Click humans and pictures to show the origin of man, analysis of human production; click on embryonic development , will present you a different person during the development of embryos; click artistic body, you will see the beauty of the human body; click on virtual human Museum, will take you to a virtual three-dimensional structure of the world body, where you can roam the body structure of the world, to explore the mysteries of the human body to stimulate the power of learning human anatomy.

Virtual Human Science Museum		
Human Science Museum Introduction	Production of human Into	Introduce historical figures …
	Embryonic development Into	… …
Body Museum Video	Human Art Into	
Plug-in Download Help	Virtual Human Science Museum Into	

Fig. 6. The main interface framework diagram

4.5.2   Interface Design Samples

When the rover roaming the body in the hall, the interest of a sample, click on the specimens into the specimen interface, the interface in this specimen the left part, viewers can sample any zoom, rotate, move, etc, carefully observed from three different angles specimens; Also in the text portion of the right side of this interface will display the name of this specimen analysis and knowledge points; because many of the museum specimens, when a special animation effects need to show the animation window will pop up animation, interactive sub-interface specimens need to be designed according to two-thirds split-screen or screen window.

5   Summaries

In this paper, Xinxiang Medical Museum on the background of human use of virtual reality project development design, Science Museum of virtual humans a detailed analysis and design, in the virtual scene, samples, interactive roaming, knowledge points and the use of virtual reality design teaching design, will fit and imagination combine to give full consideration to the friendly user interface and ease of use, the design implementation of the system has laid a good foundation, is the virtual reality technology in medical education intends to try, To carry out virtual experiments medical research on educational innovation, reform and improve teaching effectiveness and efficiency of providing new ideas and new methods.

References

[1] The Virtual Reality Modeling Language International Standard, ISO/IEC 14772-1 (January 1997)
[2] Zhang, Z., Zhang, A., Zhang, J.: Virtual reality three-dimensional network programming language, VRML, 1st edn. Tsinghua University Press, Beijing (2004)
[3] Zhang, Z., Zhang, J., Zhang, D., et al.: VRML programming training tutorial, 1st edn. Tsinghua University Press, Beijing Jiaotong University Press, Beijing (2008)
[4] Yang, T.: Virtual reality technology and its application in modern medicine. Information and Control 32(3), 251–255 (2003)
[5] Luo, Y.: Roaming design principle. Computer Skills and Knowledge 4(7), 2020–2021 (2008)
[6] Li, J., Chen, Y.-H., Fei, K.Y., et al.: Cellular network courseware design and development. China Medical Education Technology 19(2), 144–146 (2005)

ECC-Based Image Encryption Using Code Computing

Zhongjian Zhao[1] and Xiaoqiang Zhang[2]

[1] North China University of Water Conservancy and Electric Power
Zhengzhou, China
`26202939@qq.com`
[2] State Key Lab of Software Development Environment
Beijing University of Aeronautics and Astronautics, Beijing, China
`grayqiang@163.com`

Abstract. To ensure the security of image interaction, a new image encryption scheme based on code computing and elliptic curve cryptosystem (ECC) is presented. First, a code matrix A is obtained by encoding the original image. Second, a new code matrix C can be obtained by performing the code addition operation among the interval elements of A. Third, divide C into some datum packages with equal length. Each package adds the secret key encoded by codes. Fourth, transform these packages into a code matrix D and decode it, and then we can get the encrypted image. Finally, the secret key is encrypted by ECC. The process of decryption is an inverse process of encryption. The experimental results and security analyses show that our scheme cannot only achieve good encryption, but also resist the exhaustive, statistical and differential attacks.

Keywords: Image encryption, elliptic curve cryptosystem (ECC), digital envelop.

1 Introduction

Nowadays, digital image are used in many fields. To ensure the security of image interaction, people pay attention to the research of image encryption and obtain fruitful achievements, such as image encryption algorithms based on ECC (Elliptic Curve Cryptosystem) [1], chaotic theory [2], Arnold transformation [3]. ECC possesses high security at present. However, the complexity and low efficiency of ECC limits the quantity of encryption datum. Generally speaking, the low-dimension chaos possesses a few parameters and initial values, which lead to a small key space. Therefore, people always utilize the high-dimension chaos to encrypt images nowadays. Although Arnold transformation is simplicity and easy realization, for its small period, the attacker can obtain the plaintext with finite-times iterations.

In this paper, a new image encryption algorithm is proposed, which uses code computing combining with ECC. The experiment result shows that this algorithm possesses a large key space and can resist the exhaustive, statistical and differential attacks.

The rest of the paper is organized as follows. Section 2 describes basic theory of the proposed algorithm, ECC and code computing. Section 3 designs the detailed steps of new algorithm. Section 4 offers a reference experiment of the proposed algorithm. Algorithm secure analyses are discussed in Section 5. Section 6 concludes the paper.

2   Basic Theory of the Proposed Algorithm

(1) ECC
For the difficulty of elliptic curve discrete logarithm problem, Koblitz and Miller proposed ECC in 1985 [4, 5]. The research indicates that the security of 160 bit's (210 bit's) elliptic curve key equals to the security of 1024 bit's (2048 bit's) RSA key [6]. For applications to cryptography, we consider finite fields of p element, i.e., $F_p = \{0, 1, 2, \cdots, p-1\}$. An elliptic curve $E_p(a, b)$ is the set of solutions (x, y) to an equation $y^2 = x^3 + ax + b \pmod{p}$, together with an infinity point O, where p is a prime and more than 3, $a, b \in F_p$ and satisfy $4a^3 + 27b^2 \neq 0 \pmod{p}$ [7].

(2) Code Computing
The codes A, B, C, D denote 00, 01, 10, 11 respectively. Therefore, the pixel gray value can be expressed with these codes. E.g., the binary of the pixel gray value 235 is 11101011, so the corresponding code sequence is DCCD. Conversely, the code sequence ACDB can be decoded as a binary sequence 00101101, i.e., the pixel value is 45.

Two operations for new codes A, B, C, D are defined, as shown in Tab.1 and Tab.2. Code subtraction is the reverse operation of code addition. In this paper, these efficient operations are utilized to scramble the pixel gray values.

Table 1. Code addition operation

+	A	B	C	D
A	B	C	D	A
B	C	D	A	B
C	D	A	B	C
D	A	B	C	D

Table 2. Code subtraction operation

-	A	B	C	D
A	D	C	B	A
B	A	D	C	B
C	B	A	D	C
D	C	B	A	D

3   Achievement Steps and Core Technology

Information interaction on the network platform concerns both Bob (the receiver) and Alice (the sender). The proposed image encryption algorithm uses code computing combining with ECC. The algorithm steps are as follows.

(1) Alice's Encryption Steps

1) Let $I_{m\times n}$ be the interactive image. Alice encodes all the pixels of the original image with codes, and then she can obtain a code matrix $A = [a_{i,j}]_{m\times n}$;

2) Using code addition operation, compute $b_{1,1} = a_{m,n} + a_{1,1}$, $b_{1,2} = b_{1,1} + a_{1,2}$, $\cdots$, $b_{1,n} = b_{1,n-1} + a_{1,n}$; $b_{2,1} = b_{1,n} + a_{2,1}$, $\cdots$, $b_{2,n} = b_{2,n-1} + a_{2,n}$; $\cdots$; $b_{m,1} = b_{m-1,n} + a_{m,1}$, $\cdots$, $b_{m,n} = b_{m,n-1} + a_{m,n}$, where $a_{i,j} \in A$. She can obtain a new code matrix $B = [b_{i,j}]_{m\times n}$;

3) Repeating the operation of Step *2)* for the matrix B , she can get another code matrix $C = [c_{i,j}]_{m\times n}$;

4) Randomly choose a key K , whose length equals $l \times 8$ bits, $l \in Z^+$;

5) Encoding K with codes, she can get a code key K' ;

6) According to the order of pixels, transform C into small datum packages $\{P_1, P_2, \cdots\}$, whose lengths are $l \times 8$ bits. If there are not enough elements for the last package, Alice can add some codes randomly;

7) Using code addition operation, compute $K' + P_i$, $i = 1, 2, \cdots$. She can obtain new datum packages $\{P_1', P_2', \cdots\}$;

8) According to the size of original image $m \times n$ and the order of datum packages, transform $\{P_1', P_2', \cdots\}$ into a new code matrix $D = [d_{i,j}]_{m\times n}$;

9) After decode D , she can get an image I' , i.e., encrypted image;

10) To enhance the security, she encrypts the key K with ECC, and sends the cipher text and the encrypted image I' to Bob on Internet.

(2) Bob's Decryption Steps

The process of image decryption is an inverse process of encryption. After Bob receives the cipher text and the encrypted image I' , he can obtain secret key with ECC decryption. To decrypt the encrypted image according to inverse operations of the encryption steps, code addition operation is replaced by code subtraction in Steps *2), 3)* and *7)*. Other steps are almost unchanged.

4   Certification Experiment

To verify the correctness and security of our algorithm, we perform an experiment with Matlab 7.0. The detailed description is as follows.

(1) Bob's Key-generating Steps

To encrypt the key K with ECC, Bob should generate his private and public keys. The detailed steps are as follows.

1) Choose the 192 bits elliptic curve, which is recommended by National Institute of Standards and Technology (NIST) [6]. The parameters of this elliptic curve are $a = -3$, $b = $ 2455155546008943817740293915197451784769108058161191238065, and $p = $ 6277101735386680763835789423207666416083908700

390324961279. The base point is N = (6020462823756886567582134805875261119166989766368846848818, 174050332293622031404857552280219410364023488927386650641). The period of N is T =627710173538668076383578942317605901376719477318284228481;

2) Randomly choose private key d = 144733012205131907263555417602725731090381560869852245600077. Compute $Q = dN$ = (361201717529764029683746716116484914867980349004855967056909, 20807115023575731694783631186889057351897540912667620333755) ;

3) Publish his public key $[a, b, p, N, T, Q]$.

(2) Alice's Encryption Steps

1) We use the standard gray image Lena as the original image $I_{512 \times 512}$, as shown in Fig.1. Alice encodes all the pixels of $I_{512 \times 512}$ with codes, and then she can obtain a code matrix $A = [a_{i,j}]_{512 \times 512}$. E.g., $I_{1,1} = 161 = 10100001_2$, so $a_{1,1} = \text{CCAG}$;

2) Computing $b_{1,1} = a_{512,512} + a_{1,1}$ = BCDB + CCAB =ABAD, $b_{1,2} = b_{1,1} + a_{1,2}$ = ABAD + CCAC= DABC, $\cdots$, $b_{512,512} = b_{512,511} + a_{512,512}$ = DACB + BCDB=BDCD, where $a_{i,j} \in A$. She can obtain a new code matrix $B = [b_{i,j}]_{512 \times 512}$;

3) Repeating the operations of Step 2) for B, she can get another code matrix $C = [c_{i,j}]_{512 \times 512}$;

4) Let $l = 16$. Therefore, the key K length is 16×8 = 128 bits. Choose K =11110011 00111011 10011011 01111100 11100100 11000011 01110100 00000100 11010010 01110001 10011101 11001010 11101011 10111100 00101101 01100111 $_2$ =323311902158967700380 956580 802144775527;

5) Encoding K with codes, she can get a code key K' =DDAD ADCD CBCD BDDA DCBA DAAD BDBA AABA DBAC BDAB CBDB DACC DCCD CDDA ACDB BCBD;

6) According to the order of pixels, transform C into small datum packages $\{P_1, P_2, \cdots\}$, whose lengths are 128 bits. E.g., $P_1 = c_{1,1}, c_{1,2}, \cdots, c_{1,16}$ = ADBA DABC $\cdots$ DACA;

7) Using code addition operation, compute $K' + P_i$, $i = 1, 2, \cdots$. She can obtain new datum packages $\{P'_1, P'_2, \cdots\}$. E.g., P'_1 = ADCA AAAC $\cdots$ BDAA;

8) According to the size of original image 512×512 and the order of datum packages, transform $\{P'_1, P'_2, \cdots\}$ into a new code matrix $D = [d_{i,j}]_{512 \times 512}$;

9) After decode D, she can get the encrypted image I', as shown in Fig.2;

10) To enhance the security, she encrypts the key K = 323311902158967700 38095658080802144775527 with ECC. The cipher text is 73354839578360345425 858501970345303836886548634598304312 and related data $C1$ = (5335404 1684176503351069798245934075865964825268186681531265, 4862228052154 472335257606367320535670708455628941591992995). Finally, she sends the cipher text, $C1$ and the encrypted image I' to Bob on Internet.

(3) Bob's Decryption Steps

After Bob receives the cipher text, $C1$ and the encrypted image I', he can obtain the key K =32331190215896770038095658080214477557 with his private key $d =$ 144733012205131907263555417602725731090381560869852 2456077. The process of image decryption is almost an inverse process of encryption. Finally, the decrypted image is as shown in Fig.3.
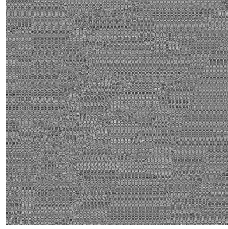
Fig. 1. Lena   **Fig. 2.** Encrypted image   **Fig. 3.** Decrypted image
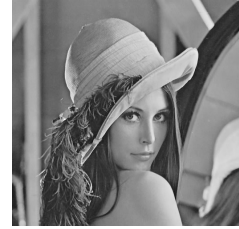
5   Algorithm Analyses

In the view of cryptography, an effective encryption algorithm should have desirable features for withstanding all kinds of known attacks, such as exhaustive, statistical and differential attacks, etc [8]. The security analyses of the new algorithm are in detail as follows.

(1) Resistance to Exhaustive Attack

   1) Key space analysis: The key K length is 128 bits in the experiment, i.e., the key space is 2^{128}. Meanwhile, we can easily enlarge the key space by increasing the value of l in the proposed algorithm. Therefore, the key space is large enough to resist exhaustive attack.

   2) ECC analysis: ECC is based on a difficult mathematic problem, i.e., ECDLP, and there isn't an effectively deciphered method at present. The best attack on ECDLP is the parallelized Pollard rho method [6].

(2) Resistance to Statistical Attack

Fig.4 and Fig.5 show the gray histograms of the original image and encrypted image, respectively. Comparing these histograms, we find that the pixel values of the original image are concentrated on some values, but the histogram of the encrypted image is very uniform, which makes the statistical attack difficult.
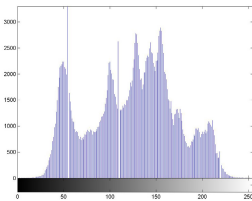
Fig. 4. The histogram of original image   **Fig. 5.** The histogram of encrypted image
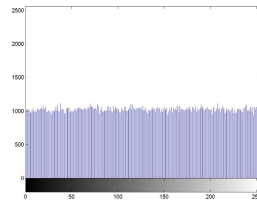
(3) Resistance to Differential Attack

Attackers often make a slight change to the original image, and use the proposed algorithm to encrypt for the original image before and after changing, through comparing two encrypted image to find out the relationship between the original image and the encrypted image. It is called *differential attack* [9]. Researchers usually utilize Number of Pixels Change Rate (NPCR) and Unified Average Changing Intensity (UACI) as two criterions to examine the performance of resisting differential attack. Their equations are as follows.

Let $I'_{m \times n}$, $I''_{m \times n}$ be the encrypted images, which correspond the original image and the original image with a slight change, respectively.

$$f(i, j) = \begin{cases} 0 & I'(i, j) = I''(i, j) \\ 1 & I'(i, j) \neq I''(i, j) \end{cases} \tag{1}$$

$$NPCR = \frac{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} f(i, j)}{m \times n} \times 100\% \tag{2}$$

$$UACI = \frac{\sum\limits_{i=1}^{m} \sum\limits_{j=1}^{n} |I'(i, j) - I''(i, j)|}{255 \times m \times n} \times 100\% \tag{3}$$

Let I be the original image. The pixel value $I(80, 40)$ is 168, and we change it to 200. According to Equations (2) and (3), we obtained NPCR=99.62% and UACI=33.68%. These results demonstrate that our algorithm possesses a strong ability to resist differential attack.

6   Summaries

By defining the code combining, we proposed a new image encryption algorithm using code combining with ECC. The experimental result indicates that the pixel gray values of the original image are completely scrambled by code addition operation. And the secret key is encrypted by ECC. The security analyses show that our algorithm possesses a large key space and can resist the most known attacks, such as exhaustive, statistical and differential attacks. Therefore, our algorithm is a good candidate for the security of image interaction.

References

[1] Zhu, G., Zhang, X.: Mixed image element encryption algorithm based on an elliptic curve cryptosystem. Journal of Electronic Imaging 17(2), 023007.1-5 (2008)

[2] Sahar, M., Eftekhari-Moghadam, Masud, A.: Color image encryption based on coupled nonlinear chaotic map. Chaos, Solitons and Fractals 42(3), 1745–1754 (2009)

[3] Zhenwei, S., Honge, R., Jian, Z.: A block location scrambling algorithm of digital image based on Arnold transformation. In: Proceedings of the 9th International Conference for Young Computer Scientists, pp. 2942–2947. Inst. of Elec. and Elec. Eng. Computer Society, Shanghai (2008)

[4] Miller, V.S.: Use of Elliptic Curves in Cryptography. In: Williams, H.C. (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 417–426. Springer, Heidelberg (1986)

[5] Koblitz, N.: Elliptic curve cryptosystems. Mathematics of Computation 48(17), 203–209 (1987)

[6] Zhu, Y., Zhang, Y.: Introduction to elliptic curve cryptosystem, vol. 10, p. 130. Science Press, Beijing (2006) (in Chinese)

[7] Stinson, D.R.: Cryptography theory and practice, 3rd edn., pp. 257–258. Chapman & Hall/CRC Press Taylor & Francis Group, London (2006)

[8] Peng, J., Jin, S., Liu, Y., et al.: A novel scheme for image encryption based on piecewise linear chaotic map. In: Proceedings of 2008 IEEE International Conference on Cybernetics and Intelligent Systems, pp. 1012–1016. Inst. of Elec. and Elec. Eng. Computer Society, Shanghai (2008)

[9] Zhang, Q., Guo, L., Wei, X.: Image encryption using DNA addition combining with chaotic maps. Mathematical and Computer Modelling 52, 2028–2035 (2010)

Biographies

Zhongjian Zhao. He is a teacher at North China University of Water Conservancy and Electric Power. His current research interests include finite element analysis, information security.

Xiaoqiang Zhang. He is a doctor candidate at State Key Lab of Software Development Environment, Beijing University of Aeronautics and Astronautics. His research interests include image encryption, video encryption, cryptology theory, and software engineering.

A Tamper-Resistant Authentication Scheme on Digital Image

Guangqi Liu, Xiaoshi Zheng, Yanling Zhao, and Na Li

Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center, Jinan, China
`liuguangqi@keylab.net`

Abstract. A digital image anti-tampering authentication scheme based on semi-fragile watermarking algorithm is given. The system implements two modes of embedding for single image and batch image. In addition, it realizes anti-tampering authentication function. Experimental results show that the algorithm has strong robustness to JPEG compression, and has high sensitivity to copy, cut, replace, and erase malicious tampering. The system not only can provide an effective technical means for copyright protection of the digital image but also can be used widely in certification authority with judicial effectiveness.

Keywords: Digital watermark, Copyright protection, Tampering authentication.

1 Introduction

With the continuous development of network technology and PS technology, it's easier for people to get more digital image resources from the network and do some modification according to their own needs. Some of them carry out the PS to meet the interests of some people, while there are some criminals who use high PS technology to tamper the image resources that they got to make it used as evidence of judicial perjury, which will bring great disadvantage to those copyright owners of image.

As an effective means of digital copyright protection, digital watermarking technology can be used as owner identification, ownership identification, verification of ownership, operation tracking, content authentication, copy control, etc. [1-2]. Scholars have proposed algorithms about the semi-fragile digital watermarking technology for digital security [3-5]. Based on this idea, this article puts forward a tamper-resistant authentication system design for digital image, using semi-fragile digital watermarking technology to protect digital image, and can effectively detect and mark the location of tampering when tampering of the digital image occurs.

2 System Design

In this paper, tamper-resistant digital image authentication system consists of the following six modules: (1) adaptive digital watermark generation module; (2) Jpeg image codec module; (3) semi-fragile watermarking module; (4) Watermark Detection Module; (5) tamper-resistant authentication module; (6) System Setup module, including the jpeg image compression factor is set, a single image or batch image processing settings, watermarking algorithm parameter settings, file path settings. Design block diagram shown in Fig.1.

Digital image Tamper-resistant
authentication system
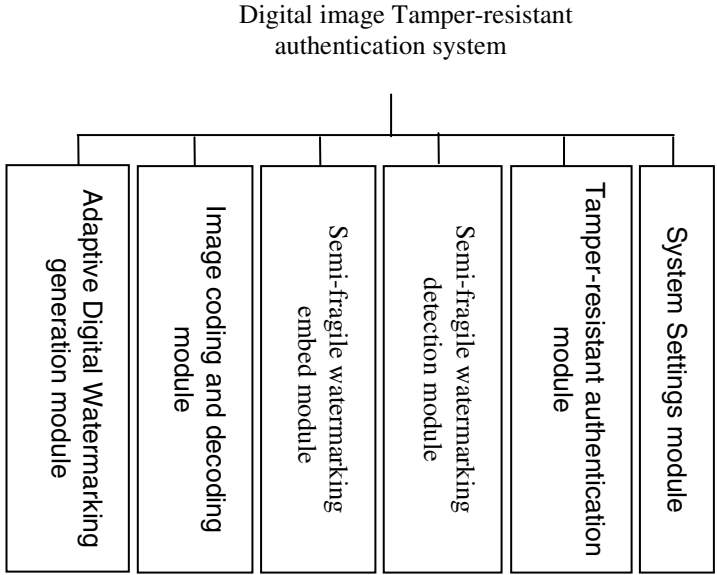
Fig. 1. System framework

2.1   Semi-fragile Digital Watermarking Algorithm

It has been verified in [6] that watermarking has been used by DC coefficient in DCT domain. In this article, we use 8×8 quantization algorithm block to make a binary watermark embedded into the DC coefficient in DCT domain of the Jpeg image, so that any 8×8-bit of the Jpeg image is embedded a watermark. The algorithm flow chart is shown in Fig.2, and the specific algorithm [7]. This algorithm has certain tolerance to the lossy compression, when excessive compression occurs, there is a slight visual deterioration; meanwhile, this algorithm can provide the space of changed region and location information related to the frequency domain, can detect a variety of malicious tampering on Jpeg watermark images, such as replacing, inserting, deleting, copying, etc.

2.2   Tamper Authentication Algorithm

Firstly, get the size of the desired extracted watermark according to the size of being detected Jpeg image; Secondly, decode the watermarked Jpeg image, and do the 8 * 8 block DCT transformation, get watermark by using the DC coefficient quantization algorithm; Then, compare the extracted digital watermark information and digital watermark information source to obtain the changed location of watermark information; Finally, according to the coordinates of the altered positions of digital watermark information which is obtained through the previous step, which can be mapped to the watermarked image to be detected, so that the altered position can be marked. Mapping rules are as follows. Tampering authentication algorithm flow chart shown in Fig.3.

Mapping rules have shown in the formula:

$$\begin{cases} x' = 8 \times x \\ y' = 8 \times y \end{cases}$$

(x, y) is coordinates location of source watermark, (x', y') is corresponding coordinates location of detected image.
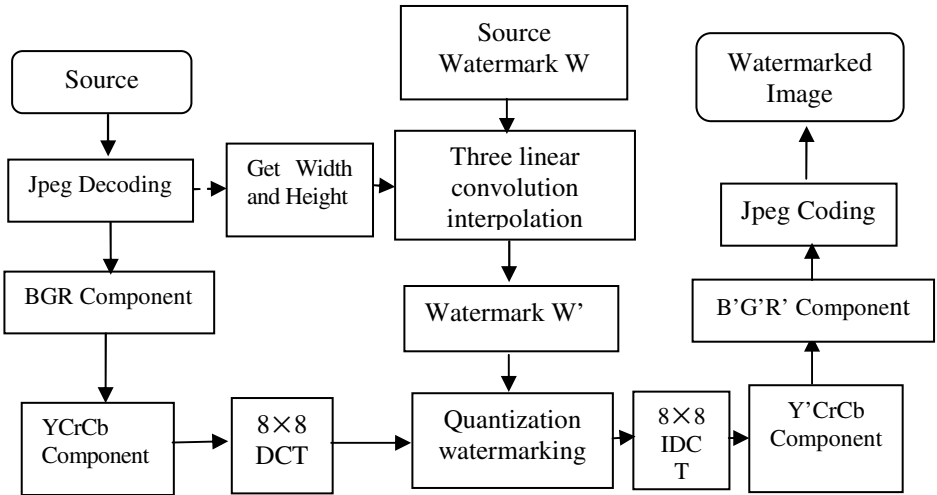
Fig. 2. Watermark flow of Jpeg image
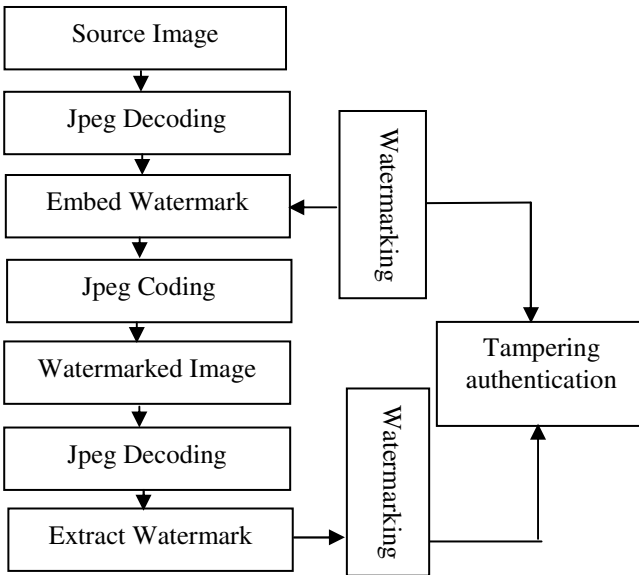
Fig. 3. Flow of tampering authentication

3   Experimental Analysis

According to the proposed semi-fragile digital watermarking algorithm and tampering authentication algorithm. We carried out JPEG compression robustness testing and tampering authentication testing. The original JPEG image of 3024×2016 is shown in Fig.4 (a), the watermark image of 273×302 is shown in Fig.4 (b), adaptive generation watermark image of 378×252 is shown in Fig.4 (c).

（a）   （b）   （c）
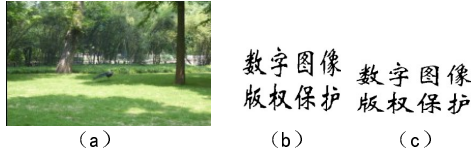
Fig. 4. Test image (a) Source image, (b) Watermark image, (c) Adaptive watermark image

3.1   JPEG Compression Robustness Test

In order to adapt to the network transmission system or some system environmental requirements, digital images often need to be compressed. The design of the proposed

Table 1. Robustness test of Jpeg compression

Compression quality	Watermarked image	Watermark image
Q = 100		
Q = 75		
Q = 50		
Q = 40		
Q = 30		
Q = 25		

scheme was robust anti-Jpeg compression test, respectively testing the quality factor Q for the 100, 75, 50, 40, 30 and 25. The compression test results were shown in Tab.1.

Through the analysis of the Jpeg compression test results by using the algorithm derived from Tab.1, the Jpeg compression algorithm has a strong anti-attack capability, thus still can ensure to effectively detect the digital watermark marks when deeply compress the image which has copyright protection.

3.2   Tampering Authentication Test

The algorithm not only has some robustness for JPEG compression, but also the watermarked image must have sensitivity to malicious tampering during transmission and distribution channels. We carried out tampering authentication testing, such as copy, cut, replace and erase attack. We select JPEG compression quality for 50 in the test, the results of the test is shown in Tab.2.

Table 2. Result of tamper authentication test

It is easy to draw from Tab.2, the algorithm to malicious attacks such as copy, cut, replace, erase and so on are highly sensitive. The system can accurately mark the malicious tampered image regions to provide reliable and effective authentication conclusion.

4   Summaries

In this paper, a tamper-resistant digital image authentication system is designed, which based on semi-fragile digital watermarking algorithm of DCT domain. The

watermarking can be detected without any of the original information and the tamper authentication location can be marked in the system. At the same time, the program has a strong robustness to Jpeg compression and it is sensitive to malicious tampering. The system provides an effective means for digital image copyright protection and has broad application prospects.

References

[1] Lin, C.Y., Chang, S.F.: Semi2fragile watermarking for authenticating JPEG visual Content. In: Proceedings of SPIE: The International Society for Optical Engineering, pp. 140–151 (2000)

[2] Winne, D.A., Knowles, H.D., Bull, D.R., et al.: Compression compatible digital watermark algorithm for authenticity verification and localization. In: Proceedings of SPIE: The International Society for Optical Engineering, pp. 357–367 (2002)

[3] Ding, K., He, C., Wang, H.X.: A chaotic fragile watermarking technique with precise localization. Acta Electronica Sinica 32(6), 1009–1012 (2004) (in Chinese)

[4] Zhang, X.H., Meng, H.S., Liu, F., et al.: A new kind of efficient fragile watermarking technique. Acta Electronica Sinica 32(1), 114–117 (2004) (in Chinese)

[5] Li, Z.H., Hou, J.J.: DCT2 domain fragile watermarking algorithm based on logistic Maps. Acta Electronica Sinica 34(12), 2134–2137 (2006) (in Chinese)

[6] Zheng, X., Zhao, Y., Li, N., Liu, G., Zhou, W.: Research of Synchronization Robustness in Video Digital Watermarking. In: Proceedings-ISECS International Colloquium on Computing, Communication, Control, and Management, CCCM 2008, vol. 1, pp. 276–279 (2008)

[7] Liu, G.-Q., Zheng, X.-S., Zhao, Y.-L.: A Tampering Authentication Algorithm for JPEG Image Based on Adaptive Digital Watermarking. In: FCC 2010, vol. 3, pp. 1–4 (2010)

Research on WLAN Planning System Based on Field Strength Measurement

Ying Li and Yuzhi Wu

Jiao Zuo University, Jiaozuo, China
{liying19771224,wuyuzhi70}@163.com

Abstract. This paper built up a wireless sensor network to do real-time WLAN field strength measurement. Then a propagation model is built using Mat lab based on the measurement result. When optimizing access points, genetic algorithm is used. The map coloring principle is used to solve the channel distribution problem. Our research found the balance between the two large scale WLAN deployment schemas. Firstly, we build up a propagation model for the indoor environment by a little site survey, i.e. field strength measurement. Based on that propagation model, then we optimized the access points locations and amount and their channels to maximize the signal coverage and minimize the signal interference.

Keywords: Wireless LAN, Field Strength Measurement, Propagation Model.

1   Introduction

Wireless local area networks specified in IEEE 802.11 standard has emerged as flexible communication systems, which widely apply indoors. For WLAN deployment, some important network parameters for the effective data transmission, for example, sufficient coverage and interference should be considered. When planning a WLAN network in small region indoors, the access points locations and their channels are generally decided by experience of planners and optimized then again and again by adjustments. For large scale WLAN deployment, there are two ways: site survey and software-aided tools had built-in wireless signal propagation models. It will cost too much labor workforce and time to do site survey, and because the existing deterministic model or empirical model used by software-aided tools is not precise enough, so computing results can not come up to expectation.

Wireless LAN is being given widespread concern since its birth; and IEEE 802.11 agreement is developing rapidly under the push of the market and the academic power. Compared with the speed of technological innovation, the application of wireless LAN is also developing rapidly, with various application occasions .Usually, Wireless LAN exists as broadband public access service, and in a environment deployed with Wireless LAN, the user can access the Internet or data sharing if they have notebook.

The use of wireless LAN is a process with disposition first and then use. It's mainly used in indoor environment. The transmission of wireless signals in indoor environment and outdoor environment is very different. Because of the complexity of

indoor environment, the transmission of wireless signals will occur multi - path – propagation and multi - path - reflection because of the obstacles blocking ,and these variables makes the strength of wireless signal and quality of signal unpredictable, which affects the application of wireless LAN.

2   Designs of Overall Project

2.1   Design Goal

The design goal of this article is to reduce human involvement, improve work efficiency, use the advantages of model prediction ways in effective way, solve problems of rapid deployment of indoor WLAN, and guarantee the good performance of network.

For large-scale deployment of WLAN network, network parameters such as signal coverage and interference will directly influence the performance of WLAN. The coverage of signal is decided by the number, location and transmit power of AP , antenna type and the spread of WLAN signal in indoor places. The disturbance of signal is decided by the channel of AP with overlapped signals. Wile these important network parameters directly depended on signal strength. There are two ways of obtaining signal strength, one is to get through measurement, and the other is to predict through propagation model. However, by measuring you can only obtain signal strength of limited position, and you can get signal strength of any point by predicting through propagation model. So this is the technical difficulties of this article: (1) the establishment system of WLAN indoor propagation model requires WLAN indoor propagation model has characteristics of less difficulty, small amount of calculation, and high accuracy. While the existing determinism model and empiricism model can't satisfy the above requirements. Therefore, need to establish WLAN indoor propagation model. (2)Optimized deployment of the number and position of AP.

2.2   Overall Design of the System

1)   Scenarios of System
In indoor environment, deploying wireless sensor network; sensor node is in charge of field of the access point running within the scope of its sensibility and of its neighboring nodes; send measured data to the gateway through multi-hop, and gateway will forward data to the server. Server is responsible for the control of operation of wireless sensor network and data collection, and determines parameters of propagation model after processing collected data, then carry out WLAN network planning based on this model.

2)   The Architecture of the System
System structure is shown in fig.1, it includes three modules, top-down: management module, gateway module and probe module. Below is a brief description of the three modules.

Management module: Management module is running on server, including components such as the user interface, controller, AP optimization, channel allocation, and document. User interface provides friendly user interface that enable the user to

easily control field measurement process and display measuring results in a dynamic way. Controller is responsible for send control commands from the user to wireless sensor network (including the establishment and delete of routing, and functions of system, namely measure field of neighboring nodes or of visual points), and is responsible for data collection and data in files.

Gateway module: responsible for send control commands from management module to sensor nodes of wireless sensor network. At the same time, forward the measured data from sensor to management module.

Probe module: Probe module, which is running on sensor node, is responsible for sense of field, package transfer or forwarding of data, and data being sent to the gateway node by means of multi-hop.
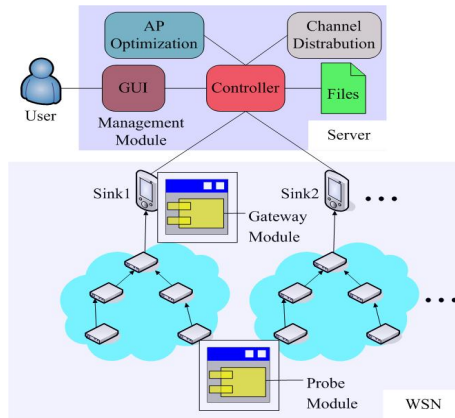
Fig. 1. System framework

3 Reallization of System

3.1 Realization of Management Module

The goal of management module is to control operation of wireless sensor network and data gathering and processing, and specific requirements are as follows:

Provide user with friendly operation interface. The operation that system supports are: modeling of network scene of wireless sensor network and configurations of location information of each node; establishment and removal of topology structure of wireless sensor network; send commands to wireless sensor network, and control the start and stop of field measurement ; dynamic display of field measurement data.

WLAN network planning According to the collected measured data, determine parameters of the propagation model of the measured indoor environment, carry out optimization of number and position of AP, then distribute channels.

In order to satisfy this module, supports of communication components, control components, and WLAN optimized deployment components are needed; component diagram of module is as fig.2 shows:
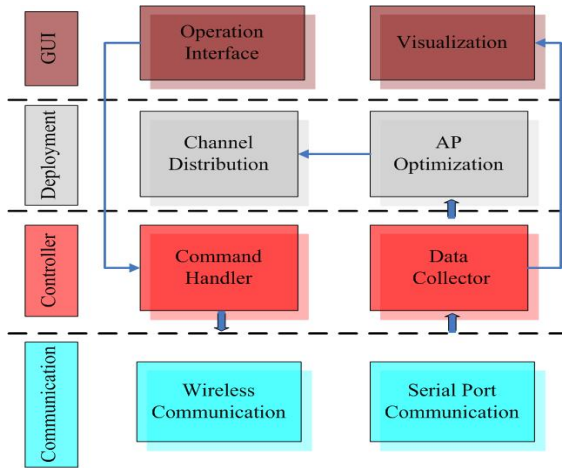
Fig. 2. Management Component module diagram

3.2   Realization of Gateway Module

Gateway module is in charge of forwarding commands from management module to each node of wireless sensor network, and at the same time can send field data from each node to the management module for further processing. Therefore, gateway is a bridge between server and sensor nodes. Gateway module and management module can communicate through wireless signals or serial, and gateway module and probe module in the nodes can only communicate through wireless signals. The main components of gateway module are USB Driver and USB host management.

USB driver：Since the two hardware equipment AT90USB1287 and RT2571WF communicate through USB interface, and as control equipment AT90USB1287 should have the ability to drive USB devices. USB Driver module drives USB bus to work, provides equipment mounted, bus control, and data transceiver, etc.

USB host management: responsible for the control of RT2571WF bottom register, read various parameters of EEPROM in RT2571WF, and properly initialized RT2571WF to control wireless card.

Gateway module is responsible for transmission of two types of data: Command and Measurement Result. Command comes from the management module, and Measurement Result comes from the sensor node.

3.3   Field Measurement

At first, management module send AP field measurement command to wireless sensor network, and the command contains parameters such as SSID of pending AP and measuring interval. After nodes receiving this command, set pending SSID and measuring interval according to parameters of the command, and node into field measurement state. The node start timer, and length of timer Set for measuring interval. When the timer expires, discontinuity will occur, and the processing of this discontinuity is as follows: broadcast Probe request message. When one or more AP of

the area of receives Probe request message, each AP will broadcast Probe request message. After node discontinuity, if receive Probe request message, then compare SSID field values in the message with pending SSID, if the same convert RSSI of this Probe response message into power, and send to the next hop node,; and in the end send it to server through gateway. Then reboot the timer, and repeat the process.

4   Perimentand Analysis of System

The main experimental process of this system is field measurement. Using field data measured by this system to verify propagation model with pending parameters.

This scene deployed 5 sensor nodes, a gateway node, a server, and two AP; and the SSID of AP are ts and Siemens, respectively. After system starts operating, import scene graph, move all nodes in real scene and gateway from the list of nodes to the scene graph, and set location information of nodes and gateway. Then add AP and set SSID of AP according to the actual scene. After finishing modeling process click the green button to send orders generated by topology. Then send AP field measurement command or neighboring nodes field measurement command to the network.

In the validation process of propagation model with pending parameters, dividing the selected area into six blocks in the modeling area of system. Place a node in each block to measure field of neighboring nodes. Because the transmission power of each node is not the same, thus field measurement value of two nodes has no symmetry. Then, using this transmission model carry out forecast and measurement of the field, and compare the forecast results with the measured results, which are shown in fig.3:
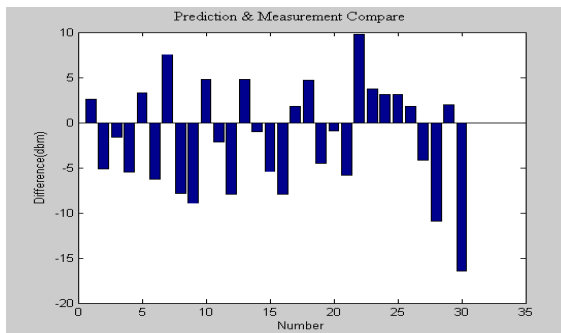
Fig. 3. Field measure comparison results

It can be seen from chart 3that the average error between use of predictive values and measured values of propagation model is less than 10db.Therefore; the accuracy of this propagation model is acceptable.

5   Conclusions

The main content of this paper is WLAN network planning based on field measurement. The paper mainly completes the following important aspects of work:

the establishment of indoor WLAN propagation model, measurement and transmission of the field, dynamic display of the field, and the optimization and deployment of AP. System uses a genetic algorithm to carry out   optimization and deployment of AP; in fact, parallel genetic algorithm can greatly improve the optimal speed, and there are many other optimization algorithm, such as simulated annealing, and ant colon algorithm,etc. Whether can use the parallel algorithm or a combination of various of optimization algorithm to increase optimal speed and optimization effect is an issue that worth studying. And in the present process of optimization and deployment of AP, only performance factors such as signal coverage and signal interference are considered. Using genetic algorithm makes the coverage reaches more than 90%, with the least number of AP. Achieving distributions of channel, and avoids intruding to improve the network performance. The system that discussed and realized in the thesis has basically satisfied the design requirements.

References

[1] IEEE STD 802.11e, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications. Amendment 8: Medium Access Control (MAC) Quality of Service Enhancements. ANSI/IEEE (2005)

[2] IEEE STD 802.11h, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 5: Spectrum and Transmit Power Management Extensions in the 5 GHz band in Europe[S]. ANSI/IEEE (2003)

[3] IEEE STD 802.11i, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications Amendment 6: Medium Access Control (MAC) Security Enhancements. ANSI/IEEE (2004)

[4] IEEE Std 802.11g, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Amendment 4: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band. ANSI/IEEE (2003)

[5] Islander, M.F., Yun, Z.: Propagation prediction models for wireless communication systems. IEEE Trans. Microw. Tech. 50(3), 662–673 (2002); Young, M.: The Technical Writer's Handbook. University Science, Mill Valley (1989)

An Automatic Video Image Mosaic Algorithm Based on SIFT Feature Matching

Fuhua Song and Bin Lu

China Jiliang University
Hangzhou 310018, China

Abstract. An automatic video image matching algorithm based on SIFT feature points was presented in this paper. The optimized implementation of the best-bin-first(BBF) algorithm based on kd-Tree was used in coarse matching, random sample consensus(RANSAC)was used in accurate matching, the last use the global information are calculated to update the match precision. with pixel brightness weighting in HIS color space was proposed in blending images. The experimental results show that the method with strong robustness performs fast and effectively and has highly valuable in practice.

Keywords: SIFT, feature matching, image fusion, image matching.

1 Introduction

Video is playing an increasingly important role in the expression of information. With the development of information technology, the research of video images developed rapidly, Such as video conferencing, video surveillance, virtual reality and so on. Owing to the limited information single camera bearing, can not satisfy user's personality needs. video image mosaics is an active area of research in photogrammetric fields, computer vision, image processing, and computer graphics[1].

Image mosaicing technique is the image processing areas of a popular research topic. Image mosaic is to combine two or more images into a new one such that the mosaic image is as little distortion from the original images as possible, It solves the contradiction between the vision and resolution. In recent years, image mosaic at home and abroad have done a lot of research, but also made remarkable achievements, for example, David Lowe[2-3] from Columbia University established a novel point characteristics extracting method due to the SIFT (Scale Invariant Feature Transform) algorithm. However, there are some drawbacks of the algorithm, the SIFT operator is very time-consuming computational complexity, which is difficult to meet real-time video image processing requirements. this article has improved SIFT operator. SIFT extracted the extreme points that were detected by harris corner, it removed the extreme point of instability, Aimed at the thresholds of the SIFT affect matching speed and accuracy, threshold early termination of technology and threshold adaptive technology for automatic and fast image matching were proposed. The optimized implementation of the best-bin-first(BBF) algorithm based on kd-Tree was used in coarse matching, random sample consensus(RANSAC)was used in accurate matching. In the end match rate information of the neighbor keypoints and the global information are calculated to

update the match precision, Optimum path best-matched-line combined with pixel brightness weighting in HIS color space was proposed in blending images.

2   The Entire Algorithm of the Automatic Stitching

The entire algorithm mainly includes: Image preprocessing, Image feature extraction, matching feature points, Image position transformation and Image fusion, which can be described as the figure1. In order to ensure consistency and space visual scene in the ties, Image preprocessing is needed. The images is shooted by the camera which is taken in turn from the horizontal direction, then put the image plane mapped to the cylindrical projection.

Image feature extraction is the key of Image mosaic. The image feature extraction algorithm which is stdeay and efficient is widely applied value in image processing. SIFT algorithm is robustness in extracting feature point. But it is also complex and inefficient. On this basis, this article has improved SIFT algorithm. It can extract feature points steady and rapidly.

And then we will match the feature, searching the feature points is very difficult, The time complexity of algorithm is high. Through the use of kd - tree can speed up the search process. The optimized implementation of the best-bin-first (BBF) algorithm based on kd-Tree was used in coarse matching, random sample consensus (RANSAC) was used in accurate matching.
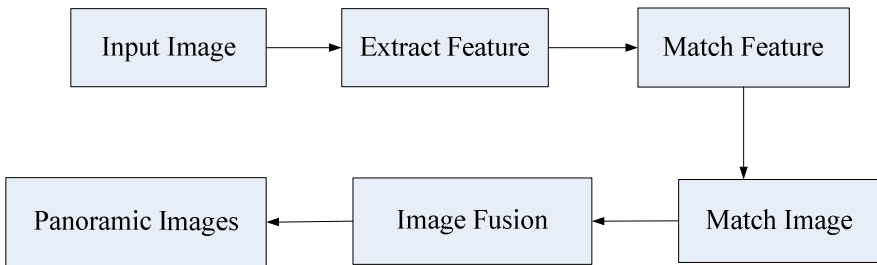
Fig. 1. The entire process of Image mosaicing

3   Extract Sift Features

3.1   Establishment a/Scale Space and Extreme Points Finding

This stage of the filtering attempts to identify those locations and scales those are identifiable from different views of the same object. It has been shown by Koenderink and Lindeberg that under a variety of reasonable assumptions the only possible scale-space kernel is the Gaussian function. The scale space of image is defined as:

$$L(x, y, \sigma) = G(x, y, \sigma) * I(x, y)$$

(1)

where $G(x, y, \sigma)$ is a variable-scale Gaussian, $G(x, y, \sigma) = \dfrac{1}{2\pi\sigma^2} e^{-(x^2+y^2)}/2\sigma^2$, $I(x, y)$ is the input image, and σ is scale space factor. A big size presents general feature of images, and a small size gives detailed feature. In order to effectively detect stable keypoints in the scale space, difference of Gaussian scale space was put forward, which was generated by the convolution of difference of Gaussian nuclear and original image.

$$D(x, y, \sigma) = (G(x, y, k\sigma) - G(x, y, \sigma)) * I(x, y) = L(x, y, k\sigma) - L(x, y, \sigma)$$
(2)

If a point in the DoG scale space and 26 areas of upper and lower layers are the maximum or minimum, the point is believed to be a feature point of an image in this scale.

3.2   Key Point Localization

In this stage a candidate keypoint detected in stage A will be refined to subpixel level and eliminated if found to be unstable.

An extreme value of a poorly-defined DoG operator would have bigger principal curvature on crossing edge, and smaller principal curvature in the direction of vertical edge. The principal curvature is calculated by one 2x2 Hessian matrix (H), as described in the following:

$$H = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$$

Trace and determinant ratios of Hessian matrix is to reduce the edge effect of DOG function.

3.3   Determination of Direction of Key Points

Assign a direction for each feature point based on image property. The feature point descriptor section will use the direction of key points and describe characteristic of feature points.

3.4   Keypoint Descriptor

To ensure rotational invariance, first turn the coordinate axis to the direction of key points, take one key point as the center, choose 8×8 windows, cut the window into 2×2 child windows, and count direction column diagram in each child window, as shown in Figure 3. The direction of each child window is determined by its 4×4 blocks based on above-mentioned computing method for directions. Directions of each key point in the figure are determined by directions of 4 (2x2) seed points. Each seed point has information about 8 directions, and then each key point has 4×8 = 32 dimensions, which are usually described using 16 (4x4) seed points, and such a key point can

generate the 128 data that ultimately form 128-dimensional SIFT feature vector. At this time SIFT feature vector, which eliminates the effect of scale, rotation and other geometric distortion factors, continue to the length of feature vector normalization, to further remove the effect of the light change.
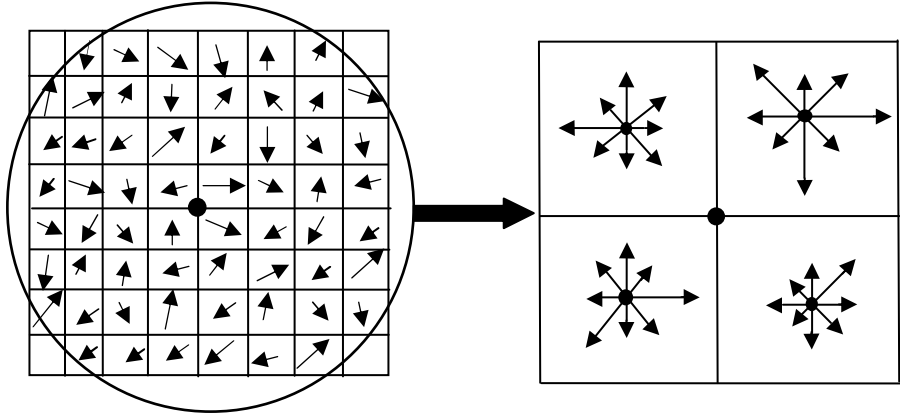
Fig. 2. Feature vector generated by keypoint neighborhood gradient information

4   Improved Matching for Sift

The traditional SIFT algorithm detected the feature points are still many unnecessary characteristics, As shown in image A, B and C, there is overlap between A and B and B and C. The feature points of the white areas of the image B and the feature points of extracting from S2 can not match the feature points of Image A. While the feature points of extracting from S2 can not match the feature points of Image C, the feature points that detected from the blank area of Image B can not match the feature points of Image A and Image B. Therefor there does not exist a correct matching keypoint in the non-overlapping area of the two matching images. In view of this situation, matching operations are performed by making full use of the location of feature points. then judge the keypoints by harris corner, if there is a significant corner, then the retention of the feature points, or removal of the point. The task consists of the following steps:

(1) the keypoints are restricted within overlapped areas.
(2) As frequency of query and kd-Tree query the total number of nodes in the same order of magnitude[8], so you can initialize the kd-Tree for calculating the boundaries of each node hypercube, stored in a predetermined space.
(3) Use queue to achieve the BBF algorithm.
(4) On M obtained matching points on the pseudo-randomly selected 3 pairs of matching points, the linear transformation matrix calculation of H;
(5) Calculated for each pair of match points and pseudo-H the vertical distance d;
(6) Within a distance less than the threshold based on the principle of calculating the value of H of the point and the point in this domain re-estimate H;
(7) Random sample N times, until you get to date the largest collection of interior point; According to the revised estimate of the maximum change in the point set.
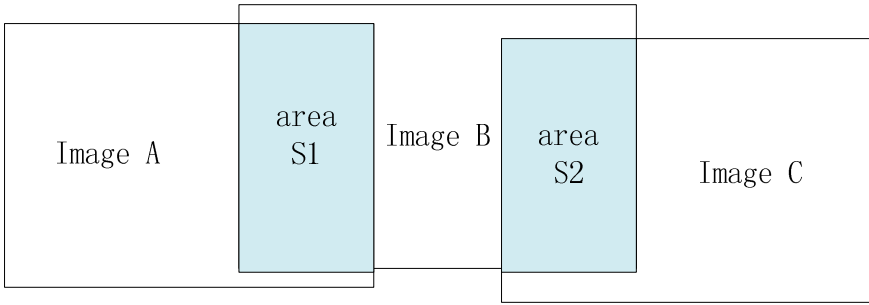
Fig. 3. Image overlap area limits

5   Image Fusion

In order to avoid and remove the ghost produced in Image fusion process and the brightness of different Images, this paper selects the method of combination of the brightness function algorithms.

In image fusion, there are still differences in brightness, in order to achieve satisfactory results splicing, This spatial integration in HIS image. Compared with the traditional weight function fusion algorithm, the brightness of weight function fusion is only to the brightness in brightness space value is weight fusion, It is simple. Experimental results indicated that the proposed method could get good recognition results. The formula from RGB color space to HSV color space is as follows:

$$
\begin{cases}
I = \dfrac{R + G + B}{3} \\[2mm]
H = \dfrac{1}{360}\left[90 - \arctan\left(\dfrac{F}{\sqrt{3}}\right)\right] + \{0, G > B; 180, G < B\} \\[2mm]
S = 1 - \left[\dfrac{\min(R,G,B)}{I}\right]
\end{cases}
\tag{8}
$$

The formula HSV from color space to RGB color space is:

$$
\begin{cases}
R' = \dfrac{I - R}{I - \min(R,G,B)} \\[3mm]
G' = \dfrac{I - G}{I - \min(R,G,B)} \\[3mm]
B' = \dfrac{I - B}{I - \min(R,G,B)}
\end{cases}
\tag{9}
$$

its implementation steps are as follows: First get the width pixels of the best path around 30 pixel, with its RGB value are calculated the I, the weighting function is used for fusing the brightness. The weighting function is as follows:

$$C(x) = \frac{\sum_m w(d(x))I_m(x)}{\sum_m w(d(x))} \tag{10}$$

Where w is monotonic function, $w(x) = x$, $d(x)$ is the distance of image m in the fusion point (x, y) and integration in the x-axis. $I_m(x)$ is the brightness values of Fusion points. the fusion with the brightness and then use the formula (9) is converted to RGB space.

6   Experimental Results

The algorithm has been implemented in VC++6.0based on OpenCV library and experiments are carried on a PC with dual-core cpu 3.0 GHz and memory 2 G.. In order to verify the effectiveness of the proposed matching algorithm, effective pictures are given as follows: Fig.4 (a) is not limited area splicing results, fig.4 (b) is limited area splicing results. extra four features points are removed from fig.4 (b).The brightness of fig.5 (a) and figu.5 (b) is different, It is a good way to solve the brightness difference image fusion with the brightness weight function. Experimental result of five images mosaics is given in fig.7. Table 1 is with the traditional algorithm based on feature matching algorithm calculated SIFT the performance comparison, experimental results show significant improvement
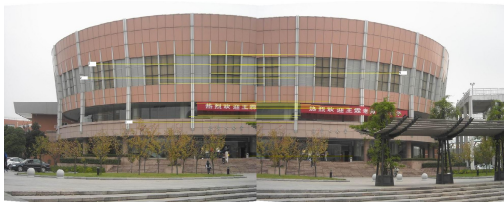
Fig. 4. (a) No limited area matching figure
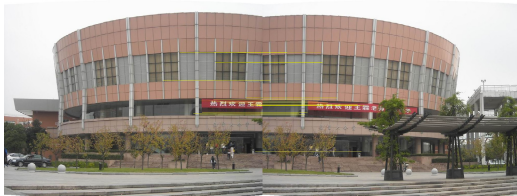
Fig. 4. (b) Limited area matching figure

Fig. 5. (a) Direct weight function integration

Fig. 5. (b) Brightness weight function integration

Fig. 6. (a)   **Fig. 6. (b)**   **Fig. 6. (c)**

Fig. 6. (d)   **Fig. 6. (e)**

Fig. 7. Mosaic image

Table 1. Image Mosaic Result Data

	matching images	feature point extracting time /s	matching time /s	matching rate
traditional method	(a) and (b)	11.23	2.13	48.6%
	(b) and (c)	10.56	1.85	47.7%
	(c) and (d)	8.03	1.52	53.7%
	(d) and (e)	13.32	2.53	48.9%
proposed method	(a) and (b)	7.12	1.56	65.4%
	(b) and (c)	6.33	1.13	67.3%
	(c) and (d)	4.95	0.97	75.8%
	(d) and (e)	6.78	1.77	70.6%

7   Summaries

We have proposed a mosaic algorithm based on SIFT, by limiting the regional feature point extraction and feature point matching, which improved matching speed and accuracy. In the image fusion process, we use the brightness function to solve the brightness of difference images. The experimental results have proved its correctness

References

[1] Szeliski, R.: Image Alignment and Stitching, Technical Report. MSR-TR-2004-92 Redmond, Washington DC, USA: Microsoft Research (2004)
[2] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(1), 91–110 (2004)
[3] Brown, M., Lowe, D.G.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision 74(1), 59--73 (2006)
[4] Mass, H.G., Schneider, D.: A geometric model for linear-array-based terrestrialpanoramic cameras. The Photogrammetric Record 21(115), 198–210 (2006)
[5] Zhang, X., Zhang, X., Ji, F., Yu, Q.: Known Landing Area Rough Locating from Far Distance Based on DDM-SIFT. IEEE Image and Signal Processing 1, 686–690 (2008)
[6] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. Journal of Computer Vision 60, 91–110 (2004)
[7] Ce, Z., Xiao, L., Chau, L.-P.: Hexagon-Based Search Pattern for Fast Block Motion Estimation. IEEE Transactions on Circuits and Systems for Video Technology 12, 349–355 (2002)
[8] Mikolajczy, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision 60(1), 63–86 (2004)

Attribute Reduction Algorithm
Based on Equivalence Matrix

Dai Xue Zhen, Ma Ying Cang, Liu Huan, and Liu Da Zhuo

School of Science, Xi'an Polytechnic University, Xi'an, Shan'xi 710048, P.R. China
daixuezhenqq@126.com

Abstract. The equivalence matrix reduction algorithm is proposed in this paper. First, the equivalence matrix under the relationship between the relation matrix and the equivalence classes are introduced, which can be instead of seeking the explicit equivalence classes in tradition algorithm. Moreover the measure on the attribute importance is established based on this equivalent matrix. Second, the corresponding attribute reduction algorithms on the information systems and decision information system are given. Finally, two corresponding examples are proposed to verify the feasibility and effectiveness of the algorithms.

Keywords: Equivalence matrix, Attribute importance, Attribute reduction.

1 Introduction

Rough set theory is proposed by professor Pawlak on early 80's in the 20th century, which studies incomplete, uncertain knowledge and data presentation, learning, generalization[1]. One important application of rough set theory is knowledge reduction, which purpose is to gain the decision or classification rules of the problem through the knowledge reduction under the condition of insurance the classification unchanged [2]. As rough set method does not need to provide any prior knowledge, and easily understood and applied, in recent years, which has so much attention and be studied extensively by the researchers.

The traditional attribute-reduction algorithms have blind search algorithm, differential matrix method, attribute importance algorithm, information entropy method [3] and some other heuristic search algorithm [4-8]. The idea of this paper stems from the relationship between the relation matrix and the equivalence classes, and use "equivalence matrix" approach instead of seeking explicit equivalence classes.

The paper is organized as follows: After the introduction in this section 1, the corresponding background and theories are summarized. We propose the equivalence matrix and a corresponding measure on the attribute importance in section 2. In section 3 and in section 4, equivalent matrix attribute reduction algorithm of information systems and decision systems are given. And in section 5, two examples are validated the feasibility and the effectiveness of these algorithms in the knowledge representation system. We make a conclusion in last section.

2   Basic Knowledge

2.1   Definition of the Equivalence Relation

In the information system under classical set theory，equivalence relation matrix R can be written in the form of

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}, \text{ which } r_{ij} = \begin{cases} 1 & x_i R x_j \\ 0 & \text{otherwise} \end{cases}$$

Then, the elements of the equivalence class of x_i is $[x_i]_R = \{x_j \mid r_{ij} = 1\}$, the elements number of the equivalence class of x_i is $\left\|[x_i]_R\right\| = \sum_j r_{ij}$, that is sum of the elements of line i in the equivalence matrix. Easy to know that it is the symmetric matrix and all the matrix elements are 0 and 1. Therefore, it needs only half storage space in the actual calculation.

According to the definition of equivalence matrix we can get the following theorems easily:

Theorem 1: In the information system: $IS = (U, C, V, f)$, $\forall \alpha, \beta \in C$, we can derive that $R_{\{\alpha\} \cup \{\beta\}} = \min\{R_{\{\alpha\}}, R_{\{\beta\}}\}$, $\min\{R_{\{\alpha\}}, R_{\{\beta\}}\}$ means the smaller element of the two matrix   corresponding elements.

Theorem 2: In the reduction definition of information systems
$U / ind(C) = U / ind(B)$ equivalent to $R_B = R_C$, which $B \subseteq C$; and C Contains all the attributes.

2.2   Definition of the Attribute Significance in the Equivalence Matrix

For $\forall \alpha \in C - B$, define the attribute significance is $\rho(\alpha, C, B) = 1 - \dfrac{\sigma(\alpha)}{n}$. which $\sigma(\alpha) = \sum_i \sum_j |t_{ij}|$ is   sum   of   all   the   elements   of   the   difference   matrix $T = R_{B \cup \{\alpha\}} - R_C$; and $n = \sum_i \sum_j 1$.

Then we consider the attribute reduction algorithm of the knowledge presentation system.

3   Attribute Reduction Algorithm of Information System Based on Equivalence Matrix

First we think over the reduction algorithm of the information system. The algorithm as following divided two parts: Before seek for the attribute reduction algorithm we count the core.

3.1   For Counting the Core

Input: information system $IS = (U, C, V, f)$.

Output: the $CORE(C)$ of the information system IS

Specific steps are as follows:

Step 1: let $CORE(C) = \Phi$;

Step 2: for $\forall \alpha \in C$, if $R_{C-\{\alpha\}} \neq R_C$,   then $CORE(C) \Leftarrow CORE(C) \cup \{\alpha\}$;

Step 3: after the traversal, output $CORE(C)$, the algorithm end.

3.2   Seek for Concrete Steps of the Attribute Reduction

Input: information system $IS = (U, C, V, f)$.

Output: one attribute reduction $B \in RED(C)$ of the information system IS

Step 1: According to the definition of equivalence matrix to calculate R_C .

Step 2: let $B = \Phi$;

Step 3: calculate $CORE(C)$, and let $B \Leftarrow CORE(C)$;

Step 4: Judge $R_B \overset{?}{=} R_C$, if equivalent, then export B , the algorithm end; else transmit to Step 5;

Step 5: for $\forall \alpha \in C - B$, calculate the attribute importance $\rho(\alpha, C, B)$ of α , set $B \Leftarrow B \cup \{\beta\}$, which $\beta = \max_{\alpha} \{\rho(\alpha, C, B)\}$, transmit to step 4.

Step 6: output B , the algorithm end.

4   Equivalence Matrix Attribute Reduction Algorithm in Decision System

As following we consider the reduction algorithm in decision system.

In the decision information system, if $R_C \leq R_D$,   then the system is consistent, otherwise the system is inconsistent. In this paper we only discuss the consistent system. We can easily get the following theorem.

Theorem 3: In the consistent decision information system ($R_C \leq R_D$), for $\forall \alpha \in C$, If $R_{C-\{\alpha\}} \leq R_D$ does not hold (that is resulting inconsistent after removal the attribute α), then α is the relative core property of the system.

4.1   For Counting the Relative Core of Decision System

Input: decision system $DT = (U, C \cup D, V, f)$.

Output: the relative core of decision system $CORE_C(D)$.

Specific steps are as follows:

Step 1: let $CORE_C(D) = \Phi$;

Step 2: for $\forall \alpha \in C$, calculate $R_{C-\{\alpha\}}$. if $R_{C-\{\alpha\}} \leq R_D$ doesn't hold, then $CORE_C(D) \Leftarrow CORE_C(D) \cup \{\alpha\}$;

Step 3: after the traversal, output $CORE_C(D)$, the algorithm end.

4.2   Seek for Concrete Steps of the Attribute Reduction in Decision System

Input: decision system $DT = (U, C \cup D, V, f)$.

Output: The relative reduction $B((B \subseteq C) \wedge (B \in RED_C(D)))$ of the condition attribute C relative to the decision attribute D.

Step 1: to calculate the R_C、R_D ($R_C \leq R_D$) according to the definition of equivalence matrix;

Step 2: let $B = \Phi$;

Step 3: calculate the $CORE_C(D)$, and $B \Leftarrow CORE_C(D)$;

Step 4: judge the $R_B \leq R_D$, if hold, then output B, the algorithm end; if not hold, transmit on step 5;

Step 5: for $\forall \alpha \in C - B$, calculate attribute property of α ($\rho(\alpha, C, B)$), let $B \Leftarrow B \cup \{\beta\}$, which $\beta = \max_{\alpha} \{\rho(\alpha, C, B)\}$, transmit on step 4;

Step 6: output B, the algorithm end.

5   Examples

Example 1: Tab.1 shows the information system, which $U = \{1, 2, 3, 4, 5\}$, $C = \{a, b, c, d\}$.

Table 1. Information system for example 1

U	a	b	c	d
1	0	1	2	0
2	1	2	0	2
3	1	0	1	0
4	2	1	0	1
5	1	1	0	2

First we count core of the information:

Calculate that $R_C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, It is said that each elements of U is divided into a class.

Then we calculate

$$R_{C-\{a\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = R_C ; \quad R_{C-\{b\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} \neq R_C$$

$$R_{C-\{c\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = R_C ; \quad R_{C-\{d\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = R_C$$

Then we get the core $CORE(C) = \{b\}$.

Second we seek for the attribute reduction:

$$R_{\{b\}\cup\{a\}} = R_{\{b\}\cup\{d\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} = R_C ; R_{\{b\}\cup\{c\}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix} \neq R_C$$

So, the two attribute reductions are $\{a,b\}$ and $\{b,d\}$. It is all the reductions of this information system.

Note: Either in the information system or decision system, we should remove the repeated elements before making reduction algorithm, which can be simplify the calculation.

Example 2: Tab.2 shows the decision system after eliminated the repeated elements. $U = \{1,2,3,4,5\}$, $C = \{a,b,c\}$, $D = \{D\}$.

Table 2. Decision system of example 2

U	a	b	c	D
1	2	1	3	1
2	3	2	1	2
3	2	2	3	2
4	1	1	4	3
5	1	1	2	3

First calculate $R_C = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$, $R_D = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 \end{bmatrix}$, then $R_C < R_D$, it

indicates the decision system is consistent.

Calculate $R_{C-\{a\}}$, $R_{C-\{b\}}$, $R_{C-\{c\}}$, respectively, we get $R_{C-\{a\}} < R_D$,and

$R_{C-\{b\}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} < R_D$ hold, however, $R_{C-\{b\}} = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} < R_D$ not

hold. So $CORE_C(D) = \{b\}$.

As $\{b,a\} = C - \{c\}$, $\{b,c\} = C - \{a\}$, that is $R_{\{b,a\}} < R_D$, $R_{\{b,c\}} < R_D$, so the

reductions of this decision system are $\{b,a\}$ and $\{b,c\}$.

6   Summaries

The paper proposed the equivalence matrix attribute reduction algorithm， and we can see the result of this method and the result of attribute importance algorithm are equal. But the advantages of this method are the ideas simple and clear, easy to operate by computer. Compare with the definition reduction algorithm, we define the equivalence matrix in stead of counting the equivalence classes, it is more convenient and feasible. It is avoiding the comparison of the properties compare to the difference matrix algorithm, and the matrix elements are simple. Furthermore, the calculation is simpler than information entropy algorithm. So this algorithm is more simple and efficient properties reduction method. In the following, we will use this method to the reduction of fuzzy rough sets.

References

[1] Pawlak, Z.: Rough sets. International Journal of Information and Computer Science 11(5), 341–356 (1982)

[2] Wang, G.Y.: Rough Sets Theory and Knowledge Acquisition. Xi'an Jiaotong University Press, Xi'an (2001)

[3] Miao, D.Q., Li, D.G.: Rough Sets Theory, Algorithm and Application. Tsinghua University Press, Beijing (2008)

[4] Miao, D.Q., Hu, G.R.: A heuristic algorithm for reduction of knowledge. Journal of Computer Research and Development 36(6), 681–684 (1999)

[5] Ye, D.Y., Chen, Z.-J.: A New Discernibility Matrix and the Computation of a Core. Acta Electronica Sinica 30(7), 1086–1088 (2002)

[6] Yang, L.Y., Wen, M., Zhuo, Q., Wang, W.Y.: Entropy-based knowledge acquisition approach on fuzzy rough set. Computer Engineering and Applications 44(23), 59–61, 126 (2008)

[7] Wang, X.Z., Ha, Y., Chen, D.G.: On the reduction of fuzzy rough sets. In: Proceeding of the Third International Conference on Machine Learning and Cybernetics, Guangzhou, vol. 18-21, pp. 3175–3178 (2005)

[8] Guan, J.W., Bell, D.A., Guan, Z.: Matrix computation for information systems. Information Sciences 131, 129–156 (2001)

An Overview of Software Architecture Description Language and Evaluation Method

Yang Lingling and Zhao Wei

School of Software Engineering
Beijing JiaoTong University, Beijing, 100044

Abstract. Software architectures shift the focus of developers from lines-of-code to coarser-grained architectural elements and their overall interconnection structure. Architecture description languages (ADLs) have been proposed as modeling notations to support architecture-based development. And a large amount of evaluation methods have been proposed since people concentrated more on the quality of the software. The aim of the software architecture evaluation is to analyzing and finding the potential risks in the architecture .There is, however, little consensus in the research community on what is an ADL's advantage and shortcoming ,what are the features of each evaluation methods, and what are the differences between different ADL/ evaluation methods. This paper attempts to provide an answer to these questions.

Keywords: ADL(Architecture Description Language), AEM (Architecture Evaluation Methods), comparison.

1 Introduction

Architecture Description Language (ADL) provides a simple notation for describing service-oriented problems. By using ADL, developers can quickly define needed services, components, and deployments. ADL is intended to separate behaviors (service contracts) from implementations (components). ADL is composed of several simple constructs: **components, roles, connectors, ports, and protocols.**
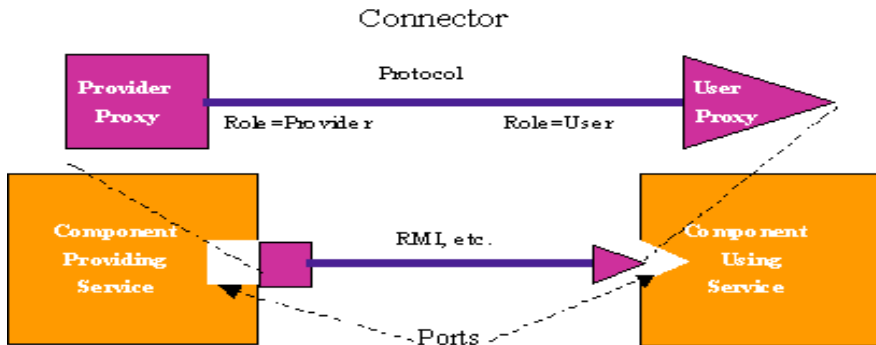
Fig. 1. (ADL composition)

An architectural description language (ADL) is used to describe a software architecture. An ADL may be a formal or semi-formal descriptive language, a graphics language, or include both. The advantage of using an ADL lies in the ability to rigorously specify an architecture so that it can be analyzed. An ADL may have associated with it a set of tools for doing useful analysis of architectures specified in the language. In recent years, there has been a considerable amount of research into developing ADLs.

Leading Candidates:

ACME (Researched by CMU/USC); **Rapide** (Researched by Stanford)；

Wright (Researched by CMU Carnegie-Mellon University); **Unicon** (Researched by CMU)

Secondary Candidates:

Aesop (Researched by CMU); **MetaH** (Researched by Honeywell)；

C2 SADL (Researched by UCI University of California, Irvine);**SADL** (Researched by SRI)

Others:

Lileanna; UML; Modechart

2   Architecture Language

2.1   ACME

An Architecture Internet Exchange Language which supports the mapping between ADL and tool integration. The goal of ACME is to sever as a common exchange format for each other of architecture design, and the unify the existing various ADL in this framework.

Language Features:

- ACME was developed   jointly by Monroe, Garlan (CMU) and Wile (USC);
- ACME is a general purpose ADL originally designed to be a lowest common denominator interchange language;
- ACME as a language is extremely simple (befitting its origin as an interchange language);
- ACME has no native behavioral specification facility so only syntactic linguistic analysis is possible;
    - there are currently efforts under consideration to define a behavioral semantics for ACME, possibly along the Wright/CSP line;
- ACME has no native generation capability;
- ACME has seen some native tool development, and there are indications of more, as well as use of other language tools via interchange;

2.2   Rapide

Rapide, a kind of architecture description language (ADL in short), provides a method for testing the consistency of component transaction. An event-driven ADL, architecture defines it as a development framework that supports component-based development. The language provides modeling, analysis, simulation and code generation capabilities, but no child will be connected explicitly represented as first-order entities.

Language Features:

- Rapide was developed by Dr. David Luckham at Stanford;
- Rapide is a general purpose ADL designed with an emphasis on simulation yielding partially ordered sets of events (posets);
- Rapide as a language is fairly sophisticated, including data types and operations
- Rapide analysis tools focus on posets ;
    - matching simulation results against patterns of allowed/prohibited behaviors
    - some support for timing analysis;
    - focus on causality;
- Rapide has some generation capability since Rapide specifications are executable;
- Rapide has a fairly extensive toolset;

2.3   Wright

One representative of the ADL, Its main feature is the CSP for the description of software architecture, in order to complete some formal reasoning (including compatibility checking and deadlock checking, etc.). But it is only a design specification language, can only be used to describe, can not support the system-generated, while the use of CSP is relatively difficult thing

Language Features:

- Wright was developed by Dr. David Garlan at CMU;
- Wright is a general purpose ADL designed with an emphasis on analysis of communication protocols;
    - Wright uses a variation of CSP to specify the behaviors of components, connectors, and systems;
        - CSP - Communicating Sequential Processes - process algebra developed by C. A. R. Hoare;
- Wright as a language focuses primarily on the basic component/ connector/ system paradigm
    - Wright is very similar syntactically to ACME and Aesop;
- Wright analysis focuses on analyzing the CSP behavior specifications;
    - Any CSP analysis tool or technique could be used to analyze the behavior of a Wright specification;
- Wright does not currently have a generation capability;
- Wright has minimal native tool support (but CSP tools could be used);

2.4   Unicon

A set of predefined sub-mapping connections to specific implementation, thus providing architecture design generated from the application possibilities. But now it only supports the connection of a sub-group has been selected, there are some limitations

Language Features:

- Unicon was developed by Dr. Mary Shaw at CMU;
- Unicon is a general purpose ADL designed with an emphasis on generation of connectors;
  - Unicon developed to support treatment of connectors as first class objects by providing for the generation of systems with explicit connectors ;
- Unicon as a language focuses primarily on the basic component/connector/system paradigm but with an emphasis on architectural styles;
  - Emphasis on styles simplifies generation efforts;
- Unicon has a generation capability;

2.5   Aesop

Language Features:

- Aesop was developed by Dr. David Garlan at CMU;
- Aesop is a general purpose ADL emphasizing architectural styles;
  - Aesop is also a toolset and a framework;
- Aesop the ADL is very similar to ACME/Wright;
  - Emphasis on styles reflected in more sophisticated hierarchical facilities centered around subtyping and inheritance;
- Wright analysis focuses on analyzing the CSP behavior specifications.;
  - Any CSP analysis tool or technique could be used to analyze the behavior of a Wright specification;
- Aesop does have limited generation capability for some styles;
- Interchange facilities to and from Aesop via ACME exist and have been used to make Aesop editing tools available to other ADLs, notably Wright;

3   Architecture Analysis Method

Analysis and evaluation of software architecture is divided into ① **qualitative analysis**: Based on checklist, questionnaire and scene analysis and evaluation techniques; ② **quantitative analysis**: based on metrics, simulation, prototyping systems and mathematical modeling techniques.

3.1   SAAM(Scenario-Based Architecture Analysis Method)

SAAM is the first one to be widely accepted method for architecture analysis and evaluation. For modifiability, scalability, and functional coverage and other quality attributes.

Scene analysis and evaluation of software architecture a technique commonly used by the user, external incentives such as initialization, usually including system events and trigger the event-specific incentives. SAAM method of using technology to build the brain storm scene, the description of the architecture using natural language or other forms said. Methods scenes are classified into direct and indirect scene, respectively, the static architecture to support structural analysis and dynamic analysis.

3.2   ATAM(Architecture Tradeoff Analysis Method)

ATAM method is based on the following assumptions: the architecture style of architecture quality attributes most important qualification, so a ABASs to limit the description of the architecture forms. ATAM concerned about how to get from the business objectives of quality attributes target architecture; the architecture style is how to help achieve these goals. ATAM method of analysis and evaluation is an iterative process, including the four phases: preparation, research and analysis, testing, reporting.

ATAM with a questionnaire to collect quality attributes of the elements of the architecture, describing the characteristics of quality attributes, and the scene is divided into three categories: use case scenario, the use of a system for information access; growth scenario, the expected coverage of the system change; exploratory scenario, covering the system may lead to "pressure" of the major changes, from a different angle characteristics of the detection system to help improve the integrity of the architecture scenario and the risk of decision-making support. ATAM analysis of the interaction between quality attributes and dependencies, and to explore trade-offs between quality attributes of different mechanisms.

3.3   ALPSM(Architecture Level Prediction of Software Maintenance)

ALPSM maintainability of the framework is defined, that a group of maintenance tasks that change the scene. A scenario describing the system may occur in one or a series of moves, the scene changes to describe a specific maintenance tasks. ALPSM to change the size of the scene as a predictor of maintainability, the scene changes through the analysis of maintenance costs required for the architecture maintainability.

ALPSM method combines design experience and historical data to validate the framework of maintaining the line, and the effective introduction of change to predict system maintainability.

3.4   ALMA(Architecture Level Modifiability Analysis)

Maintainability prediction method is based on the cost and risk assessment metrics; change the scene through the construction of evaluation for the analysis can be modified, assuming that the scale of change of the major cost factors can be modified to construct a modified prediction model.

ALMA is introduced quantitative metrics; support from the risk assessment, cost projections, architecture selection evaluation system the angle of the structure can be modified and built to stop the scene to provide guidelines. ALMA disadvantage is the lack of accuracy of results and risk assessment to determine the integrity of the judge.

3.5   SACAM(Software Architecture Comparison Analysis Method)

The core of SACAM is an analytical framework, the distinction between function-driven and quality-driven criteria, the quality attributes of the functional part of the scene map into design elements (modules, subsystems, etc.) of the act; will be non-functional part of the map structure to model that , which provides an architecture to meet the comparison criteria is the reasoning framework.

SACAM method according to organizational business goals, by comparing the different architecture of the system to meet the level of choice provided the basic principles of architecture.

4   Summaries

4.1   Comparison of Different Languages

Component: Overall, surveyed ADLs provide comprehensive support for modeling components. All of them regard components as first-class entities. Furthermore, all model interfaces and distinguish between component types and instances. On the other hand, a majority of the ADLs do not support evolution or non-functional properties. It is illustrative that Aesop is the only ADL that provides at least some support for each of the six classification categories and that, of the four ADLs that support five of the categories, C2 and Rapide do not model non-functional properties, and MetaH and UniCon do not support evolution.

Connector: The support provided by the ADLs for modeling connectors is considerably less extensive than for components. Three ADLs (Darwin, MetaH, and Rapide) do not regard connectors as firstclass entities, but rather model them in-line. Their connectors are always specified as instances and cannot be manipulated during design or reused in the future. Overall, their support for connectors is negligible. It is interesting to note that, as in the case of components, support for evolution and non-functional properties is rare, and that Aesop is again the only ADL that provides at least some support for each classification category.

Configuration: It is at the level of configurations that the foci of some ADLs can be more easily noticed. For example, SADL's particular contribution is in architectural refinement, while Darwin mostly focuses on system compositionality and dynamism. No single ADL satisfies all of the classification criteria, although Rapide comes close. Coverage of several criteria is sparse across ADLs: refinement and traceability, evolution, dynamism, and non-functional properties. These are good indicators of where future research should be directed. On the other hand, most ADLs allowor also provide explicit support for understandability, compositionality, and heterogeneity.

4.2 Comparison of Different Methods

Table 1. Comparison of different methods

Evaluation Methods	Architecture Description	Quality Attribute	Main Stakeholders	Analysis and Evaluation Techniques	Supporting Tools
SAAM	Description easy to be understood	Modifiability	Requirement Analyzer, Architect, Analyzer, Evaluator	Scenario	SAAMTOOL
ATAM	Described from multi-views and ABASs style	Performance, Availability, Security, Modifiability	Designer	Scenario, Questionnaire	None
ALPSM	No special requirement	Maintainability	Designer, Analyzer, Evaluator	Scenario	None
ALMA	4+1 view	Modifiability	Designer	Scenario	None
SACAM	Description of candidates architectures are consistent	Multi Quality Attributes	SACAM Group, Designer	Scenario, Metric	None

References

[1] Wolf, A.L.: Succeedings of the Second International Software Architecture Workshop (ISAW-2). ACM SIGSOFT Software Engineering Notes, pp. 42–56 (January 1997)

[2] Wolf, A.L. (ed.): Proceedings of the Second International Software Architecture Workshop (ISAW-2), San Francisco, CA (October 1996)

[3] IEEE, Inc. IEEE Standard VHDL Language Reference Manual. IEEE Standard 1076-198. IEEE CS Press, Los Alamitos (1987)

[4] Tracz, W.: Parameterized Programming in LILEANNA. In: Proceedings of ACM Symposium on Applied Computing (SAC 1993) (February 1993)

[5] Taylor, R.N., Medvidovic, N., Anderson, K.M., Whitehead Jr., E.J., Robbins, J.E., Nies, K.A., Oreizy, P., Dubrow, D.L.: A Component- and Message-Based Architectural Style for GUI Software. IEEE Transactions on Software Engineering, 390–406 (June 1996)

[6] Shaw, M., DeLine, R., Klein, D.V., Ross, T.L., Young, D.M., Zelesnik, G.: Abstractions for Software Architecture and Tools to Support Them. IEEE Transactions on Software Engineering, 314–335 (April 1995)

[7] Perry, D.E., Wolf, A.L.: Foundations for the Study of Software Architectures. ACM SIGSOFT Software Engineering Notes, 40–52 (October 1992)

[8] Purtilo, J.: The Polylith Software Bus. ACM Transactions on Programming Languages and Systems, 151–174 (January 1994)

[9] Moriconi, M., Qian, X., Riemenschneider, R.A.: Correct Architecture Refinement. IEEE Transactions on Software Engineering, 356–372 (April 1995)

The Two-Stage Filtering Algorithm Based on Local Similarity

Genghao Zhou and Shiping Lin

Dept. of Computer Science, Fuzhou University, Fuzhou, China, 350108

Abstract. In the field of denoising methods, traditional filters are still widely used because of their simple programming and high processing efficiency. However, they have been less accepted since their low efficiency. For example, Median filters cannot denoise Gaussian noise effectively, while mean filters would lead to fuzziness of images due to the loss of lots of details. Against this backgrounds, we proposed a two-stage filtering algorithm based on local similarity, which can preserve better of the structural details and contrast of images. In the first stage, we use adaptive median filter algorithm to filter out the impact of impulse noise; And then we do the second image denoising according to the principle of local similarity of the image. Extensive simulation results under various experimental settings demonstrated the effectiveness of the proposed algorithm.

Keywords: Traditional filters, Adaptive median filter, Local similarity, The two-stage filter algorithm, The peak signal to noise ratio.

1 Introduction

The principal sources of noise in digital images comes mainly from two aspects. One is that during image acquisition the performance of imaging sensors is affected by a variety of factors, such as environmental conditions,and by the quality of the sensing elements themselves. The other is that during transmission images are corrupted principally due to interference in the channel used for transmission. Digital images are inevitably affected by noise during image acquisition or transmission, leading to the decline in the quality of images, and are not unsuited to practical application. Therefore, before images are used in application, they usually are processed with a valid pre-denoising algorithm to obtain more reliable results. There are many types of image noise, such as gaussian noise, rayleigh noise, ireland (gamma) noise, uniform noise and impulse noise. Impulse noise (also known as salt-and-pepper noise), is the most common in the image.

Traditional filters mainly contain mean filters and order-statistics filters. Because of the effect of image denoised by traditional filters whose contrast is not obvious and border details are not clear enough, the techniques and models of image denoising have been being developed. Recently, a lot of excellent filters arise at home and abroad, such as bilateral filer[1], lots of improved order-statistic filters[2], a number of fuzzy filters[3-4], total variation filter[5] and wavelet-based method filters [6]and so on.

Considering mean filters and order-statistic filters can not effectively denoise images and retain the image details, this paper proposes the two-stage filtering algorithm based on local similarity. The algorithm is better than mean filters and order-statistics filters in not only denoising effect but also edge retention.

2   The Principle of Mean Filters and Order-Statistics Filters

Mean filters are the linear spatial filters and mainly include arithmetic mean filter and geometric mean filter. The basic idea of mean filter is that using a template successively slip over the whole image to denosie image. Given the point place (x,y) in a image, then its the gray value is $f(x,y)$, and Sxy represent the set of coordinates in a rectangular subimage neighborhood of size $m \times n$, centered at it. It is the template. $g(x,y)$ is the gray value of the restored image at point (x,y). The value of $g(x,y)$ of arithmetic mean filter is simply the arithmetic mean computed using the pixels in the template defined by Sxy. The equation (2-1) represents arithmetic mean filter.

$$g(x, y) = \frac{1}{mn} \sum_{(s,t) \in S_{xy}} f(s,t) \tag{2.1}$$

The equation (2-2) represents geometric mean filter.

$$g(x, y) = [\prod_{(s,t) \in S_{xy}} f(s,t)]^{\frac{1}{mn}} \tag{2.2}$$

Mean filters are well suited for random noise like gaussian noise and uniform noise, but unsatisfactory effect in impulse noise. They corrupt image and decrease the contrast.

Order-statistics filters are the nonlinear filters. The best-known order-statistic filter is the median filter, which, as its name implies, replaces the value of a pixel by the median of the intensity levels in the neighborhood of that pixel:

$$g(x, y) = \underset{(s,t) \in S_{xy}}{median} \{f(s,t)\} \tag{2.3}$$

In addition to median filter, there are many order-statistics filters, such as max and min filters, midpoint filter, alpha-trimmed mean filter and so on.

Order-statistics filters have well effect in denoising impulse noise, but unsatisfactory effect in gaussian noise and uniform noise. The denser noise density in images are, the worse the effect obtained by order-statistics filters is.

3   The Two-Stage Filtering Algorithm Based on Local Similarity

In order to perserve image details as much as possible, the paper designs the two-stage filtering algorithm based on local similarity(TSFLS). At the first stage adaptive median filter is used to denoising image. Because impulse corruption usually is large compared with the strength of the image signal, impulse noise generally is digitized as extreme values in an image and largely corrupt the image. According to this feature, the pixel can easily be determined whether it is the impulse noise, and be addressed.

Adaptive median filter:

Step1: If $Z_{med} - Z_{min} > 0$ AND $Z_{max} - Z_{med} > 0$, go to Step2.
Else increase the window size.
If window size $\leqslant S_{max}$, repeat Step1;
Else output Z_{med}.
Step2: If $Z_{xy} - Z_{min} > 0$ AND $Z_{max} - Z_{xy} > 0$, output Z_{xy};
Else output Z_{med}.

The symbol means: Z_{xy} is the gray value at point (x,y); S_{xy} represent the set of coordinates in the rectangular subimage window of size m$\times$n, centered at point (x,y); Z_{min} is the minimum gray value in S_{xy}; Z_{max} is the maximum gray value in S_{xy}; Z_{med} is the median of gray values in S_{xy}; S_{max} is the maximum allowed size of S_{xy}.

In contrast to median filter, adaptive filter has two advantages. First, it can have more effective in dealing with high denser impulse noise. Secondly, it can save more details during processing non-impulse noise.

At the second stage, filtering algorithm based on local similarity is used.Given a image corrupted by noise. f(x,y) is the gray value at point (x,y) in the image. S_{xy} represent the set of coordinates in the rectangular subimage window of size m$\times$n, centered at point (x,y) in the image. A weighted average of all pixels in S_{xy} is computed as the output g(x,y).

$$g(x, y) = \sum_{(s,t) \in S_{xy}} w(N_{st}, N_{xy}) f(s,t)$$

(3.1)

Where the weight $w(N_{st}, N_{xy})$ depend on the similarity between the pixels (x,y) and (s,t) and satisfy the usual conditions $0 \leqslant w(N_{st}, N_{xy}) \leqslant 1$ and $\sum_{(s,t) \in S_{xy}} w(N_{st}, N_{xy}) = 1$.

The similarity between two pixels (x,y) and (s,t) depend on the similarity of the intensity grey-level vectors $Z(N_{xy})$ and $Z(N_{st})$, where N_{xy} represent the set of coordinates in the rectangular subimage window of fixed size, centered at point (x,y) and is the similarity window. The restriction of Z to a neihborhood N_{xy} is denoted by $Z(N_{xy})$. In order to compute the similarity of the intensity grey-level vectors $Z(N_{xy})$ and $Z(N_{st})$, a Gaussian weighted Euclidean distance is computed.

$$d(N_{st}, N_{xy}) = \left\| Z(N_{st}) - Z(N_{xy}) \right\|_{2,a}^{2}$$

(3.2)

Then, the weight $w(N_{st}, N_{xy})$ can be computed like this:

$$w(N_{st}, N_{xy}) = \frac{\exp(-d(N_{st}, N_{xy})/h^2)}{\sum_{(s,t) \in S_{xy}} \exp(-d(N_{st}, N_{xy})/h^2)}$$

(3.3)

where the parameter h controls the decay of the exponential function, and therefore the decay of the weights, as a function of the Euclidean distances.In the Eq.(3-3),

when $(x, y) = (s, t)$, the weight $w(N_{st}, N_{xy})$ is too large to fully consider the role of other point. To solve this problem, the weight $w(N_{st}, N_{xy})$ can be computed like this:

$$w(N_{xy}, N_{xy}) = \max(w(N_{st}, N_{xy}), \forall (x, y) \neq (s, t)) \tag{3.3}$$

Core algorithm of the second stage: ls(f)

Input: *f*
Output: *g*
1:    for *i* = 1 to size$_x$ do   % the size of f is size$_x$×size$_y$
2:      for *j* = 1 to size$_y$ do
3:        for *s* = 1 to *m* do
4:          for *t* = 1to *n* do
5:              d(i,j,s,t)=GwEd(f(i,j),f(s+i-m/2,t+j-n/2))   %Gaussian weighted Euclidean distance
6:            end for
7:          end for
8:        end for
9:      end for
10:   for *i* = 1 to size$_x$ do
11:     for *j* = 1 to size$_y$ do
12:       for *s* = 1 to *m* do
13:         for *t* = 1to *n* do
14:             w(i,j,s,t)=conputeW(i,j,s+i-m/2,t+j-n/2)
15:             g(i,j)= w(i,j,s,t)*f(s+i-m/2,t+j-n/2)+g(i,j)
16:           end for
17:         end for
18:       end for
19:   end for

---

4   Experiments

In order to compare the effects of previous methods, the peak signal to noise ratio (PSNR) [7] is used to make the quantitative analysis of denoising effect. The PSNR:

$$PSNR = 10\log(255^2 MN / \sum_{i=1}^{M} \sum_{j=1}^{N} [f_0(i, j) - g(i, j)]^2) \tag{4.1}$$

Where $f_0(i,j)$ is the gray value at point (i,j) in the original noise-free image of size m× n, and g(i,j) is the gray value at point (i,j) in the denoised image. The bigger the PSNR is, the better the effect of denoising image is.

In the experiment, the parameters of algorithms are adjusted to get better results.

Tab.1 represent the average of PSNR of 100 images that was corrupted by additive gaussian noise of zero mean and variance of different values and impulse noise with a probability of 0.05 obtained by previous methods.

Table 1. The average of PSNR of 100 images

d=0.05 Var=	Mean filter	Median filter	Mean filter + median filter	TSFLS
0.02	55.4528	59.4925	64.3506	65.6379
0.06	49.2595	51.5111	53.4166	53.8726
0.10	42.9127	43.9426	44.7791	44.9682
0.14	37.9974	38.2603	38.7301	38.8289

Tab.2 represent the average of PSNR of 100 images that was corrupted by additive gaussian noise of zero mean and variance of different values and impulse noise with a probability of 0.1 obtained by previous methods.

Table 2. The average of PSNR of 100 images

d=0.10 Var=	Mean filter	Median filter	Mean filter + median filter	TSFLS
0.02	50.5981	58.4626	63.6669	65.0235
0.06	46.2322	50.7494	52.9215	53.3662
0.10	41.5052	43.8094	44.7995	44.9966
0.14	37.1593	38.1546	38.7048	38.8209

(a)   (b)   (c)

(d)   (e)   (f)

Fig. 1. Result of denoising with different methods

The lady image selected from the 100 images corrupted by additive gaussian noise of zero mean and variance of 0.1 and impulse noise with a probability of 0.1 is denoised by the previous method. The result is as Fig.1.
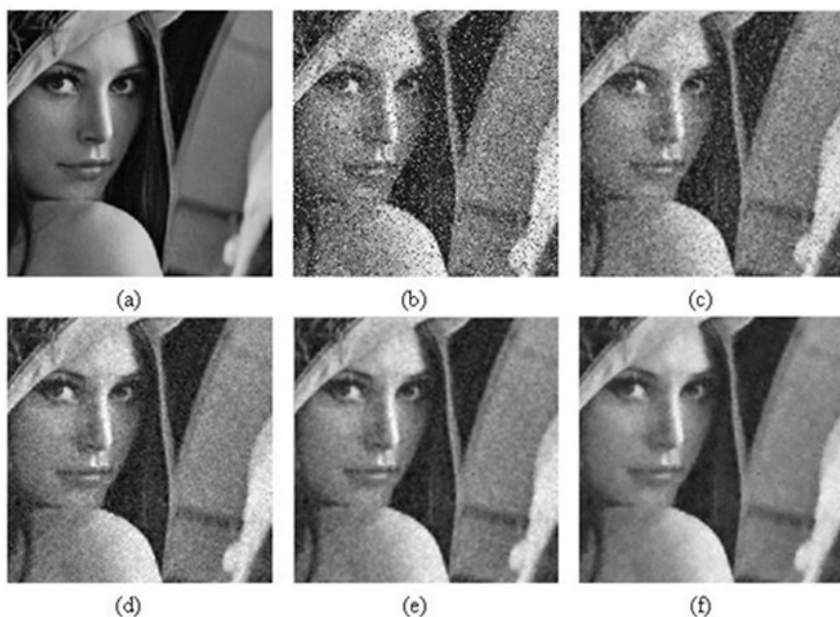
Fig.1(a) shows the lady image, and fig.1(b) shows the same image, but corrupted by additive gaussian noise of zero mean and variance of 0.1 and impulse noise with a probability of 0.1.Fig.1(c) and (d) show, respectively, the result of filtering the noisy image with an arithmetic mean filter of size 3×3 and a median filter of the same size. Fig.1(e) shows the result of denoiseing the noisy image with arithmetic mean filter of size 3×3 and a median filter of the same size. Fig.1(f) shows the result of denoiseing the noisy image with TSFLS.

It is easy to see that the effect of Fig.1(c) and (d) is not satisfactory, and there are still lots of noise. Compared with Fig.1(e), the smoothing effect of Fig.1(f) is better. In Fig.1(f), the noise is eliminated greatly, and there is in the absence of over-smoothing. Fig.1(f) is more close to the original image.

From the previous experment results, mean filter and median filter independntly can't effectively denoise the image that was corrupted by gaussian noise and impulse noise. Although the image corrupted by gaussian noise and impulse noise can be denoised by both mean filter and median filter, the processed image is blurred, and decreases the contrast, and lose structural detail. TSFLS can not only effectively denoise image corrupted by gaussian noise and impulse noise as more as possible, but also obtain a clear image and retain the structure details.

References

[1] Tomasi, C., Manduchi, R.: Bilateral filtering for gray and color image. In: Proceedings of International Conference on Computer Vision, Bombay, India, pp. 839–846 (1998)

[2] Chen, D., Xue, D., Pan, F.: An improved median filter based on automatic parameter tuning approach. In: IEEE International Conference Proceedings on Mechatronics and Automation, pp. 1305–1309 (2007)

[3] Kumer, M., Stoll, N., Kaber, D., et al.: Fuzzy filtering for an intelligent interpretation of medical data. In: IEEE international Conference on Automation Science and Engineering, pp. 225–230 (2007)

[4] Tseng, C.-S.: Robust Fuzzy Filter Design for a Class of Nonlinear Stochastic Systems. IEEE Transactions on Fuzzy Systems 15(2), 261–274 (2007)

[5] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60(2), 259–268 (1992)

[6] Coifman, R., Donoho, D.L.: Translation-invariant de-noising. In: Waveltes and Statistics, pp. 125–150. Springer, New York (1995)

[7] Turaga, D.S., Chen, Y., Caviedes, J.: No reference PSNR estimation for compressed pictures. Image Communication (19), 173–184 (2004)

[8] Gonzalez, R.C., Woods, R.E.: Digital Image Processing, 3rd edn. Prentice-Hall, Inc., Upper Saddle River

[9] Buades, A., Coll, B., Morel, J.M.: A review of image denoising algorithms, with a new one. Multiscale Modeling and Simulation (SIAM Interdisciplanary Journal) 4(2), 490–530 (2005)

Study of Voice Conversion Based on the 405LP

Xiaoning Li[1], Xiaofeng Li[2], and Zhuo Zhang[3]

[1] College of Computer Science and Technology of Chang Chun Normal University,
Chang Chun, China
[2] Jilin University Chang Chun, China
[3] Changchun City Experimental High School, Chang Chun, China

Abstract. voice conversion is based on the combination of proposed independent target speaker pitch between the source of the transfer Because, although the conversion is in the field by modifying the residual signal IoSE Time interval to achieve, it is not on cloud / unvoiced decisions need And make the same silent treatment, which, in the expression part of the, IoSE Equivalent pitch information, and in the silent period.

Keywords: Voice conversion, Prosody transformation, Power PC 405LP.

1   Introduction

Parameters in the speaker, in addition to personality characteristics reflect channel Spectral envelope information parameters[1], the rhythm of the speech signal functions pack Speaker of the status and wealth. Prosodic parameters include Pitch contour, phoneme duration and energy and other parameters. In this study, Consider the main speaker and target speaker pitch between the source of the transfer Because, although the conversion is in the field by modifying the residual signal IoSE Time interval to achieve, it is not on cloud / unvoiced decisions need And make the same silent treatment, which, in the expression part of the, IoSE Equivalent pitch information, and in the silent period, IoSE reflected signal suddenly Development time. In the past, voice conversion algorithm, frequent change of pace Only part of the pitch period of voiced and voiceless, including neglect Prosodic information. Prosodic features as a parameter IoSE, more fully into account Useful information on the speech characteristics[2] of speech signal conversion rhythm.

Residual signal is the minimum phase signal, the minimum phase signal, when the Group uses the characteristics of the promotion, the average slope, extracts the residual signal.

Advances in low-power components and system design have brought general purpose computation into watches, wireless telephones, PDAs and tablet computers. Power management of these systems has traditionally focused on sleep modes and device power management. In a system we normally operate the CPU/SDRAM at 266/133 MHz above 1.65 V and at 66/33 MHz above 0.9 V, typically providing a 13:1 SOC core power range over the 4:1 performance range[3].

From the discrete form of matching function has been through continuous process, the matching function is based on early vector quantization (VQ's) code of this, is a discrete form. It will be a source and target feature space quantization of the feature

space, respectively, from the histogram, each code word between the two codes match probability matrix formation, thus completing the source to the target feature space mapping feature space.

2   Voice Machine Based on the 405LP

From a system design and active power management perspective the most interesting facets of the 405LP SOC design concern the way the clocks are generated and controlled. These features of the processor are described in the remainder of this Section. A. PLL The 405LP PLL exists in a separate voltage island, internally regulated to the minimum voltage specified for the CPU core. As long as the core voltage slew rate is below 10 mV/µ s the PLL is immune to relocking requirements [4] during voltage scaling. Hence, the 405LP continues to execute code and respond to interrupts even during voltage scaling. With low-cost power supply designs we find voltage scaling times typically to be on the order of several hundred µ s, so it is advantageous to be able to continue execution while waiting for voltages to stabilize.

The CPU clock is generated by a clock divider outside of the PLL feedback path, and the circuit design allows divide ratios to be changed without glitches on the output clock. This allows CPU frequency scaling over a wide range with low latencies. B. System Clocking The 405LP[5] CPU clock is further divided down to generate clocks for the internal and external busses: the high-speed Processor Local Bus (PLB, also SDRAM frequency), and lower speed On-board Peripheral Bus (OPB) and External Peripheral Bus (EBC). These clock dividers are glitch-free, and the divide ratios can be changed at any time as long as the resulting overall frequency settings are lower than the maximum frequency allowed by the current operating voltage.

Clocking flexibility is important for optimal power management. In a CPU-intensive application like media decoding we find that lowering bus frequencies from their maximum performance points can result in system-wide energy savings without impacting real-time performance. However, power management policies are typically defined to only scale the CPU and PLB (memory) frequencies, leaving the lowerspeed busses at fixed frequencies. This avoids the need to reprogram bus controller parameters and IP cores that have some sensitivity to peripheral bus frequencies. In the 405LP design changing the CPU and bus clocking scheme requires updating anywhere from one to three control registers and may also require reprogramming the SDRAM controller and other bus controllers.

The traditional power management technology is only lower power consumption during idle time, very effective, and in the run time can not do anything. At present, some operating system kernel for dynamic voltage management (DPM) support or adjust the core voltage is not limited, mainly by adjusting the CPU frequency and power supply support switch peripheral devices to achieve. But in recent years, with the development of semiconductor technology, the dynamic changes in the state of the processor are running the operating voltage and / or clock frequency has become a reality. Many commercial embedded processing support dynamic voltage scaling DVS this technology, such as TI's OMAP family of processors, Intel StrongARM processor, the company, Transmeta Crusoe processor, the company, as well as IBM's PowerPC 405LP processor.

Wavelet transform time-frequency analysis, chirplet transform and fractional Fourier transform of time trying to get the signal frequency information. At the same time the ability to resolve frequency and time mathematically limited by the uncertainty principle. The following table lists the family of Fourier transform Fourier transform family. It is asy to find the function in time (frequency) domain as a discrete function in the frequency corresponding to its (time) domain periodicity.

On the contrary it means a continuous in the corresponding domain of aperiodic signals. transform time-frequency continuous Fourier transform continuous, aperiodic continuous, aperiodic Continuous Fourier series, periodic discrete, aperiodic discrete-time discrete Fourier transform, aperiodic continuous, periodic discrete Fourier transform, periodic discrete, periodic Fourier transform of the basic ideas first put forward by the French scholar Fourier system, so named in its honor. From the perspective of modern mathematics, Fourier transform is a special integral transform. It will be a function satisfying certain conditions, expressed as a linear combination of sinusoidal basis functions, or points.

In different fields of study, Fourier transform has many different variations, such as the continuous Fourier transform and discrete Fourier transform. Fourier transform harmonic analysis of the content belongs. "Analysis" the word can be interpreted as depth. Literally, "of" the word actually is "Tiaofenlvxi" only. It functions by "Tiaofenlvxi" to achieve a deep understanding of complex functions and research. From a philosophical point of view, "Analysis of doctrine" and "reductionism", that is, through proper analysis of the internal things to the purpose of promoting understanding of its nature. Such as the modern atomic theory tries to the origin of all matter in the world of the atom, but only hundreds of atoms, however, the relative wealth of the material world of unlimited, no doubt this analysis and classification of various nature of cognition provides good means.

3   Voice Conversion

From the discrete form of matching function has been through continuous process[6], the matching function is based on early vector quantization (VQ's) code of this, is a discrete form. It will be a source and target feature space quantization of the feature space, respectively, from the histogram, each code word between the two codes match probability matrix formation, thus completing the source to the target feature space mapping feature space. But codeword VQ codebook based on the form of matching limited, so limited to limited set parameters, resulting in discontinuous parameters, greatly reducing the voice quality. Stylianou, who proposed based on Gaussian mixture model (GMM) of the matching function of the continuous, the minimum mean square error criterion to estimate the parameters of the matching parameters to maintain the continuity of space and improve the quality of synthetic speech, and demonstrates how the matching function VQ is based on a special form.

DPM policies are data structures registered with and interpreted by the DPM implementation in the kernel. Policy activation is controlled by an application-specific, executable policy manager, also provided by the system designer. The policy manager is optional, as some systems may be effectively power managed by a single policy installed at system initialization.

It shows the different components and their interactions in a DPM-enabled system. Several researchers have arrived at the conclusion that an optimal power management policy will require the active participation of power-aware tasks in an operating system capable of task-/event-specific power and performance management [8]. However, we believe that in many systems only a small number of application programs will ever be modified to be power-aware. Therefore, the system must be able to effectively handle mixed workloads of power-aware and conventional tasks. We also do not believe that a single task running on a general purpose system should be able to unilaterally set the system operating point, unless that task also has the authority to act as a complete power-policy manager for the system.

The DPM task state mechanism for taskspecific power management was arrived at as a compromise between competing concerns for simplicity, flexibility, performance and optimum power efficiency. Under DPM, the various task states are recognized as separate operating states of the system. The task state index of each task is stored in the task structure, and is not interpreted by the core OS. Instead, DPM policies map operating points to the various task states, and whenever a task is scheduled the context switch invokes the DPM layer to actuate the operating point associated with the task's DPM task state. The task state mechanism allows privileged, power-aware tasks to indirectly set their own operating point or the operating points of other tasks by changing task state assignments[9].

Only the policy manager determines the task state to operating point assignment, however, by choosing to activate an appropriate DPM policy. IV. DPM STRATEGIES we argue that DPM is a useful abstraction because of its ability to easily implement a wide variety of effective power management strategies. In the following we describe several strategies that we have implemented and tested on experimental platforms based on the 405LP. Some of the strategy descriptions refer to Table I, which abstracts the key mappings of several possible DPM policies for the platforms. A. Single-policy Strategies The simplest DPM strategies require only a single policy and no run time policy manager. An example is the idle scaling (IS) strategy. We are particularly interested in strategies that combine load scaling, which has proved to be "good enough" for generalpurpose applications without real-time constraints, and taskspecific requirements. We refer to these hybrid systems as application scaling (AS) strategies. Video decoding is commonly used as an example of an important workload that is difficult to power manage without application participation, so we developed an AS strategy for an experimental[10], multi-threaded MPEG4 video/audio decoder for Linux.

Fourier transform belongs to harmonic analysis and the transformation of form and are very similar; basis functions are eigenfunctions of differential operation, which makes the linear differential equation with constant coefficients can be transformed into algebraic equations the nature of the frequency is unchanged, so the system can respond to incentives for the complex by combining its response to sinusoidal signals of different frequencies to obtain; convolution theorem states that: Fourier transform can convolution of the complex product of a simple operation, thus providing a simple means of calculating the convolution; discrete Fourier transform can be calculated using a fast digital computer (the algorithm is called fast Fourier transform algorithm.

4   Summaries

This power-aware video decoding approach[11] does not require extensive analysis and prediction of processing requirements or the implementation of new scheduling mechanisms, but still achieves good results. We present this AS example to illustrate an interesting use of DPM's policy mechanisms without which the application would need significant additional complexity to make it power-aware. We made two simple changes to the video decoding threads to make it easier to power manage this application with our LS policy manager. We modified the video decoder to begin processing the next frame immediately upon completion of the current frame, rather than idling.

References

[1] Wu, C.H., Hsia, C.C., Liu, T.H., Wang, J.F.: Voice conversion using duration-embedded biHMMs for expressive speech synthesis. IEEE Transactions on Audio, Speech, and Language Processing 14(4), 1109–1116 (2006)

[2] Morrison, D., Wang, R.-l., De Silva, L.C.: Ensemble methods for spoken emotion recognition in call-centres. Speech Communication 49(2), 98–112 (2007)

[3] Rao, K.S., Yegnanarayana, B.: Prosody modification using in-stants of significant excitation. IEEE Transaction on Audio, Speech and Language 14(3), 972–980 (2006)

[4] Elovitz, H.S., Johnson, R., McHugh, A., et al.: Letter-to-soundrules for automatic translation of English text to phonetics. IEEE Transactions on Acoustics, Speech and Signal Processing 24, 446–459 (1976)

[5] Wang, G., Liang, W., Liu, J., et al.: Moderate vocabulary English speech recognition system embedded on a chip. Qinghua Daxue Xuebao/Journal of Tsinghua University 45(10), 1393–1396 (2005)

[6] Ververidis, D., Kotropoulos, C.: Emotional speech recognition: resources, features, and methods. Speech Communication 48(9), 1162–1181 (2006)

[7] Li, H.-X., Li, L.-X., Wang, J.-Y., et al.: Fuzzy Decision Making Based on Variable Weights. Mathematical and Computer Modelling, 163–179 (2004)

[8] Ramamohan, S., Dandapat, S.: Sinusoidal model-based analysis and classification of stressed speech. IEEE Trans. on Audio, Speech, and Language Processing 14(3), 737–746 (2006)

[9] Nwe, T.L., Foo, S.W., De Silva, L.C.: Speech emotion recognition using hidden Markov Models. Speech Communication 41(4), 603–623 (2003)

[10] Hasan, M.M., Nasr, A.M., Sultana, S.: An approach to voice conversion using feature statistical mapping. Applied Acoustics 66(5), 513–532 (2005)

[11] Marchand, Y., Damper, R.I.: A multi-strategy approach to improving pronunciation by analogy. Computational Linguistics 26(2), 195–219 (2000)

Poisson Noise Immunity Analysis of the Improved Fractional Differential Algorithm

Jia Changyun[1] and Jin Liang[2]

[1] Huaihai Institute of Technology, Lianyungang, 222005, China
[2] The City Vocational College of Jiangsu, 222005, China

Abstract. A fractional template was developed based on the settlement of drifting problem in edge detection for fractional differential operator. The edge information of the image was extracted accurately. By way of applying to the edge information extraction of an image containing poisson noise, this algorithm was proved to be more accurate, and more noise immunity than the traditional algorithm.

Keywords: Fractional order, Edges, Poisson Noise, Noise Immunity.

1   Introduction

The object of image information extraction is for image recognition. Edge information is one of the most important information for object recognition, but is not all. For most of natural images, their texture information has same importance for recognition. Nowadays, algorithms of image information extraction are all focused on edge information detection, such as Sobel, Canny and Prewitt. They have good detection results of smoothly gray scale region's edge, but can not satisfy the edge detection request for image with weakening edge and not smoothly gray scale edge. It was very essential to extract texture information under that condition.

Differential operation could enhance the high frequency signals. As the increasing of frequency and differential order, it increased rapidly in non-linear. For low frequency signals, fractional differential operation[1-3] filtered more and more as increasing of differential order. The signal with high frequency would be enhanced by fractional differential, also the very low frequency components would be reserved in non-linear form[4-6]. Image was one kind of signal, its signal processing with fractional differential would be favorable to enhance its high frequency edge, and also reserve its low frequency texture information. [7].

2   The Principle of the Algorithm

Although fractional differential had some unique properties from integer order differential, it could not be applied to extract image texture information directly for its difficulty of peak drift detection as differential order varying.
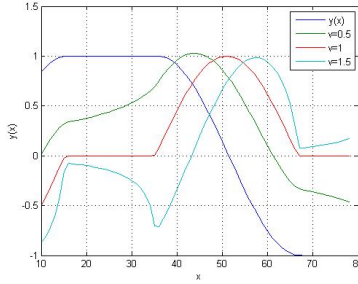
Fig. 1. Signal detection results of differential of different orders

The signal detection result of different differential order was showed in fig1. The peak value position of fractional differential was not consistent with the position which signal varied most dramatically. The improved smoothing method was used to restrain the drift of peak value detection, as follows.

Assume T_s is sampling period, if shift operators Q_1 and Q_2 are acted on discrete time signal $x(nT_s)$, what can be got is $Q_1 x(nT_s) = x(nT_s - T_s) = x_1$, $Q_2 x(nT_s) = x(nT_s + T_s) = x_2$

Let $D_{x_1}^1 x(nT_s) = \dfrac{x(nT_s) - x(nT_s - T_s)}{T_s}$, $D_{x_2}^1 x(nT_s) = \dfrac{x(nT_s) - x(nT_s + T_s)}{T_s}$

Then we can get $D_{x_1}^1 = \dfrac{1 - Q_1}{T_s}$, $D_{x_2}^1 = \dfrac{1 - Q_2}{T_s}$,   And also we can get.

$$D_{x_1}^v = \left(\frac{1 - Q_1}{T_s}\right)^v , \quad D_{x_2}^v = \left(\frac{1 - Q_2}{T_s}\right)^v$$

$\left(1 - Q^{-1}\right)^v$ is done with Taylor expansion, and we can get.

$$\left(1 - Q_1\right)^v = 1 + (-1)vQ_1^1 + (-1)^2 \frac{v(v-1)}{2!}Q_1^2 + ... + (-1)^k \frac{v(v-1)...(v-k+1)}{k!}Q_1^k + ...$$

$$\left(1 - Q_2\right)^v = 1 + (-1)vQ_2^1 + (-1)^2 \frac{v(v-1)}{2!}Q_2^2 + ... + (-1)^k \frac{v(v-1)...(v-k+1)}{k!}Q_2^k + ...$$

$$D_{x_1}^v - D_{x_2}^v = T_x^v = \sum_{k=1}^{\infty}(-1)^k \frac{v(v-1)...(v-k+1)}{k!}(Q_1^k - Q_2^k)$$

It is acted on signal x (t), then the difference expression is.

$$T_x^v x(t) = D_{x_1}^v x(t) - D_{x_2}^v x(t) = \sum_{k=1}^{\infty}(-1)^k \frac{v(v-1)...(v-k+1)}{k!} \bullet (x(t - kT_s) - x(t + kT_s)) \qquad (1)$$

With difference expression，difference coefficient of fractional order differential can be derived.

$$m_0 = 0, m_1 = -v, m_2 = \frac{(-v)(-v+1)}{2}, m_3 = \frac{(-v)(-v+1)(-v+2)}{6}, ..., m_n = \frac{\Gamma(-v+1)}{n!\Gamma(-v+n+1)}$$
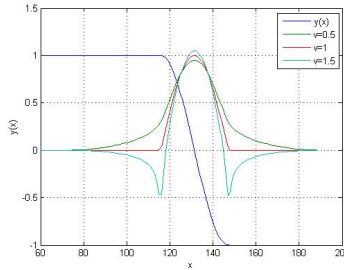
Fig. 2. The fractional order differential after improvement of different orders of signals

Fig.2 illustrated signal processing results with improved fractional differential. The problem of drift has been resolved. We constructed differential template of image based on this principle.

3   Fractional Differential Template

The minimal distance of gray scale of digital image signal was the distance between two adjacent pixels. The distance of digital image with two dimension could only be measured with pixel in x and y coordination. The smallest isotonic interval of image signal could only be 1. Pixel interval of image was replaced by sample interval of equation (1), then the differential equation in x and y coordination was.

$$T_x^v I(x, y) = \sum_{k=1}^{\infty} (-1)^k \frac{v(v-1)...(v-k+1)}{k!} \bullet (I(x-k, y) - I(x+k, y)) \qquad (2)$$

$$T_y^v I(x, y) = \sum_{k=1}^{\infty} (-1)^k \frac{v(v-1)...(v-k+1)}{k!} \bullet (I(x, y-k) - I(x, y+k)) \qquad (3)$$

As image signal's number of terms is also limited, the differential value of image is just an approximate value with high series and error, being impossibly completely equal. The size of fractional order differential model dimension is equivalent to the degree of approximation of its resolution of fractional order differential. However,

considering that the oversize operator model may costs long computing time of computer, according to equation (2) and equation (3), the differential model of image in the x axis and y axis is constructed (as fig 3)

-m(3)*m(3)	-m(2)*m(3)	0	m(2)*m(3)	m(3)*m(3)
-m(2)*m(3)	-m(2)*m(2)	0	m(2)*m(2)	m(2)*m(3)
-m(3)	-m(2)	0	m(2)	m(3)
-m(2)*m(3)	-m(2)*m(2)	0	m(2)*m(2)	m(2)*m(3)
-m(3)*m(3)	-m(2)*m(3)	0	m(2)*m(3)	m(3)*m(3)

a

m(3)*m(3)	m(2)*m(3)	m(3)	m(2)*m(3)	m(3)*m(3)
m(2)*m(3)	m(2)*m(2)	m(2)	m(2)*m(2)	m(2)*m(3)
0	0	0	0	0
-m(2)*m(3)	-m(2)*m(2)	-m(2)	-m(2)*m(2)	-m(2)*m(3)
-m(3)*m(3)	-m(2)*m(3)	-m(3)	-m(2)*m(3)	-m(3)*m(3)

b
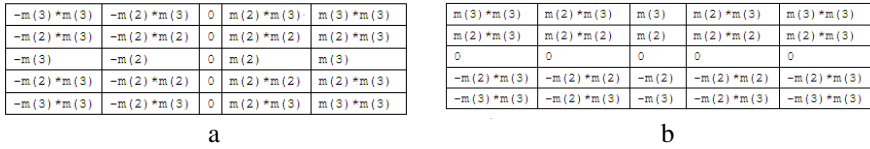
Fig. 3. (a) differential model m_1 in the x axis; (b) differential model m_2 in the y axis.

Enable differential to be D_x^v in x axis, and D_y^v in y axis.

$$D_x^v = I(x, y) * m_1(x, y) \qquad D_y^v = I(x, y) * m_2(x, y)$$

Calculating peak value R . $R = \sqrt{(D_x^v)^2 + (D_y^v)^2}$

The threshold value was set to $0.8 * \max_{R \in I} R$, points which had higher value than threshold value were defined as extraction information. The order was set to 1.5 for texture information extraction during marginal detection.

4   Experimental Analysis of Noise Immunity

Fig 4 showed that the comparison between the testing results of square signal with random noises by this algorithm and traditional differential algorithm. It could be seen from the figure that the common differential methods were not able to detect jumping points of square signal with random noises. This experiment evidenced that this algorithm was equipped with the ability of noise immunity. Fig. 5 showed the experimental images. Fig. 5 (a) (c) were original images and Fig. 5 (c) (d) were images with poisson noise respectively.

a   b   c
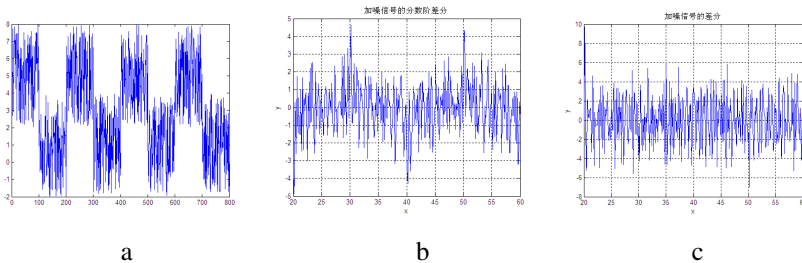
Fig. 4. (a) square signal with random noises (b) testing results of square signal with random noises (c) testing results of square signal with random noises by differential methods

a   b

c   d
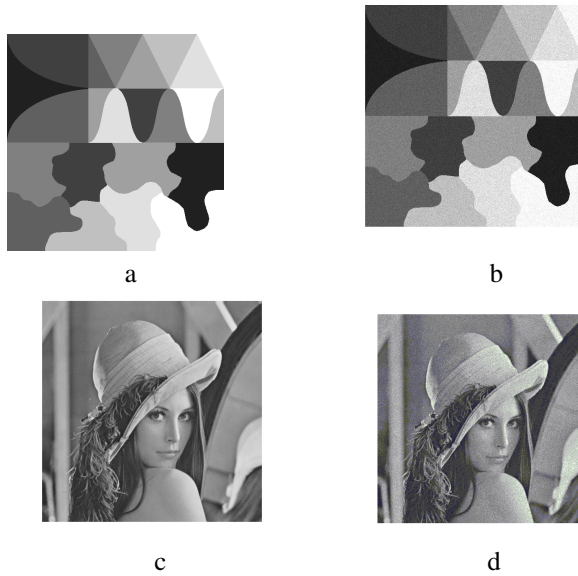
Fig. 5. (a) synthetic image, (b) synthetic image with poisson noise, (c) lena image, (d) lena image with poisson noise

a   b   c

d   e   f
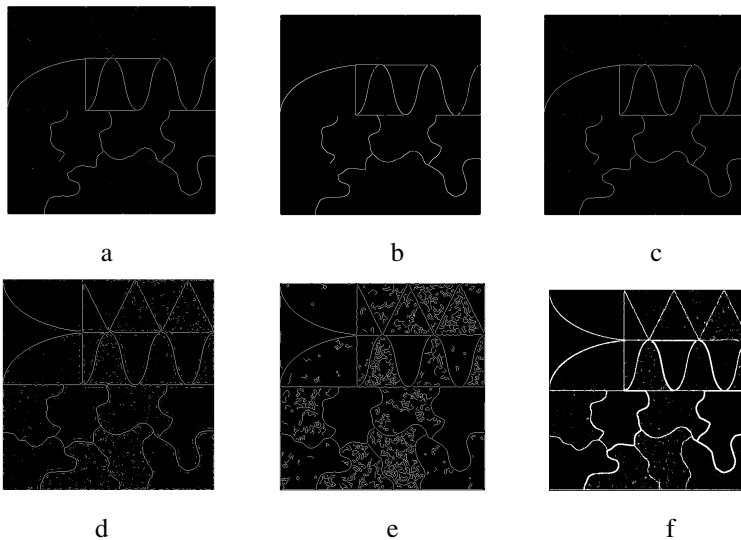
Fig. 6. (a) The testing results of sobel operator on synthetic image with poisson noise; (b) The testing results of roberts operator on synthetic image with poisson noise; (c) The testing results of prewitt operator on synthetic image with poisson noise; (d) The testing results of log operator on synthetic image with poisson noise; (e) The testing results of canny operator on synthetic image with poisson noise; (f) The testing results of this algorithm on synthetic image with poisson noise.
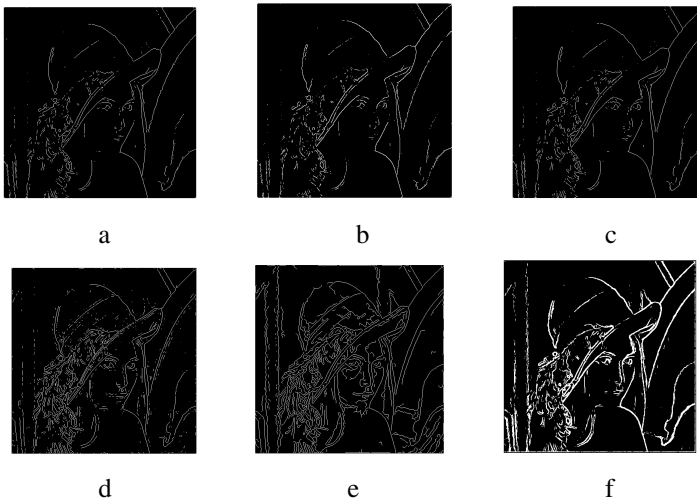
Fig. 7. (a) Testing results of sobel operator on lena image with noise; (b) Testing results of roberts operator on lena image with noise; (c) Testing results of prewitt operator on lena image with noise; (d) Testing results of log operator on lena image with noise; (e) Testing results of canny operator on lena image with noise; (f) Testing results of this algorithm on lena image with noise

Fig. 6 showed the testing results of sobel, roberts, prewitt, log, canny, and our algorithm on synthetic images with poisson noise. This figure indicated that sobel, roberts and prewitt operator could not work properly in the case of noise. Canny operator introduced more noises although they could detect the edge. Log operator and our algorithm discussed in this paper performed better in edge detection and had certain suppression ability to noise. Fig.7 showed the detection results of sobel, roberts, prewitt, log, canny, and the algorithm discussed in this paper on lena image with poisson noise. It could be seen from the figure that sobel, roberts and prewitt operators have missed some edge information under the influence of the noise. The detection results of log and canny operators even could not be used for the purpose of identification. The algorithm discussed in this paper performed better in edge detection and had certain suppression ability to noise.

5 Summaries

To sum up, in this paper, a fractional template was developed based on the settlement of drifting problem in edge detection for fractional differential operator. The edge information of the image was extracted accurately. By way of applying to the edge information extraction of an image containing poisson noise, our algorithm was proved to be more noise immunity than the traditional algorithm.

References

[1] Engheia, N.: On the role of fractional calculus in electromagnetic theory. IEEE Transactions Antennas and Propagation Magazine 39(4), 35–46 (1997)

[2] Miller, K.S.: Derivative of Noninteger Order. Mathematics Magazine 168(3), 183–192 (1995)

[3] Zhou, Z.: Existence and uniqueness for a nonlinear fractional volterra integro Differential Equation. Journal of Huaihua University 26(8), 1–4 (2007)

[4] Capus, C., Brown, K.: Fracional fourier transform of the Gaussian and fractional domain signal support. Image Signal Process. (12), 99–106 (2002)

[5] Tenreiro Machado, J.A., Jesus, I.S., Galhano, A., Boaventura Cunha, J.: Fractional order electromagnetics. Signal Processing 86(10), 2637–2644 (2006)

[6] Pu, Y., Wang, W., Zhou, J., Wang, Y., Jia, H.: Fractional differential approach to detecting textural features of digital image and its fractional differential filter implementation. Science in China Series F: Information Sciences 51(9), 1319–1339 (2008)

[7] Wang, Y., Pu, Y., Zhou, J.: 1/2 Order fractional differential tree type circuit of digital image. Congress on Image and Signal Processing (2008)

Solving Bi-level Programming Problem Based on Electromagnetism-Like Algorithm

Yang Xiaoling, Qiu Dishan, and Shen Jianwei

Science and Technology on Information Systems Engineering Laboratory
National University of Defense Technology
Changsha, China, 410073

Abstract. Bilevel programming problem is NP-hard and not easy to be solved. In the paper, we studied and designed an optimization algorithm for bilevel programming problems based on Electromagnetism-like mechanism. The experiment results showed that the proposed algorithm is feasible and more efficient that existing algorithms based on evolutionary mechanism.

Keywords: bilevel programming, electromagnetism-like mechanism, NP-hard, evolutionary mechanism.

1   Introduction

The general formulation of a bilevel programming problem (short of "BLPP") is

$$\min_{x \in X, y} F(x, y)$$

$$s.t. \quad G(x, y) \le 0$$

$$\min_{y} f(x, y)$$

$$g(x, y) \le 0$$

where the upper-level variables $x \in \mathbb{R}^{n_1}$ and the lower-level variables $y \in \mathbb{R}^{n_2}$. Similarly, the functions F and G are the upper-level and lower-level objective functions respectively. A model of BLPP can be decomposed into two programming models called the upper-level and lower-level model as illustrated. The basic idea of a bilevel programming problem is that the subset of the variables have to be the optimal solution of the lower-level model, i.e. a solution of BLPP, (x^*, y^*), is optimal such that $y^* = \arg\min_{y} f(x^*, y)$ and $(x^*, y^*) = \arg\min_{x \in X, y} F(x, y)$. Essentially, BLPP describes the hierarchical relationship between two different decision levels in which the lower-level decision making is influenced by the upper-level decision.

Many classical combinatorial optimization problems are equivalent to bilevel programs, such as traveling salesman, nonconvex quadratic programming, bilinear disjoint programming, etc. The research for BLPP is highly significant in both theory and practice.

Due to the property that BLPP is NP-hard and even for the linear-linear case, solving the problem is a difficult job. Colson et al.(2007) gave an overview of bilevel

optimization, enumerated six kinds of approaches for solving BLPP with some special restriction (the linear case , the quadratic convex case, etc) [1]. LI Hong et al. (2005) applied genetic mechanism for solving non-linear BLPP [2]. Through the definition of the feasible degree, Osman et al. (2009) constructed the fitness function of the Upper-level model and also presented a genetic algorithm but for linear BLPP [3]. Zhao et al. (2007) introduced a modified particle swarm optimization (short of "PSO"), and a universal effective solution algorithm was extended[4]. Population operation mechanism has exhibited a new direction and prospect of algorithms for solving BLPP.

Birbil and Fang (2003) proposed a stochastic population-based method named electromagnetism-like mechanism (short of "EM") algorithm, which simulates the electro-magnetism theory of physics by considering each particle in the population associated with an electrical charge and uses an attraction-repulsion mechanism to move particles towards optimality [5, 6, 7].

Followed up on the idea of the application of genetic mechanism and PSO for BLPP, the paper presents an new algorithm based on an modified EM algorithm for solving BLPP. The experiment studies show that the availability and performance of the algorithm.

The paper is organized as follows. In Section2 we represent the main content of EM algorithm. Section3 introduces the new algorithm based on EM for solving BLPP. The test example and simulation results are represented in section 4, and finally conclusion and future work are represented in section 6.

2   Electromagnetism-Like Mechanism

For global optimization problem, each solution can be represented by a point in a multi-dimensional solution space, and a charge is associated with each point and related to the objective function value corresponding to the solution. Each solution point will exert attraction or repulsion on other points that simulates the electromagnetism theory of physics. By forces that obey the rules of electro-magnetism, each solution point moves and better solutions are generated.

Each sample point can be regarded as a charged particle which attends the population iteration. EM algorithm consists of four phases. They are initialization of population, calculation of the total force exerted on each particle, movement along the direction of the force, and neighbor- hood search to exploit the local minima (the strategy can be cancelled according to the computation demand).

Initialization of population aims to randomly generate a population of particles inside the feasible domain decided by the constraints. Calculation of the total force exerted on each particle is based on the charge. In the original version of EM, the charge of each particle i , q^i , is evaluated as,

$$q^i = \exp\left(-n \bullet \frac{f(x^i) - f(x^{best})}{\sum_{k=1}^{m}(f(x^k) - f(x^{best}))}\right), \forall i$$

and the total force F^i exerted on particle i is computed by the following equation:

$$F^i = \sum_{j \neq i}^{m} \begin{cases} (x^j - x^i)\dfrac{q^i q^j}{\|x^j - x^i\|^2}, & if \ f(x^j) < f(x^i) \\ (x^i - x^j)\dfrac{q^i q^j}{\|x^j - x^i\|^2}, & if \ f(x^j) \geq f(x^i) \end{cases}, \forall i$$

After computing the total force vector F^i, the paritcle i moves along the direction of the force by a random step length limited by the upper and lower bounds of x^i, and the position update of particle i follows the rule as follows:

$$x^i = x^i + \lambda \frac{F^i}{\|F^i\|}(RNG)$$

where λ is the random step length and $\lambda \in U(0,1)$, $F^i / \|F^i\|$ is a unit vector representing the direction of total force exerted on particle i, and RNG is the feasible step length of particle i towards the upper bound or the lower bound.

A complete version of EM algorithm may be found in Birbil and Fang (2003).

Notice that EM algorithm originally solve global optimization of continuous functions with box constraints, i.e. the variable $x \in [l, u]$ $\triangleq \{x \in \mathbb{R}^n \mid l_k \leq x_k \leq u_k, k = 1, \cdots, n\}$, but general bilevel problems have more complicated constraints. So when we use EM algorithm for solving BLPP, the movement rule of particle should be modified correspondingly.

3   The Description of Solving BLPP Based on EM

BLPP is intrinsically hard, due to the nonconvexity, the nondifferentiablity and the complicated restriction between the upper-level problem and the lower-level problem. Recently, population-based evolutionary algorithms have been introduced for BLPP. One of the basic idea using these algorithms is: firstly, choose the initial population satisfying the constraints for the upper-level model, then solve the lower-level problem for a given individual, thereby the given individual can be evaluated. Do the same operation for every individual. Based on the evolutionary mechanism, update the whole population. Repeat the process above (exclude the initialization of population) over and over until the optimal solution is searched.

Zhao et al. (2007) presents a PSO-based algorithm following up on the idea of the last paragraph [4]. The performance of the algorithm was verified through computing several test problems and comparison with other algorithms.

Similarly, we proposed an EM-based algorithm with the same train of thought and the similar process. The difference is that EM has a higher performance for searching in the solution space and the global convergence has been proved.

The new algorithm named EM-BLPP is as follows[4]:

ALGORITHM EM-BLPP
1: *Initialize algorithm parameters and the population*
2: **while** iteration < MAXGEN **do**
3:      **for** $i = 1$ to m **do**

4:     *Run EM for the upper-level model to update the coordinates of particle i denoted by* (x^i, y^i) *;*

5:     *According to the partial solution of the upper-level model,* x^i *, solve the lower-level model by traditional optimization method with EM, and denote the optimal solution by* $(y^i)^*$ *, i.e.*

$$(y^i)^* = \arg \min_y f(x^i, y) \; ;$$

6:     *Compute the fitness of particle i,* $F(x^i, (y^i)^*)$ *;*

7:     *If* $F(x^i, (y^i)^*) < F(x^{best}, y^{best})$ *, then the best particle,* (x^{best}, y^{best}) *is substituted by* $(x^i, (y^i)^*)$ *;*

8:     **end for**

9:     *Add disturbance factor to the best particle and update the coordinates of best particle, such that* $(x^{best}, y^{best}) = (x^{best}, y^{best}) \times (1 + \eta)$ *, where* $\eta \sim N(0,1)$ *.*

10: **end while**..

4   Numerical Experiments and Results

In this section, we report the numerical results obtained by running EM-BLPP algorithm on the eight bilevel programming problems referred in [4]. We won't repeat the detail of the problems for testing because of the limit of paper length. The ability of finding the optimal solutions is compared between EM-BLPP and PSO-BLPP in the respect of solving accuracy.

We set the population size $m = 40$, the maximum number of iterations *MAXGEN* = 50, the maximum number of local search iterations *LSITER* = 10, the local search factor $\delta = 0.001$. All the computations were conducted on a Pentium-IV 2.40GHz PC. The algorithm is coded in C++ .

The numerical results are given in Tab.1 and Tab.2 including the data of the upper-level goal value and the optimal solution for every problem in [4].

Table 1. Comparison between EM-BLPP and PSO-BLPP on the optimal objective value

Problem	The upper-level goal value		The lower-level goal value	
	EM-BLPP	*PSO-BLPP*	*EM-BLPP*	*PSO-BLPP*
1	0	0	100	100
2	-12.678707	-12.6787	-1.015624	-1.01563
3a	-8.9173	-8. 9172	-6.15723	-6.157
3b	-7.57851	-7.578458	-0.573918	-0.57192
3c	-11.9985	-11.9985	-193.061	-178.07
3d	-3.6	-3.6	-2	-2
4a	1.86044	1.860462	-10.9316	-10.93147
4b	0.897472	0.897460	-14.9289	-14.92894

Tab.1 and Tab.2 show that for problem 3a, 3b and 3c , our algorithm gained better result. For problem 1 and problem 3d, our result is as excellent as PSO-BLPP due to the nature of the problems. For problem 2 and problem 4b, though the lower-level goal value is worse, we gained the better optimal solution. EM-BLPP firstly got

Table 2. Comparison between EM-BLPP and PSO-BLPP on the optimal solution

Problem	The optimal solution (x, y)	
	EM-BLPP	*PSO-BLPP*
1	(0,30,-10,10)	(0,30,-10,10)
2	(0,2,1.874999,0.906249)	(0,2,1.875,0.90625)
3a	(1.0381,3.0986,2.597345,1.79383)	(1.03867,3.09868,2.5973,1.7938)
3b	(0.2834,0.4891,2.3449,1.0359)	(0.27878,0.47498,2.34383,1.03253)
3c	(52.9088,36.9642,2.9985,2.9985)	(14.9086,69.9642,2.9985,2.9985)
3d	(2,0,2,0)	(2,0,2,0)
4a	(3.4562,1.7071,2.568464)	(3.45616,1.70709,2.56846)
4b	(2.998534,1.6658,3.887223)	(3.998535,1.665009,3.887223)

the better lower-level solution, though the upper-level solution has a little gap with the data of PSO-BLPP for problem 4a. We can conclude that EM-BLPP has a higher performance than PSO-BLPP for general bilevel programming problems very probably.

5   Summaries and Future Work

In this paper we have developed a new heuristic algorithm based on electro-magnetism-like mechanism for solving bilevel programming problems. The main idea of the algorithm comes from PSO-BLPP, but the numerical results showed our algorithm owns a higher performance in the aspect of solution accuracy.

It has been proved that EM algorithm is quite powerful for global search. In this paper, we test the solving accuracy of EM-BLPP and we should make further efforts to consider the speed. The difficulty rests in lacking of test function set.

References

[1] Colson, B., Marcotte, P., Savard, G.: An overview of bilevel optimization. Ann. Oper. Res. 153, 86–88 (2007), doi:10.1007/s10479-007-0176-2
[2] Li, H., Wang, Y.-P., Jiao, Y.-C.: A new genetic algorithm for nonliear bilevel programming problems and its global convergence. Systems Engineering-Theory & Practice 25, 62–71 (2005) (in Chinese)
[3] Osman, M.S., Abd El-Wahed, W.F., El Shafei, M.M.K., Abd El-Warhab, H.B.: A solution methodology of bi-level linear programming based on genetic algorithm. Journal of Mathematics and Statistics 5, 352–359 (2009)
[4] Zhao, Z.-G., Gu, X.-Y., Li, T.-S.: Particle Swam Optimization for Bi-level Programming Problem. Systems Engineering-Theory & Practice 8, 92–98 (2007) (in Chinese)
[5] Birbil, S.I., Fang, S.C.: An electromagnetism–like mechanism for global optimization. Journal of Global Optimization 25, 263–282 (2003)
[6] Birbil, S.I., Fang, S.C., Sheu, R.L.: A On the convergence of a population-based global optimization algorithm. Journal of Global Optimization 30, 301–318 (2004)
[7] Han, L.-X., Wang, Y.-P.: Electromagnetism-like mechanism algorithm for unconstrained optimization problem. Chinese Journal of Electronics 37, 664–668 (2009) (in Chinese)

Research and Realization on Dynamic Path Construction Algorithm

Guoxing Peng, Bei Li, and Qi Tong

Computer and Communication School of Hunan University of Technology
`pengguoxing168@163.com`

Abstract. Certificate path Construction (CPC) plays an important role in cross certification. Criterion and literature available deal little with path construction algorithm based theoretically on the shortest path construction algorithm widely adopted in certification network graph(CNG). As a result, the current algorithms are not in a position to satisfy the wide-scale and dynamic development of cross certification. The thesis makes a thorough analysis of the limitations of the current algorithms and designs a cooperative and dynamic path construction algorithm between cross certification and certificate authority(CA), with a purpose to improve the efficiency of multi-paths cross certification and arguing that the path construction requesting information between CAs should be dynamic.

Keywords: cross certification, path construction, PKI.

1 Introduction

Cross certification among independent CAs undertakes to connect the two independent trust domains in the PKI technology, ensuring smooth interactions between PKIs[1]. With cross certification, trust can be established between independent CAs, ensuring that a user in one CA can validate the certificate of a user in another CA. If entity A in one PKI trust domain expects to obtain the permission of public key of entity B in another trust domain, it must obtain entity B's public key certificate, which can be realized through certificate path process consisting of the two steps of certificate path construction and certificate path validation[2][3]. For current path constructions are designed to establish a certification network graph theory, on which search algorithms are based to construct certificate path. These path constructions are thought imperfect and improvements are badly needed to make[4].

2 Limitations of Certificate Path Construction Algorithm Based on Certification Network Graph Theory

Limitations of Certificate Path Construction Algorithm based on Certification Network Graph theory can be illustrated with the introduction of depth-first search algorithm adopted in cross-certification model in Fig.1.

Depth-first search works in the same way as preorder traversal. Supposing all the tree-roots are not traversed in Figure G, and depth-first search starts from the initial node V and then traverses the neighboring node[5]. Depth-first search will not stop until all the nodess are traversed. If some nodess are not traversed, one will be taken as the initial node and depth-first search will be repeated until all the un-traversed nodes are traversed. With depth-first search algorithm, the depth-first tree of CA1 constructed in Figure 1 can be illustrated in Figure 2 as its certificate path being CA1→CA2→CA5→CA6. Certification network graph is inefficient. The limitations of path construction based on dept-first search algorithm are illustrated as follows:

(1)   Multi-paths cross-validation can not be realized. In Fig.2, there are two paths from CA1 to CA7: CA1→CA2→CA5→CA6 and CA1→ CA3→CA7→CA4→CA6. With dept-first search, only the shorter path can be found. If the path fails to be validated, communications can not be gone through successfully between Bob and Alice, though the certificate path CA1→CA3→CA7→CA4→CA6 is validated.

(2)   The path construction available is not dynamic. If cross-validation goes so far that it takes a long time to construct the certification network graph, which therefore can not reflect the dynamic changes of the validation construction. For example, omission or addition of a CA or cancelation of reestablishment of a cross-validation may result in failure of certificate path validation based on certification network graph. A cross-validation between CA4 and CA2 is added for the purpose of constructing a new path to meet the requirements of the extensions of PKI.
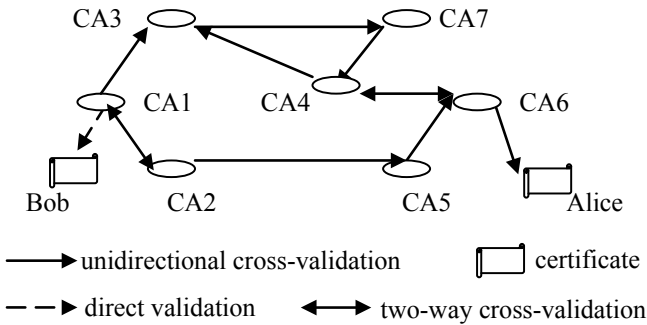
(3)   Certification Network Graph is inefficient.
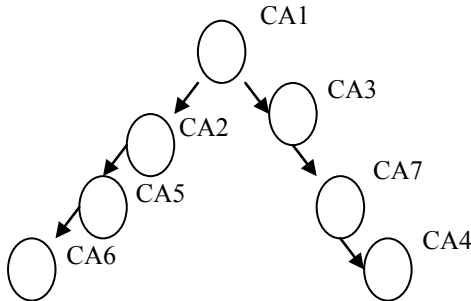

Fig. 1. Cross-certification model

Fig. 2. CA1 of the depth-first tree

3   Cooperative and Dynamic Path Construction Algorithm

3.1   Definition of Data Structure

In order to improve the efficiency of cross validation, the thesis puts forward a new algorithm----cooperative and dynamic path construction algorithm. Before touching upon the new algorithm, the thesis attempts to make clear several concepts. To a CA, certificates issued by other CAs to it are called forward certificates while certificates issued by it are called reverse certificates. An third concept introduced for a good understanding of the new algorithm is Cross certificate to the directory data construction stored in forward certificates and reverse certificatse, as illustrated in Table 1 and a fourth concept introduced is certification path construction information data structure.

Table 1. Cross-certificate to the directory

Forward Certificate List	Reverse Certificate List
Subject: CA1	Subject: CA2 Issuer: CA1
Issuer: CA2	Subject: CA3 Issuer: CA1

Certification Path Construction Information Data Structure is Defined as:

typedef   stru   PathConInfo{

        char Infortype;

        char DestinationCA;

        char SourceCA;

        char   *CAList;

        }NodeList;

Where Information type means requesting information or responding information of path construction; destination CA(CA is classified into two types: source CA and destination CA) is the distinguished name(DN) of destination CA. Path CA list means all the DNs of the CA or the DNs of certificate chains in serial numbers, forming the certificate path. When the source CA constructs its path construction requesting information, it adds all its DNs to the path CA list.

3.2   Algorithm Descriptions

When receiving the requests of the terminal entity, the source CA creates path construction requesting information and transmits it to all the subject CAs of the reverse certificate in the cross certification to the pair directory.

Receiving a piece of path construction requesting information, a CA responds as follows:

(1)   If the CA is the destination CA itself, it adds all its DNs to the path CA list, changing the requesting information information type volume into responding information, then transmits it to the source CA;

(2)   If the CA is not the destination CA itself, it checks the path CA list to see whether it is in the list or not. If it is in, the requesting information should be crossed out in order to avoid circuit. If it is not, it transmits the requesting information to all the CAs except the one previously transmitting the requesting information into it. The process of the formation of the multi-path construction from CA1 to CA7 is illustrated as in Figure 3.

When the process is over, CA1 receives two pieces of path construction responding information. Path construction responding information in the path CA list means the complete path from the source CA to the destination CA. The responding information indicates two paths: CA1→CA2→CA5→CA6 and CA1→CA3→CA7→CA4→CA6.

CA1 creates a path construction and transmits the requesting information to CA2 and CA3.

CA2 transmits the requesting information to CA5; CA3 transmits the requesting information to CA7.

CA4 transmits the requesting information to CA6; CA6 transmits the responding information to CA1; CA7 transmits the requesting information to CA4.

CA5 transmits the requesting information to CA6; CA6 transmits the responding information to CA1; The requesting information from CA4 to CA3 is given up for a circuit has been checked.

Fig. 3. The construction process of dynamic path from ca1 to ca6

3.3   The Analysis on the Performance of the Algorithm

Based on the avove mentioned analysis, it can be concluded that if a path exists between the source CA and the destination CA, the path construction requesting information is sure to reach the destination CA. As a result, all the possible paths can be constructed. Additionally, the CA's cross certificate to the pair directory can truly reflects the current certification network structure---when the CA changes, the cross certificate to the pair directory changes accordingly. And the CA transmits the path construction requesting information based on the cross-certificate to the pair directory. To sum up, the path construction algorithm is dynamic.

Multi-paths existing, path validation being under way while it is constructed (for example, checking whether the certificate satisfies the length restriction, naming restriction, strategy restriction, etc.), those unqualified paths can be crossed out beforehand. In other words, it is time-saving and economical, compared with the traditional methods, for those unqualified paths have been checked before the path validation is under way. Consequently, path management is more quickly and efficient.

A comparison is made between the cooperative and dynamic construction algorithm and the path construction algorithms based on certification network graph theory in terms of special complexity and time complexity.

(1) In terms special complexity, multi-path construction algorithm is superior to the path construction algorithm based on validation network graph theory for it needs space memory validation network graph and calculation based on the validation network graph, which is not vacant in the multi-path construction algorithm.

(2) The time complexity the depth-first search experiences by searching from the first node to the last one can be expressed with the following formula O $\left(\sum_{i=1}^{m} di + 1 \right)$, where di means the number of the subnodes of the ith certificate and L means the layers of the path, with validation network G（V，m） indicating the PKI system has several trust domains, where V means certificate and m means the scale of the system or the number of the certificates of the system. The time the dynamic path construction algorithm costs depends the transmission speed and the performance of the network.

(3) The dynamic path construction algorithm is more efficient for it does not need validation network graph.

4   Summaries

Based on the introduction of the relevant path construction algorithm, the thesis attempts to design the cooperative and dynamic path construction algorithm with the purpose of putting forward solutions to multi-paths. Taking relevant restrictions into consideration and simplifying the process of certificate validation, the cooperative and dynamic path construction algorithm can be further optimized. And the new algorithm can be designed as functional components to be commercially applied software for users, which is the future research focus. The new algorithm is expected to be of help to research on fundamental theory of quick charge of high gain li-ion battery.

References

[1] Gon, J., Liu, J.: Research on the Methods of PKI Extention Based on Path Findings. Computer Engineering and Science 22(4), 1–3 (2000)
[2] Bassham, L., Polk, W., Housley, R.: Algorithms and Identifiers for the Internet X.509 Public Key Infrastructure Certificate and Certificate Revocation List, Profile (2002)

[3] LGhosh, R.K., Bhattachar Jee, G.P.: Parallel breadth first Search algorithms for trees and graphs. Inter. J. Computer (12), 255–268 (1996)
[4] Housley, R., Ford, W., Polk, W.: Internet X.509 Public Key Infrastructure: Certificate and CRL Profile. Internet Request for Comments 3280 (2002)
[5] Wang, X., Wang, S., Xiang, C.: The Path Construction Algorithm in PKI Infrastructure. Computer Engineering 28(6), 69–70 (2002)

Particle Swarm Optimization Based Active Contour Model

Liu Xiaogang, Ren Xuemei, and Xue Guangyue

School of Automation, Beijing Institute of Technology, Beijing, China
`liuxiaogang@bit.edu.cn`

Abstract. In this paper, an active contour model is proposed for contour extraction based on a new particle swarm optimization (NPSO) to overcome the traditional active contour model's sensitivity to noise and initial position. The traditional snake is used to define the initial region and the fitness function for the swarm. Then, the contour of the object can be found using NPSO. Experiments are provided to show that the proposed model can not only extract the accurate contour but also reach good anti-noise performance.

Keywords: Particle swarm optimization, Active contour model, Contour extraction.

1   Introduction

Active contour models, popularly known as snakes, are first proposed by Kass [1]. Because of their good performance, they are often utilized in detection of contours, motion tracking [7] and stereo matching. But the tridition models are very sensitive to the initial position of the snakes and noise. Many kinds of modified models are proposed to improve the performance of the snakes. The Gradient Vector Flow (GVF) model in [2] has a good performance in solving these two problems. Some other models based on B spline [4] and fishing swarm algorithm [8] are also used to improve the traditional model.

In 1990, a population-based, self-adaptive and stochastic optimization technique called particle swarm optimization (PSO) is proposed in [6]. PSO has been widely used in parameters optimization [5], [9]. But the traditional PSO is prone to converge to local minimum. Recently, a new particle swarm optimization (NPSO) is proposed in [3]. In the NPSO, the bad experience component is also counted. With the consideration of the bad experience, the searching capability of the swarm has been enhanced.

In this paper, a model combining the traditional active model and the NPSO is proposed for contour extraction. Based on the global searching capability of the NPSO, the proposed model can extract the contour of the object more accurately and be robust to Gaussian and uniform distributed noises.

The rest of the paper is organized as follows. The basic mathematical description of the NPSO-snake and the details of the process of the algorithm are given in Section 2. The analysis of the experiments is given in Section 3. Finally, Section 4 concludes the paper.

2   Active Contour Model Based on NPSO

2.1   Active Contour Model

The traditional active contour model (snakes) is a controlled continuity spline under the influence of image forces and external constraint forces. The internal spline forces provide a smoothness constrain. The image forces push the snake toward salient image feature like lines, edges and subjective contours. The external constraint forces put the snake near the desired local minimum.

The position of a snake parametrically is represented by

$$v(s) = (x(s), y(s)) \tag{1}$$

where s is the contour curve parameter between 0 and 1. The energy function of the snake can be written as:

$$E_{sanke} = \int_0^1 [E_{int}(v(s)) + E_{image}(v(s) + E_{con}(v(s))] ds \tag{2}$$

where E_{int} is the internal energy of the spline due to bending, E_{image} represents the image forces, and E_{con} gives rise to the external constraint forces.

The internal spline energy is written as

$$E_{int} = (\alpha(s)|v_s(s)|^2 + \beta(s)|v_{ss}(s)|^2)/2 \tag{3}$$

where E_{int} is the internal energy. The first-order term $v_s(s)$ makes the snake act like a membrane controlled by $\alpha(s)$. The second-order term $v_{ss}(s)$ makes the snake act like a thin plate controlled by $\beta(s)$.

2.2   New Particle Swarm Optimization

The velocity update equation and position update equation of NPSO are given by

$$V_{ij}^t = \omega \times V_{ij}^{t-1} + C_{1g} \times r_1 \times (Pbest_{ij}^{t-1} - X_{ij}^{t-1})$$
$$+ C_{1b} \times r_2 \times (X_{ij}^{t-1} - Pworst_{ij}^{t-1}) \tag{4}$$
$$+ C_2 \times r_3 \times (Gbest_i^{t-1} - X_{ij}^{t-1})$$

$$X_{ij}^t = X_{ij}^{t-1} + V_{ij}^t \tag{5}$$

$$i = 1, 2, ..., N_D, \quad j = 1, 2, ..., N_{par}$$

where t is the iteration count; V_{ij}^t and X_{ij}^t respectively represent the ith dimension of the velocity and position in particle j at tth iteration; ω represents the inertia weight; $Pbest_{ij}^t$ and $Pworst_{ij}^t$ are the ith dimension of the own best and worst position of particle j at tth iteration; $Gbest_i^t$ represents ith dimension of the best particle in the swarm at tth iteration; N_D is dimension of the optimization problem; N_{par} represents

the number of particles in the swarm; C_{1g} and C_{1b} are weight components; r_1, r_2, r_3 are three separately generated uniformly distributed random numbers in the range [0, 1].

It should be noted that the inclusion of the worst experience component, in the behavior of the particle, gives additional exploration capacity to the swarm.

2.3   Active Contour Model Based on NPSO

It has been proved that the traditional contour model [1] is not robust to noise. The NPSO based active contour model in this paper is able to reduce the influence of the noise in the contour extraction.

For the proposed active contour model, every snake point is treated as a set of particles. First, the initial contour of the snake is defined; then swarms are trained for every snake points by (6) and (7). Once all the swarms converge to their best position, the snake extracts the contour of the object.

$$
\begin{aligned}
V_i(t) = \omega \times V_i(t-1) &+ C_{1g} \times r_1 \times (Pb_i(t-1) - X_i(t-1)) \\
&+ C_{1b} \times r_2 \times (X_i(t-1) - Pw_i(t-1)) \\
&+ C_2 \times r_3 \times (Gb(t-1) - X_i(t-1))
\end{aligned}
\tag{6}
$$

$$
X_i(t) = X_i(t-1) + V_i(\text{t}), \quad i = 1, 2, \dots, N
\tag{7}
$$

where $V_i(t)$ and $X_i(t)$ are respectively the velocity and position of particle i at time t; $Pb_i(t)$ and $Pw_i(t)$ are the own best and worst position of particle i until time t respectively; $Gb(t)$ is the best particle at tth iteration; N is the number of the swarm.

The fitness function of the particles is defined by choosing the external energy of the snake energy function as

$$
f(x, y) = |\nabla I(x, y)|^2
\tag{8}
$$

where $f(x, y)$ is the fitness value at the point (x, y).

Since the particles have the trend to extend their searching space, they can search new area and avoid the influence of the noise. The first part of (6) can make particles have the ability of certain global search. The third part of (6) gives an additional exploration capacity to the swarm.

The process of the algorithm in this paper can be summarized as follows.

1)   Define the initial position of the snake points;
2)   Find out the center point C of the image region determined by

$$
x_c = \frac{m_{10}}{m_{00}}, \quad y_c = \frac{m_{01}}{m_{00}}
\tag{9}
$$

where $m_{p,q}$ is the central moment of order (p+q)

$$
m_{p,q} = \sum_{x=1}^{M} \sum_{y=1}^{N} x^p y^q f(x, y)
\tag{10}
$$

3) Define the searching regions of the swarms;
4) Calculate fitness value for each snake point;
5) For every particle of the swarm corresponding to each snake point, update its best fitness value Pb and worst fitness value pw;
6) For every swarm of snake points, update its fitness value gb;
7) Evolve the speed and the position of the particles according to (6) and (7);
8) If do not fulfill the termination condition, return to 4);
9) If every snake point converges to the image edge, then the algorithm is finished.

3   The Result and the Analysis of the Experiments

The experiments are realized by VC++6.0 under the Window XP operating system. Two pictures of about 400 by 400 pixels are used in the following experiments. The first one is a yellow ball whose contour is simple and shown in Fig.1 and the second one is a blue pentagram with a more complex contour described in Fig.2. It can be seen from the figures that the red points are the snake points and the number of the snake point of the contour model is set ten. In the traditional model the α, β are set to 0.2 and γ is set to 0.4.

In Fig.1 and Fig.2, (a) is the initial image. (b)-(d) show the convergence performances of the NPSO-snake for image (a) without noise, with Gaussian noise and with uniform distributed noise, respectively. (e) and (f) are the convergence performances of the traditional snake for image (a) with Gaussian and uniform distributed noises, respectively.

(a) Initial image

(b) Convergence of NPSO-snake

(c) Convergence of the NPSO-snake with Gaussian noise

(d) Convergence of the NPSO-snake with uniform distributed noise

(e) Convergence of the traditional active contour model with Gaussian noise

(f) Convergence of the traditional active contour model with uniform distributed noise
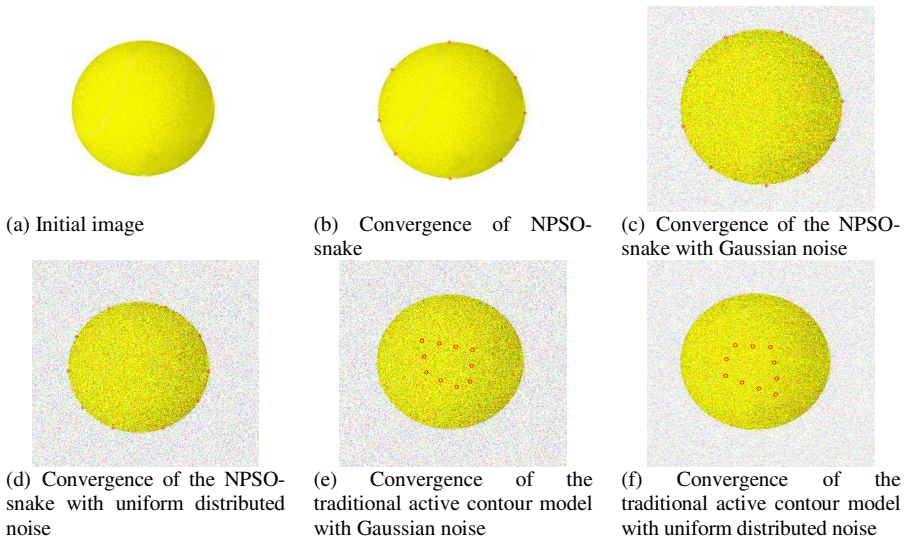
Fig. 1. The performance of the NPSO-snake and the traditional active contour of the object with simple contour

(a)Initial image

(b) Convergence of NPSO-snake

(c) Convergence of the NPSO-snake with Gaussian noise

(d) Convergence of the NPSO-snake with uniform distributed noise

(f) Convergence of the traditional active contour model with Gaussian noise

(e) Convergence of the traditional active contour model with uniform distributed noise
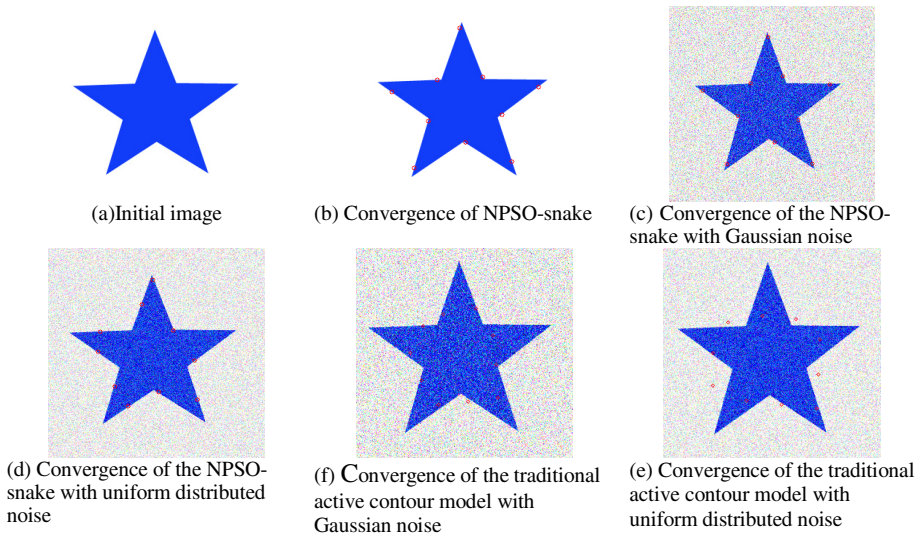
Fig. 2. The performance of the NPSO-snake and the traditional active contour of the object with concave contour.

From Fig.1 and Fig.2, we can see that compared with the traditional snake, the proposed NPSO-snake model can obtain a higher accuracy and be more robust to Gaussian and uniform distributed noises. Furthermore, NPSO-snake model can converge to concave contours, which is more convenient to practical application.

4   Summaries

An active contour model (snake) is proposed based on a NPSO in this paper. The initial region and the fitness function for the swarm are defined by the traditional snake. Then the NPSO algorithm is utilized to find the object contour. The proposed model is less sensitive to noise and initial position than the traditional active contour model. Experiments are given to show the effectiveness and good anti-noise performance of the presented NPSO-snake model.

References

[1] Kass, M., Witkn, A., Terzopulos, D.: Snakes: active contour models. Int. J. Comput. Vis. 1, 321–331 (1987)
[2] Xu, C., Prince, J.L.: Snakes, Shapes, and Gradient Vector Flow. IEEE Trans. Image Processing 7, 359–369 (1998)

[3] Immanuel Selvakumar, A., Thanushkodi, K.: A New Particle Swarm Optimization Solution to Nonconvex Economic Dispatch Problems. IEEE Trans. Power Systems 22(1), 42–51 (2007)

[4] Su, Y., Chen, H.: Active Contour Model Based On B-spline. Sience Technology and Engineering 7(9), 2010–2014 (2007)

[5] Qiao, W.: Study of short-term power load forcasting based on improved PSO-BP neural network model 35(17), 17–21 (2007)

[6] Chowdhury, B.H., Rahman, S.: A review of recent advances in economic dispatch. IEEE Trans. Power Syst. 5(4), 1248–1259 (1990)

[7] Lv, X., Huang, X.: Visual Tracking of Robotic Man ipulator Motion Trajectory Using Dynamic Active Con tour Model. Control and Decision 21(10), 1143–1147 (2006)

[8] Yang, X., Li, S.: Artificial Fish Swarm Algorithm of Snake Model. Computer Technology and Development 20(9), 61–65 (2010)

[9] Liu, D., Feng, Q., Jiang, Q.: Parameter Optimization of Maglev PID Controller Based on Improved PSO algirithm. Journal of Southwest Jiaotong University 45(3), 405–410 (2010)

Research on Mobile IPV6 Technology and Handover Performance Optimization

Mo Lin-Li

School of Software, East China Jiao Tong University, Nanchang, 330013, China
`mllzhao@163.com`

Abstract. The working principle and handover process of MIPv6、FMIPv6 and HMIPv6 are analyzed firstly, some problems existing in handover process and crucial factors impacting on handover performance are pointed out next. Then several optimization programs about mobile handover are proposed. FHMIPv6 combines the advantages of FMIPv6 and HMIPv6, and is more suitable for the multimedia real-time requirement. A handover algorithm of HMIPv6 based on fast DAD can reduce the intra-domain and inter-domain delay effectively. An optimization program based on optimal routing functions can obviously reduce the binding registration delay and packet loss, but it still need to be validated in complex and real-time network environment. Finally, the research outlook of MIPv6 technology is proposed.

Keywords: Handover delay, Care-of-Address, Duplicate address detection, Signaling interactions.

1   Introduction

With the rapid development of Internet and mobile communication technology, the combination of wireless mobile access technology and IPv6 has been a new research focus. The Internet Engineering Task Force (IETF) Mobile IP Working Group standardized RFC3775 for supporting Mobile IPv6 (MIPv6) in 2004[1]. In MIPv6 protocol, when a Mobile Node (MN) moves to new subnet, its communication may be disconnected for a few seconds, which is called handover delay. If the time of handover delay is more, the packets will be lost, network resource will be wasted, which will seriously affected network performance. Therefore, how to enhance the handover performance in MIPv6 is a research focus.

2   MIPv6 Handover Performance Analyses

In MIPv6, when in local subnet, MN communicates normally with other nodes; when moving to a new subnet, MN will disconnect with previous access router(PAR), then connects with a new access router(NAR). This procedure is called mobile handover, which includes link-layer procedure and network-layer procedure. The network-layer handover procedure is initiated only after the link-layer handover procedure comes to

end. And the network-layer handover process includes Movement Detection, care-of address configuration, Duplicate Address Detection and Binding Update.

DL2 denotes the link-layer delay in this paper, which is different because of using different devices. For example, DL2 is 200ms~1500ms in 802.11b.

Movement Detection (MD) procedure is to determine whether or not it moves to a new subnet, the delay of MD is marked DMD. In MIPv6 the value of DMD is 0.5-1.5s.

After completing MD, MN configures its New Care-of Address (NCoA), and this delay is marked DNCoA. If using stateful address auto-configuration, the value of DNCoA equals to the time of configuring DHCP server. If using stateless address auto-configuration, MN generates a NCoA through adding its interface ID to the router prefix information, and the delay of this way can be ignored.

To verify the uniqueness of the NCoA, MN should run Duplicate Address Detection(DAD) procedure, before assigning the address to its interface. D_{DAD} which denote the delay of DAD procedure in this paper is a higher percentage of the whole delay of MIPv6.

Through the Binding Update (BU) procedure, MN registers its temporary location to its Home Agent (HA) in its home network and Correspondent Node (CN) every time it moves, which brings a lot of signaling messages, and increases network load. $D_{BUHA/CN}$ denotes the delay of BU in this paper.

Fig.1 shows that $D_{MIPv6}= D_{L2}+D_{MD}+ D_{NCoA} + D_{DAD} +D_{BUHA/CN}$, D_{MIPv6} denotes the whole delay of MIPv6, among which D_{MD}, D_{DAD}, and $D_{BUHA/CN}$ can be optimized. Because of long delay of handover in MIPv6, MN can not receive packets when in handover procedure, and the ratio of data loss is high.
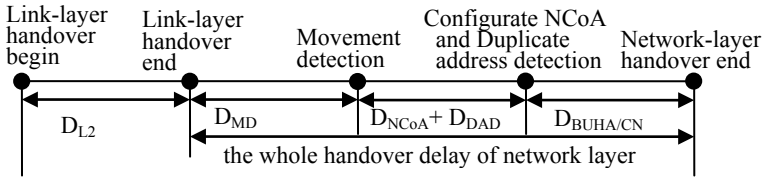

Fig. 1. Handover delay of MIPv6

3   FMIPV6 Handover Performance Analyses

IETF standardized Fast Handover for Mobile IPv6 (FMIPv6) for supporting IPv6 mobility in 2005[2]. In FMIPv6, before moving to a NAR, MN generates its NCoA which can be used in the region of NAR through signaling messages. If MN hasn't generated its NCoA after connecting to the NAR, it should establish a tunnel from the PAR to the NAR to forward packets.

FMIPv6 technology can be divided into predictive handover and reactive handover. When connecting with PAR MN receives FBACK message, which is called predictive handover. When disconnecting with PAR MN hasn't received FBACK message, which is called reactive handover. Fig.2 shows the procedure of predictive handover.

In predictive handover, MN communicates in the region of PAR before handover. When detecting a new network, MN sends a Router Solicitation for Proxy (RtSolPr) to
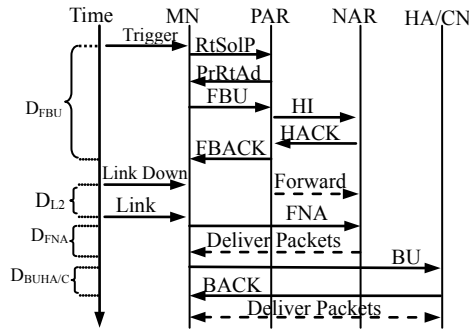
Fig. 2. Predictive handover procedure

the PAR, the PAR will responds to MN a Proxy Router Advertisement (PrRtAdv) message including the information of the NAR's network premix and IP address. Then MN generates an NCoA and sends a Fast Binding Update (FBU) message to the PAR, the PAR immediately sends a Handover Initiate (HI) message with MN's NCoA to the NAR. The NAR verifies that the NCoA can be used on the NAR's link, and responds to PAR a Handover Acknowledge (HACK) message. Then PAR sends a Fast Binding Acknowledgement (FBACK) message to MN, and begins forwarding packets to the NAR by bidirectional tunnel between PAR and NAR. In this paper, D_{FBU} denotes the time from MN's beginning to detect the NAR to this moment. Since then MN disconnects with PAR, and can not receive packets. Link-layer handover begins.

After moving to the region of NAR, MN sends a Fast Neighbor Advertisement (FNA) message to NAR asking for packets. Then NAR starts to forward buffering packets to MN, and then the communication between MN and CN is restored. This paper use D_{FNA} denoting the above time.

Finally, MN sends BU message to HA and CN, then HA and MN responds to MN by Binding Acknowledgement (BACK), the whole handover procedure comes to end. Therefore $D_{FMIPv6} = D_{FBU} + D_{L2} + D_{FNA} + D_{BUHA/CN}$, D_{FMIPv6} denotes the whole delay of FMIPv6.

There are several problems in FMIPv6. Because MN registers its NCoA to HA and CN every time it moves, the registration signaling overhead is still high; when MN moves repeatedly and fast between two adjacent routers, the handover performance of FMIPv6 will be deteriorated, and more data lost; when MN has completed the network-layer handover and does not binding update with HA and CN, MN moves again from current NAR to another NAR, more data lose.

4   HMIPV6 Handover Performance Analyses

In MIPv6 and FMIPv6, when MN far from its HA and moving through more subnets, frequent handover brings more signaling messages, and increases delay. IETF standards Hierarchical Mobile IPv6 (HMIPv6) protocol[3], which using local registration mechanism to update binding update procedure, reduce the number and delay of MN's registration to HA and CN.

HMIPv6 uses a local anchor point called Mobility Anchor Point(MAP). The MAP's function is equal to the local agents of foreign network; it deals with MN's intra-domain movement. The domain managed by MAP can be divided into several subnets. The Fig.3 shows the system architecture of HMIPv6.

In HMIPv6, MN obtains Regional Care-of Address(RCoA) from MAP and uses RCoA to communicate with HA and CN; MN also obtains Link Care-of Address(LCoA) from access router and communicates with MAP using LCoA. When moving in a MAP's intra-domain, MN only registers LCoA to MAP (this paper use D_{LBU} denoting the delay of this registration), and needn't send BU message to HA and CN because of RCoA having no change. This intra-domain handover is very fast, so $D_{IHMIPv6} = D_{FBU} + D_{L2} + D_{FNA} + D_{BUHA/CN}$, $D_{IHMIPv6}$ denotes the whole delay of intra-domain HMIPv6.

When performing inter-domain movement, MN needs to select a new MAP as its regional agents, then sends binding update message to the new MAP to bind LCoA and RCoA. Then MN also sends binding update message to HA and CN to bind HA and RCoA. Let us denote by $D_{OHMIPv6}$ the whole delay of inter-domain movement, it can be concluded that $D_{OHMIPv6} = D_{IHMIPv6} + D_{BUHA/CN}$.

HMIPv6 protocol reduces effectively the times of MN's registration to HA and CN through local registration mechanism, and has obvious improvement in signaling cost. But the map discovery protocol in HMIPv6 still needs to improve; and the latency of MD and NCoA's configuration is also not been shortened.
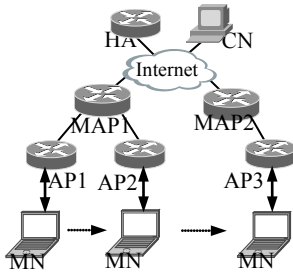

Fig. 3. System architecture of HMIPv6   Fig. 4. FHMIPv6 handover process
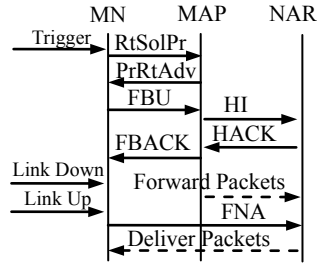
5   Optimization and Improvement of Mobile Handover Technology

5.1   FHMIPv6 Technology and Handover Performance Analysis

Fast Handover for Hierarchical MIPv6 (FHMIPv6) protocol combines the advantages of HMIPv6 and FMIPv6 [4]. The network architecture of FHMIPv6 is similar to HMIPv6's architecture which shown in Fig.3.

According to the Fig.3, in the simple combination way, HMIPv6 protocol is used to connect the MAP and access router, and FMIPv6 protocol is used to realize handover between the two access routers. The packets CN sending to MN will be delivered to PAR through MAP, and then PAR will use tunnel to follow the packets to NAR. Therefore, it will bring triangle routing problem, and increase handover delay, waste communication bandwidth.

In the optimized combination method, the Fig.4 shows the handover procedure of the figure 3. ① MN sends a RtSolPr message to MAP to ask for the NAR's information. ② The MAP replies a RtSolPr message, including the NAR network prefix and IP address. ③ MN generates a New Link Care-of Address (NLCoA) according to the NAR's network prefix and interface address, and sends FBU message to MAP. ④ The MAP sends a HI message including MN's NLCoA to the NAR. ⑤ After receiving HI message, the NAR firstly verifies that the NLCoA can be used on the NAR's link, and then responds with a HACK message to MAP. ⑥ The MAP responds a FBACK message to MN, and begins forwarding packets to the NAR by bidirectional tunnel between MAP and NAR. ⑦ MN begins link-layer handover. After completing link-layer handover, MN sends a FNA message to NAR. Then NAR forwards the buffering packet to MN. The whole handover procedure comes to end.

Based on the above analysis we can see that the FHMIPv6 protocol can shorten the registration delay, and reduce the disruption in the communication process. But it requires higher device preference; MN needs periodically send location information when moving to a coverage area covered by several routers. Thus it will generate more signaling, and increase the processing load of MAP.

5.2   Duplicate Address Detection Optimization

The delay of DAD is a large proportion in the whole MIPv6 delay. The every DAD operation consumes a random time which is 1s in IETF RFC2462. But the delay of multimedia real-time business is less than 600ms. In order to reduce the DAD delay's impact on handover performance, RFC3315 and RFC 4429 optimize the DAD operation.

Through NS2 simulation, a handover algorithm of HMIPv6 based on fast DAD can reduce the intra-domain handover delay to 202ms, and the inter-domain handover delay to 339ms [5]. Therefore it can meet the mobile real-time requirement of multimedia business. But it should be authenticated in more complex experimental environment, and it also brings the load of address pool management to NAR and MAP.

5.3   Binding Registration Optimization

In HMIPv6 protocol, the intra-domain binding registration delay of MN can be reduced effectively, but the inter-domain binding registration delay has no improvement. A new enhanced forwarding mechanism reconstructs FBU message from MN to PAR, it can rapidly establish the tunnel between PAR and NAR, shorten binding delay, and reduce the handover delay [6].

An optimization program based on optimal routing functions has been proposed [7]. In this optimization program, when switching intra-domain, MN just needs to find the optimal functions routing FR* between the two MAP according to optimal routing algorithm, and registers to the FR*. Obviously, the distance between MN and FR* will be less than the distance between MN and HA. The results of experimental simulation show that this new method can reduce greatly the binding delay and packet loss.

6   Summaries

Reducing the handover delay is one of key issues of MIPv6 technology. The FMIPv6 mechanism reduces address configuration delay to speed handover process; the HMIPv6 mechanism uses MAP management to reduce the binding update delay and network signaling load. The FHMIPv6 mechanism is more suitable for the multimedia real-time requirement, but it increases the MAP's load dealing with signaling interactions. This paper also proposes several optimization programs about the delay of DAD and BU, but these programs just past simulation, and still need to be validated and improved in real network environment. With increasing quality requirements of mobile Internet services, the future research should focus on address allocation policy based on FHMIPv6, and reduce or cancel DAD operation. Therefore it is necessary for future research to find a balance point to reduce the handover delay, improve network performance, and reduce the complexity of protocol and network load.

References

[1] Johnson, D., Perkins, R., Arkko, J.: RFC3775 Mobility Support in IPv6[S], IETF (2004)
[2] Koodli, R.: Fast Handovers for Mobile IPv6[S]. RFC 4068 (June 2005)
[3] Soliman, H., Catelluccia, C., et al.: Hierarchical Mobile IPv6 Mobility Management (HMIPv6), RFC4140 (August 2005)
[4] Hee, Y.J., Seok, J.K., Soliman, H., El-Malki, K.: Fast Handover for Hierarchical MIPv6(F-HMIPv6). draft-jung-mobileip-fastho-hmipv6-04.txt (June 2004)
[5] Chen, W.-X., Han, G.-D., Liu, H.-B.: Handover Algorithm using Fast DAD Mechanism for Hierarchical Mobile IPv6. Jouranl Communications, 115–120 (January 2008)
[6] Gwon, Y.: Enhanced Forwarding from the Previous Care-of Address(EFWD) for Fast Handovers in Mobile IPv6. In: WCNC (March 2004)
[7] Liu, Z., Cai, M.: A Hierarchical Mobile IPv6 Optimization Scheme. Computer Engineering & Science, 27–29 (July 2010)

An Efficient Variant of the Batch RSA Cryptosystem

Guang Zhao[1] and Hengbo Li[2]

[1] Sports Information &Technology Department
Wuhan Sports University
Wuhan, China
`smalllittlecat@126.com`
[2] School of Software
Nanyang Institute of Technology
Henan, China
`kaixinyezi@126.com`

Abstract. This paper aims at speeding up Batch RSA decryption. The performance of Batch RSA decryption has directs relationship with the efficiency of the full-sized modular exponentiation implementation in the Exponentiation Phase [2]. This paper proposes a variant (BS1PRSA-Batch RSA-S1 Multi-Power RSA) to improve the Batch RSA decryption performance. The experimental results show that the speed of the decryption has been substantially improved.

Keywords: Batch RSA, decryption, acceleration, modular exponentiation.

1   Introduction

The RSA cryptosystem [1] is one of the widely used public key systems. It is the main operation of RSA to compute modular exponentiation. Since RSA is based on arithmetic modulo large numbers, it can be slow in constraining environments. Especially, when RSA decrypts the ciphertext and generates the signatures, more computation capacity and time will be required. So some variants of RSA were proposed to speed up RSA decryption. Fiat [2] is the first to propose speeding up RSA decryption via batching. Suppose there are two messages to encrypt M_1 and M_2 respectively with the small public keys $e_1 =<N,3>$ and $e_2 =<N,5>$, Fiat showed that it is possible to decrypt the ciphertexts $V_1 = M_1^3 \bmod N$ and $V_2 = M_2^5 \bmod N$ for approximately the price of a single RSA decryption. Let $A = (V_1^5 \cdot V_2^3)^{1/15}$, and then compute:

$$M_1 = V_1^{1/3} = \frac{A^{10}}{V_1^3 \cdot V_2^2} \qquad \text{and} \qquad M_2 = V_2^{1/5} = \frac{A^6}{V_1^2 \cdot V_2}$$

Hence, at the cost of computation a single 15 'th root both V_1 and V_2 can be decrypted. Note that the public exponents e_1 and e_2 have to be chosen small.

Otherwise, the extra arithmetic required is too expensive. Also, notice that one can only batch-decrypt ciphertexts encrypted using distinct public exponents [2] [3].

Batch RSA waits for more than one RSA decryption and performs one full-sized exponentiation for all decryptions, it can spare a lot of running time capacity. But, the full-sized exponentiation is the main cost operation in the Batch RSA. Reducing modulus in modular exponentiation is a technique to speed up the RSA decryption. Multi-Power RSA is based on modifying the structure of the RSA modulus to the form $N = p^{b-1}q$ with p and q $\lfloor n/b \rfloor$-bits each [3-4]. It which speeds up the RSA decryption reduces the size of the moduli.Another method which speeds up the RSA decryption is to shift some work to the encryption and the exponents of decryption become small numbers. This paper proposes a new variant of Batch RSA based on the reduction of the modulus and private exponents in modular exponentiation of the decryption. The variant can not only speed up Batch RSA decryption but it also guarantees the security of Batch RSA cryptosystem.

2   The New Proposed Variant

In this section, a new variant of RSA is proposed and is called BS1PRSA (Batch RSA-S1 Multi-Power Improved RSA) in the paper.The variant effectively combines Multi-Power RSA [3-4] and RSA-S1 system [5-6] based on Batch RSA. It can obtain a higher speedup than the Batch RSA and the above two RSA variants. Before the proposals of optimizing the RSA cryptosystem [1] is presented, the RSA basic algorithms will be reviewed.

2.1   Basic RSA system

In general, there are three phases in RSA cryptosystem.

Key generation: The key generation algorithm takes a security parameter n as input and chooses two distinct random $(n/2)$-bit primes p and q. their product $N = p \times q$, is called the RSA modulus. Next, the algorithm picks some small value e that is relatively prime to $\varphi(N) = (p-1)(q-1)$. The e value is usually chosen as e=65537. Finally, the integer d is computed as the multiplicative inverse of e modulo $\varphi(N)$. The pair $<N,e>$ is the public key and the pair $<N,d>$ is the private key. The integer e is called the encryption exponent while the integer d is called the decryption exponent. Typically, one sends the public key $<N,e>$ to a Certificate Authority to obtain a certificate for it.

Encryption: A plaintext message $M \in Z_N$ is encrypted by raising it to the eth power modulo N. The ciphertext is computed as $C \leftarrow M^e \bmod N$.

Decryption: A ciphertext message $C \in Z_N$, for a given plaintext message M, is decrypted by raising it to the dth power modulo N. This follows since $C^d (\bmod N) \equiv M^{ed} (\bmod N)$ and $M^{1+k\varphi(n)} (\bmod N) \equiv M (\bmod N)$. k is the unique integer satisfying $ed = 1 + k \cdot \varphi(N)$ called the key equation.

The standard practice of the basic RSA employs the Chinese Remainder Theorem (CRT) [7] to speed up the decryption. When the RSA employs the CRT for RSA decryption, the private key becomes the $< N, d_p, d_q, p, q >$, where $d_p \equiv d \bmod p - 1$ and $d_q \equiv d \bmod q - 1$. The d_p and d_q are called the CRT-exponents. The decryption algorithm employing the CRT is as follows:

For $M_p = M^{d_p} \bmod p$ and $M_q = M^{d_q} \bmod q$, the algorithm gets the $C^d = M_p \times n_p + M_q \times n_q (\bmod N)$, where $n_p = q \times (q^{-1} \bmod p)$ and $n_q = p \times (p^{-1} \bmod q)$. Because n_p and n_q can be pre-computed, since exponentiations mod p or q are more efficient to compute than those mod N, using the CRT can obtain an approximate factor of 4 speed-up of private –key operations.

Using the criterion presented in [3], one makes a theoretical cost estimation in terms of the number of exponentiations executed by the two methods. Basic algoritms to compute exponentiations of the form $C^d \bmod N$ take time $O(\log d \log^2 N)$. When d is on the order of N the running time is $O(\log^3 N)$. Theoretical speedup of RSA using CRT with relation to the original RSA is simply: $\dfrac{\log^3 N}{2(\log N / 2)^3} = 4$.

2.2   RSA-S1 Systems

The RSA-S1 system [5] was originally proposed as a way to reduce load on small devices (smartcards) by shifting some heavy-weight cryptographic computation to more powerful server-host computers equipped with smartcard readers. The system was improved in [6][8-14]. The detail of the RSA-S1 system is as follows:

- The RSA private exponent d is represented as
  $d = d_1 \cdot f_1 + ... + d_k \cdot f_k (\bmod \phi(N))$, where the d_i's and f_i 's, $1 \le i \le k$, are random vector elements of c and $|n|$ bits, respectively.
- Servers sends vector $D = (d_1, d_2, ..., d_k)$ to client.
- Client computes vector $Z = (z_1, z_2, ..., z_k)$ and sends Z back to client, where $z_i = x^{d_i} (\bmod N)$, for $1 \le i \le k$.
- Finally, server computes $\prod_{i=1}^{k} z_i^{f_i} = \prod_{i=1}^{k} x^{f_i d_i} = x^d \pmod{N}$.

The [10] adopts RSA-S1 system to speed up the decryption by assigning the roles: "client" becomes encryption and "server" becomes decryption. The main idea is to shift some computational burden from the decryption to the encryption. The resultant technique is called Encrypt Assistant RSA (EARSA) in this paper.

2.3   BS1PRSA

In this subsection, BS1PRSA is proposed and it is based on RSA-S1 system. It can obtain a higher decryption speedup than the original Batch RSA. BS1PRSA is

described as Batch RSA with four phases: Setup, Percolate-Up, Exponentiation-Phase and Percolate-Down[9-11].

Setup: Given a security parameter n and four additional parameters k, c and b as input. b is the batch size.

- Compute two distinct primes p,q each one $\lfloor n/3 \rfloor$ bits in length and generate
  $N = p^2 q$ and $\phi(N) = (p-1)(q-1)$.

- Let $e_1, \cdots, e_b$ be b different encryption exponents, relatively prime to $\phi(N)$ and to each other. The public exponent e_i should be very small. Otherwise, the extra arithmetic required is too expensive. Each e_i computes, $1 \leq i \leq b$, $d_{ei} = e_i^{-1} \bmod \phi(N)$ and $E = \prod_{i=1}^{b} e_i \pmod N$. Compute private exponent $d = d_{e1} \times d_{e2} \cdots \times d_{eb} \bmod \phi(N)$. Compute $e\_inv\_p = e^{-1} \bmod p$ and $p^2\_inv\_q = (p^2)^{-1} \bmod q$.

  Represent the private exponent d as $d = f_1 d_1 + ... + f_k d_k \bmod \phi(N)$, where the d_i's and f_i's, $1 \leq i \leq k$, are random vector elements of c and $|n|$ bits, respectively. The choice and security of parameters: k and c are discussed later. In the paper, k is chosen 2 and d's representation becomes $d = f_1 d_1 + f_2 d_2 \bmod \phi(N)$ storing it for use in Exponentiation-Phase.

- The private key is $< N, d_1, d_2 >$ and the public key is $< N, e_i, f_1, f_2, e\_inv\_p, p^2\_inv\_q >$ for each encryption, for $1 \leq i \leq b$.

- Given messages $m_1, ..., m_b$ and v_i is computed by $v_i = m_i^{e_i} \pmod N$, for $1 \leq i \leq b$.

In the Step phase, the private exponent d of original Batch RSA is divided to two d_i's vectors and two f_i's vectors, for $k=2$. The elements of f_i's vector are big numbers computed in the encryption and the decryption computes the small elements of the d_i's vectors. The decryption load is transferred to encryption by above methods and Batch RSA decryption is speeded up.

Percolate-Up: Computation process is as a binary tree with b leaves, and every inner node has two children. Nodes in the tree deal with pairs (V, E) where the component called V is a value, and the component called E is an exponent. Because the private exponent d is divided to the d_1, d_2, f_1 and f_2 vectors, for $k=2$, and the f_1 and f_2 vectors are computed in the phase. So, the original goal of getting $V = \prod_{i=1}^{b} V_i^{E/e_i} \pmod N$ in Percolate-Up becomes getting the $V_{f1} = \prod_{i=1}^{b} (V_i^{E/e_i})^{f_1} \pmod N$, where $E = \prod_{i=1}^{b} e_i \pmod N$, and the $V_{f2} = \prod_{i=1}^{b} (V_i^{E/e_i})^{f_2} \pmod N$. The Percolate-Up phase includes two processes. The first process is to get the (V,E) values that are stored to use in Percolate-Down. Another

process is to compute the V_{f1} and V_{f2}, and computing steps of the process will generate middle computations values (V,E) that do not need to store and the root node only need to store the V_{f1} and V_{f2}. There are some same computation steps in the two processes. Firstly, the first process is described as follows:

- Step1: Assign to each leaf node a public exponent $E = e_i$ and a ciphertext $V = v_i$ and pass (V, E) upward, for $1 \le i \le b$.
- Step2: Each interior node combines results passed upward from children into a result. The left-hand and right-hand children yield the results (V_L, E_L) and (V_R, E_R), and the interior nodes compute $V = V_L^{E_R} \cdot V_R^{E_L} \pmod N$, $E = E_R \cdot E_L$ and passes (V,E) upward. The process continues to carry out and is ended in the previous layer nodes of root node. The process generates (V_L, E_L) and (V_R, E_R) and they are stored for use during Percolate-Down.

The first process has been completed by above steps and it doesn't need to compute the V final value and the final V will be computed in the second process. The steps of the second process are as follows:

- The most contents of the second process are the same as those of the first process, except that the second process should compute the final two values $V_{f1} = (V_L^{E_R} \cdot V_R^{E_L})^{f_1} \pmod N$, for $k=2$, and $V_{f2} = (V_L^{E_R} \cdot V_R^{E_L})^{f_2} \pmod N$ in the root node. The final values $V_{f1} = V^{f_1} \pmod N$ and $V_{f2} = V^{f_2} \pmod N$, where $V = \prod_{i=1}^{b} V_i^{E/e_i} \pmod N$ and $E = \prod_{i=1}^{b} e_i \pmod N$, are obtained. The f_i modular exponentiation, $1 \le i \le k$, can be firstly done in the second process so that every encryption is able to share computation load, and the V_{f1} and V_{f2} are stored for use in the Exponentiation-Phase.

Exponentiation-Phase: In the exponentiation phase, the Eth root of V is the d and it has been computed in the Setup phase. The original goal of the computing $m = V^{1/E} \pmod N = V^d \pmod N$ in the phase becomes the getting $m = V^d = (V_{f1})^{d_1} \pmod N \cdot (V_{f2})^{d_2} \pmod N \bmod N$, where the $V_{f1} = V^{f_1} \pmod N$ and $V_{f2} = V^{f_2} \pmod N$ have been computed in the Percolate-Up phase and $d = d_1 \cdot f_1 + d_2 \cdot f_2 \pmod{\varphi(n)}$. In the phase, computation m is divided into two parts and they are $(V_{f1})^{d_1} \pmod N$ and $(V_{f2})^{d_2} \pmod N$. The two parts are respectively computed using the Chinese Remainder Theorem (CRT) [9]. Finally, m is the product of these two parts. BEARSA improves the performance of the Batch RSA decryption by transferring the $V^d \pmod N$ computation to encryption in the Percolate-Up phase. Firstly, the steps of computing the $(V_{f1})^{d_1} \pmod N$ by using CRT are as follows:

- Compute C_{fp} and C_{fq} by the $C_{fp} = V_{f1} \bmod p$ and $C_{fq} = V_{f1} \bmod q$, where V_{f1} has been computed in Percolation-Up.

- Compute M_{fq} by the $M_{fq} = C_{fq}^{d_1} \bmod q$, where d_1 has been computed in the Setup phase.

- Compute M'_{fp} by, where $C\_inv\_P = C^{-1} \bmod p$, the $M'_{fp} = (C\_inv\_P) \cdot C_{fp}^{d_1} (\bmod p)$ and C is the V_{f1}.

- Compute M_{fp} by using Hensel lifting [4][5] and the algorithm steps with modulus $p^2 q$ are as follows:

  1. $A = M'_{fp} \times C \bmod p$
  2. $p^2 = p \times p$
  3. $F = A^e \bmod p^2$
  4. $E = C - F \bmod p^2$
  5. $B = E \times M'_{fp} \times (e\_inv\_p) \bmod p^2$
  6. $M_{fp} = A + B \bmod p^2$

  The M_{fp} value can be computed by above steps.

- According to CRT, compute $V = M_{fq} - M_{fp} \bmod q$ and $V = V \times (p^2\_inv\_q) \bmod q$.

- Using the CRT to combine the $M = (V_{f1})^{d_1} (\bmod N)$'s to obtain $M = V \cdot p^2 (\bmod N)$ and $M = M + M_{fp} \bmod N$.

The value of $(V_{f2})^{d_2} (\bmod N)$ can also be obtained by the above same steps. The final m is computed by $m \equiv (V_{f1})^{d_1} \cdot (V_{f2})^{d_2} \bmod N$. The exponentiation yields the $m = V^{1/E} (\bmod N)$ and stores m for using in the Percolate-Down phase. In the phase, the CRT is used twice to compute $(V_e)^{d_1} (\bmod N)$ and $(V_f)^{d_2} (\bmod N)$. A full exponentiation of original Batch RSA, for $k=2$, is divided into two small exponentiations and the Batch RSA decryption performance is improved.

Percolate-Down: During the Percolation-Down phase, nodes in the tree deal with values called m.

- The root node passes the m that has been computed during Percolation-Up downward to the single child.
- An interior node has the m passed downward and needs to break it up to pass down further and computes $m_R = m^t / (V_L^{t_R} \cdot V_R^{t_L})(\bmod N)$ and

$m_L = m / m_R \pmod{N}$, where t is a value obtained via

CRT based on $t \equiv 0 \pmod{E_L}$ and $t \equiv 1 \pmod{E_R}$. $t_L = t / E_L$ and $t_R = (t-1) / E_R$. Pass m_L and m_R downward to the left-hand and right-hand children.

- The leaf node terminates the process and the b-th node has $m = M_i = C_i^{1/e_i} \pmod{N}$. All plaintext messages are obtained in the leaf nodes of binary tree.

At the end of the Percolate-Down process, each leaf's m is obtained by the above steps.

3 Performance and Experimental Results

3.1 Experiment Set-Up

The speedup is measured by the execution time of RSA decryption in the paper. These test algorithms were written by using the OpenSSL[15] cryptographic library (version 0.9.8 k). The hardware platform was 1.73GHz Intel Pentium Dual Core with 1 GB RAM running Windows XP Professional. In the experiment, the algorithms were called different names according to different parameters and values of the parameters. B6S1PRSA, B1S1PRSA and B2S1PRSA in the paper respectively indicate the BS1PRSA of the parameters $k=2$, $b=4$ and $c=64$-bit, the parameters $k=2$, $b=4$ and $c=128$-bit and the parameters $k=2$, $b=4$ and $c=256$-bit. Four small public key exponents are used and they are $e_1 = 3$, $e_2 = 5$, $e_3 = 7$ and $e_4 = 11$ in the experiments.

3.2 Experimental Results

Tab.1 lists the average decryption time for the four moduli with four Batch RSA variants.

Table 1. Decryption time For batch size=4

Variant	Decryption Time (msec)			
	1536	*2048*	*2560*	*3072*
B6S1PIRSA	15	16	26	32
B1S1PIRSA	16	18	30	40
B2S1PIRSA	25	28	39	54
Batch RSA	32	63	94	125

Table 2. Speedup related to decryption

Variant	Speedup to Batch RSA				
	1536	*2048*	*2560*	*3072*	*Average*
B6S1PIRSA	2.13	3.94	3.61	3.91	3.40
B1S1PIRSA	2.00	3.50	3.13	3.12	2.94
B2S1PIRSA	1.28	2.25	2.41	2.31	2.10

Tab.2 shows the speedups for the four moduli with three variants to the Batch RSA using CRT. The B6S1PRSA got the highest speedup in all variants. The results show the average speedup of B6S1PRSA is 3.40 is from 1536- to 3072-bit.

4   Conclusion

In this paper, the new Batch RSA variant which can improve the performance of the decryption was proposed. The variant can obtain high performance by transferring the decryption computations to encryption and reducing the modulus and private exponents. The next study will focus on: How to optimize the parameters of the new variant to make the variant get higher performance and security.

References

[1] Rivest, R., Shamir, A., Aldeman, L.: A Methoed for Obtaining DigitalSignatures and Public-key Cryptosystems. Communications of the ACM 21(2), 120–126 (1978)
[2] Fiat, A.: Batch RSA. In: Brassard, G. (ed.) CRYPTO 1989. LNCS, vol. 435, pp. 175–185. Springer, Heidelberg (1990)
[3] Boneh, D., Shacham, H.: Fast Variants of RSA. RSA Laboratories Cryptobytes 5(1), 1–8 (2002)
[4] Takagi, T.: Fast RSA-type Cryptosystem Modulo $p^k q$. In: Krawczyk, H. (ed.) CRYPTO 1998. LNCS, vol. 1462, pp. 318–326. Springer, Heidelberg (1998)
[5] Matsumoto, T., Kato, K., Imai, H.: Speeding up Secret Computations with Insecure Auxiliary Devices. In: Goldwasser, S. (ed.) CRYPTO 1988. LNCS, vol. 403, pp. 497–506. Springer, Heidelberg (1990)
[6] Castelluccia, C., Mykletun, E., Tsudik, G.: Improving secure server performance by re-balancing SSL/TLS handshakes. In: Proc of the 2006 ACM Symposium on Information, Computer and Communications Security, pp. 26–34. ACM, New York (2006)
[7] Quisquater, J.-J., Couvreur, C.: Fast decipherment algorithm for RSA public-key cryptosystem. Eletronic Letters 18, 905–907 (1982)
[8] Li, Y., Liu, Q., Li, T.: Design and Implementation of an Improved RSA Algorithm. In: EDT 2010, pp. 390–393 (2010)
[9] Li, Y., Liu, Q., Li, T.: 'Two efficient methods to speed up the batch RSA decryption. In: IWACI 2010, pp. 469–473 (2010)
[10] Li, Q., Liu, T.: Design and implementation of two improved batch RSA algorithms. In: ICCSIT 2010, vol. (4), pp. 156–160 (2010)
[11] Liu, Q., Li, Y., Li, T., Hao, L.: The Research of the Batch RSA Decryption Performance. Journal of Computational Information Systems 7(3), 948–955 (2011)
[12] Li, Y., Liu, Q., Li, T.: Efficient variant of RSA cryptosystem. Journal of Computer Applications 30(9), 255–293 (2010)
[13] Liu, Q., Li, Y., Zhou, B.: Research of batch RSA based on multi-prime. Application Research of Computers 28(2), 714–716 (2011)
[14] Li, Y., Liu, Q., Li, T.: Parallel accelerated method of batch RSA based on multi-core processor. Journal of Yunnan University (Natural Sciences Edition) 33(1), 22–26 (2011)
[15] Viega, J., Messier, M., Chandra, P.: Network Security with OpenSSL. O'Reilly (2002)

Image Segmentation
Based on Multiscale Initialized Gaussian Mixtures

Tao Guan[1] and Tao Xue[2]

[1] Department of Computer Science and Application,
Zhengzhou Institute of Aeronautic Industry Management,
Zhengzhou, Henan, China
timm.guan@gmail.com
[2] College of Computer Science
Xi'an Polytechnic University
Xi'an, Shanxi, China
xthappy@gmail.com

Abstract. Image segmentation is a key step for image processing and Gaussian Mixture Models(GMMs) are the common models for segmentation. The EM algorithm is usually used to estimete the parameters of GMMs, which is opt to get stuck at local minimum. In this paper we propose a new initialized shceme, multiscale online learning, for EM to aviod local minima and for GMMs to decide the optimal initial number of components. Experimental results have shown that this scheme can effectively improve the precision of segmentation compared to classical EM algorithm.

Keywords: Image segmentation, clustering, Gaussian mixture models(GMMs), online multiscale learning.

1   Introduction

Image segmentation is a process of partition an image into several non-overlapped regions and a preprocessing step in image processing. There have been many approaches for segmentation and many of them are based on   clustering techniques. Among them, Gaussian mixture models(GMMs) are a class of effictive methods[1-4]. In image segmentation, one assumes that data comply with Gaussian mixture distributions with unknown parameters and then determines the optimal parameter values by using EM algorithm. By far, there have existed some extened GMMs for image segmentation and fearture selection. Recently, researchers have extended GMMs by using Markov Random Fields as the spatial constraints of pixels[5-11], such as SVMMs, DCM-SVFMMs. However, the elementary GMMs with EM algorithm still have some drawbakcs in image applications. First, the initial number of clusters must be prespecified. This is not practical in many situations. Second, EM is apt to get stuck into the local minima, which may make the model unable to find all clusters.

   To overcome the drawbacks of GMMs and EM algorithm, we propose in this paper a new online learning algorithm for EM initialization and the reduction of conponents

of GMMs, and it simultaneously avoids EM algorithm get stuck into local minima. Experimental comperisons on standard test image set have shown our algprithm takes advantage over the common EM algorithm in segmentation quality.

2   Multiscale Online Learning(MSOL)

Borrowed from the basic principle of the online competitive learning approach, the online multiscale learning rule for ith prototype has the following form[12],

$$P_i^{(k+1)} = P_i^{(k)} + \frac{pr_{ij}^{(k)}}{k}\left(x_j - P_i^{(k)}\right), \tag{1}$$

where $pr_{ij}^{(k)} = G\left(x_j, P_i^{(k)}\right) = \exp\left\{-d^2\left(x_j, P_i^{(k)}\right)\middle/2\sigma_i^2\right\}$. $d(\cdot)$ is commonly Euclidean distance and σ_i^2 is estimated by

$$S^2 = \frac{1}{K-1}\sum_{x \in N(P_i)}\left(x - P_i\right)^2 , K = \left|N\left(P_i\right)\right| . \tag{2}$$

Suppose that the numbers of sample and prototypes are n and m, respectively, then the computation complexity of MSOL is approximate $O(mn)$ when $K \ll n$. This is superior to multiscale spectrul clustering algorithm whose computation complexity is $O(n^2)$. Figure 1 demostrates the clustering result on an irregular dataset. MSOL correctly divides the data into two classes.
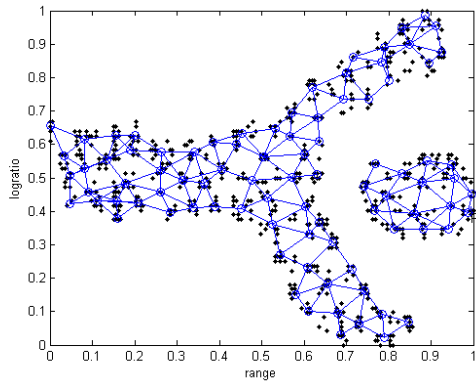

Fig. 1. The cluster result of MSCL on an irregular dataset

3   Gaussian Mixture Models and EM Algorithm

Given sample X and unknown parameter Θ, GMMs have the following form[1-4],

$$G(X \mid \Theta) = \sum_{i=1}^{M}\alpha_i G_i\left(X \mid \theta_i\right), \tag{3}$$

where $\theta_i = (\mu_i, \Sigma_i)^T$ and

$$G_i(X \mid \theta_i) = \frac{1}{2\pi |\Sigma_i|^{1/2}} \exp\left\{-\frac{1}{2}(X - \mu_i)'\Sigma_i^{-1}(X - \mu_i)\right\}. \tag{4}$$

In image segmentation, GMMs use a combination of Gaussian functions with unknown parameters to approximate the unknown distribution. The optimal values of parameters are computed via maximal likelihood estimation and EM algorithm proposed by A. P. Dempster et al in 1977. The genral iterative formulas of EM algorithm for parameter estimation are presented as follows[3],

$$w_i^{j(k)} = \frac{\alpha_i^{(k)} G\left(x_j \mid \theta_i^{(k)}\right)}{\sum_{i=1}^{M} \alpha_l^{(k)} G\left(x_j \mid \theta_l^{(k)}\right)},$$

$$\alpha_i = \frac{1}{N} \sum_{i=1}^{N} w_i^{j(k)}, \tag{5}$$

$$\mu_i = \frac{1}{N\alpha_i^{(k)}} \sum_{i=1}^{N} w_i^{j(k)} x_j, \tag{6}$$

$$\Sigma_i = \frac{1}{N\alpha_i^{(k)}} \sum_{i=1}^{N} w_i^{j(k)} (x_j - \mu_i)(x_j - \mu_i)^T. \tag{7}$$

To make sure the number of components, two schemes are frequently used in image segmentation. First scheme is to preset the number of components, in which K-means algorithm is usually used to initialize the EM, and second is to decide it by online learning way. First scheme can not satisfy the reqirements of online analysis and thus adopted in real time environments. On the other hand, most of online learning algorithms is oriented to single scale of data and may produce more components for the cluster with larger scattering degree. To settle this problem and decrease the number of components in GMMs, this paper presents a new algorithm based on the multiscale online learning algorithm and GMMs and then apply it to image segmentation.

4   Algorithm

The detail steps of our algorithm is present as follows.

Initialize
① initialize P_0, K, ε, δ, pr for MSOL;
② locate the centers μ_i of components by MSOL;
③ compute the local variance Σ_i of components;

Main loop
④ compute the α_i for each component by (5);

⑤ update μ_i and Σ_i separately by (6) and (7);

⑥ if $\left|Q^{(k+1)} - Q^{(k)}\right| \leq \varepsilon$, then stop; else goto ④.

This algorithm has the approximate linear complexity and suit to large scale data analysis.

5   Experiments on Image Segmentation

We ran our algorithm on standard test image sets and compared the results with GMMs+EM model. The image sets can be freely obtained from web and have been widely applied to many segmentation problems. We select the figures with 256×256 pixels and the segmentation results of our algorithm and GMMs+EM are shown in Figure 2. From these figures we can know that k-means can not adaptively decide the number of clusters and thus may induce to loss clusters, such as the blue balloon in the figure of pallon. Moreover our algorithm can clearly show the detail parts in original figures.
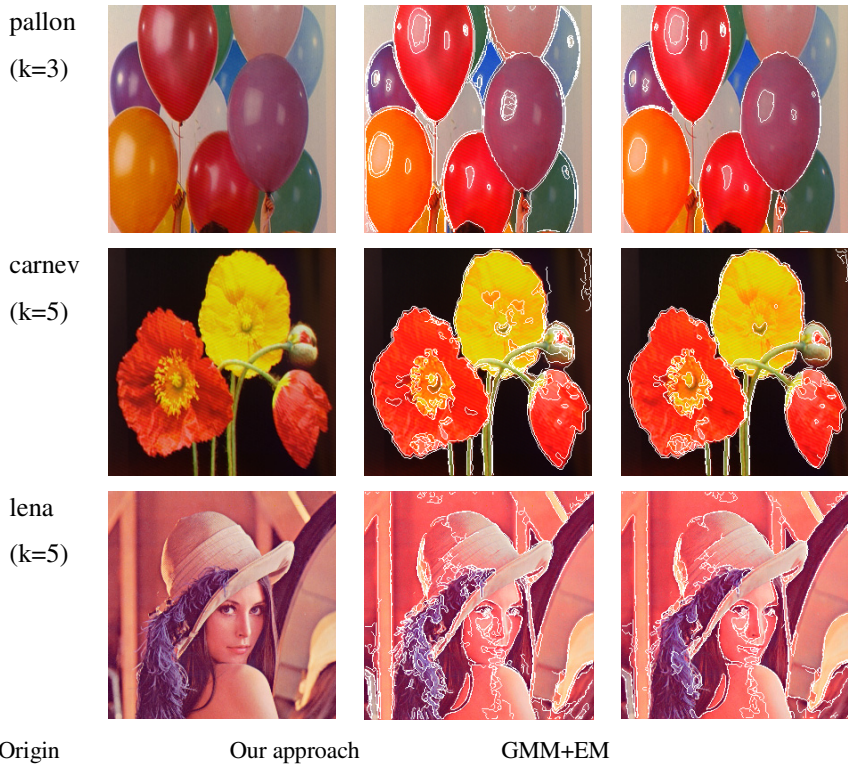

pallon (k=3)

carnev (k=5)

lena (k=5)

Origin   Our approach   GMM+EM

Fig. 2. The result comparisons of our approach and GMMs+EM

References

[1]   Mclachlan, G., Peel, D.: Finite Mixture Models. John Wiley & Sons, New York (2000)

[2]   Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. R. Stat. Soc., ser B 39, 618–633 (1977)

[3]   Bilmes, J.A.: A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models, U. C. Berkeley, CA, TR-97-021 (1998)

[4]   Titterington, D.M., Smith, A.F.M., Makov, U.E.: Statistical analysis of finite mixture distributions. Wiley, New York (1985)

[5]   Sanjay-Gopal, S., Hebert, T.J.: Bayesian pixel classification using spatially variant finite mixtures and the generalized EM algorithm. IEEE Transactions on Image Processing 7(7), 1014–1028 (1998)

[6]   Sfikas, G., Nikou, C., Galatsanos, N., Heinrich, C.: Spatially varying mixtures incorporating line processes for image segmentation. Journal of Mathematical Imaging and Vision 36(2) (2010)

[7]   Stauffer, C., Eric, W., Grimson, L.: Adaptive background mixture models for real-time tracking. In: CVPR 1999, pp. 2246–2252 (1999)

[8]   Nikou, C., Likas, A.C., Galatsanos, N.P.: A Bayesian framework for image segmentation with spatially varying mixtures. IEEE Transactions on Image Processing 19(9), 2278–2289 (2010)

[9]   Blekas, K., Likas, A., Galatsanos, A.P., Lagaris, I.E.: A spatially constrained mixture model for image segmentation. IEEE Transactions on Neural Network 16(2), 494–498 (2005)

[10] Green, P.J.: Bayesian reconstructions from emission tomography data using a modified EM algorithm. IEEE Transactions on Medical Imaging 9(1), 84–93 (1990)

[11] Nguyen, T.M., Jonathan Wu, Q.M., Ahuja, S.: An extension of the standard mixture model for image segmentation. IEEE Transactions on Neural Networks 21(8), 1326–1338 (2010)

[12] Guan, T., Yu, Y., Xue, T.: An online multiscale clustering algorithm for irregular data sets. In: 2011 International Conference on Future Computer Sciences and Application, HK (2011)

[13] Guan, T., Li, L.-L.: Self-Branching Competitive Learning for image segmentation. In: Proc. International Conference on Bio-Inspore Computing: Theories and Applications, Changsha, China, pp. 652–656 (September 2010)

A Logic Based Framework
for Multi-Objective Decision Making

Qing Zhou[1] and Donglei Chen[2]

[1] The Software Institute,
Zhongshan University,
Guangzhou, Guangdong 510275, P.R. China
lnszq@mail.sysu.edu.cn
[2] The Water Control Center, Guangzhou Water Supply Co.,
Guangzhou, Guangdong 510600, P.R. China
cdonglei@gmail.com

Abstract. In this paper a framework for multi-objective decision making, *FMODM*, is proposed. The framework finds the best option for our goal based upon the knowledge and the facts in our hands and hence it is quite close to the "what ... if ..." analysis often used by human beings making decision. The whole framework is described in details and the theoretical analysis is discussed comprehensively. Computability of the framework is discussed through a few theorems proven in the paper. Examples in the paper clearly illustrate the whole working of the framework.

Keywords: Multi-objective, supported degree, expected value.

1   Introduction

In this paper a framework for multi-objective decision making is proposed.

We may describe the decision-making problem that we face in life as following: Assume that we have a problem P and have a few options $c_1,...,c_n$ which are proposed as the possible options. The decision making problem is rooted in the need to find the best option from the option set $\{c_1,...,c_n\}$. It might happen that our goal may consist of a few objectives $G_1, ..., G_m$. The multi-objective decision making problem is to find a c_i in the option set which is the best option for the problem P when all the objectives $G_1, ..., G_m$ are considered. Multi-objective decision making problems appear everywhere in our real life, such as military, utility, management, finance, etc.

For the problem P we want to solve, there are usually some corresponding domain knowledge. Such knowledge can be divided into two parts: one is the group of basic rules in the area (called "knowledge set"); the other are facts we connected (called "evidence set"). Of course, we want to use such knowledge in our hands to analyze the situation on the problem. The principle here is obviously: if an option can brings us the most advantages and the least disadvantages that is the one we want. Therefore we need domain knowledge to analyze what advantages and disadvantages an option will bring to us for each option. So the more knowledge we have used, the more reliable our option is.

In most cases, those advantages and disadvantages can be derived by deductions with an option is assumed, and are supported by evidences (i.e. they can be derived by evidences, too). So the common consequences of both the evidences in our hands and an option give supports to that option. It is also quite obvious that the more such common consequences an option has, the more possible for the option to be true. We call the number of such common consequences the "supported degree" of the option.

Most of our knowledge can be expressed by first-order languages and first-order logic has strong deduction power. So *FMODM* is built within first-order logic so that we can use our domain knowledge and the facts we gathered for the problem to make reasoning needed for calculating the supported degree.

For the decision making with multi-objectives we want the results of our decision is the best as a whole. Then we have to use our knowledge not only to analyze what advantages and disadvantages an option will bring to us for each objective but also to make a balanced consideration of the affects of all the objectives. This also needs to use our knowledge to analyze. With this our decision is easy to make. Then we need a way to express these impacts. In many cases, such impacts can be expressed by functions which can be obtained either from theories or from experience. We give an example on this situation.

***Example 1*:** Suppose that we want to do some investments. We think that two objectives must be considered: "business-familiarity" (means how much we know about the business we are going to invest) and "development-affect" (means how much it affects our own company).   We think that "business-familiarity" impacts the investment *70%*, "development-affect" impacts the investment *-30%*. Then once we know all the supported degrees about these two objectives from the options we have to use such impact factors to composite functions to calculate the impact of the whole problem of investment. We called them "Impact degree".

Once we have the supported degree for each objective, and the impact degree of each option, we can compose them to a value, called "expected value", for each option. The higher expected value an option has, the more benefits we can have from that option if we take it as our decision. This is the idea of this paper.

The remaining sections of this paper include: Outlined in Section 2 discusses how to calculate expected value to solve multi-objectivess decision making. Then make conclusion in section 3.

2   Basic Definitions

In this paper, we assume the logical reasoning system *T* containing our knowledge is fixed and we use the following notations:

1)   *L* is the language of *T*;
2)   The knowledge set *K*, a finite set of formulae of *L*;
3)   The evidence set *E*, a finite set of closed formulae of *L*;
4)   The objective set *G*, a finite set of formulae of *L*.

We assume that $K \cup E$ is consistent.
Now we define:

Definition 1 (*Multi-Objectives Decision Making Problem*): A multi-objectives decision making problem P is: For P, we have objectives $G_1, G_2, \ldots, G_m \in G$, where P and $G_1, G_2, \ldots, G_m$ are predicates. Each objective corresponds to real number $v_j \in [-1,1]$, and v_j is called the impact degree for G_j, where $j \in N$. Then the function $F = f(s_1 \times v_1, s_2 \times v_2, \ldots, s_m \times v_m)$ is called the expected value of P, where $f: [-1,1]^m \rightarrow [-1,1]$ is a function, which is called the "expectation function". The symbols $s_1, s_2, \ldots, s_m$ will be defined later.

Example 2: The objectives in *Example 1* are G_1=business-familiarity, G_2=development-affect. Once we have the impact degrees of "business-familiarity" and "development-affect" given, we can use the expectation function $f(s_1 \times v_1, s_2 \times v_2) = s_1 \times 0.7 - s_2 \times 0.3$ to get the expected value F from the description in *Example 1*.

Given P, we usually have a few options which are supposed to be possible options of P. By decision-making problem, we mean that we have to choose an option from them. Let us call them possible options. This lead to:

Definition 2 (*Option Set*): A option set C is a finite set of formulae $\{C_1, C_2, \ldots, C_n\}$ of L such that for all $i \in \{1, 2, \ldots, n\}$, $K \cup E \nvdash C_i$ and $K \cup E \nvdash \neg C_i$.

Example 3: In *Example 2* we consider three investment portfolios as options: investment portfolio 1 and investment portfolio 2. These portfolios consist of some businesses: investment portfolio 1 consists of business a_1, business a_2, and business a_3; investment portfolio 2 consists of business a_2 and business a_4. Then $C = \{C_1 = $ *investment portfolio 1*, $C_2 = $*investment portfolio 2*$\}$ is the option set, and our purpose is to find the best one in C.

To know what we will have from a option for each objective, we observe that what we will get have to have something in common with both the evidences in our hands and the objective considered. Then we have:

Definition 3 (*Compatible Result Set*): Suppose $C_i \in C$ and $G_j \in G$. A formula A is called a compatible result of C_i to G_j if:

- $K \cup E \vdash A$
- $K \cup \{C_i\} \vdash A$
- $K \cup \{G_j\} \vdash A$
- $K \nvdash A$

Then we use D_{ij} to denote the set of all compatible results of C_i to G_j, that is:

$$D_{ij} = \{A \mid A \text{ is a compatible result of } C_i \text{ to } G_j\}.$$

Example 4: In the *Example 3* the language L has the following formulas:

1) Objectives: $G_1(t)$, which indicates business-familiarity; $G_2(t)$, which indicates development-affect, where t is a set of investment projects;
2) Options: C_1, which indicates investment portfolio 1; C_2, which indicates investment portfolio 2;
3) $Prj(x)$, which indicates investing to the business x;
4) $BfHigh(x)$, which indicates the degree of business-familiarity is high; $DAHigh(x)$, which indicates the degree of development-affect is high.

The corpus of knowledge set *K* consists of the following axioms:

1) $G_1(\{a_1, a_2, a_3, a_4\}) \rightarrow BfHigh(a_1) \wedge \neg BfHigh(a_2) \wedge BfHigh(a_3) \wedge BfHigh(a_4)$, which indicates that to the objective G_1, the degree of the business-familiarity is high to business a_1, and is not high to business a_2, and is high to business a_3, and is high to business a_4;

2) $G_2(\{a_1, a_2, a_3, a_4\}) \rightarrow DAHigh(a_1) \wedge \neg DAHigh(a_2) \wedge DAHigh(a_3) \wedge \neg DAHigh(a_4)$, which indicates that to the objective G_2, the degree of the development-affect is high to business a_1, and is not high to business a_2, and is high to business a_3, and is not high to business a_4;

3) $C_1 \rightarrow Prj(a_1) \wedge Prj(a_2) \wedge Prj(a_3)$, which indicates that if investment portfolio 1 chosen, it has business a_1, a_2, and a_3;

4) $C_2 \rightarrow Prj(a_2) \wedge Prj(a_4)$, which indicates that if investment portfolio 2 chosen, it has business a_2, and a_4;

5) $Prj(a_1) \rightarrow BfHigh(a_1) \wedge DAHigh(a_1)$, which indicates that business a_1 has high degree of business-familiarity and development-affect;

6) $Prj(a_2) \rightarrow BfHigh(a_2) \wedge \neg DAHigh(a_2)$, which indicates that business a_2 has high degree of business-familiarity and has not high degree of development-affect;

7) $Prj(a_3) \rightarrow \neg BfHigh(a_3) \wedge DAHigh(a_3)$, which indicates that business a_3 has not high degree of business-familiarity and has high degree of development-affect;

8) $Prj(a_4) \rightarrow \neg BfHigh(a_4) \wedge DAHigh(a_4)$, which indicates that business a_4 has not high degree of business-familiarity and development-affect.

We assume that the system has two deduction rules: 1. if $\alpha \rightarrow \beta$ and α, then β; 2. if $\alpha \rightarrow \beta$ and $\beta \rightarrow \gamma$, then $\alpha \rightarrow \gamma$.

The evidence set *E* consists of:

$BfHigh(a_1)$, $\neg BfHigh(a_2)$, $BfHigh(a_3)$, $BfHigh(a_4)$, $DAHigh(a_1)$, $\neg DAHigh(a_2)$, $DAHigh(a_3)$, $DAHigh(a_4)$, $BfHigh(a_2)$, $\neg BfHigh(a_3)$, $\neg BfHigh(a_4)$, $\neg DAHigh(a_4)$.

Then we have made up the construction of the system, Let us look at how can we get the compatible result set:

To get the compatible result set of C_1 to G_1, for example, we can get $\{BfHigh(a_1)$, $\neg BfHigh(a_2)$, $BfHigh(a_3)$, $BfHigh(a_4)\}$ by $K \cup \{G_1\}$, get $\{BfHigh(a_1)$, $BfHigh(a_2)$, $\neg BfHigh(a_3)$, $DAHigh(a_1)$, $\neg DAHigh(a_2)$, $DAHigh(a_3)\}$ by $K \cup \{C_1\}$, and get E by $K \cup E$. Then the compatible result set of C_1 to G_1 is $\{BfHigh(a_1)\}$. The other compatible result sets will be gotten like this.

Now as we know that although the presentation of some conclusions is not the same, but the meaning of them is the same, such as "A's father is B", has the same meaning with "A is B's son and B is man". Then our compatibility result set must merge these results. In logic, if the conditions for the establishment of two conclusions identical to that of these two conclusions are equivalent, such as "A's father is B" implies itself and "A is B's son and B is man" implies "A's father is B", then we take them as the same, and that means these two are equivalence. The equivalence relation thus defined the equivalence class to essentially the same relationship can be combined into one. Therefore, the following statement :

Set $A_1, A_2 \in D_{ij}$, if for all $E' \subseteq E$, $K \cup E' \vdash A_1$ if and only if $K \cup E' \vdash A_2$, called A_1, A_2 is equivalent in D_{ij}, denoted by $A_1 \sim A_2$.

Definition 4 (*Equivalence Class Set*): B_{ij} is the set of equivalence class of D_{ij} to $\sim$, which $B_{ij} = \{U : U$ is the maximum sub-set of D_{ij} which satisfy that for all $A_1, A_2 \in U$, $A_1 \sim A_2\}$.

Example 5: In *Example 4*, we can see that the equivalence class of the compatible result set such as $\{BfHigh(a_1)\}$ is the equivalence class of $BfHigh(a_1)$.

The equivalence class set B_{ij} have been merged the repeated compatibility results in D_{ij}, then the size of B_{ij} is just the possibility of option C_i to objective G_j. In other words, it is how deep the option supports the objective, just what we called "supported degree". Therefore defined as follows:

Definition 5 (*Supported Degree*): $s(i, j)$ is the supported degree of C_i to G_j. And $s(i, j)$ calculated according to the following:

$$s(i, j) = |B_{ij}| / (\sum_{i=1}^{n} |B_{ij}|)$$

Example 6: So we can calculate the supported degrees of all the options as the following:

Table 1. Supported degree of all the options

	G_1	G_2
C_1	1/1	3/4
C_2	0/1	1/4

Supported degree $s(i, j)$ is the degree the option C_i supports the objective G_j. And then we can merge all the objects as a whole by the expectation function mentioned in *Definition 1*:

Definition 6 (*Expectations*): Given a option C_i, F_i is the expected value of C_i to the problem P, and F_i is calculated by the following:

$$F_i = f(s(i, 1) \times v_1, s(i, 2) \times v_2, \ldots, s(i, m) \times v_m)$$

Then we can get the option which has the highest expected value as the best option to the problem P.

Example 7: All the expected values are calculated as following:

$$F_1 = 0.7 \times 1 - 0.3 \times 0.75 = 0.475$$

$$F_2 = 0.7 \times 0 - 0.3 \times 0.25 = 0.075$$

We can see that the option C_1 ("investment portfolio 1") is the best.

Theorem 1 (*Computability of Expected Value*): $(s(i, j) \times v_j) \in [-1, 1]$
    From this theorem, that $s(i, j) \times v_j$ is finite, and their summation, the expected value, is finite too, and it is computable.

3   Summaries

In this paper a framework *FMODM* of decision making for multi-objectives is proposed. The framework is built within first-order logic and is completely based upon our knowledge on the decision we want to make. This makes *FMODM* is reasonable and has a wide range of applications. Competing approaches in literature mainly contains those based upon probability theory or fuzzy logic. In these approaches decisions are making only relying on static data which may or may not be available or be reliable. Computability of *FMODM* is discussed through a few theorems. The whole working of *FMODM* is illustrated via a unified example. A significant advantage of *FMODM* is that some natural words related to decision making, such as "advantage" or "disadvantage", are precisely defined and hence all the knowledge on the decision we are going to make can be well utilized so that it makes our decision more reliable.

Many works on various directions can be done from our work. Here we only mention a couple of them. In the field of decision making a lot of work has been done by using technologies, such as genetic algorithm, neural network. The work in this paper can be combined with such technologies so that knowledge can be well used in these technologies. We believe that this will expand the range of applications of computer-aided decision making and make it more reliable. Another direction can go along the direction on computational efficiency. Although it is proven "*FMODM*" is computable, but there are rooms of improving the efficiency of searching right results in a logical system.

Given the state of art on decision making with multi-objectives we believe that this is a very nice step forwards. The suggestions mentioned above can be greatly increasing the power of our proposal and improve the research in this field.

References

[1] Shen, X., Guo, Y., Chen, Q., Hu, W.: A multi-objective optimization evolutionary algorithm incorporating preference information based on fuzzy logic. Computational Optimization and Applications 46(1), 159–188 (2008)

[2] Rollón, E., Larrosa, J.: Bucket elimination for multiobjective optimization problems. Journal of Heuristics 12(4-5), 307–328 (2006)

[3] Pappa, G.L., Freitas, A.A.: Evolving rule induction algorithms with multi-objective grammar-based genetic programming. Knowledge and Information Systems 19(3), 283–309 (2009)

[4] Zhou, Q., Peng, W.: Uncertain Reasoning and Decision Making. In: Proceedings of International Conference on Intelligent Information Processing (IIP 2004). Springer (2004)

[5] Zhou, Q., Deng, X., Jones, J.D., Zhang, K.: A diagnostic system based upon knowledge and experience. The First Supporting Degree, pp. 280–283 (2008)

[6] Datta, D., Deb, K., Fonseca, C.M.: Multi-objective Evolutionary Algorithms for Resource Allocation Problems. Class Timetabling and Land-Use Management Problems as Multi-objective Optimization Problems, pp. 403-405; NSGA-II-UCTO and NSGA-II-LUM, pp. 405-409 (2007)

[7] Cai, Z., Gong, W., Huang, Y.: A Novel Differential Evolution Algorithm Based on ε-Domination and Orthogonal Design Method for Multiobjective Optimization. Our Approach: ε-ODEMO 2007 (2007)

[8] Zhou, Q., Ju, S.: The Concept of Uncertain Reasoning and its Supporting Degree. In: Proceedings of the ISCA 9th International Conference (2000)

Underground Goaf Filling and Management Sequence Optimization Research

Gan Deqing[1] and Zhang Yabin[2]

[1] College of Mining Engineering,
Safety Technology Key Lab. of Hebei Province,
Hebei United University, Tangshan 063009 China
[2] College of Mining Engineering, Hebei United University, Tangshan 063009 China
Civil & Environment Engineering School,
University of Science & Technology Beijing, Beijing, China

Abstract. According to the current characteristics of underground stranded goaf in Xishuangcheng Iron Mine and numerical simulation technology, the stress field and displacement field in roof and surrounding rock areas were calculated. The dangerous degree of underground goaf was graded and filling sequence was optimized. It is of great guiding significance for underground goaf management, pillar mining, the formulation of deep mining scheme and safety.

Keywords: Goaf, Numerical simulation, Dangerous degree, Filling sequence.

1 Introduction

Without processing of goaf timely, widespread underground stranded goaf brought huge potential safety problems in the underground mining process. Roof collapse of large areas would cause caving and underground casualties and property losses. Therefore, the management of underground goaf was of important significance to sustainable development of the mine [1]. In recent years, researchers attached more importance to detection technology research, detection and treatment of underground complex goaf, goaf stability analysis and evaluation, etc and made certain progress [2-6]. For underground stranded goaf filling, however, optimized sequence must be adopted so that better filling effect could be obtained. Current research in this aspect is still rare. This essay took Xishuangcheng Iron Mine in Zunhua City as example and studied filling sequence optimization of underground stranded goaf by the means of numerical simulation.

2 Underground Stranded Goaf Status

The construction of Xishuangcheng Iron Mine began from 2003 with underground mining methods. To 2009 sixteen shafts had been dug. Most of them did not run through and mined independently. Their bottom elevations were inconsistent and the goaf was disodered. There were more than 800,000 cubic meters underground goaf, the size of goaf ranging from 500 to 119,000 cubic meters. Due to not standardized

mining, there were 11.1291 million tons of underground ore under recovery. For safety mining of the deep ore, filling and management in goaf area should be conducted soon.

3   Stability Analysis of Present Goaf Situation

3.1   The Establishment of the Numerical Simulation Model

Typical profile of mine was selected to calculate. There are 10 goaf areas on the profile and 32120 units in the model. Numerical simulation model is shown in Fig.1, the numbers of goaf areas were 1-10.

3.2   Boundary Condition and the Selection of Calculation Parameters

Fixed constraint conditions were imposed on bottom and sides of the model while calculation, calculation parameters were shown in Tab. 1.
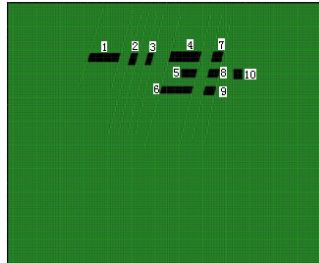
Fig. 1. Numerical simulation model

Table 1. Numerical calculation parameters

Title	γ(kg/m³)	E(GPa)	μ	C(MPa)	Φ(°)
ore	3300	4.8	0.21	2.4	38
rock	2700	4.31	0.22	2.29	36
soil	1600	0.015	0.25		32

3.3   Stability Analysis of Goaf

3.3.1   Stress Field Analysis

Numerical calculation results of stress field as shown in Fig. 2-3.

According to the results, maximum stress value of roof in horizontal direction was around 8MPa, vertical direction around 12.0Mpa, stress values decreased in turn from the both sides of the roofs to surrounding rocks. It is mainly compressive stress in surrounding rocks. Among them No.1, 4, and 6 goaf roof stress values in vertical direction approached the compressive strength value of surrounding rocks, the other stress values in vertical direction were less than the compressive strength value.
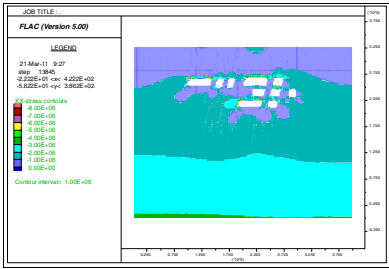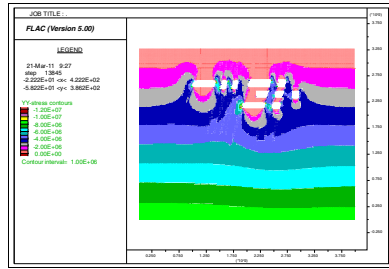
Fig. 2. Horizontal stress field

Fig. 3. Vertical stress field

3.3.2   Displacement Field Analysis

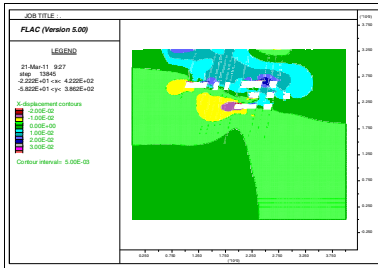Numerical calculation results of displacement field were shown in Fig. 4-5.

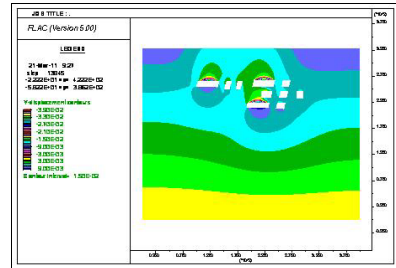**Fig. 4.** Horizontal displacement field

Fig. 5. Vertical displacement field

It could be concluded from the figures that the maximum vertical displacement of goaf roof was 35mm, which all happened in the middle of the roof. The maximum horizontal displacement was up to 30 mm, which was in the middle of side walls of goaf areas. The higher were the goaf areas, the larger were the value of horizontal displacements.

4   The Filling and Management Sequence Optimization of Goaf

4.1   Goaf Danger Classification

Dangerous degree of underground goaf could be classified into four levels, corresponding dangerous state was shown below. Ⅰ Supermajor danger Ⅱ Major danger Ⅲ significant danger and Ⅳ generally danger[7]. According to the calculating results on stress distribution of the roof and surrounding rock together with displacement, the dangerous degrees can be graded. Goaf maximum stress and maximum displacement line charts in vertical direction were shown in Fig. 6-7.
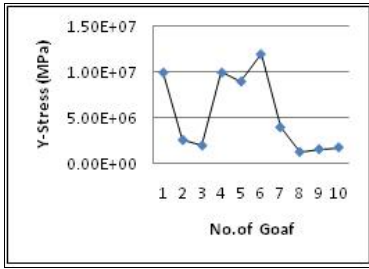
Fig. 6. Goaf maximum stress in vertical direction
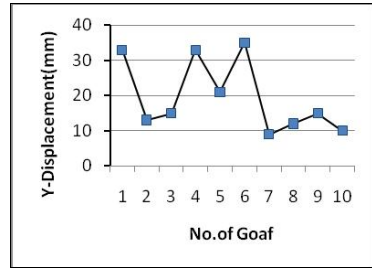
Fig. 7. Maximum displacement in vertical direction

It can be concluded from the figures that the sizes of 1,4 and 6 goaf were relatively larger, their maximum roof stress and displacement values were larger which were classified into I level ; the No.5, 7, 8, 9 goaf were medium in size and close to No. 4 and No. 6 goaf , they were classified into II level. The No. 2, 3 and 10 were classified into IV level for they were smaller in roof stress and displacement values as well as in size.

4.2   Determination of Goaf Filling and Management Sequence

While determining filling sequence, the highest dangerous degree goaf should be filled prior to the lower ones. For the same danger degree, the bottom and middle section should be filled firstly and then the upper and middle section. According to the above principles, the filling sequence of Xishuangcheng Iron Mine should be No. 6 the first, and then No. 1and 4 goaf, followed by No.9, 5, 8, 7 goaf and finally No. 10, 2, 3 goaf.

5   Summaries

FLAC software was used to conduct numerical simulation of the stability for underground goaf in Xishuangcheng Iron Mine, stress field and displacement field distribution of roof and surrounding walls in goaf was obtained, the filling sequence was determined, which played the guiding role in the goaf management, pillar mining, the formulation of deep mining scheme and safety.

References

[1] Chen, W., Zhu, W., Wang, B., et al.: Numerical simulation and testing study of large deformation for cavern's surrounding jointed rock mass. Chinese Journal of Rock Mechanics and Engineering 17(3), 223–229 (1998)
[2] Liu, D., Xu, G., Huang, R., Gu, D.: A New Technique for Prospecting Exhausted Areas in Metalmines. China Mining 9(4), 34–37 (2000)

[3] Zhang, J., Zhang, Y.-F., Yuan, T.-Z., Chen, X.-Q.: Analysis and Appraisal for Stabiliy of Goaf. Journal of Water Resources and Architectural Engineering 8(2), 145–146, 155 (2010)

[4] Yang, B., Luo, Z., Liu, X., Lu, G., Lu, H.: Dangerous degree estimation of complicated goaf group with FEM simulation. Mineral Engineering Research 25(1), 4–8 (2010)

[5] Luo, Z.-Q., Lu, H., Liu, X.-M., Li, C., Zhang, B.: Numerical simulation method for mined-out area group stability basedon Cavity Monitoring System. Industrial Minerals and Processing (1), 18–21 (2008)

[6] Lin, J.: Simulation of stability data of goaf in gentle-inclined ore body. Mining Engineer 8(1), 17–20 (2010)

[7] Luo, Y.: The major hazard sources identification in widespread mined-out area. Central South University, Changsha (2005)

A Fast Free-Surface Tracking Algorithm

Chen Jing and Liu Fuyan

School of Computer Engineering and Science, Shanghai University
Shanghai, China

Abstract. Interface tracking has long been an important issue in a wide variety of areas. Based on the level set method, this paper introduces a fast free-interface tracking method which is suitable for the computer simulation. By correcting the level set function with particles, the method can avoid the loss of details and improve the computational efficiency. Moreover, the method is unconditionally stable. These results are verified by simulations.

Keywords: Free surface tracking, Level set, Particle Level Set, Unconditionally Stable.

1   Introduction

Free surface tracking has always been an important study in different fields[1], such as computer graphics, computer vision and computer fluid, etc. In computer graphics, the key of liquid simulation is to get the liquid surface at any time by using the interface tracking method. However, the movement of liquid surface is very complex and merging or pinching phenomena often happen, which can lead to a great change of the original topological structures. As a result, how to accurately track the liquid surface is a challenging task and has been widely studied.

The level set method is one of the famous surface tracking methods. It was first proposed by Osher and Sethian[2]. Because it can deal with complicated topological deformations and is easy to be implemented, it has been drawing much attention. It is a typical Euler method, the main idea of which represents the surface implicitly using a signed distance functions $\varphi(x,t)$. $\varphi(x,t)$ is defined as the shortest distance from position x to the surface, so the surface is the zero contour of the distance field. Positive $\varphi(x,t)$ shows that x is outside of the object while negative $\varphi(x,t)$ shows that x is inside of the object. $\varphi(x,t)$ has the following property:

$$|\nabla \varphi(x,t)| = 1. \tag{1}$$

The governing equation of level set is a PDE, as follows:

$$\frac{\mathrm{D}\varphi}{\mathrm{D}t} = \frac{\partial \varphi}{\partial t} + u \cdot \nabla \varphi = 0 , \tag{2}$$

where u(x,t) is the velocity field of the surface and the surface will be changed by the velocity. In the fundamental level set method, the initial value of function $\varphi(x,t)$ should be given and the value of $\varphi(x,t)$ other than the initial time can be got by solving the fundamental level set equations.

There are a lot of calculations in the fundamental level set method, so it's difficult to satisfy the needs of computer simulations. In order to improve the speed of calculation, Sethian[3] has proposed a partial and rapid level set method, which limits the calculation in a narrow band around the surface of the zero level set. This method reduces the complicated calculation of one iteration from $O(N^3)$ to $O(N^2K)$, where N is resolution and K is the width of the narrow band.

The discreteness of the level set equation causes a lot of numerical consumption, which further leads to the loss of mass and capacity. To improve the accuracy of level set method, Enright[4] and Wang[5] used a particle level set method, which is the mixture of Euler method and Lagrangian method. The method taking into account the MAC property can correct the level set function with particles. The surface captured by the method is smooth and no loss of mass[6].

What's more, the nature of signed distance from the grid to the evolution surface must be kept when the fundamental level set equation is solved but the character can't be satisfied even with several iterations because of the inner effect of value calculation. As a result, it becomes an important step to re-initialize. However, the usual initialization methods to calculate the distance between the point and curves will consume a long time.

This paper analyses the problems of the level set method and then proposes a level set method which combines the advantages of many other methods and is suitable for the computer simulation. Although the method only has one order accuracy, it can avoid the loss of details and improve the efficiency by correcting the level set function with particles. Furthermore, it is simple to be implemented and is fit for the fluid animation.

2   Partial Particle Level Set

2.1   Narrow Band Level Set

The basic idea of Narrow Band Level Set [3] is to found an adaptive narrow band around the surface. When evolved, only updating the functions of grids is necessary. During the evolution, to prevent the grids of the surface from spanning the narrow band, it is necessary to store the boards of the narrow band. When the grids come close to the boards, a narrow band whose centre is the present curve and whose width is k must be founded. By this way, the complicated calculation is decreased to $O(Nk)$, in which k is the width of   the narrow band.

2.2   Particle Level Set Method

Based on the particle level set method, some trace particles are put. The particles solve and correct the level set functions with Lagrangian method. The result can avoid the serious smooth effect at the point of the level set surface. Its algorithm steps for:

(1) Initializing particles;
(2) In every time step:

- Update the position of particles;
- Update narrow band   level set;
- Correct narrow band level set using particles;

- Re-initialize narrow band level set;
- Correct narrow band level using particles again;
- Re-adjust radius of particles;

(3) Particle reseeding after several time steps.

2.2.1 Initializing Particles

Particle level set method puts some random positive and negative particles in the grids close to the surface, in which some positive particles are put in the $\varphi(x,t)>0$ area and negative particles are put in the $\varphi(x,t)<0$ area. The number of the particles in every grid can be changed as a controls parameter that can be used to control the amount of resolution, e.g. we use 32 particles per cell for most of our examples. Every particle has a radius between $0.1\min(\Delta x,\Delta y,\Delta z)$ and $0.5\max(\Delta x,\Delta y,\Delta z)$. It is defined by the following formula:

$$\begin{cases} r_{\max} & s_p\varphi(x,t) > r_{\max} \\ s_p\varphi(x,t) & r_{\min} < s_p\varphi(x,t) < r_{\max} \\ r_{\min} & s_p\varphi(x,t) < r_{\min} \end{cases} , \tag{3}$$

Where s_p is the sign of particles and $\varphi(x_p,t)$ is the level set value of the particle, which can be got by the linear interpolation. This can make particles be tangent with the surface.

2.2.2 Updating the Level Set Value and Particle Position

Level Set Equation is a Hamilton-Jacobi Equation. In generally, this equation be solved with fifth order accurate WENO integration methods, along with third order accurate TVD Runga Kutta time integration. Though it is very accurate, the method is complicated and takes a long time. It is limited by time and not suitable to computer simulation. A semi-Lagrangian method[7], which introduced by Stam[8] to the computer graphics community, is adopted to update the level set in the paper. It isn't limited by CFL condition. Unconditional stable solutions can be got by using it.

The level set equation is a convectional equation. It shows the value of the level set is fixed when it moves along the feature line of the velocity field. According to the idea, we suppose the certain particle is moving in the velocity field V, and we know that it will reach this grid cell after a time step. To find where it started we run backwards through the velocity field from this grid point. The new value of level set in this grid cell exactly is the old value of this start-point in the previous time step.

Known that the particle will end up at X_{Grid} in the $t +\Delta t$. So let particles move toward the opposite direction which start movement from X_{Grid} under reverse velocity field -V, after Δt they get to the position which is the starting position of particles movement: X_P

$$X_P = X_{Grid} - V\Delta t .\tag{4}$$

Then the level set value of X_{Grid} is:

$$\varphi(X_{Grid}, t+\Delta t) = \varphi(X_{Grid} - V\Delta t, t) .\tag{5}$$

In most cases, X_P is not on the sampling point of the grid, and doesn't store their accurate values, but the .approximation can be got by interpolating the value of φ(x,t). Bilinear interpolation and trilinear interpolation are the usual interpolating ways.

The method of updating the value of the level set is a rapid first order method. To ensure the accuracy of simulation, the second order TVD Runge-Kutta method is used to update the position of the trace particles.

2.2.3   Correcting the Level Set Functions with Particles

In regions with sharp features, implicit surface has suffered inaccuracies due to the coarseness of the computational grid, so after the centers of particles are set, the level set functions will be corrected with the particles. When particles cross the surface in the distance of a radius, they are considered as escaped particles. The level set value given by an escaped particle with a radius r_p and a position vector x_p is defined as:

$$\varphi_p(x,t) = s_p\left(r_p - |x - x_p|\right). \tag{6}$$

If the level set equations function values φ(x,t) and φ(x,t) are not equal, show that φ(x,t) may exist errors. Error correction is performed using the φ⁺ and φ⁻.For each positive escaped particle, φ_p is found for each corner of the cell that contains the particle. The value for each cell is then set to:

$$\varphi^+(x,t) = \max\left(\varphi_p, \varphi^+(x,t)\right). \tag{7}$$

For each negative escaped particle, φp is similarly defined for each corner of the cell that contains the particle. The value for each cell is set to:

$$\varphi^-(x,t) = \min\left(\varphi_p, \varphi^-(x,t)\right). \tag{8}$$

The level set function is then reconstructed using φ(x,t)⁺ and φ(x,t)⁻ by choosing the value with minimum magnitude at each grid cell.

$$\varphi(x,t) = \begin{cases} \varphi^+(x,t) \text{ if } |\varphi(x,t)^+| \leq |\varphi(x,t)^-| \\ \varphi^-(x,t) \text{ if } |\varphi(x,t)^+| \triangleright |\varphi(x,t)^-| \end{cases}. \tag{9}$$

2.2.4   Reseeding the Level Set of Particles

With the surface constantly changing even breaking and combining, the trace particles in some areas will disappear, so it's difficult to reform the movement surface accurately. At that time, it is necessary to reseed the level set of particles and spread them. If particles are not in the grids of the narrow band, they will be deleted. If the number of the particles in the grids is less than the beginning, some particles will be put in. If the number of the particles in the grids is more than the beginning, particles will be deleted.

3   Re-initializing the Level Set Functions with Fasting Marching

In order to make the level set functions satisfy Equation (1), it is necessary to initialize them. In generally, the Godunov method or corrected Godunov method is

used to solve it. Both of them are super accurate methods but their iterative speed is low and they are limited by CLF conditions, so they are not fit for computer simulation. Sethian[3] proposed the fasting marching method to solve boundary value problems of free interface and this method can re-initialize the level set function. We discretize the equations (1) with the following first-order upwind difference (2d):

$$[\max\left(D_{ij}^{-x}\varphi,0\right)^2 + \min\left(D_{ij}^{+x}\varphi,0\right)^2 + \max\left(D_{ij}^{-y}\varphi,0\right)^2 + \min\left(D_{ij}^{+y}\varphi,0\right)^2] = 1$$

$$D_{ij}^{-x} = \frac{\varphi_{i,j} - \varphi_{i-1,j}}{h} \quad D_{ij}^{+x} = \frac{\varphi_{i+1,j} - \varphi_{i,j}}{h}$$

$$D_{ij}^{-y} = \frac{\varphi_{i,j} - \varphi_{i,j-1}}{h} \quad D_{ij}^{+y} = \frac{\varphi_{i,j+1} - \varphi_{i,j}}{h} \tag{10}$$

According to the equation above, the spreading direction of the board is from the smaller point of $\varphi(x,t)$ to the bigger point. Based on this character, Sethian proposed fasting marching to solve this kind of problems. But in this method, $\varphi(x,t)$ must be positive, that is to say, the spreading direction of the board is unidirectional. If $\varphi(x,t)$ needs to be negative, the calculation will become discommodious. Based on the character of fasting marching, the paper extend Equation(10) to Two way Fast Marching, making fasting marching can spread not only positive $\varphi(x,t)$, but also negative $\varphi(x,t)$.

The main idea of fasting marching method is that step by step calculate the value of the narrow-band level set begin from the containing interface grid. The points in grids are divided into three kinds when initialized: Know, Trial, and Far away. The steps of the solution are as follows:

(1) Initialization

A. Mark all of the point in the initial motion interface as Know points which constitute the set of Know and set the value of $\varphi(x,t)$ as 0;

B. Mark the points which adjacent with the points of the Know set as Trial point. These points constitute the set Trial, set the $\varphi(x,t)$ value of inside point as -1, the outside points as 1. Set up a positive and a negative fast sort heap, insert the Trial points into the sort heap according to its sign: negative Trial points into the negative sort heap and positive Trial points into the positive sort heap, and ensure that the point which has minimum $|\varphi(x,t)|$ is always at the top of the heap.

C. Other grid points constitute the set of Faraway, the φ value of point is $+\infty$.

(2) Iteration

a) Forward Marching

A. Select the point p^+ which has minimum $|\varphi(x,t)|$ from the positive sort heap, insert it into the know set and delete it from the positive sort heap;

B. Select the point which is one of the four points adjacent to the point p^+ but not in the know set, according to equation (10), update the $\varphi(x,t)$, if the point currently marked as Faraway, its mark should be changed to Trial. Then insert the point into the sort heap, otherwise according to the current value of $\varphi(x,t)$, adjust the sort heap to meet the minimum point at the top of the heap always.

b) Backward Marching

A. Select the point p⁻ which has the minimum |φ(x,t)| from the negative sort heap, insert it into the Know set and delete it from the negative sort heap.

B. The same process with marching forward, according to equation (10), update φ(x,t) of the non-Know set point which is one of the four points adjacent to the point p- and adjust the negative sort tree.

(3) Examine the φ(x,t) of p⁺ and p⁻ in the forward marching and backward marching whether over the boundary condition that |φ(x,t)|> k/2 (k is the narrow width), or whether each sort heap is empty. If a point has reached the boundary or sort heap is empty, the marching should over, or else repeat the step (2).

4   Results

Here we tracked the surface of the disc zalesak with the trough which has the constant flow field. The rotating speed is:

$$u(x, y, z) = -\pi(y - 0.5), v(x, y, z) = -\pi(x - 0.5). \tag{11}$$

In the simulation, the resolution of the grid is set to 100*100*100 and the radius of the disc is 0.15. Figure 1 gives the result of the 3D simulation.

The simulation results in the first line are obtained by the original level set method and the results in the other line are obtained by the particle level set method proposed in the paper.
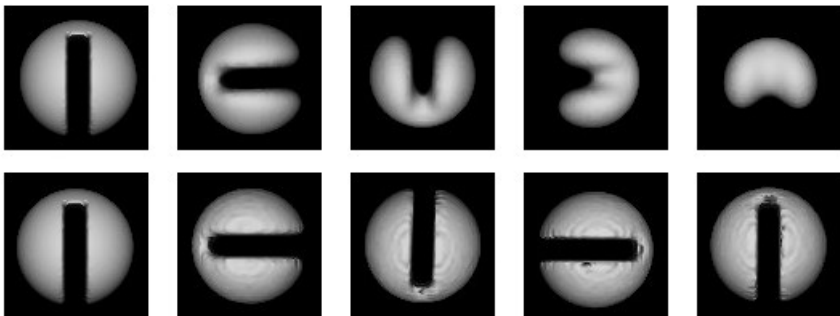
Fig. 1. Tracked the surface of the disc zalesak which is in rotating velocity field with level set method and particle level set method

According to the picture, the method proposed in the paper obviously improved details. Furthermore, our particle level set method can also satisfy the fast calculation demands in many practical systems.

5   Summaries

The method mentioned in the paper is effective, which has been proved by the simulation. Compared with many other surface tracking methods, the tracking plan proposed in the paper can satisfy fast calculation and can improve the details. As a result, it is a promising method.

In the future, we want to further increase the accuracy of our method and verify the performance, usability, and scalability of our method by real data.

References

[1] Osher, S., Sethian, J.A.: Fronts Propagating with Curvature-dependent Speed: Algorithms Based on Hamilton -Jacobi Formulations. Journal of Computational Physics 79, 12–49 (1988)

[2] Osher, S., Fedkiw, R.: Level-set methods and dynamic implicit surfaces. Springer Verlag New York, Inc. (2003)

[3] Sethian, J.A.: Level Set Methods and Fast Marching Methods: Evolving interfaces in computational geometry. Fluid Mechanics, Computer Vision, and Materials Science (1999)

[4] Enright, D., Fedkiw, R., Ferziger, J., Mitchell, I.: A Hybrid Particle Level Set Method for Improved Interface Capturing. Journal of Computational Physics, 92–116 (2002)

[5] Wang, Z., Yang, J., Stern, F.: An improved particle correction procedure for the particle level set method. Comp. Phys. 228(16), 5819–5837 (2009)

[6] Enright, D., Marschner, S., Fedkiw, R.: Animation and rendering of complex water surfaces. In: Proc. ACM SIGGRAPH 2002, vol. 21, pp. 736–744 (2002)

[7] Enright, D., Losasso, F., Fedkiw, R.: A fast and accurate semi-Lagrangian particle level set method. Comput. Struct. 83, 479–490 (2005)

[8] Stam, J.: Stable Fluids. In: Proceedings of SIGGRAPH. ACM Press, New York (1999)

A Synchronization Acquisition Algorithm for UWB Signal Based on Double-Diamond Window

Kangnian Wang, Hongde Zhang, Changchun Li, and Hongyun Liu

ChongQing Communication Institute, ChongQing China
Wkn890127@sina.com

Abstract. Synchronization acquisition is a difficult problem in UWB radio communication. A novel synchronization acquisition method of UWB signal based on double-diamond window is proposed. The average of acquisition time and probability in synchronization acquisition are carried out by a mathematical analysis in detail. Computer simulation results show that acquisition probability is greatly improved, and the synchronization time is shortened obviously.

Keywords: UWB wireless communication, Synchronization acquisition, Acquisition probability, Double diamond window.

1 Introduction

Because UWB radios can offer high user capacity with low-complexity and low-power transceivers [1]-[3], they have attracted an increasing importance in short-range high-speed wireless communications. However, to harness these benefits, one of the most critical challenges is how to obtain an accurate timing synchronization and more specifically timing offset estimation.

The traditional methods of UWB timing acquisition include sliding window, match related and searching for variable regions [4]-[6]. Although some traditional methods solved problem partly, they still can not meet the needs of practical applications. Matched filter can reduce the mean acquisition time, but need complex hardware. Channel characteristics must meet the conditions for the linear search in the method of using variable regions. Characteristics of the TH code has special requirements [7]-[9], though the method of using differential characteristics of time-hopping code is fast and only need simple hardware implementation. In this paper, a novel synchronization acquisition algorithm using double-diamond window is tested and developed for synchronize UWB signals. It can achieve synchronization rapidly with high probability and simple hardware implementation.

2 The Task of UWB Synchronization

2.1 Mathematical Description of Synchronization Acquisition

The waveform of TH-PPM signal is described by the following model:

$$S_{PPM}^{(k)}(t) = \sum_j w(t - j \cdot t_f - C_j^{(k)} \cdot t_c - \delta \cdot d_{[j/N_s]}^{(k)}) , \tag{1}$$

where $w(t)$ is single-cycle pulse, w_b represents duration, k denotes the k-th user in multi-user system, t_f is the repeat cycle of pulse, $w_b \ll t_f$, d is information sequence, δ represents the time delay of sending pulse controlled by information sequence, a single binary symbol is send during each Ns single-cycle pulse, d is the binary sequence which include "0" or "1", " [ ] " represents rounding operation, t_c is time delay of sending pulse controlled by PN code, $j \cdot t_f + C_j^{(k)} \cdot t_c + \delta \cdot d_{[j/Ns]}^{(k)}$ is the start of the j-th pulse in impulse pulses, $C_j^{(k)} = C_{j+i.Np}^{(k)}$ is j-th symbol of PN sequence of k–th user, its maximum is N_h, and the cycle of PN code is N_p. Then, the received signal is given by:

$$r^{(k)}(t) = \sum_j w(t - j \cdot t_f - C_j^{(k)} \cdot t_c - \delta \cdot d_{[j/N_s]}^{(k)} - \tau) + n(t) , \tag{2}$$

where τ is propagation delay, $n(t)$ presents gaussian white noise.

Local correlation signal is given by：

$$S^{(k)}(t) = \sum_j w(t - j \cdot t_f - C_j^{(k)} \cdot t_c) , \tag{3}$$

The task of UWB synchronization acquisition is searching the receiver phase whose error less than specified accuracy. The uncertainties of received signal come from pseudo-random TH code $C_j^{(k)}$ and signal propagation delay mainly. Let the number of uncertainty phase during repeat cycle t_f is $M = t_f / \Delta t$, where Δt is interval of each phase, N notes the number of uncertainty phase of PN code, then the number of uncertainty phase which needs detection is $C = MN$.

2.2   Efficiency Analysis of Sliding Window Algorithm

Let (n, j, k) denotes a correct acquisition event [9], where n is the position of the right phase. For a point (X, Y), $n = Y \times M + X$. j and k present the number of missed and false detection respectively. $T(n, j, k)$ and $P(n, j, k)$ are the used time and probability of (n, j, k) event. Let T_i denotes the consumed time of each position, and T_{fa} is the consumed time of false acquisition. So,

$$T(n, j, k) = nT_i + jCT_i + kT_{fa} . \tag{6}$$

The probability is $1/C$ that the correct phase lies the n-th position, and the detection probability is P_d, the probability is $P_d(1 - P_d)^j$ that reachs the correct phase after j missed detection. The probability of false detection is P_{fa}, and altogether $(n + jC)$ positions are detected during the whole process, where $(j + 1)$ positions is the right

position, others $K = n + jC - j - 1$ position is the wrong position, $0 \le k \le K$. In K positions, the probability of k false detection is $\binom{K}{k} P_{fa}^k (1 - P_{fa})^{K-k}$, so the probability of the correct acquisition event (n, j, k) is

$$P(n, j, k) = \frac{1}{C} P_d (1 - P_d)^j \binom{K}{k} P_{fa}^k (1 - P_{fa})^{K-k}. \tag{7}$$

Let $\overline{T}_A$ denotes the Mean acquisition time, then

$$\overline{T}_A = \sum_{n,j,k} P(n, j, k) T(n, j, k), \tag{8}$$

where $n \in (1 \sim C)$, $j \in (0 \sim \infty)$, $k \in (0 \sim K)$, so (8) becomes

$$\begin{aligned}
\overline{T}_A &= \frac{1}{C} \sum_{n=1}^{C} \sum_{j=0}^{\infty} \sum_{k=0}^{K} [(n + jC)T_i + kT_{fa}] \binom{K}{k} P_{fa}^k (1 - P_{fa})^{K-k} P_d (1 - P_d)^j \\
&= \frac{1}{C} \sum_{n=1}^{C} \sum_{j=0}^{\infty} (n + jC) T_i [\sum_{k=0}^{K} \binom{K}{k} P_{fa}^k (1 - P_{fa})^{K-k}] P_d (1 - P_d)^j \\
&\quad + \frac{1}{C} \sum_{n=1}^{C} \sum_{j=0}^{\infty} T_{fa} [\sum_{k=0}^{K} \binom{K}{k} P_{fa}^k (1 - P_{fa})^{K-k}] P_d (1 - P_d)^j
\end{aligned} \tag{9}$$

Then,

$$\overline{T}_A = (C - 1) T_{da} \left(\frac{2 - P_d}{P_d} \right) + \frac{T_i}{P_d}, \tag{10}$$

where $T_{da} = T_i + P_{fa} T_{fa}$ is the mean consumed time on each wrong position. According（10）, the acquisition time has relation with the detection time T_i, the detection probability P_d, the false detection cost T_{fa} and the acquisition length C.

3 Double-Diamond Window Acquisition Algorithm and Efficiency

3.1 Algorithm Description

Acquisition phase of Ultra-wideband signal in the two-dimensional space can be described as follows:

$$I : (C_x, C_y) \qquad x = 0,1,2...M; y = 0,1,2...N, \tag{11}$$

where Cx is the phase during the repeat-cycle of pulse, and Cy is the phase of TH code. The target of acquisition is searching a best match point in two-dimensional

space. In a traditional acquisition algorithm, it is assumed when the detection point away from the global optimum point, the correlation value decreases monotonously along [10]. This assumption is essentially required in the capture window graphic error monotonic. In fact, because of the multi-path effects, noise and other factors, it does not necessarily satisfy the monotonicity conditions in practice. However, if the above assumptions limit to the scope of the global near optimal point, it is possible. That is, "near the global optimum in a small range, measurement of error plane can satisfy the monotonicity conditions". From the mathematical point of view, it can be expressed as the optimization objective function as follows:

$$\overline{d} = \sum_{x_i \in S} \left\| s_i - x_i \right\| P(x_i) , \tag{12}$$

where S denotes the set of point within the capture area, S_i is the nearest point to x_i, $\|\ \|$ presents Euclidean distance norm, $P(x_i)$ is probability that point x_i is the global extreme point.

The acquisition area of double diamond algorithm is a circle with radius of two points, as shown in Fig.1. There are 13 possible detection points in acquisition area. Two capture patterns are used in this algorithm. One is LDSP (Large Diamond Search Pattern) which includes 9 detection points, the other is SDSP (Small Diamond Search Pattern) which includes 5 detection points, as shown in Fig.2. Probability distribution of the global best match point is considered in this strategy [11].

(a) LDSP   (b) SDSP
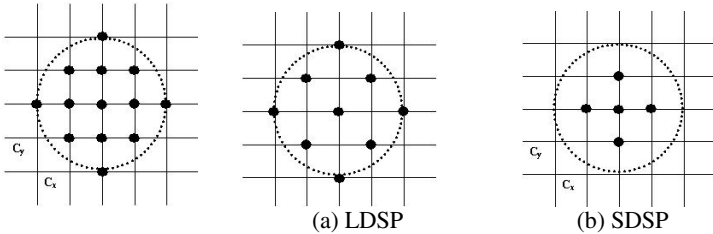
Fig. 1. Diamond detection   **Fig. 2.** Acquisition pattern

The steps of the double-diamond window algorithm can be summarized as follows:

1) According to the PN TH sequence, Δt and other parameters to determine the search space I:$\left(C_x, C_y \right)$.

2) Constructing LDSP centered at the origin of acquisition window, then detect the 9 points of LDSP. If the minimum phase error point lies at the center of the LDSP, then go to step 4; otherwise go to step 3.

3) Reconstructing LDSP centered at the minimum phase error point obtained in the last step. If the new optimal synchronization point lies at the center of LDSP, then go to step4; otherwise repeat this step, until the edge of the capture window is detected.

4) Capture pattern becomes the SDSP. The minimum phase error point in this step is just the best final synchronization points.
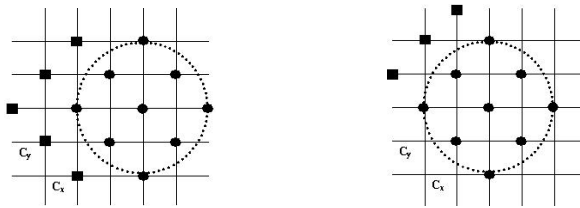
Fig. 3. The possible points when the best match point lies at the corner or edge of LDSP

If the best match point occurs at the center of LDSP, only 4 points need to be detected when the LDSP pattern turns to the SDSP pattern. Only 3 or 5 points need to be detected when it lies at the corner or edge of LDSP because of the detection points will overlap partly, as shown in Fig.3. So, the capture speed will improved greatly.

3.2 Analysis of Acquisition Efficiency

In the double-diamond search algorithm, there are 25 points in each diamond, and only 9 points need to be detected in LDSP. So the max times of missed detection is $\frac{9}{25}C$. When the best match point occurs at the center of LDSP, there the largest number point (5 points) need to be detected. The coverage point will be increased by 2 points in both two directions each new detection of LDSP. Considering the worst case (coverage all the possible phase points), if the right phase lies at the n-th unit, and coordinates are（C1, C2）, the number of point need to be detection in the right search can be described as follows:

$$n' \le 9 + \frac{C_2 - 5}{2} \times 5 + \frac{C_1 - 5}{2} \times 5 = \frac{C_1 + C_2}{2} - 16. \tag{13}$$

If the right phase still lies at the n-th unit, and the coordinates are (C1, C2), $n = C_2 \times C_x + C_1$, then

$$T(n, j, k) = n'T_i + jC'T_i + kT_{fa}$$
$$n' \le \frac{C_1 + C_2}{2} - 16 \tag{14}$$
$$C' \le \frac{9}{25}C$$

$$P(n, j, k) = \frac{1}{C} P_d (1 - P_d)^j \binom{K'}{k} P_{fa}^k (1 - P_{fa})^{K-k}, \tag{15}$$

where $K' = n' + jC' - j - 1$.

In order to facilitate analysis, the calculation of n remains unchanged in formula (12) and (13), then

$$\overline{T}_A = (C' - 1)T_{da}(\frac{2 - P_d}{P_d}) + \frac{T_i}{P_d}. \tag{16}$$

If $P_d = 1$, $P_{fa} = 0$, (that is, there is no missed and false detection), it becomes

$$\bar{T}_A = \frac{(C' + 1)T_i}{2}. \qquad (17)$$

Comparison of formula (1) with (2), formula (10) with (14), and the change of n, the capture probability can be improved greatly with less time cost.

4   Simulation Results and Conclusion

To analyze the capture probability and acquisition time, simulation was done under additive white gaussian noise environment. Single pulse is the quadratic differential gaussian pulse in [12], and its width is $w_b = 1.25e\text{-}8$, the repeat cycle is $t_f = 1.25e\text{-}6$. There are $N_s = 10$ pulses during each bit, and there are only '0' code in information sequence. The m-sequence with 1023 bit length and the simplest polynomial of m-sequence with n=10 are used. 10 shift registers are applied in this m-sequence in which the feedback is introduced at the 3-th and 10-th shift register. The results of simulation is shown in Fig.4 and Fig.5.

It shows that acquisition probability is greatly improved, and the synchronization time is shortened obviously compared with traditional sliding window algorithm.
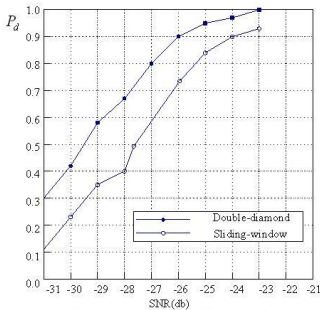
Fig. 4. Comparison of acquisition capability   Fig. 5. Comparison of acquisition time
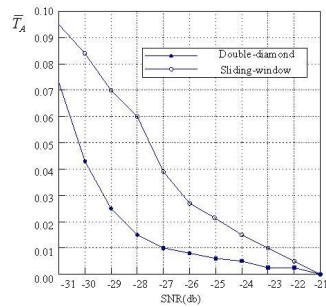
References

[1]   Win, M.Z., Scholtz, R.A.: Ultra Wide Bandwidth Time-hopping Spread-spectrum Impulse Radio for Wireless Multiple Access Communications. IEEE Transactions of Communications 48(4), 679–691 (2000)
[2]   Qiu, R.C., Liu, H., Shen, X.: Ultra-wideband for Multiple Access Communications. IEEE Communications Magazine 43(2), 80–87 (2005)

[3] Yang, L., Giannakis, G.B.: Ultra-wideband Communications: an Idea Whose Time has Come. IEEE Signal Process. Mag. 21(6), 26–54 (2004)

[4] Fleming, R., Kushner, C., Roberts, G., Nandiwada, U.: Rapid Acquisition For Ultra-Wideband Localizers. In: IEEE Conference on Ultra Wideband Systems and Technologies, pp. 4497–4500 (March 2002)

[5] Goeckel, D.L., Zhang, H., Kose, C., Wei, S., Win, M.: Rapid Code Acquisition in UWB Systems (PPT) (2002)

[6] Homier, E.A.: Rapid Acquisition of Ultra-Wideband Signals in the Dense Multipath Channel (PPT) (May 2002)

[7] Erseghe, T.: Two Novel Time-Hopping Sequence Constructions for Impulse Radio. In: Proceedings of The IEEE International Conference on Communications, New York, USA, p. 28 (May 2002)

[8] Erseghe, T.: Ultra Wide Band Pulse Communications. PhD, Thesis. University of Padova, Italy, pp. 26–28 (2001)

[9] Simon, M.K., Omura, J.K., Scholtz, R.A., Levitt, B.K.: Spead Spectrum Communications Handbook. McGraw-Hill Companies Inc. (2002)

[10] Li, R., Zeng, B., Liou, M.L.: A New Three Step Search Algorithm for Block Motion Estimation. IEEE Trans. on Circuits and Systems for Video Technology 4(4), 438–442 (1994)

[11] Zhu, S., Ma, K.K.: A New Diamond Search Algorithm for Fast Block Matching Motion estimation. IEEE Trans. Image Processing 9(2), 287–290 (2000)

The Multiple Linear Programming and Time Series Prediction Model Foundation and Analysis of Passenger Transport Quantity

Zhang Aixia and Wang Zheng

Hebei United University Tangshan, China, 063009

Abstract. Passenger traffic forecasting is an analysis and judgment for future period. It is a method on the basis of passenger service market research and statistical data by scientific method and means. In this paper multiple linear programming is used for forecasting passenger traffic by analysis GDP, public transport fares, urban population and the number of bicycles. Using quadratic moving average method, scientific prediction is made out for next years of passenger traffic, and a prediction model is set up. The forecasting precision of two type prediction models is compared. The fitted value based on the two models and the actual value is compared. The application of two models is also inspected.

Keywords: Passenger traffic, Forecasting model, Multiple linear programming, Quadratic exponential smoothing method.

1   Introduction

Transportation is an important link for human society and economy, living, and so on. Passenger traffic forecasting is a kind of the transport demand forecasting. Passenger traffic forecasting has critical implications for enterprises' prosperity. Market forecasting can provide the dynamic information for enterprises. The policy of the enterprise may be changed by knowing the relevant information in advance, such as the market will provide to enterprises any chance, or will cause danger, so a plan or measure may be made in advance. Market forecasting is not only the basis for enterprise management but also the basis of strategic plan. Passenger traffic forecasting is a guidepost for transport enterprise to development new markets. It can make technology and new line, carry out the new service.

   Passenger traffic forecasting is an analysis and judgment for future period. It is a method on the basis of passenger service market research and statistical data by scientific method and means.

2   Urban Passenger Traffic Forecast Models Design and Analysis

By investigation, passenger traffic has relation with GDP, public transport fares, the city's population and the quantity of bicycles. Last ten years' data are shown in Tab. 1.

Table 1. Last ten years' data

ordinal number of year	item				
	passenger traffic (10,000 /day)	GDP (100 million ¥)	public transport fares (0.01 ¥/km)	the city's population (10,000)	the quantity of bicycles (10,000)
1	122.06	123.96	2.5	201.9	87.9
2	120.84	139.27	2.5	205.92	98.38
3	117.66	161.29	2.5	210.5	108.7
4	125.78	180.59	2.5	216.99	118.70
5	128.18	191.14	2.5	222.54	119.73
6	126.53	203.32	4.00	226.98	132.64
7	124.14	222.93	4.00	230.85	125.90
8	121.32	244.56	7.00	236.45	133.58
9	127.91	332.84	9.00	240.85	147.67
10	135.27	390.07	9.00	248.35	153.15
11	147.73	421.31	9.00	255.71	160.91

2.1   Analysis for Passenger Traffic and Related Data

From the Tab.1, it can be seen as rapid growth of the GDP and urban population that the passenger traffic has also increased. Based on the data in last ten years the passenger traffic prediction model can be established. According to the model the demand of passenger traffic in future several years can be predicted. Set y for dependent variable (the overall passenger traffic), independent variable as x_1 (GDP), x_2 （public transport fares）, x_3 （the city's population） and x_4 （the quantity of bicycles）, the model can be established.

2.2   Multiple Linear Programming Model

(1) Multiple linear programming equation

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_4 x_4 \tag{1}$$

Using least square method and the data of Tab.1, the parameter values can be found.

$$b_0 = 94.924 \quad b_1 = 0.178 \quad b_2 = -3.24 \quad b_3 = 0.315 \quad b_4 = -0.195$$

(2) Multiple linear programming model

$$\hat{Y} = 94.924 + 0.178 X_1 - 3.24 X_2 + 0.135 X_3 - 0.195 X_4$$

(3) Model testing

residual sum of square： $S_c = 94.184$

r-square: $r^2 = 0.863$,

variance of bi:

$$cov(b_1, b_1) = 0.002\,9 \qquad cov(b_2, b_2) = 1.677\,9$$
$$cov(b_3, b_3) = 0.190\,5 \qquad cov(b_4, b_4) = 0.084\,9$$

Statistics of F:

$$F = \frac{n-p-1}{p} \frac{r^2}{1-r^2} = 9.52$$

If a=0.01 then the value of Fa can be calculated.

$$F_a(p,n-p-1)=F_{0.01}(4,6)=9.15 \quad , \quad F=9.52>F_{0.01}(4,6)=9.15$$

It can be seen that the regression effect is notable. It is to be remarked that all of the independent variable in the regression equation is prominent, but there is not notable for every independent variable. So it is necessary to test all the regression coefficient.

$$t_1 = \frac{b_1}{\sqrt{Cov(b_1,b_1)}} = \frac{0.178}{\sqrt{0.00296}} = 3.27 \qquad t_2 = \frac{b_2}{\sqrt{Cov(b_2,b_2)}} = \frac{-3.24}{\sqrt{1.6779}} = -2.50$$

$$t_3 = \frac{b_3}{\sqrt{Cov(b_3,b_3)}} = \frac{0.135}{\sqrt{0.1905}} = 0.31 \qquad t_4 = \frac{b_4}{\sqrt{Cov(b_4,b_4)}} = \frac{-0.195}{\sqrt{0.0824}} = -0.67$$

When a is equal to 0.05, it can be obtained the value of t_a by bilateral critical table of t-distribution.

$$t_a(n-m-1)=t_{0.05}(6)=2.43 .$$

Then the absolute value of t_i can be gained.

$$|t_1|=3.27>t_{0.05}(6)=2.43; \qquad |t_2| = 2.50 > t_{0.05}(6) = 2.43;$$

$$|t_3| = 0.31 < t_{0.05}(6) = 2.43; \qquad |t_4| = 0.67 < t_{0.05}(6) = 2.43;$$

It can be seen that x_1 and x_2 perform important function, but x_3 and x_4 have not prominent function. So much as a=0.04, it can be obtained by bilateral critical table of t-distribution that $t_{0.04}(6)=0.91$ and the absolute value of t_3 and t_4 are:

$$|t_3|=0.31<t_{0.04}(6)=0.91; \qquad |t_4|=0.67<t_{0.04}(6)=0.91;$$

So x_3 and x_4 should be deleted from the regression equation. The new regression equation is :

$$Y = b_0 + b_1 X_1 + b_2 X_2$$

Using least square method and the data of Tab.1, the parameter values is calculated again.

$$\hat{Y}=105.675\ 1+0.155\ 8X_1-3.148\ 7X_2$$

Test model:

residual sum of square : $S_c=103.512\ 4$

r-square: $r^2=0.850\ 4$

variance of bi: $cov(b_1,b_1)=0.001\ 1 \qquad cov(b_2,b_2)=1.355\ 4$

Statistics of F:

$$F=\frac{n\text{-}p\text{-}1}{p}\frac{r^2}{1\text{-}r^2}=22.75$$

When a is equal to 0.01, it can be obtained the value of F_a by critical table of F-test.

$$F_a(p,n\text{-}p\text{-}1)=F_{0.01}(2,8)=8.65 \quad .$$

So it can be seen the value of F.

$$F=22.75>F_{0.01}(2,8)=8.65$$

The regression effect is notable. It is to be remarked that all of the independent variable in the regression equation is prominent. Is every independent variable notable? So it is necessary to test all the regression coefficient again.

$$t_1 = \frac{b_1}{\sqrt{Cov(b_1,b_1)}} = \frac{0.1558}{\sqrt{0.0011}} = 4.70 \qquad t_2 = \frac{b_2}{\sqrt{Cov(b_2,b_2)}} = \frac{-3.1487}{\sqrt{1.3454}} = -2.71$$

When a is equal to 0.05, it can be obtained the value of t_a by bilateral critical table of t-distribution.

$$t_a(n\text{-}m\text{-}1)=t_{0.05}(8)=2.31$$

$$|t_1|=4.70>t_{0.05}(8)=2.31; \qquad |t_2| = 2.71 > t_{0.05}(8) = 2.31;$$

It can be seen that x_1 and x_2 are prominent linear with y. So the regression function can reflect the relation of GDP and public transport fares with passenger traffic.

From the regression function, fitted value of passenger traffic can be calculated. Aactual value and fitted value are statistic and show in Tab. 2.

Table 2. Actual value and fitted value

sequence number	item		
	actual value	fitted value	error
1	122.6	118.95	3.165
2	120.84	120.123	0.780
3	117.66	122.632	-4.972
4	125.78	125.046	0.734
5	128.18	127.484	0.696
6	126.53	122.859	3.671
7	124.14	128.169	-4.092
8	121.32	121.5	-0.18
9	127.91	128.479	-0.569
10	135.27	138.588	-3.318
11	147.73	143.636	4.096

2.3   Set Up Forecasting Model by Quadratic Moving Average Method

(1) Quadratic moving average method
Quadratic moving average method is a method that counterpoises simple moving average again and compares actual value with the value of simple moving average

and sets up forecasting model. It is a high-grade form for moving average method. It can overcome simple moving average method's shortcomings and enhance forecasting effact. It is applied to slantingly forecasting.

First, according to the lag relationship among the actual, the simple and the quadratic, quadratic moving average method calculates the differentials between the simple moving average value with the quadratic moving value. Next, the differentials is added to the simple moving average values and trend value should be taken into account. Last, a near actual value may be seeked.

(2) Quadratic Moving Average method forecasting model

$$\hat{x}_{t+T} = a_t + b_t T$$

Formula:

$\hat{x}_{t+T}$ ——forecasting value for number t+T

a_t , b_t ——params

$$\begin{cases} a_t = 2S_t^{(1)} - S_t^{(2)} \\ b_t = \dfrac{2}{n-1}(S_t^{(1)} - S_t^{(2)}) \end{cases}$$

$S_t^{(1)}$ is the simple moving average value and $S_t^{(2)}$ is the quadratic moving average value.

$$S_t^{(1)} = (x_t + x_{t-1} + \cdots + x_{t-n-1})/n$$

$$S_t^{(2)} = (S_t^{(1)} + S_{t-1}^{(1)} + \cdots + S_{t-n-1}^{(1)})/n$$

(3) result

There, n is assumed for 3. Using the last ten year's data, forecasting vaule can be calculated by quadratic moving average method. The results show in Tab.3.

Table 3. The results of quadratic moving average method

sequence number	passenger traffic (10,000 people /day)	$S_t^{(1)}$	$S_t^{(2)}$	a_t	b_t	$\hat{x}_t$ T=1
1	122.06					
2	120.84					
3	117.66	120.16				
4	125.78	121.42				
5	128.18	123.87	121.81	125.93	-0.17	
6	126.53	126.83	124.04	129.62	2.79	125.76
7	124.14	126.28	125.66	126.90	0.62	132.41
8	121.32	123.99	125.70	122.28	-1.71	127.52
9	127.91	124.45	124.90	124.00	-0.45	120.57
10	135.27	128.16	125.53	130.79	2.3	123.55
11	147.73	136.97	129.86	144.08	7.11	133.09
12						151.19

The value of a_{11} and b_{11} has been gained from the Tab.3, so the forecasting model can be set up.

$$\hat{x}_{11+T} = a_{11} + b_{11}T$$
$$\hat{x}_{11+T} = 144.08 + 7.11T$$

Formula:

$\hat{x}_{t+T}$ ——forecasting value for number 11+T, T are equal to 1,2,3.

Table 4. Actual value and fitted value

sequence number	item		
	actual value	fitted value	error
1	122.60		
2	120.84		
3	117.66		
4	125.78		
5	128.18		
6	126.53	125.76	0.77
7	124.14	132.41	-8.27
8	121.32	127.52	-6.2
9	127.91	120.57	7.34
10	135.27	123.55	11.72
11	147.73	143.636	4.096

2.4   Two Prediction Models Analysis and Comparison

First, Multiple linear programming equation and Quadratic moving average equation are shown below:

$$\hat{y} = 105.6751 + 0.1558x_1 - 3.1487x_2 \qquad \hat{x}_{11+T} = a_{11} + b_{11}T$$

In the multiple linear programming equation Y is the function of x_1 and x_2. Y is shared by two variables. Quadratic moving average equation is a formula that the variable x_{11+T} is changing with T.

Second, from the Tab.2 and Tab.4, multiple linear fitted value is close to the actual. That is to say multiple linear regression model has a higher accuracy than quadratic moving average. For example, the error for the forecasting value of the number 11 is 4.96, but the quadratic moving average is 14.64. Because only 11 year's passenger traffic is taken into account and other factors dose not be considered. The forecasting precision is lower than multiple linear regression. But multiple linear regression is more complex than quadratic moving average method.

3   Summaries

The transport demand forecast model has been investigated in this paper. From the multiple linear programming prediction model, a conclusion can be drawn that passenger traffic is the funcion of GDP and public transport fares. When GDP will not

change with the public transport fares increasing, the passenger traffic shows decreasing trend. Similarly, when the public transport fares will not change with the GDP increasing, the passenger traffic shows increasing trend. The city passenger traffic demand of the next few years can easily be forecasted. For example, by using quadratic moving average forecasting model, passenger traffic of number 12 is 151.19(10,000 people /day). So the manager can easily forecast the passenger traffic demand of the next few years, which can provide scientific data for passenger transportation busywork plan formulated. During the enterprise management process, enterprises can develop healthily by utilizing effectively forecasting method. The manpower, material and financial resources can be reasonable and the implementation of resources can be optimized. In the practical application, more forecasting methods should be utilized and more factors should be taken into account. Using combination methods to predict the result must be amended and improved prediction of precision.

References

[1] Shen, Z.-Y., Deng, X.-J.: Transportation Engineering, pp. 22–23. China Communication Press, BeiJing (2003)
[2] Zhai, Z.-M.: Road Traffic Optimize, pp. 17–19. China Communication Press, BeiJing (2004)
[3] Zhao, J.-Y.: Road Transport System Engineering, pp. 33–34. China Communication Press, BeiJing (2004)
[4] Luo, Y., Yu, R.-D.: Road Transport Organization, pp. 69–71. China Communication Press, BeiJing (2005)

A Fast Matching Algorithm for Content-Based Publish/Subscribe Systems

Tao Xue and Qi Jia

College of Computer Science, Xi'an Polytechnic University, Xi'an, Shaanxi, China
xthappy@gmail.com, jiaqi19861215@163.com

Abstract. How to efficiently match high volumes of events against large numbers of subscriptions is a key issue for large-scale content-based publish/subscribe systems. In this paper we present an efficient and applied matching algorithm that uses multi-dimensional indexing mechanism to speed up constraints query and exploits the covering relations between constraints to reduce unnecessary matching. Experiments show that our algorithm is significantly more efficient and scalable than other common used matching algorithms.

Keywords: Publish-subscribe, Content-based matching, Multi-dimensional indexing.

1 Introduction

Currently, the content-based publish/subscribe system is a hot research area. In this system the event is no longer dependent on an external standard (such as channel, theme, etc.) to classify, but the content in accordance with the event itself. Subscribers subscribe to the event according to the contents of the event, not to the subject of pre-defined themes by the system. When designing efficient content-based publish/subscribe systems, one of the key issues to be resolved is what strategy should be used to efficiently match large numbers of events and subscriptions.

IBM's Gryphon [1] system uses a parallel search tree to solve the matching between the event and subscription. Each event agent has a complete parallel search tree, and determines the forwarding direction of the incident by means of the tree. But the algorithm can only deal with equal determination. Siena [2][3] is a event notification service in the WAN, which uses the covering relation of the subscriptions and merging subscriptions to reduce the matching frequency, thus reducing the matching complexity. Yan and Garcia-Molina [4] describe the way of solving the matching of the keywords and the document object applying the counting and tree model algorithm, and achieved good results. Gough and Smith [5] proposed a match based on automata theory, it transfers all the constraints relying on a property to a finite state. It's been proved theoretically feasible, but not given a specific algorithm.

2 Subscription Language and Event Model

For content-based publish/subscribe systems, on the one hand sufficient expression languages must be provided so that users can describe the interesting events

conveniently, on the other hand the cost that it takes to match the event and subscription should be balanced. As a compromise, we use a subset of SQL language of relational database as the subscription language.

● Event

Event consists of a set of attributes {A1, A2, ... An}, and each attribute is a triple (data type, attribute name, value). Events can be expressed as: Event = $\cup$ attribute.

● Constraint

Constraint is a stateless Boolean expression. It's expressed as quad (type, attribute, operator, value).

Definition 1. If attribute A ($type_a$, $name_a$, $value_a$) of the event matches against a constraint C ($type_c$, $attribute_c$, op_c, $value_c$), the condition must be: ($type_a = type_c$) $\wedge$ ($name_a = attribute_c$) $\wedge$ { op_c ($value_a$, $value_c$) = true}, denoted by A$\in$ _match_ C.

Definition 2. Given two constraints c1 and c2, we say that c2 covers c1, denoted by c1$\subseteq$c2, namely: ($\forall$ a: a$\in$ _match_ c1 $\rightarrow$a$\in$ _match_ c2)$\Rightarrow$c1$\subseteq$c2.

● Subscription

Subscription is composed of a constraints set {C1, C2, ... Cn}. The relation between constraints is the conjunction relations, expressed as Subscription = $\cup$ constraint.

Definition 3. For each constraint c in the subscription S, there is at least one attribute a in the event E, making a$\in$ _match_ c, then we affirm that the event E and subscription S match, namely: $\forall$ c$\in$S, $\exists$ a$\in$E$\wedge$a$\in$ _match_ c$\Rightarrow$E$\in$ _match_ S.

3   Matching Algorithm

● Data structure

The main data structure is shown in Figure 1. First, we break all the subscriptions into the constraint sets, remove duplicate constraints. Then these constraints are organized into a multidimensional indexing structure, in accordance with its data type, attribute name, comparison operators. Index is made level by level, and the final index entry comes into a list of constraints.

To maintain the subordination relation of the constraints and subscriptions, a constraint-subs link is taken, where each constraint points to a list that stores all the subscriptions identification that contains the constraint.
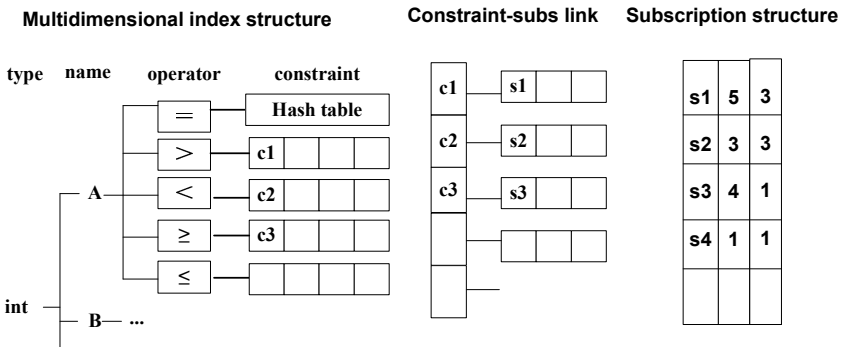
Fig. 1. Data Structure of index count matching algorithm

● Algorithm steps

The first stage is to find matching constraint for each attribute of the event according to their types and supported comparison operators; while the comparison operator is "=", do the hash table lookup directly, and for other comparison operators, get the first match of the constraints, thanks to the coverage relationship between constraints, its subsequent binding will also meet the match.

In the second phase, for each constraint has been successfully matched, the subscriptions belongs to it is searched. Then these subscriptions' corresponding count value will plus 1. If the subscription's count value is equal to the target value, the subscription matches.

Figure 2 shows the steps of the algorithm in detail.

```
Input: Event e = (T₁,A₁,V₁),...,(Tₙ,Aₙ,Vₙ)
Output: matchedSubscriptions: the set of
subscriptions matched event e
{ matchedSubscriptions:={};
 matchedConstraints:={};

// Stage I
for each Aᵢ in e
    for each operater op support by e.Tᵢ{
        if op is "=" then
            constraint con
                :=getFromConstraintHT(e.Vᵢ);
        else {
                search first matched constraint
                con in op.CL;
                if con ≠ null then
                    for each constraint
                        c∈getSucceeder(con,op.CL)
                        matchedConstraints
                            :=matchedConstraints ∪ {c};
    }
        if con ≠ null then
          matchedConstraints
              :=matchedConstraints ∪ {con};
    }

    // Stage II
    for each constraint c∈matchedConstraints
        for each subscription s∈getRelatedSubs(c){
        s.count:=s.count + 1;
        if s.count = s.target then
          matchedSubscriptions
              :=matchedSubscriptions ∪ {s};
     }
    return matchedSubscriptions;
}
```

Fig. 2. Index count matching algorithm

4   Experiment and Evaluation

We implement the sequence comparison method, the traditional counting method and the proposed count matching algorithm with multi-dimensional indexing and constraints covering characteristics. The prototype system includes a data generator, which automatically generates appropriate events and subscriptions according to the configuration.

Experiment is conducted on the PC of a Pentium IV 3.0GHz processor, 2G memory, Windows2003 operating system. Each data point of the results is the average value obtained by running the same parameters 10 times.

Due to space limit, we only illustrate the number of subscribers' effects on the algorithm. The experiment result is shown in Figure 3.
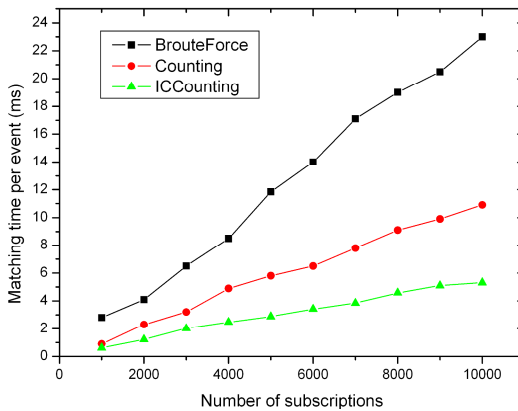
Fig. 3. The number of subscribers' effects on the algorithm

From the experimental results, it can be seen that linear comparison method has the least efficiency, and due to the characteristic of multiple index and constraints coverage, our algorithm is more efficient than the traditional counting method. In addition, as the number of subscribers increases, the efficiency of the algorithm declines much less than the linear comparison method and the traditional counting method, which indicates that its scalability is very good and it's ideal to be used in large-scale systems.

5   Summaries

In this paper, an efficient matching algorithm based on subscription language and event model is proposed, which uses multi-level indexing for high-speed matching of the event and related constraints and reduces the number of matching by the use of the cover relation of the constraints. We have implemented the algorithm and tested it with a variety of configurations, the experimental results show that the algorithm has better performance and strong scalability. Our future work is to integrate the algorithms into our prototype content-based publish subscribe system, and do further research on content-based routing algorithm.

References

[1] Aguilera, M., Strom, R., Sturman, D., Astley, M., Chandra, T.: Matching events in a content-based subscription system. In: PODC: 18th ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing (1999)

[2] Carzaniga, A., Rosenblum, D.S., Wolf, A.L.: Design and Evaluation of a Wide-Area Event Notification Service. ACM Transactions on Computer Systems 19(3), 332–383 (2001)

[3] Carzaniga, A., Wolf, A.L.: Forwarding in a Content-Based Network. In: Proceedings of ACM SIGCOMM 2003, Karlsruhe, Germany, pp. 163–174 (August 2003)

[4] Yan, T.W., Garcia-Molina, H.: The SIFT information dissemination system. ACM Transactions on Database Systems 24(4), 529–565 (1999)

[5] Gough, K., Smith, G.: Efficient recognition of events in distributed systems. In: Proceedings of 18th Australasian Computer Science Conference (ACSC) (February 1995)

Optimal Formation Shape Control of Wheeled Robots

Yixin Cao, Qinghe Wu, and Huang Huang

School of Automation, Beijing Institute of Technology, Beijing, China
`cao.yi.xin@sohu.com`

Abstract. This paper considers the shape control of a group of mobile robots. The desired formation is only constrained by a shape with a flexible scale. An algorithm is proposed that optimizes the geometric performance during formation attainment. It is shown that the introduction of time varying scale function on the real time positions of robots improves the reliability of estimation in sensor-target localization. The algorithm is validated on the Pioneer wheeled robots platform where feedback linearization is introduced to overcome the nonlinearity of the robots kinematics.

Keywords: Formation control, Optimal geometry, Wheeled robots, Feedback linearization, Multi-agent system.

1   Introduction

Nowadays, cooperative control of multi-agent systems became one of the leading research areas in artificial intelligence and for better performance in these applications, various cooperative control capabilites need to be developed , including formation control, consensus, rendezvous, obstacle avoiding, flocking, and attitude alignment[1].

Formation control is an important issue in coordinated control of a group of robots, various methods are proposed, such as behavior based, aritificial pentential field, or leader following. Generally speaking, most, if not all of these approaches required a pattern for robots before formation control, including both of edges and degrees, although the pattern might be less impressive in performance, eventually giving rise to additional cost for keeping system stable. Actually, early in the 1950s, a similar argument had been discussed in sensor networks, and it's widly shared that geometric shape is the main factor of sensor location accuracy while the distance between sensors and target is much larger than that among sensors[2]. Therefore, to improve work efficiency is considerable by designning a algorithm that drive robots reaching an agreement in a desired shape as well as maintaining a high DOS to the optimal shape, rather than individuals move independently, then during convergence the sensors' measurement would be more reliable, and sequentially formation errors decrease, either. Refenrences [3] proposed a fixed structure nonlinear   control law, instead of assigning a set of desired side lengths, the only requiement for the desired geometry is a triangle without any location, rotation and most importantly, scale constraints.

This paper considered a mobile robot's kinematic model and using graph theoretical notions as detailed blow and improved a algorithm that optimize the formation shape, finally validated on three Pioneer wheeled robots and achieved favorable results. In process of these formation control experiments, multi-agent

system chose a optimal one from an ocean of triangles similarly calculated by a pre-defined degrees of triangle so as to improve the cooperative performance of system, and consequently the reliability of sensor-target localization. Additionally, implementing the algorithm on wheeled robots, both maximum velocity and maximum acceleration should be taken into account, otherwise, while formation control law exceeds the limit speed of robot, it's easy to lead to a slice of unpredictable problems, such as harmful for robots, or influence the stability of system in a adverse way and then result in formation failure. So, after using feedback linearization, which combined the robot's kinematic model and their input/output equation for applicable inputs acting on vehicles directly, this paper also limitted the expected speed in a permissive range of vehicles.

2   Graph Theory Notations

Let $G = (V, E)$ be a directed graph to model interaction among vehicles, where V is a finite nonempty set of nodes, and $E \in V \times V$ represents the edge set, which is a set of ordered pairs of nodes, moreover, $|E|$ is used for the number of edges in this set. Each edge (i, j) means that the jth vehicle can access the state information of the ith vehicle, but not necessarily vice versa. On the contrary, the edge (i, j) in an undirected graph are unordered, that is to say, the ith vehicle and the jth vehicle can exchange information from each other. Note that an undirected graph can be considered a special case of a directed graph.

The incidence matrix $H(G) = [h_{ij}] \in R^{n \times |E|}$ (the following brief write down for H)of a graph G is the 01-matrix with rows and columns indexed by the nodes and edges of G, and if the node i is in the edge j, h_{ij} is defined as 1, otherwise 0.

3   Fixed Structure Control Law

In recent years, the key issues in multi-agent optimal control focus a lot on the control strategies using minimum energy dissipation, fastest convergence rate or shortest path. However, the object function in this section is an accumulation of formation and the desired geometries degree of similarity, and a fixed structure nonlinear control law that ensure system stable is designed to find a optimal pattern matching strategy.

Considering the optimal pattern matching problem, this section compared between a system with a time varying scale function relative to edges and the one with a contant scale. Furthermore, the fixed-structure nonlinear control law is a fuction relative to the fromation scale and edges.

a) Similar   b) Dissimilar   c) Dissimilar
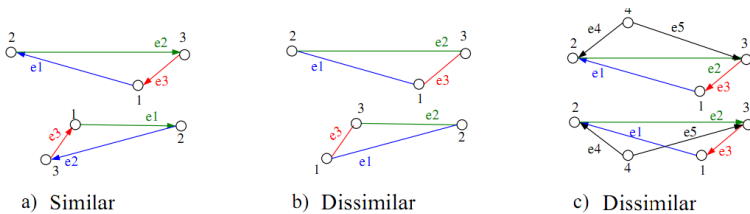
Fig. 1.

Let a graph $G = (V, E)$, and then $z(G) \in \mathbb{R}^{2|V|}$ (the following brief write down for z)is a representation of G, where the ith vector denotes the cartesian coordinates of the ith vertex. If each edge is given a direction, correspondingly $|E|$ edge vectors $e_i \in \mathbb{R}^2$.

Figure 1a and 1c represent, respectively, similar geometries and dissimilar ones. Despite that it is universally acknowledged that the two triangles in figure 3-1b are similar, but this paper regards them as dissimilarity due to the swapping vertexes.

A pre-defined $e_d(G)$ represents but one geometries, denoted by $S(e_d)$:

$$S(e_d) = [s_1\ s_2\ \cdots\ s_l]^T \in \mathbb{R}^2, l = |E|, s_i \in \mathbb{R}^+ \tag{1}$$

Where $s_i = \|e_{d_i}\|$, called S.

By introducing the extended incidence matrix: $\widehat{H} = H \otimes I_2$

Where H denotes the oriented incidence matrix for the triangular formation system and $\otimes$ means the Kronecker product, we have rigidity matrix:

$$R(e) = \frac{\partial \bar{z}}{\partial z} = \Lambda(e)^T \widehat{H} \tag{2}$$

With rigidity function:

$$\bar{z} = r(e) = \frac{1}{2}[\|e_1\|^2 \cdots \|e_m\|^2]^T \tag{3}$$

Where $\Lambda(e)^T = \text{diag}(e_k)$, $e_k \in \mathbb{R}^2$;

The degree of similarity(DOS) between a geometry e and a shape S is:

$$\text{dos}_\Theta(e, S(e')) = \|\rho_\Theta(e, e')\|_2^{-2} \tag{4}$$

A geometry with a higher DOS to S is said to be more resemble to S. When e is similar to S, $\text{dos}_\Theta(e, e') \to \infty$; The details with DOS function don't present here because of the limitation of space, which has been explained explicitly in the former work[3].

Consequently the cooperative performance of the three agents is the integration of geometries DOS over the entire convergence:

$$J_v(e_0, u, \Theta) = \int_0^\infty \text{dos}_\Theta^{-1}(e, S(e'))\, d\tau = \int_0^\infty \|R(e)^T[r(e) - \Theta\bar{S}]\|^2\, d\tau \tag{5}$$

Where $\bar{S} = [\bar{s_i}]$ and $\bar{s_i} = s_i^2$. Notice that smaller J_v means better cooperative performance in the convergence of multi-agent system.

Lemma 3.1: Given a desired shape S, a triangular formation system with control law:

$$u(s_c, e) = -R^T(e)[r(e) - s_c^*\bar{S}] \tag{6}$$

converges to the invariant set $\{e|S\}$ exponentially at the minimal cost value $J_v = J_v(e_0, u, s_c^*)$ if and only if

$$s_c^* = \frac{\sum_{i=1}^n \|e_i(0)\|^2 \bar{s_i}}{2 \sum_{i=1}^n \bar{s_i}^2} \tag{7}$$

Lemma 3.2: Consider a triangular formation system with three agents where each agent is modeled by a single integral $\dot{z} = u$, consider the following control law on $\tilde{s}(e)$ and e:

$$u(\tilde{s}(e), e) = -R(e)^T M(e)^T [r(e) - \tilde{s}(e)\bar{S}] \tag{8}$$

The triangular formation system will be exponentially stable and converge if:

$$M(e) = s_D^2 I_3 - s_D \bar{S} r(e)^T (\bar{D}^T + \bar{D}) + s_N \bar{S} \bar{S}^T \bar{D} \tag{9}$$

Where

$$D = \begin{bmatrix} \bar{s}_2 + \bar{s}_3 - 4\bar{s}_1 & \bar{s}_2 - \bar{s}_3 & \bar{s}_3 - \bar{s}_2 \\ \bar{s}_1 - \bar{s}_3 & \bar{s}_1 + \bar{s}_3 - 4\bar{s}_2 & \bar{s}_3 - \bar{s}_1 \\ \bar{s}_1 - \bar{s}_2 & \bar{s}_2 - \bar{s}_1 & \bar{s}_1 + \bar{s}_2 - 4\bar{s}_3 \end{bmatrix}$$

$$\tilde{s}^*(e) = \frac{r(e)^T \bar{D} r(e)}{\bar{S}^T \bar{D} r(e)} : \triangleq \frac{s_N}{s_D}$$

Moreover, $J_v(e_0, u(\tilde{s}(e), e), \tilde{s}(e)) < J_v(e_0, u(s_c, e), s_c)$.

4   Feedback Linearization

Before applying them on vehicles, both of the two control laws above need to be converted to applicable inputs: linear velocity v_i and lateral velocity ω_i. So feedback linearization is introduced,   which combined the robot's kinematic model and their input/output equation to linearize the two nonlinear control laws.

Figure 2 shows Pioneer robot amigobot's kinematic model:
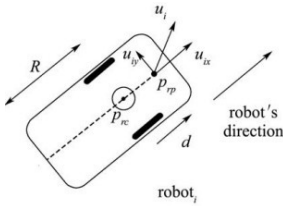
Fig. 2.   **Fig. 3.**

Where p_{rp} and p_{rc} represent, respectively, amigobot's refenrence point and autogyration center, d is offset between them with the heading of robot as positive direction.

Let (x_i, y_i) denotes the Cartesian position, θ_i is the orientation with a range between π and $-\pi$. Then:

$$\begin{cases} x_i = x_{p_{rp}} = x_{p_{rc}} + d \cos \theta_i \\ y_i = y_{p_{rp}} = y_{p_{rc}} + d \sin \theta_i \end{cases} \tag{10}$$

and $\dot{x}_i = u_{i_x}, \dot{y}_i = u_{i_y}, \dot{\theta}_i = \omega_i$.

According to[4], we feedback linearize for (x_i, y_i), letting:

$$\begin{cases} v_i = u_{xi} \cos \theta_i + u_{yi} \sin \theta_i \\ \omega_i = -\dfrac{1}{d} \sin \theta_i + \dfrac{1}{d} \cos \theta_i \end{cases} \tag{11}$$

Which are simplified kinematic equation and feedback linearization but is sufficient for the purpose of this paper.

In our experiments, three amigobots of the Pioneer family of moblie robots as shown in Fig. 3 have been employed. They can communicate with each other through ethernet with TCP/IP protocols and rely on encoder data for their position and orientation information.

Also, these amigobots has a limitation in maximum linear velocity 1m/s and maximum angle velocity 300deg/s. Let d=0. 15m, then:

$$u^* = Ku(\Theta, e) \tag{12}$$

Where K=15 is the proportional gain of system. And then:

$$v_i = \begin{cases} v_{max}, & v_i(t + \Delta t) > v_{max} \\ v_i(t + \Delta t), & -v_{max} > v_i(t + \Delta t) > v_{max} \\ -v_{max}, & v_i(t + \Delta t) < -v_{max} \end{cases} \tag{13}$$

$$\omega_i = \begin{cases} \omega_{max}, & \omega_i(t + \Delta t) > \omega_{max} \\ \omega_i(t + \Delta t), & -\omega_{max} > \omega_i(t + \Delta t) > \omega_{max} \\ -\omega_{max}, & \omega_i(t + \Delta t) < -\omega_{max} \end{cases} \tag{14}$$

5   Experimental Results

In this section, we experimentally validate the proposed control law on three amigobots, we conduct experiments with time-varying and constant scale, respectively, and then compare them.
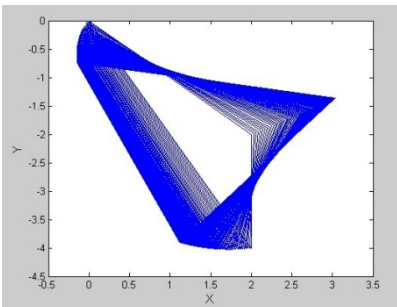
Fig. 4.
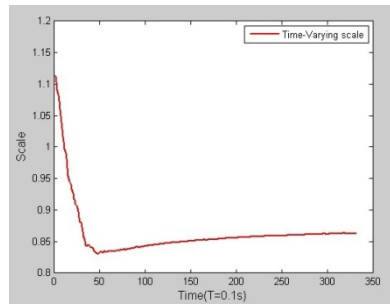
Fig. 5.
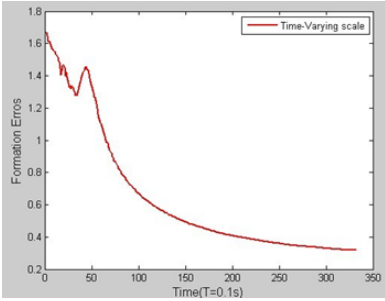
Fig. 6.
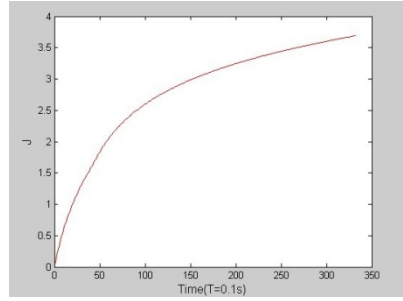
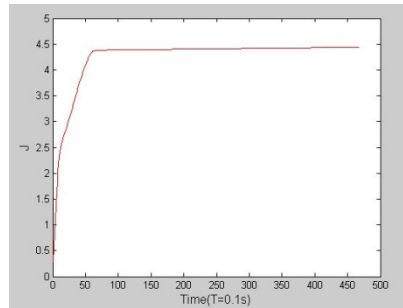Fig. 7.
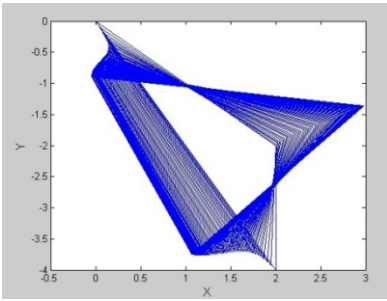
Fig. 8.
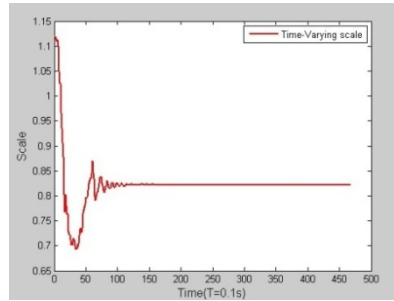
Fig. 9.
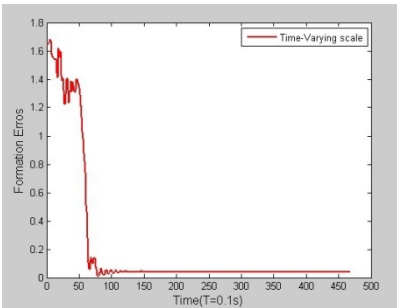
Fig. 10.

Fig. 11.

Letting the initial position of robots (*0, 0*), (*2, -2*), (*2, -4*) in all experiments, thus, the triangular angles are $135°$, $26.6°$, $18.4°$.

Fig. 4 to 7 shows the process of formation with time-varying scale. Fig. 4 is position of amigobots during every sampling, as it shown that amigobots move to the optimal triangular in a smooth way, and finally achieve steady that means convergence. Fig. 5 and 6 are, respectively, scale function and formation errors, in fact, the formation errors have not converged to zero due to the alignment errors. Fig. 7 shows that the cost function J_v, which is smooth curve correspond to the dynamic modulation character analysised above, increases but increments decrease to zero.

Fig. 8 to 11 shows the process of formation with constant scale. Fig. 8 indicates that during amigobots' movement to convergence, compared with the situation of time-varying scale, vehicles approach in a unsmooth way, which leads to the increase of cost fuction. As shown in fig. 9, the initial and final value of the scale function is alike that in the experiment above. Fig. 10 shows that a better performance of final formation errors is achieved in this case. The cost function in fig. 11 represents the performance of system, and it shows formation control law with constant scale costs more relative to a greater overshoot, corresponding to analysis above.

To sum up, despite of the final formation errors are smaller in the latter experiment compared with the former due to alignment errors on amigobot, but considering the major function that represents performance of system, the introduction of time varying scale according to real time positions of robots increase the performance of multi-agent system.

6   Conclusion and Future Research

This paper considered a mobile robot's kinematic model and using feedback linearization and improved an algorithm that optimize the formation shape, finally validated on three Pioneer wheeled robots and achieved favorable results. In the experiments, using the same initial postions and desired geometry, we compared control law with time-varying scale and with constant one in the process of geometries modification, formation scale and errors, what's more, the cost function. Experiments results shown that the introduction of time varying scale according to real time positions of robots increase the performance of multi-agent system. Future research will experimentally decrease the final formation errors by using a second order integral, which increases the complexity of the problem.

References

[1] Ren, W.: Consensus Tracking Under Directed Interaction Topologies: Algorithms and Experiments. IEEE Transactions on Control System Technology, 370–373 (2010)
[2] Bishop, A.N., Fidan, B., Anderson, B.D.O., Dogancay, K., Pathirana, P.N.: Optimality analysis of sensor-target localization geometries. Automatica 46(3), 479–492 (2010)
[3] Huang, H., Yu, C., Wang, X.: Control of Triangular Formations with a Time-Varying Scale Function. In: American Control Conference (2011)
[4] Ren, W.: Nathan Sorensen. Distributed coordination architecture for multi-robot formation control. Roboties and Autonomous Systems, 324–333 (2008)

An Adaptive Semi-fragile Watermarking Algorithm Based on Wavelet Low Frequency Coefficients

Ying Jun[1], Liu Qing[2], and Liu Chao[1]

[1] School of Computer and Information
Hefei University of Technology
HeFei, 230009, China
[2] School of Physics & Information Science
Tianshui Normal University
Tianshui 741001, China

Abstract. In this article, a new adaptive semi-fragile watermarking algorithm based on low frequency coefficients of discrete wavelet transform is proposed. With the wavelet transform, the low frequency coefficients are classified into two classes, one class corresponds to the image of the strong texture region, the other class corresponds to the weak texture region. According to the result of classify, A binary watermarked logo is scrambled by Arnold transform and embedded into the low frequency coefficients through adaptive quantization modulation. Experimental results show that high imperceptibility is provided as well as higher robustness against common signal processing attacks, such as JPEG compression, adding noise. At the same time, it can also detect the location of any minute changes for image authentication; can be widely used in practice.

Keywords: Digital watermark, Wavelet transform, Wavelet tree, Authentication.

1 Introduction

Along with the explosive growth of the computer and internet technology, it provides a great convenience for available access to information databases around the world, but also brings a lot of security issues, Such as sensitive information is illegal copying and tampering. Therefore, information security, intellectual property protection and authentication are becoming important more and more. Digital watermarking as a new practical information hiding and copyright protection technology has attracted much attention and has become an international academic research focus in recent years.

Watermarking algorithms can be categorized in different ways. One classification is based on domain which the watermark is embedded. The earlier watermarking algorithms were almost works in spatial domain. Spatial domain techniques are not resistant enough to image compression and other normal image processing [1]. Transform domain watermarking schemes can provide higher image imperceptibility and are much more robust to image manipulations. Now transform domains are mainly focus on the discrete cosine transform (DCT) [2,3] and the discrete wavelet transform (DWT) [4,5]. In transform domain, DWT has been used more widely in digital image watermarking due to its time/frequency decomposition characteristics, which resemble

to the theoretical models of the human visual system (HVS). Another typical classification is to classify the watermark into robust, fragile and semi-fragile watermark. The robust watermark survives when the watermarked digital image is severely attacked and thus can be applied in copyright protection. On the other hand, the fragile watermark will be destroyed even if the change is minute in the marked digital image. As a tradeoff of robustness and fragility, semi-fragile watermark is proposed that can resist "content preserving" operations (such as JPEG compression, filtering, etc.) and be sensitive to "content altering" transforms (such as feature replacement) is more practicable than fragile watermark in image authentication [6-10].

An adaptive semi-fragile watermarking algorithm based on low frequency coefficients of DWT is proposed in this article. Experimental results show that the algorithm is robust against normal image processing and kinds of attacks, for example JPEG compression, adding noise, and has good invisibility. Simultaneously, it can detect any spiteful tamper on the marked image, and locate the tampered spatial regions accurately.

2   Strategy of Watermark Embedding

Imperceptible and robustness of common image processing are the two basic elements of authentication watermarking. The robustness is that the watermark information is still able to be detected or extracted after some signal processing or channel attacks. Obviously, the watermark embedding strategy plays a crucial role for the robustness of the watermark. The image of the low-frequency coefficients are always clearly excluded in the previous wavelet-based digital watermarking algorithm. Daren Huang, Rong Pan[9,10] et al proposed a new watermark embedding strategy on the basis of studied the distribution and amplitude of wavelet coefficients for qualitative and quantitative analysis, that is the watermark should be embedded into the low frequency coefficients of wavelet image firstly, and then embedded into the other frequency band according to the importance of wavelet bands if left. Meanwhile, the robustness of the watermark is also related with the intensity of the embedded watermark expected for the location of the place. The effective way to improve the robustness of watermark is to give a reasonable allocation of the watermark energy and improving the strength of the local component of embedded watermark as much as possible to meet the premise of invisibility taking advantage of the visual characteristics of human visual system.

Each pixel block of the image spatial domain corresponds to a few blocks of wavelet transform domain after the decomposition of discrete wavelet transform that the location relationship can be presented by a quad-tree. They are called the Level, Vertical and Diagonal sub-tree according the direction, Collectively known as wavelet sub-tree. The roots of wavelet sub-tree in the low-frequency band, the wavelet coefficients of the same spatial location and different scales corresponding to the same direction are called its children. At the low frequencies band, if the location of parents is (i,j), and then the location of their children are $(i,j+N)$, $(i+N,j)$, $(i+N,j+N)$, Where N is the width of low-frequency band. At the other frequencies band except the low-frequency, if the location of parents is (i,j), and then the location of their children are $(2i,2j)$, $(2i,2j+1)$, $(2i+1,2j)$, $(2i+1,2j+1)$. All the quad-tree have the same roots of the three different directions organized together to form a fixed size block, All of these

fixed-size blocks are called as wavelet block. The numbers of wavelet block are the numbers of low-frequency band coefficients that means a low frequency coefficient corresponding a wavelet block. The wavelet coefficients in the different scales and directions according to the same location in different spatial domains are grouped together by the wavelet block in order to make the wavelet coefficients associated with spatial domain information. Among them, the low frequency coefficients corresponding to most of the energy concentration present the average brightness of wavelet block in the image, as shown in Figure 1.
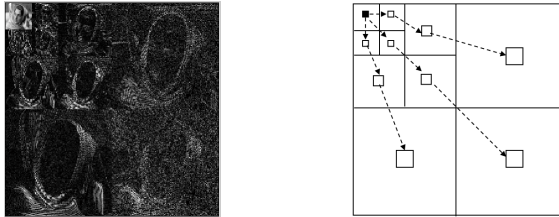
Fig. 1. Three-Level DWT decomposition and Wavelet trees

In this paper, the low frequency wavelet coefficients are classified into two classes according to the tree relationship between the different layers and the characteristics of wavelet coefficients in order to maximize the watermark strength and the energy of a rational allocation of the watermark signal.

3   The Proposed Algorithm

3.1   Watermark Preprocessing

In this paper, a meaningful binary image was chosen as a watermark. It should be scrambled firstly in order to improve the watermark's robustness against clipping, re-sampling and so on. Arnold transform is chosen as preprocess method of watermark for its intuitionist, simple, periodic and easy to use. For N × N digital square image, discrete Arnold transform[10] can be done as formula .1:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \left[\begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \right] \mod N \tag{1}$$

After traversing all pixels in the image, a scrambled image will be gained. In addition, this transform can be repeated for the image to generating different results until reaching the expectation.

3.2   Watermark Embedding

The embedding procedure includes the following step:

Step 1: Perform 3-levels DWT on the host image, it produces 3×n detail sub-bands as LH_n、 HL_n、 HH_n(n=1,2,3) and a low frequency approximation sub-band LL_3, Shown in Figure 1.

Step 2: Give the classification of low-frequency coefficients according to the texture of the sub-image: With stronger texture is the first category, denoted as S1, the other is the second category, denoted by S2. The absolute value of high-frequency coefficients is greater indicating that the corresponding position of the texture is stronger because of the high-frequency coefficients contains the texture or edge information. Therefore, the strength of the texture positions can be measured by the numbers of the high frequency coefficients of wavelet sub-tree is greater than the absolute value of a threshold T, defined as T_f. With this threshold processing, the operations is simplified. That is, if

$number\{|F(u,v) > T_f, (u,v) \in LH_n, HL_n, HH_n\} \geq T, \ n = 1, 2, 3$ then $LL_3(u,v) \in S_1$ Other

wise, $LL_3(u,v) \in S_2$, Where T_f and T are for the pre-set thresholds. Through a large number of experiments, we take $T_f = 0$ and T = 42 for different types of images have better robustness for common signal process.

Step 3: First, the binary watermark will be scrambled by Arnold transform to eliminate pixel spatial correlation and to improve robustness of the watermarking algorithm further. Then the watermark will be embedded into the host image according to the classification results by quantifying the low frequency coefficients, and the strength of the watermark is adjusted adaptively by the classification of the coefficients. It will produce a classification symbol matrix after the classification of low-frequency coefficients, this matrix can be treated as a secret key for increasing the security of the watermark. We use the DC-QM(distortion-compensated quantization index modulation) [11] algorithm presented by Chen and Wornell to embed the watermark. The DC-QM algorithm has been Prove is the optimal of capacity in several important probability models and less susceptible interference from the host signal. Compared with other quantization algorithm, the DC-QM algorithm has more advantages such as the image has smaller distortion and more robust to image process inmsame quantization step. The function of DC-QM is shown as formula.2

$$Q(i,j) = \begin{cases} round\left[\dfrac{f(i,j)}{\Delta}\right] \times \Delta & w(i,j) = 0 \\ round\left[\dfrac{f(i,j) + \frac{\Delta}{2}}{\Delta}\right] \times \Delta - \dfrac{\Delta}{2} & w(i,j) = 1 \end{cases} \quad (2)$$

Where, $f(i,j)$ is a coefficient in the LL_3 sub-band before embedding, $Q(i,j)$ is a coefficient in the LL_3 sub-band after embedding, $w(i,j)$ is the value of pixels at (i,j) of the binary watermark. Δ is the quantization step. Here, the quantization step is different according to the classification of the host image.

If $w(i,j) = 0$, then $Q(i,j) = \mu \times \Delta$; if $w(i,j) = 1$, then $Q(i,j) = \mu \times \Delta + \dfrac{\Delta}{2}$,

where,. $\mu = round[\dfrac{f(i,j)}{\Delta}]$

Step 4: Perform the inverse DWT (IDWT) on the DWT transformed image, including the modified coefficient sets, to produce the watermarked image. The watermark embedding procedure is represented in Figure 2.
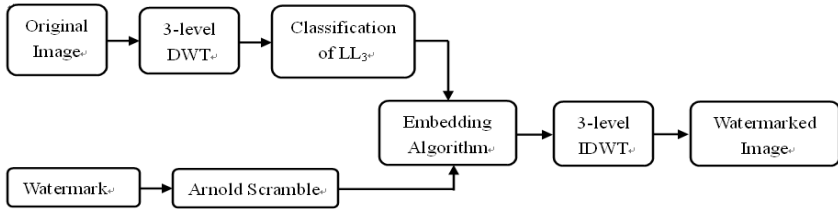
Fig. 2. Watermark embedding procedure

3.3   Extracting Procedure

The algorithm in this paper is a blind watermarking algorithm. First, perform 3-levels DWT on the watermarked image to gain the low-frequency coefficients, and then extract the watermark by the extraction algorithm according to Classification symbol matrix of the low frequency coefficients.

In this step, set the test data be $\hat{f}(i,j)$, according to the distance of $\hat{f}(i,j)$ between $(\mu \times \Delta)$ or $(\mu \times \Delta + \frac{\Delta}{2})$ to determine the information of $\hat{w}(i,j)$, if $\hat{f}(i,j)$ is near $(\mu \times \Delta)$, then $\hat{w}(i,j)=0$, otherwise, $\hat{w}(i,j)=1$. That is shown as formula.3

$$\hat{w}(i,j) = \mod[round[\frac{\hat{f}(i,j)}{\frac{\Delta}{2}}],2] \tag{3}$$

Finally, scramble the extracted watermark with Arnold algorithm with the same key times and gain the restore watermark.

4   Experiments and Results

We have performed our scheme on *lena* image (512×512×8 bits) in Matlab7.0 for testing the performance of our algorithm. The Original image and watermark, watermarked image and extracted watermark are shown in Figure3. The PSNR of the watermarked image is 42.87dB, which show that the proposed algorithm has good invisibility. Meanwhile, the extracted watermark is identical with the original watermark.

PSNR=42.87   NC=1
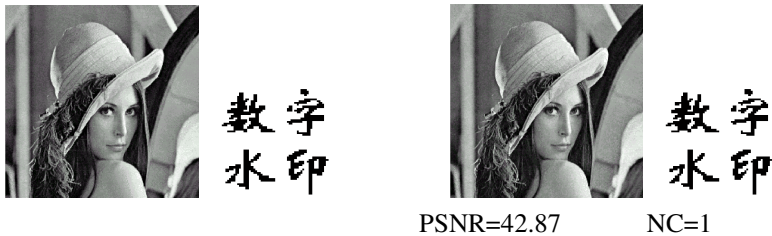
Fig. 3. The original image and watermark. The watermarked image and extracted watermark.

4.1   Robustness to JPEG Compression

The watermark algorithm robust against JPEG compression is the most essential for use because that JPEG image compression is the most common image processing operations. In order to test the robustness against the JPEG compression of the proposed algorithm, a lot of JPEG compression tests are done with different quality factor are shown in figure 4. Experimental results show that the proposed algorithm has good robustness against JPEG compression.
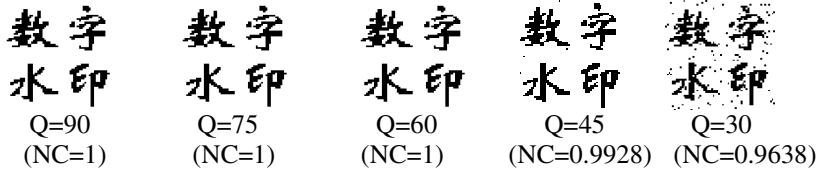

Q=90   Q=75   Q=60   Q=45   Q=30
(NC=1)   (NC=1)   (NC=1)   (NC=0.9928)   (NC=0.9638)

Fig. 4. The watermark extracted from the watermarked image compressed by JPEG with different quality factor

4.2   Robustness to Filtering and Adding Noise

Fig.5(a), Fig.5(b), Fig.5(c) are extracted watermark from the filtered image of median, Weiner and Gaussian low-pass filter, respectively and Fig.5(d) is extracted watermark from the image of adding noise. It is obviously show that the watermark extracted from the image can clearly recognized.

(a) Median filter   (b)Weiner filter   (c)Gaussian low-pass filter   (d)add Noise
(NC=0.9821)   (NC=0.9524)   (NC=0.9918)   (NC=0.8828)
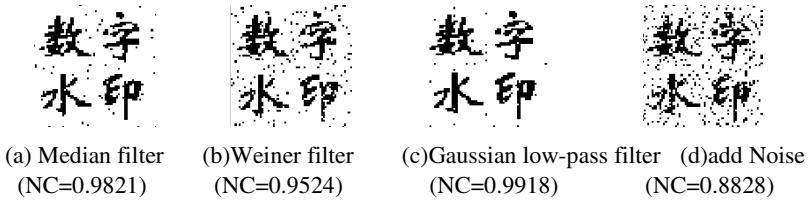
Fig. 5. The watermark extracted from filtering and adding noise

4.3   Tamper Detection and Location

Fig.6 (a),(d)are the tampered images, which some part of the image are changed after compressed by JPEG compression at the quality factor of 45, Fig.6(b),(e) are the difference image between the extracted watermark and the original watermark, Fig.6(c),(f)are the location of the original images tempered image.
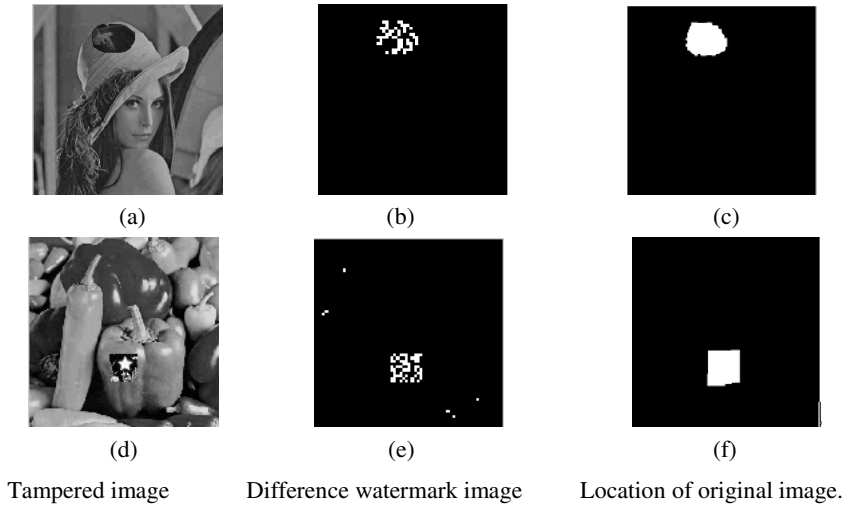
(a)	(b)	(c)
(d)	(e)	(f)
Tampered image	Difference watermark image	Location of original image.

Fig. 6. Tamper Detection and Location

5   Summaries

A secure semi-fragile watermarking algorithm for image authentication based on the low frequency of discrete wavelet transform has been proposed. It is highly secure and efficient due to the high energy focus on the low frequency of the image to make the watermark more robust to t normal image processing. And the accuracy in tamper detection due to the wavelet coefficients in the different scales and directions according to the same location in different spatial domains are grouped together by the wavelet block to make the wavelet coefficients associated with spatial domain information. Presented method is tested by most of the common image processing attack such as: JPEG compression, gaussian filtering, adding salt and pepper noise. Our future research will focus on how to recover the information from the tampered image effectively.

References

[1] Potdar, V.M., Han, S., Chang, E.: A Survey of Digital Image Watermarking Techniques. In: Proc. of the IEEE International Conference on Industrial Informatics, Perth, Australia, pp. 709–716 (2005)
[2] Chen, L.-W., Li, M.-F.: An effective blind watermark algorithm based on DCT. IEEE Trans. on Intelligent Control and Automation, 6822–6825 (2008)

[3] Zhang, B., Fan, L.-N., Song, B.-T.: A Watermark Embedding Algorithm Based on Low Bit Planes Fusion of DCT Space. IEEE Trans. on Computer Science and Engineering (2), 349–353 (2009)

[4] Tay, P., Havlicek, J.: Image Watermarking Using Wavelets. In: Proc. of the IEEE Midwest Symposium on Circuits and Systems, Oklahoma, USA, pp. 258–261 (2002)

[5] Pun, C.-M., Yuan, X.-C.: Geometric invariant digital image watermarking scheme based on histogram in DWT domain. Journal of Multimedia 5(5), 434–442 (2010)

[6] Che, S.-B., Ma, B., Huang, Q.-B., Wu, X.-D.: Semi-fragile Image Watermarking Algorithm Based on Double-Step. In: Information Engineering and Electronic Commerce, pp. 3–10 (2009)

[7] Wang, Z.-X., Zhao, X.-Y.: Recoverable Semi-fragile Watermark for Image Authentication. Journal of Image and Graphics 13(7), 1258–1264 (2008)

[8] Hu, H.-P., Zhou, C.-L.: A semi-fragile digital watermarking approach to authenticate an image. Huazhong Univ. of Sci. & Tech. (Nature Science Edition) 33(6), 4–7 (2005)

[9] Huang, D.-R., Liu, J.-F., Huang, J.-W.: An Embedding Strategy and Algorithm for Image Watermarking in DWT Domain. Journal of Software 13(7), 1290–1295 (2002)

[10] Pan, R., Gao, Y.X.: Image Watermarking Method Based on Wavelet Transform. Journal of Image and Graphics 7A(7), 667–671 (2002)

[11] Chen, B., Wornell, G.W.: Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory 47(5), 1423–1433 (2001)

Research on Chunking Algorithms
of Data De-duplication

Cai Bo, Zhang Feng Li, and Wang Can

School of Computer Science and Engineering, University of Electronic
Science and Technology of China, Chengdu 611731, China
`caibouestc@163.com`

Abstract. Data de-duplication is a technology of detecting data redundancy, and is often used to reduce the storage space and network bandwidth. Now it is one of the hottest research topics in the backup storage area. In this paper, five representative chunking algorithms of data de-duplication are introduced and their performance on real data set is compared. The experiment result shows that the performance of these methods is improved obviously from the whole-file chunking to the TTTD chunking. According to the analysis of their features, we can provide some references for backup storage systems to choose the best chunking algorithm for eliminating data redundancy.

Keywords: Data de-duplication, Eliminate data redundancy, Backup storage, Chunking algorithms.

1   Introduction

With the development of storage and computer technology, digital data has taken more and more space in recent years. According to statistics, 60% of these data is redundant, and the traditional data compression can only eliminate the intra-file redundancy. So in order to solve these problems, data de-duplication has been proposed. There are several chunking strategies of data de-duplication, such as whole-file chunking, fixed-sized chunking and content-defined chunking. The key idea of these chunking strategies is to break a file into small chunks and then find out the redundancy by fingerprint comparison. At present, these chunking methods are mainly used in the disk-based backup system because of its cost-effective space utilization [1, 2]. In general, fixed-sized chunking is mainly used in Venti [3] system and content-defined chunking is usually used in LBFS [4] and Pastiche [5].

In this paper, we mainly introduce whole-file chunking, fixed-sized chunking, basic sliding window chunking, improved sliding window chunking and TTTD chunking. Then compare their performance by experiments on real data set. At last, we will figure out the new problems of current situation and present the research direction of future works in section 4.

2   Chunking Methods

In this section, we mainly introduce whole-file chunking [6], fixed-sized chunking [3], content-defined chunking [4] and its two famous variants.

2.1   Whole-File Chunking

Whole-file chunking calculates the hash value of the whole file contents, using it as the file's identifier. We can use MD5 or SHA-1 here since they are the most widely used hash algorithms at present [7, 8]. Due to the collision resistant of these hash functions, we consider that two files with the same hash value usually have a very high probability to own the identical contents. So once a duplicate is found only one of them is stored. Otherwise, consider the file as new data. Whole-file chunking is simple and fast, but it can only detect exact file duplicate [9] since the entire file is viewed as a whole part.

2.2   Fixed-Sized Chunking

In fixed-sized chunking algorithm shown in Fig.1, all files need to be partitioned into blocks with a fixed size [10], 8Kbytes for example, and use MD5 or SHA-1 to calculate the hash value of all the chunks as their identifiers. During the process of hash comparison, once a same hash is found, consider the chunk as redundancy; otherwise, store the chunk and its hash.
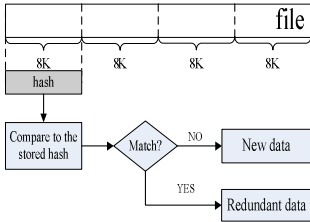
Fig. 1. Fixed-sized chunking   **Fig. 2.** Basic sliding window chunking
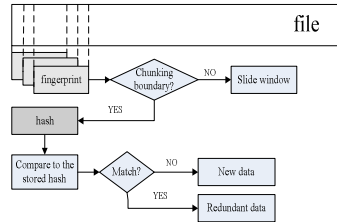
According to statistics, fixed-sized chunking can effectively reduce 30% of the storage space [11]. But since chunking boundaries are decided by the offset, this method is very sensitive to the insertion and deletion operation.

2.3   Content-Defined Chunking

Content-defined chunking also named variable-sized chunking is different from the fixed-sized chunking. Chunking boundaries in this strategy are determined based on the contents, so it is more resistant to the insertion and deletion [12]. Now it is believed as the best algorithm for backup system [13]. In this part we mainly introduce the basic sliding window chunking and its two variants, the improved sliding window chunking and the TTTD chunking.

2.3.1   Basic Sliding Window Chunking

Chunking algorithm usually consists of three steps: chunking, hash value generation and redundancy detection [14]. Basic sliding window chunking shown in Fig.2 first identifies a specific marker as the constraint. The marker may be a sequence of bytes

such that some mathematical function of the data results in a certain bit pattern or as simple as a full stop [15]. Second, it slides a fixed size window from the file beginning one byte each time. Rabin's fingerprint algorithm [16] is often used to calculate fingerprint [17] because of its high efficiency. Each time the fingerprint satisfies the marker, chunking boundary is set here [18], otherwise the window moves one byte and a new fingerprint is generated. At last, calculate the hash between the two boundaries with MD5 or SHA-1 and determine whether this chunk is redundant.

Due to the using of basic sliding window chunking, the modifications only impact the changed chunks and others remain unaffected. This algorithm highly improves duplicates detected rate, but produces variable-sized chunks.

2.3.2   Improved Sliding Window Chunking

Improved sliding window chunking shown in Fig.3 has the advantages of fixed chunk size k bytes and low sensitivity to the insertion and deletion. First set a window at file beginning, compute the checksum of the data, if a same one is found, use MD5 or SHA-1 to calculate the hash. There are several possibilities here: first, if a same hash is found, consider it as redundancy and move the window forward over this data chunk; second, if the checksum or the hash is not matched, the method proposes an assumption that the chunk may be duplicate that has been offset changed by insertion or deletion [15]. In this case, the window moves one byte forward the file, if a matched checksum and hash are found and bytes from previous chunking boundary are less than k, consider it as redundancy, the residue data between two chunks is viewed as a new chunk; last, if no matched checksum and hash value are both found when the window has passed by k bytes, output these k bytes as a new chunk.
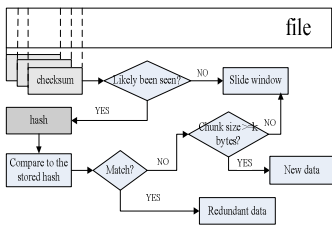

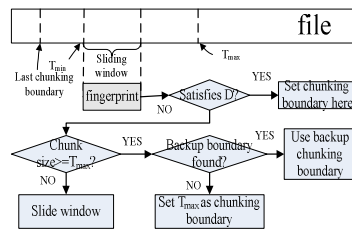**Fig. 3.** Improved sliding window chunking   **Fig. 4.** TTTD chunking

The improved sliding window chunking algorithm can detect more redundancy, having a high efficiency of processing data insertion and deletion.

2.3.3   TTTD Chunking

TTTD chunking is short for two thresholds two divisors algorithm. This method has four parameters: D, the main divisor, D', the backup divisor, T_{min}, the minimum chunk size threshold, and T_{max}, the maximum chunk size threshold. D' is smaller than D, having a higher chance of finding a fingerprint match than D [19]. First, move on the window until reaching T_{min} and calculate the fingerprint of current window data. Here

are four possibilities in this step: if fingerprint satisfies D, set chunking boundary here. If fingerprint satisfies D', remember the position as backup boundary, move on the window till a D-match is found before reaching to T_{max}; if the position satisfied D is not found during moving towards T_{max}, set chunking boundary at the backup position. If the window has reached to T_{max} without any backup or main boundary being found, set chunking boundary at the position of T_{max}.

TTTD chunking shown in Fig.4 limits the chunk size between T_{min} and T_{max}. So this algorithm is proposed to solve the problem existing in the basic sliding window chunking method of too large or too small chunk size.

3   Experiment

This section compares the performance of five chunking algorithms by experiments on real data set. Literature [20] utilizes different correlation data sets to assess the effect of data de-duplication which is shown in Fig.5. For different data sets, content-defined chunking algorithm can detect 60% of redundancy in the data set with high correlated, as whole-file chunking and fixed-sized chunking can get 20% to 30% of the rate. But in the data set with high diversity, their redundancy eliminated rate can only reach 17% to 30%.
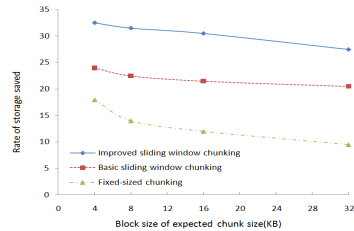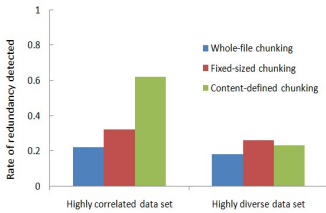

Fig. 5. Rate of redundancy detected on different data set

Fig. 6. Rate of storage saved on the same data set

In addition, according to the experiments by IBM [21], a comparison between chunk-based chunking methods is shown in Fig.6. Improved sliding window chunking performs best in redundancy detecting and storage saving, followed by basic sliding window chunking and fixed-sized chunking, but the best one is the worst one in the aspect of extra storage cost because of the metadata. As for the last method TTTD, it does almost as well on real data as it does on random data and validates the assumption that using backup breakpoints in combination with minimum and maximum thresholds can effectively remove the weak points of the other algorithms [19].

Via the statements and comparison above, we can summarize the features of these chunking algorithms. Whole-file chunking is simple and fast, using the hash technology to implement data mining [11]. But it does not perform as well as chunk-based chunking algorithms on large-scale data set because it can only detect exact file duplicate [9].

Fixed-sized chunking is often used in archival storage system. Actually it has reduced 30% of storage space [11] in Venti system. In addition, this chunking method enforces a write-once policy, preventing accidental or malicious destruction of data [3]. But because of its fixed chunk size, identical data may not be stored at the same offset after the insertion or deletion, so duplicates cannot be completely identified [15] by using this algorithm.

Content-defined chunking is widely used in LBFS [4], Deep Store [22] and Pastiche [5]. Due to the algorithm strategy, modifications only impact the changed chunks, but chunks of this method always have variable sizes. To avoid this problem, the TTTD chunking presented by HP Laboratories imports two thresholds and two divisors, limiting the chunk size between T_{min} and T_{max}. So it is now considered to be the best method in content-defined chunking.

From introduction above, we can see that in most cases content-defined chunking algorithm performs better than fixed-sized chunking algorithm and chunk-based strategy performs better than whole-file strategy. But there is not a general chunking algorithm that performs the same well on all the data sets. So, in practical application, features of data set, advantages and disadvantages of each chunking methods all need to be taken into account when we choose a chunking algorithm to attain the optimal de-duplication effect.

4   Summaries

In this paper, we have introduced five representative chunking algorithms of data de-duplication, compared their performance of redundancy detected. By using these algorithms we can effectively reduce the storage space and network bandwidth [2, 20] because duplicates are identified by the hash value comparison [23] and the same data is stored and transmitted only once.

However, data de-duplication also brings about new problems. First, there is not a general chunking algorithm both having high rate of redundancy detected and little chunking time cost. So we still need to research on the data features and present a better chunking algorithm. Second, as identical data is stored only once, different files depend on each other by sharing the same data. If a fraction of data is lost, all the files with this data fragment will be affected. So how to improve the system reliability and the rate of redundancy detected at the same time should be taken into consideration in the future works.

References

[1] Meister, D., Brinkmann, A.: Multi-level Comparison of Data Deduplication in a Backup Scenario. In: Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference, pp. 8–19. ACM (2009)

[2] Zhu, B., Li, K., Patterson, H.: Avoiding the Disk Bottleneck in the Data Domain Deduplication File System. In: Proceedings of the 6th USENIX Conference on File and Storage Technologies, pp. 269–282. USENIX Association (2008)

[3] Quinlan, S., Dorward, S.: Venti: a New Approach to Archival Storage. In: Proceedings of the 1st USENIX Conference on File and Storage Technologies, pp. 89–101. USENIX Association (2002)

[4] Muthitacharoen, A., Chen, B., Mazieres, D.: A Low-bandwidth Network File System. In: Proceedings of the 18th ACM Symposium on Operating Systems Principles, pp. 174–187. ACM (2001)

[5] Cox, L.P., Murray, C.D., Noble, B.D.P.: Making Backup Cheap and Easy. In: Proceedings of the 5th Symposium on Operating Systems Design and Implementation (OSDI 2002), pp. 285–298. USENIX Association (2002)

[6] Douceur, J.R., Adya, A., Bolosky, W.J., Simon, D., Theimer, M.: Reclaiming Space from Duplicate Files in a Serverless Distributed File System. In: Proceedings of the 22nd International Conference on Distributed Computing Systems, pp. 617–624. IEEE Computer Society (2002)

[7] Douglis, F., Iyengar, A.: Application-specific Delta Encoding via Resemblance Detection. In: Proceedings of the 2003 USENIX Annual Technical Conference (USENIX 2003), pp. 113–126. USENIX Association (2003)

[8] Kulkarni, P., Douglis, F., Lavoie, J., Tracey, J.M.: Redundancy Elimination within Large Collections of Files. In: Proceedings of the 2004 USENIX Annual Technical Conference (USENIX 2004), pp. 59–72. USENIX Association (2004)

[9] Wang, C., Qin, Z.-G., Peng, J., Wang, J.: A Novel Encryption Scheme for Data Deduplication System. In: Proceedings of International Conference on Communications, Circuits and Systems (ICCCAS 2010), pp. 265–269. IEEE Computer Society (2010)

[10] Bobbarjung, D.R., Jagannathan, S., Dubnicki, C.: Improving Duplicate Elimination in Storage Systems. ACM Transactions on Storage (TOS), 424–448 (2006)

[11] Ao, L., Shu, J.-W., Li, M.-Q.: Data Deduplication Techniques. Journal of Software, 916–929 (2010)

[12] Jin, K., Miller, E.L.: The Effectiveness of Deduplication on Virtual Machine Disk Images. In: Proceedings of SYSTOR 2009: The Israeli Experimental Systems Conference, p. 7. ACM (2009)

[13] Lillibridge, M., Eshghi, K., Bhagwat, D., Deolalikar, V., Trezise, G., Camble, P.: Sparse Indexing: Large Scale, Inline Deduplication Using Sampling and Locality. In: Proceedings of the 7th USENIX Conference on File and Storage Technologies, pp. 111–123. USENIX Association (2009)

[14] Won, Y., Kim, R., Ban, J., Hur, J., Oh, S., Lee, J.: Prun Eliminating Information Redundancy for Large Scale Data Backup System. In: Proceedings of International Conference on Computational Sciences and Its Applications (ICCSA 2008), pp. 139–144. IEEE Computer Society (2008)

[15] Hsu, W.W.S., Ong, S.: System and Method for Dividing Data into Predominantly Fixed-sized Chunks So That Duplicate Data Chunks May Be Identified. US Patent, US7281006B2 (2007)

[16] Rabin, M.O.: Fingerprinting by Random Polynomials. Center for Research in Computing Technology, Harvard University (1981)

[17] Bhagwat, D., Pollack, K., Long, D.D.E., Schwarz, T., Miller, E.L., Paris, J.-F.: Providing High Reliability in a Minimum Redundancy Archival Storage System. In: Proceedings of the 14th International Symposium on Modeling, Analysis, and Simulation of Computer and Telecommunication Systems (MASCOTS 2006), pp. 413–421. IEEE Computer Society (2006)

[18] Kruus, E., Ungureanu, C., Dubnicki, C.: Bimodal Content Defined Chunking for Backup Streams. In: Proceedings of the 8th USENIX Conference on File and Storage Technologies, pp. 18–31. USENIX Association (2010)

[19] Eshghi, K., Tang, H.K.: A Framework for Analyzing and Improving Content-based Chunking Algorithms. Hewlett-Packard Labs (2005)

[20] Policroniades, C., Pratt, I.: Alternatives for Detecting Redundancy in Storage Systems Data. In: Proceedings of the 2004 USENIX Annual Technical Conference (USENIX 2004), pp. 73–86. USENIX Association (2004)

[21] Denehy, T.E., Hsu, W.W.: Duplicate Management for Reference Data. IBM (2003)

[22] You, L.L., Pollack, K.T., Long, D.D.E.: Deep Store: an Archival Storage System Architecture. In: Proceedings of the 21st International Conference on Data Engineering (ICDE 2005), pp. 804–815. IEEE Computer Society (2005)

[23] Yang, T.-M., Jiang, H., Feng, D., Niu, Z.-Y.: Debar: a Scalable High-performance De-duplication Storage System for Backup and Archiving. In: Proceedings of International Parallel & Distributed Processing Symposium (IPDPS 2010), pp. 1–12. IEEE Computer Society (2010)

A Multi-Objective Genetic Algorithm Based
on the Uniform Design Method and Logistic Mapping

Ma Xiao-Shu[1], Liu Qing[1], and Ma Ning[2]

[1] School of Physics & Information Science, Tianshui Normal University,
Tianshui 741001, China
[2] China State Construction Engineering Corporation, Bejing 100000, China

Abstract. Many optimization problems in Scientific research and engineering practice can be modeled as multi-objective optimization problems. Thus, designing effective algorithms for them is of not only the great importance in scientific research, but also the great value in applications. In this paper, by incorporating uniform design and logistic mapping, a new genetic algorithm for multi-objective optimization problems called UCMOGA is proposed. The initial population is firstly produced by uniform design principles, and then, a novel crossover operator that can be dynamically adjusted crossover scale to enhance the searching ability is presented, in which the traversal of the location of the intersection is determined according to the pseudo-random and chaotic sequences. The new scheme makes up the defects of weak search capabilities of simulated binary crossover operator and the weak search capabilities of simulated binary crossover operator. The computer simulations for five difficult standard benchmark functions also verify this fact.

Keywords: Genetic algorithms, Multiobjective optimization, Uniform design, Chaos.

1   Introduction

Multi-objective optimization problem is not unique, it is a set of compromise solutions and called non-dominated Pareto optimal solution set. Current research on MOGA entered a rapid development stage, a lot of good multi-objective genetic algorithms have been proposed. There are VEGA, NSGA-II [1] and so on. Chaos is a common phenomenon in nonlinear systems. It is the application of genetic algorithm, including create chaos optimization genetic operators, the continuous generation of the chaotic search to better guide the population evolution of individuals. The initial population is firstly produced by uniform design principles, and then, a novel crossover operator that can be dynamically adjusted crossover scale to enhance the searching ability is presented, in which the traversal of the location of the intersection is determined according to the pseudo-random and chaotic sequences.

2   The Method of Initial Population Based on Uniform Design

Uniform design is a mathematical test method, Kaitai Fang and Yuan Wang [5] were first proposed in 1978, its essence is to give representatives of selected points of

uniform method of walking in the experimental range. It is very effective in solving multi-factor, multi-level test issues. During the test, factor has n, The level of each factor has q. When given after n and q, then all possible combinations have q^n. Uniform design is selects q combination from q^n possibility combination, which make q combinations uniformly ditributed in space of all possible combinations.

With a matrix $U(n, q) = [U_{i,j}]_{q \times n}$, that the selected q combinations are evenly distributed, Where $U_{i,j}$ is the level of the i factors in the j combinations, σ parameter can be retrieved. When q is a prime number and $q > n$, it can prove

$$U_{i,j} = (i \cdot \sigma^{j-1} \bmod q) + 1 \tag{1}$$

Pareto solution for the position in the solution space without any known information, we have to solve multi-objective optimization problem. If we can produce in the solution space uniformly distributed initial population, then the genetic operators to help in the solution space exploration. Therefore, we can quantify the space of the possible solutions to many points, and then, construct the initial population by making use of the idea of uniform design.

When the solution space is large, we hope to produce more uniform distribution of points in the solution space, but maximum level of 31 factors in uniform design, and often larger than the size of population. To overcome this difficulty, the solution space is divided into multiple subspace, solutions have some representative in these subspace by uniform design. Therefore, the search space $X = \{x = (x_1, x_2, ..., x_n)$

$\tilde{a}_i \leq x_i \leq \tilde{b}_i, i = 1, 2, ..., n\}$ is divided into S small

n-dimensional rectangular domain:

$$X_1 = \{x = (x_1, x_2, ..., x_n) \big| \tilde{a}_i^1 \leq x_i \leq \tilde{b}_i^1, i = 1, 2, ..., n\}$$

$$X_2 = \{x = (x_1, x_2, ..., x_n) \big| \tilde{a}_i^2 \leq x_i \leq \tilde{b}_i^2, i = 1, 2, ..., n\}$$

$$\dots\dots\dots\dots\dots\dots$$

$$X_s = \{x = (x_1, x_2, ..., x_n) \big| \tilde{a}_i^s \leq x_i \leq \tilde{b}_i^s, i = 1, 2, ..., n\},$$

that $X = X_1 \bigcup X_2 \bigcup ... \bigcup X_s$. Parameter s can be set to $2, 2^2, 2^3$ etc. First, this variable along the direction of the solution space is divided into two equal subspace based on the maximum range of variables, then the two subspace is divided into four sub-space in the same way, repeat the above steps $\log_2^s$ times, the solution space can be divided into S subspace. For any n-dimensional rectangle region a small subspace X_j. Let $A^j(\tilde{a}_1^j, \tilde{a}_2^j, ..., \tilde{a}_n^j)$ and $B^j(\tilde{b}_1^j, \tilde{b}_2^j, ..., \tilde{b}_n^j)$ vertices of this rectangle, using the uniform design, several points can be obtained, with these points constitute the initial population.

3   The Definition of Fitness Function

$x_1^t, x_2^t, \cdots, x_i^t, \cdots, x_N^t$ is a feasible solution to the individual, it is the population in in the t th generation, N is the population size, x_i^t is the i th individual in the t th generation, p_i^t is the number of individuals, which is better than pareto solutions of the population than the individual x_i^t $(s_i^t = p_i^t + 1)$ is sequence & value of the individual x_i^t in the t th generation, individual x_i^t fitness function for the $F_i^t = \dfrac{1}{s_i^t}$, if the pareto solution better than the individual x_i^t more in the t th generation and x_i^t sequence & value more greater, then lower fitness, that is the worse the quality of the individual x_i^t; Conversely, the higher the fitness, the better the quality of the individual x_i^t [7].

4   The New Crossover Operator Design

Simulated binary crossover operator is relatively weak performance of search used by NSGA-II, and there is some lack of diversity in maintaining population. Therefore, the combination of genetic algorithm for multi point crossover and arithmetic crossover ideas, determine the intersection point with the position of chaotic sequence, then designed a new crossover operator.

According to Logistic mapping:

$$k_{i+1} = \mu \times k_i \times (1 - k_i) \tag{2}$$

among them, the initial value $k_0 \in (0,1)$ and $k_0 \notin \{0, 0.25, 0.5, 0.75, 1\}$,so $k_{i+1} \in (0,1)$ .when $3.571448 \cdots \le \mu \le 4$, Logistic map showing the expected chaotic characteristics, which produces chaotic sequence $k_i \in (0,1)$, i is a positive integer.

Step1 The number of chaotic sequence according to formula (3) into the array A after the operation

$$A(j) = [n \times k_i] \quad i = (t-1) \times d + 1, \cdots, t \times d , \quad j = 1, 2, \cdots, d \tag{3}$$

n is the length of individual coding, t is evolution algebra, $[\alpha]$ rounded expressed α . $k_i \in (0,1)$, k_i and length of individual coding after multiplying the rounded, some values may be equal in the A array, we pick the values that are unequal pairwise from array A and store them in array $site$,to $site(k)$ ($k = 1, 2, \cdots, |site|$) as the cross-bit, the following formula is used in these cross-bit computing:

$$\overline{x_{i,site(k)}^{(t+1)}} = \alpha_k x_{i,site(k)}^{(t)} + (1 - \alpha_k) x_{j,site(k)}^{(t)} \tag{4}$$

$$\overline{x_{j,site(k)}^{(t+1)}} = \alpha_k x_{j,site(k)}^{(t)} + (1 - \alpha_k) x_{i,site(k)}^{(t)} \tag{5}$$

$x_{i,j}^{(t)}$ is the j th bit of the i th individual in the t th generation, $\alpha_k \in (0,1)$ ($k = 1, 2, \cdots, |site|$) is random number, bit in the last crossover, the principle of single point crossover. Because the size of the array $site$ different, we construct a novel crossover operator that can be dynamically adjusted crossover scale.

For example, we have identified the location of cross-participation were the first and the fourth from the array $site$, random numbers are $\alpha_1 = 0.3$, $\alpha_2 = 0.5$:

Parent individual 1:0.14　　$\boxed{0.46}$　　0.25　　$\boxed{0.54}$　　0.56　0.32　0.83

Parent individual 2:0.21　　$\boxed{0.62}$　　1.03　　$\boxed{0.98}$　　0.68　　0.65　　0.76

　　　　　　　　　　crossing point　　crossing point

Offspring individual 1:0.14　$\boxed{0.58}$　　0.25　　$\boxed{0.76}$　　0.68　0.65　0.76
Offspring individual 2:0.21　$\boxed{0.52}$　　1.03　　$\boxed{0.76}$　　0.56　0.32　0.83

　　　　　　　　　　crossing point　　crossing point

Step2 After crossover operation, empty A and $site$;

5　Mutation Operator

Let the search space for $[L, U] = \{(x_1, x_2, \cdots, x_n)^T \mid l_i \leq x_i \leq u_i, 1 \leq i \leq n\}$, the i-dimensional subspace is divided into several sub-space $[l_i, u_i]$ of the individual variation in x, they are $[w_1^i, w_2^i]$, $[w_2^i, w_3^i]$, $\cdots$, $[w_{n_i-1}^i, w_{n_i}^i]$, makes $|w_j^i - w_{j-1}^i| < \varepsilon$, $w_1^i = l_i$, $w_{n_i}^i = u_i$. Mutation probability set to p_m, generate a random number $\delta_i \in [0,1]$. If $\delta_i < p_m$ ($i = 1, 2, \cdots, n$), then randomly selected a $w_{j_i}^i$ instead of x_i in the $\{w_1^i, w_2^i, \cdots, w_{n_i}^i\}$; Otherwise, x_i remains unchanged.

6　A New Genetic Algorithm

The algorithm is designed to search for Pareto front from close and more uniform distribution of non-dominated solution, and then, a novel crossover operator that can be dynamically adjusted crossover scale to enhance the searching ability is presented, in which the traversal of the location of the intersection is determined according to the pseudo-random and chaotic sequences. Algorithm idea is as follows:

Step1(Initialization)Given hybrid probabilit p^c mutation probability p^m and population size N, maximum evolution generation T, uniform size of hybrid offspring Q, randomly generated initial population size $p(0)$, $t = 0$;

Step2 According to crossover probability p^c, randomly selected parent individuals from the t th generation $p(t)$, these parent individuals were hybridized according to crossover by 3. Each of the individual parent, hybrid offspring resulting from a pair of individuals, The parent individuals did not participate in hybrid offspring as their own. All the offspring, denoted $q'(t)$.

Step3 (Variation) To step2 cross offspring, according to the variation defined in 4 ways to participate variation, so future generations into a collection of $q''(t)$.

Step4 (Select) Sort of each individual in $p(t) \cup q'(t) \cup q''(t)$, and accroding to fitness function defined by 1, according to descending order, the selected N individuals to form the next population $p(t+1)$, order $t = t + 1$.

Step5 when $t > T$, stop, output of individual from $p(t)$; Otherwise, the switch to the step2.

7 Numerical Experiment

7.1 Test Function

Test function is used in NSGA-II, they are ZDT1,ZDT2,ZDT3,ZDT6.

ZDT1:
$$\begin{cases} f_1(x) = x_1 \\ f_2(x) = g(x)[1 - \sqrt{(x_1 / g(x))}], \quad n = 30, 0 \le x_i \le 1, i = 1, 2, \cdots, 30; \\ g(x) = 1 + 9(\sum_{i=2}^{n} x_i) / (n-1) \end{cases}$$

ZDT2:
$$\begin{cases} f_1(x) = x_1 \\ f_2(x) = g(x)[1 - (x_1 / g(x))^2], n = 30, 0 \le x_i \le 1, i = 1, 2, \cdots, 30; \\ g(x) = 1 + 9(\sum_{i=2}^{n} x_i) / (n-1) \end{cases}$$

ZDT3:
$$\begin{cases} f_1(x) = x_1 \\ f_2(x) = g(x)[1 - \sqrt{(x_1 / g(x))} - \dfrac{x_1}{g(x)} \sin(10\pi x_1)], \\ g(x) = 1 + 9(\sum_{i=2}^{n} x_i) / (n-1) \end{cases}$$

$n = 30, 0 \le x_i \le 1, i = 1, 2, \cdots, 30;$

ZDT6:
$$\begin{cases} f_1(x) = 1 - \exp(-4x_1) \sin^6(6\pi x_1) \\ f_2(x) = g(x)[1 - (f_1(x) / g(x))^2], \quad n = 30, 0 \le x_i \le 1, i = 1, 2, \cdots, 30; \\ g(x) = 1 + 9[(\sum_{i=2}^{n} x_i) / (n-1)]^{0.25} \end{cases}$$

7.2   Simulation Results Analysis

In order to verify the validity of algorithm UCMOGA, we tested the function and compared with the algorithm NSGA-II, parameters are as follows: Let population size 100, the crossover probability of 0.9; mutation probability 0.07; ZDT1, ZDT2, ZDT3 maximum evolution generation is 300; ZDT6 maximum evolution generation is 800; $d = 5$ Run independently for each function 20 times, records of which run from a typical non-dominated solution set, and draw into the non-dominated objective space frontier. Figure 1~Figure 4 which is the algorithm NSGA-II and algorithms UCMOGA function of the different non-dominated front obtained.

Comparing the results from the above chart we can see, the algorithm UCMOGA search to access non-domination is relatively evenly distributed, indicating that the algorithm has good search capabilities, basic to achieve the desired goals, and obtain a relatively uniform distribution of non-dominated solution set.
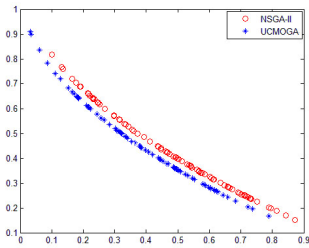
Fig. 1. ZDT1 function of the results after running
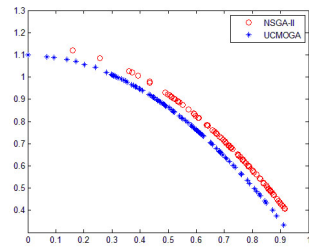
Fig. 2. ZDT2 function of the results after running
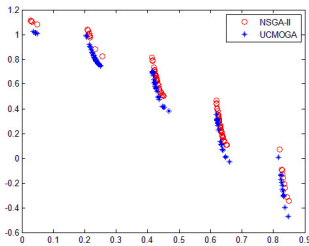
Fig. 3. ZDT3 function of the results after running
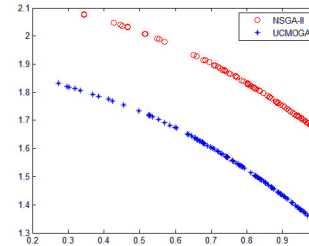
Fig. 4. ZDT6 function of the results after running

8   Summaries

Solving multi objective optimization problem using the algorithm designed in this paper, experimental results show that the proposed algorithm is effective for obtain better quality, more uniform distribution of non-dominated solutions. But how to find a more rational multi-objective optimization algorithm, that will be the next step needs to be done.

References

[1] Deb, K., Agrawal, S., Pratap, A., Meyarivan, T.: A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. In: Proc. Conf. Parallel Problem Solving from Nature VI, pp. 849–858 (2000)

[2] Wei, J.-X.: Evolutionary Algorithms for Single-Objective and Multi-Objective Optimization Problems. China Doctor Dissertation Full-text Database (08) (2009)

[3] Determan, J., Foster, J.A.: Using chaos in genetic algorithms. In: Proceedings of the 1999 Congress on Evolutionary Computation, pp. 2101–2104. IEEE Press, Piscataway (1999)

[4] Wang, M.-L.: A Study of Dynamic Optimization Problems Based on Chaotic Evolutionary Algorithm. Journal of WuYi University (Natural Science Edition) 24(1), 44–47 (2010)

[5] Lei, D.-M., Yan, X.-P., Wu, Z.-M.: Multi-Objective Chaotic Evolutionary Algorithm. Chinese Journal of Electronics 34(6), 1142–1145 (2006)

[6] Gong, M.-G., Jiao, L.-C., Yang, D.-D., Ma, W.-P.: Research on Evolutionary Multi-Objective Optimization Algorithms. Journal of Software 2(20), 271–289 (2009)

[7] Schott, J.R.: Fault tolerant design using single and multicriteria genetic algorithm optimization. Massachusetts Institute of Technology, Cambridge (1995)

An Improved Algorithm of the Shortest Path Search Problem in GIS Field

Zhao Na

Taiyuan Normal University, Taiyuan 030012, China

Abstract. Using Dijkstra algorithm for searching shortest path problem is an important content of the application of GIS. Dijkstra algorithm can not find all the shortest paths. The article based on the Dijkstra algorithm, added some data structure and proposed an algorithm that calculate all the shortest path of one vertex to others, the algorithm data structure is relatively simple, while effectiveness of the algorithm is verified by an example.

Keywords: GIS, Dijkstra algorithm, shortest path.

1   Introduction

GIS (Geographic Information System) is a collection of modern computer science, geography, information science, management science and mapping science as one of a new discipline. It uses a database, computer graphics, multimedia, and other new technology for geographic information for data processing, in real time and accurately capture, edit and update geospatial data and attribute information for decision makers to provide visual support[1]. Shortest path search problem is an important part of the study in GIS applications. Relative to the simple shortest path, GIS in the shortest path problem is more complicated. After years of research, there are many kinds of algorithms to solve the problem have been proposed, such as Genetic Algorithm, ACO(Ant Colony Optimization), SA(Simulated Annealing), and the Dijkstra algorithm and so on. Among them, majority use Dijkstra algorithm Improve and optimize the shortest path[2]. This paper introduced Dijkstra algorithm ,analyzed the shortcomings of the algorithm, and on this basis propose a new optimization method.

2   Dijkstra Algorithm

2.1   Dijkstra Algorithm Theory

Dijkstra algorithm was proposed by Dick Stella, a computer scientist from the Netherlands, in 1959. From one vertex to the rest of the vertices of the shortest path algorithm, the solution is to map the shortest path problem.

For a directed graph or undirected graph G, each edge add a real number ω (e), called ω (e) right on the edge e, G, together with an additional edge in the real number is called a weighted graph. Usually expressed as a weighted graph: G = (V, E, W),

where, V is the set of vertices, E is the set of edges, W is the set of weight of corresponding edge. That is the meaning of weight: Such as highway mileage between cities, or the costs of building roads, or travel needed to fuel consumption. Obviously, in a weighted graph G = (V, E, W), in the weight of each edge ei, W (ei) must be greater than zero. All channels from vertex u to v, the smallest sum of the weight path is the shortest path from u to v, find a given shortest path between two points is called the shortest path problem[3].

2.2   Mathematical Description of the Dijkstra Algorithm[4]

Let G = (V, E, W) is a non-negative weight network, V = (v_1 , v_2 ,..., v_n). The length of the shortest path in G(v_i, v_j)∈E satisfy the equation:

$$u_1 = 0$$
$$u_j = \min (u_k + w_{kj}) (j = 2 , 3 ,..., n) \qquad (1)$$

If G,   from the vertex v1 to the rest of the vertices, the shortest path length Sorting by size:

$$u_{i1} \leq u_{i2} \leq ... \leq u_{in}$$

Here, $i_1 = 1$, $u_{i1} = 0$,   Then from (1) have:

$$u_{ij} = \min_{k \neq j} \left\{ u_{ik} + w_{ikij} \right\}$$
$$= \min \left\{ \min_{k<j} \left\{ u_{ik} + w_{ikij} \right\} , \min_{k>j} \left\{ u_{ik} + w_{ikij} \right\} \right\}$$

$$(j= 2 ,3 , \dots ,n)$$

When $k > j$ 时1, $u_{ik} \geq u_{ij}$,   and $w_{ikij} \geq 0$,   thus
$$u_{ij} \leq u_{ik} + w_{ikij}$$

that is $$u_{ij} \leq \min_{k>j} \{ u_{ik} + w_{ikij} \}$$

therefore $$u_{ij} = \min_{k<j} \{ u_{ik} + w_{ikij} \}$$

Easy to prove: $$u_{i1} = 0$$
$$u_{ij} = \min_{k<j} \{ u_{ik} + w_{ikij} \} \qquad (j= 1 ,2 ,3 , \dots ,n)$$

u_{ij} ,one of the solution (u_{i1}, u_{i2}, ..., u_{in}) is the shortest path length of G (u_i, u_j) .

2.3   Calculation Steps and the Problems of Dijkstra Algorithm

Weighted graph can be expressed as adjacency matrix cost[] [],which states: if between vi and vj have no direct path, the cost [i] [j] = ∞;if Between vi and vj have direct path, the cost [i] [j] = wij;if i = j, then the cost [i] [j] = 0.The set of S storage initial source of shortest path, during the process of calculating , the vertex have determined the shortest path added to the S. Final dist[i] storage the shortest path length ,the source point to the vertex, specific steps are as follows：

(1) Initialize S and dist. S= {v0}, dist[i]=cost[0][i], i=0, 1, …, n- 1.

(2) Select v_j, so dist(j)=min{dist (i) | v_i ∈(V- S) } ; S= S∪vj.

(3) Modify the length of the shortest path from v_0 to v_k∈(V- S) .If dist[j] + cost [j] [k] < disk[k], so dist (k)= dist[j] + cost [j] [k].

(4) Repeat steps (2), (3), until obtain the shortest path that from the source point v_0 to the rest of vertices v_j .

However, Dijkstra algorithm can only find one shortest path from v_0 to v_j, and can not calculate all the shortest paths.In this paper, based on the Dijkstra algorithm is slightly improved, more intuitive way to solve a vertex to other vertices of all shortest paths[5].

3   Improved Algorithm and Examples

(1) The introduction of Dijkstra algorithm, that is, step (1) to step (4) in the 2.3;

(2) According to the cost and the dist, creating a correction matrix corr. Method: If 0< cost [i] [j] <∞, then corr[i] [j] =cost[i] [j] + dist [i] ; otherwise, corr[i] [j] = cost[i] [j] ;

(3) Create a successor node set succ (v_i). Method: By succ (i) = {J | corr [i] [j] = dist [j] and i ≠ j} find the set that the successor node for each vertex ;

(4) According to the corr and succ output the shortest path from the source to all other vertices.

The algorithm can find from one vertex to all other vertices the shortest path. The following detailed description of the algorithm implementation through the example of digraph and   undirected graph.

Example 1 Fig.1 shows a directed graph G1, find the vertex v_0 to all other vertices of the shortest path. Specific steps are as follows:
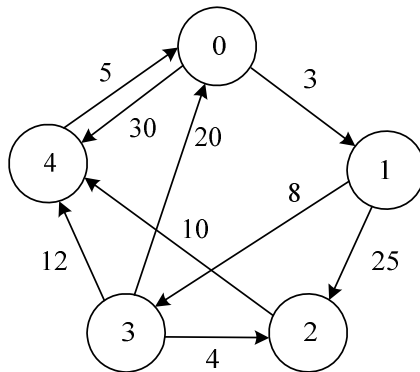

Fig. 1. Weighted directed graph G1

(1) Adjacency matrix is given:

$$\cos t = \begin{bmatrix} 0 & 3 & \infty & \infty & 30 \\ \infty & 0 & 25 & 8 & \infty \\ \infty & \infty & 0 & \infty & 10 \\ 20 & \infty & 4 & 0 & 12 \\ 5 & \infty & \infty & \infty & 0 \end{bmatrix}$$

Initialize S and dist [i] $(0 \le i \le 4)$: S= $\{v_0\}$, dist [i]= $\{0, 3, \infty, \infty, 30\}$;

(2) based on dist [i] = min{ dist [i] |vi∈ V- S}, to find dist[1] = 3; S={ v_0, v_1} ;

(3) modify the shortest path length, starting from v_0 to any node on the V-S v_k (then $2 \le k \le 4$) ,dist[i] = $\{0, 3, 28, 11, 30\}$;

(4) dist[3] = min{ dist[i] |$v_i \in$ V- S} = 11; S={ v_0, v_1, v_3 } ; dist [i] = $\{0, 3, 15, 11, 23\}$.

(5) dist[2] = min{ dist[i] | $v_i \in$ V- S} = 15; S={ v_0, v_1, v_2, v_3} ; dist[i] = $\{0, 3, 15, 11, 23\}$.

Although at this point there is a vertex not incorporated in set S, but it is the shortest distance have been determined, so the whole operation is over.

(6) created corr matrix through the cost and dist :

$$corr = \begin{bmatrix} 0 & 3 & \infty & \infty & 30 \\ \infty & 0 & 28 & 11 & \infty \\ \infty & \infty & 0 & \infty & 25 \\ 31 & \infty & 15 & 0 & 23 \\ 28 & \infty & \infty & \infty & 0 \end{bmatrix}$$

According to corr, dist and succ(i) = { j| corr[i] [j] = dist[j] and i≠j}, find the collection from the successor node for each vertex,there: succ(v0) = { v1 } , succ(v1) = { v3 }, succ(v2)= {null}, succ(v3) = { v2 , v4 }, succ(v4) = {null}. Then the vertices v0 to all other shortest path is v0 v1 ; v0 v1 v3 v2 ; v0 v1 v3 ; v0 v1 v3 v_4.

4   Summaries

In this paper, based on the Dijkstra algorithm, increased the number of data structures. Proposed an algorithm to calculate directly from one vertex to other vertices of all shortest paths. Data structure of the algorithm is relatively simple, while the corresponding C program through the design to verify the validity of the algorithm.

References

[1] Wang, K.-Y., Zhao, C.-J., Xu, G.-X., et al.: A High-efficiency Realization Way of the Shortest Path Search Problem in GIS Field. Journal of Image and Graphics 8(8), 951–956 (2003)

[2] Cheng, S., Li, Y., Chen, X.-G.: Research and Application of Dijkstra Algorithm in GIS. Journal of Anyang Institute of Technology 9(2), 43–45 (2010)
[3] Tang, J.-W.: Improving Dijkstra Algorithm about the Shortest Path in GIS Field. Journal of Weinan Teachers College 21(2), 51–54 (2006)
[4] Zhang, F.-H., Liu, J.-P., Li, Q.-Y.: A New Way of Network Analysis Based on Dijkstra. Technology of GIS 2, 38–41 (2004)
[5] Liu, P., Bai, C.-M.: An improved algorithm of shortest paths based on dijkstra algorithm. Journal of Qinghai Normal University (Natural Science) 1, 79–81 (2008)

Research and Implementation on Detection Algorithm of Objects Left Based on Video Analysis

Caiyan Yu[1,2] and Xiaoshi Zheng[1]

[1] Shandong Computer Science Center, Shandong Provincial Key
Laboratory of Computer Network
[2] Shandong Polytechnic University, Jinan, China
`yucy@keylab.net`

Abstract. A new detection algorithm of objects left based on background subtraction and temporal difference is proposed in this paper. Experiments prove that this aalgorithm has higher stability, and it can detect unknown items from monitoring areas quickly and accurately.

Keywords: Background subtractions, Temporal difference, Detection algorithm of objects left.

1   Introduction

Computer network and image processing technology have achieved rapid improvement with the development of modern society technology. Because people's security consciousness strengthen ceaselessly, intelligent video surveillance is widely used in public areas, such as police, transportation, finance, military, electric power, community buildings and so on. Intelligent video surveillance can detect unknown items and activities effectively by anglicizing video image sequences.

As the key technology of intelligent video surveillance, detection of moving targets is the basic of classification, tracking and behavior understanding and other post processing. How to extract the moving object quickly and accurately is the main content of researchers [1]. At present, there are mainly three several main common algorithms in the static background which are background subtraction, temporal differences and optical flow [2]. However, every algorithm has its own limitation. In view of this situation, an improvement detection algorithm of objects left based on background subtraction and temporal difference is proposed in this paper. This method can detect unknown items intelligently and alarm timely in public or other sensitive areas.

2   Image Conversion Technology

Usually speaking, moving images collected from the environment are digital images which are composed by red, green and blue three colors, every pixel is made up of the

three colors according to different proportion. However, there are various kinds of color images that occupied large storage space. Because they can't be handled easily, we should convert them to gray images [3]. There are only 256 level gray values in gray images which reduce the computational complexity and storage space, and improve the quickness of image processing.

The following well-known linear transform is applied to the RGB components of the decoded frames [4]:

$$Y = k_r R + k_g G + k_b B_\cdot$$
(1)

Where the known variables include: Y indicates luminance value of image, k is weighting factor, R is red color, G is green color, B is blue color.

In this paper, we define k as constant variable.

$$k_r = 0.299 \, , \quad k_g = 0.587 \, , \quad k_b = 0.114$$

So the linear transform is given as the following rules:

$$Y = 0.299R + 0.587G + 0.144B_\cdot$$
(2)

3   Detection Algorithm of Objects Left analysis

3.1   Background Subtraction Algorithms

In this method, firstly, we subtract current each frame image from background image. Then we can judge changes of gray values and other characteristics. If the absolute value of deviation is beyond a certain threshold we defined, moving area can be confirmed [5].

Calculation formula [6] can be defined as follows:

$$d = \left| I_t(x, y) - B_L(x, y) \right|$$
(3)

$$M_L(x, y) = \begin{cases} d, if & d \geq threshold \\ 0, if & d < threshold_\cdot \end{cases}$$
(4)

Where $B_L(x,y)$ is the luminance value of background image. $M_L(x,y)$ is binary image from subtraction current image $I_i(x,y)$ and background image $B_L(x,y)$, $threshold$ is the change detection threshold.

The principle of this method is simple and easily understood. The design of this algorithm is also simple relatively. It can describe the size, shape or site of moving target completely. Basically, moving targets can be defected accurately.

The deficiency of this method is that it can be affected by light, weather and other external conditions.

3.2   Temporal Difference Algorithm

This method can extract moving area from two or three consecutive frames through the threshold of absolute value. If the value is larger than threshold we defined, we can judge that there are moving targets in monitoring areas.

The expression is given by

$$d = \left| I_n(x, y) - I_{n-1}(x, y) \right| \tag{5}$$

$$M_n(x, y) = \begin{cases} d, if & d \geq threshold \\ 0, if & d < threshold \end{cases} \tag{6}$$

Where $M_n(x, y)$ is binary image that subtract current image $I_n(x, y)$ from prev-image $I_{n-}(x, y)$, $threshold$ is the change detection threshold.

The advantage of this method is that it is able to adapt to every dynamic environment. Moreover, it has low complexity and high stability.

The deficiency of this method is that it cannot extract complete areas of moving targets, just only the boundary part, so it is not suitable for detecting fast moving.

3.3   A New Algorithm Based on Background Subtraction and Temporal Difference

In this paper we propose a new detection algorithm of objects left based on background subtraction and temporal difference. Firstly, we establish a set included all pixels of background image as follows:

$$D = \left\{ d_{i,j} \mid 0 < i < width \, 0 < j < height \right\} \tag{7}$$

Where i, g indicate horizontal ordinate and vertical ordinate, $width$、$height$ is the width and height of image, $d_{i,j}$ is the RGB value of image, count and countb are two variables.We can describe the algorithm as follows:

Step 1: Initialize video files collected with camera.

Step 2: Save the first frame of video file image as background image called pBkImgb.

Step 3: Convert every pixel of current frame pFrame, prev-frame pBkImg and background frame image pBkImgb to gray image respectively.

Step 4: Subtract every pixel of pFrame from pBkImg with temporal difference algorithm, $d = \left| Y_{i,j} - Y1_{i,j} \right|$, we set $threshold1$ for 50. We can judge that there is no objects left if the threshold is less than 50. In this algorithm, variable count is used to counting.

Step 5: If $count / height * width$ is more than 0.99，turn to step 6 ，the value of N in this paper is 16.

Step 6: Subtract every pixel of pFrame from pBkImgb with temporal difference algorithm, $d = \left| Y_{i,j} - Y1b_{i,j} \right|$, we set $threshold2$ for 90. We can judge that

there is object left if the threshold is more than 50. In this algorithm, variable countb is used to counting.

Step 7: It indicates that there are objects left in the monitoring area if $countb/height*width$ is more than 0.3, then the system will alarm to remind workers to deal with the activity timely. This step can avoid mistaking sill background or objects in the first image for objects left behind.
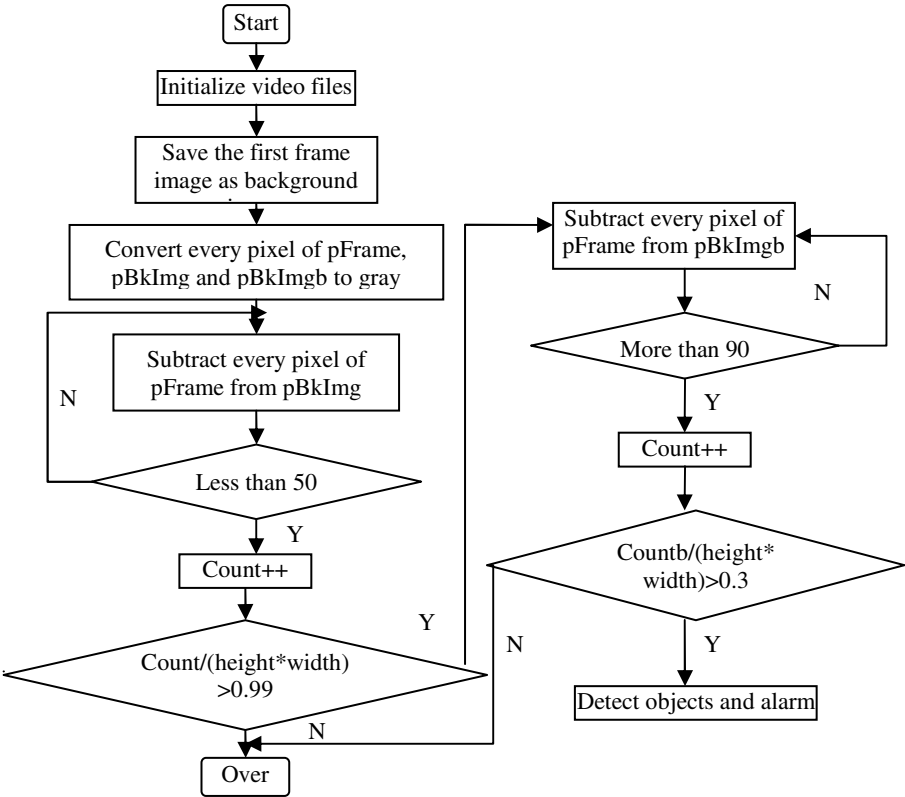
The Fig.1 is the flow chart of this algorithm.

Fig. 1. The flow chart of algorithm

4   Result of Experiments

In order to test accuracy of the algorithm proposed in the paper, we do some experiments on two groups of video sequences, and the experimental results shows as Fig.4. The Fig.2 is the first frame of video sequence, which is the background image. The Fig.3 is the scene that staff leaves. In the Fig.4, bags are detected .The final experiment results show that unknown items can be detected correctly and alarm timely with the algorithm proposed in this paper.

Fig. 2. The background image of video sequence
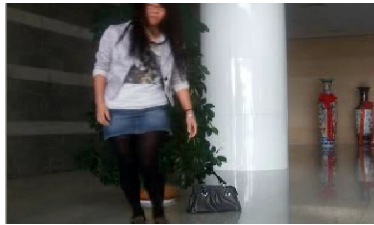
Fig. 3. The scene that staff leaves
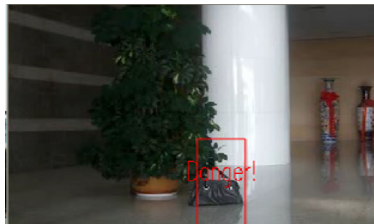
Fig. 4. The image of object detected

5   Summaries

In this paper, an improvement detection algorithm of objects left based on background subtraction and temporal difference is proposed. We make gray-scale process to color images using image conversion technology firstly, and then do some experiment to verify the accuracy of the algorithm. This method can detect objects left quickly and has high stability. Finally, we can get effective moving targets.

References

[1] Hampapur, A., Brown, L., Connell, J., et al.: Smart surveillance: applications, technologies and implication. In: Proceedings of Fourth International Conference all Information, Singapore, vol. 2, pp. 1133–1138 (2003)

[2] Shen, L., Zhuo, L.: Wavelet Coding and Network Video Transmission. Science Press, Beijing (2005)

[3] Kong, X.: Research on Intelligent Video Surveillance Technology. Fudan University, Shanghai (2008)

[4] Sacchi, C., Granelli, F., Regazzoni, C.S., Oberti, F.: A real-time algorithm for error recovery in remote video-based surveillance applications. Image Communication 17, 165–186 (2002)

[5] Kim, K., Chalidabhongse, T., Harwood, D., et al.: Real-time foreground-background segmentation using codebook model. Real Time Imaging 11(3), 172 (2005)

[6] Li, W., Zhou, X., Wang, C.: Detection algorithm of moving objects based on background subtraction method. Journal of University of Science and Technology 30(2), 212–216 (2008)

Solving 0-1 Knapsack Problem Based on Immune Clonal Algorithm and Ant Colony Algorithm

Zhao Fang[1], Ma Yu-Lei[1], and Zhang Jun-Peng[2]

[1] Department of Computer and Information Engineering, Xinxiang University,
Xinxiang 453003, China
[2] Henan Zhumadian Power Supply Company, Zhumadian 463000, China

Abstract. The 0/1 Knapsack Problem is a very well known and it appears in many real life world with different application. In this paper, we propose a new hybrid algorithm which inspired from Ant Colony Algorithm (ACA) and Antibody Immune Clonal Algorithm (AICA) to tackle 0-1 knapsack problem. ACA algorithm which benefit from its positive strategy has a great performance in local search, but it is prone to stagnation; AICA algorithm can search in a greater range, but it is lack of sufficient utilization of feedback information of system. So, we take full advantages of the ability of searching and the diversity which they provide respectively. Some parameters of the algorithm and the comparison with other algorithms have been performed. The experimental results show that the proposed algorithm is a perfect hybrid algorithm with higher performance.

Keywords: 0-1 knapsack problem, Ant colony algorithm, Antibody immune clonal algorithm.

1 Introduction

The Knapsack Problem(KP) is a typical NP complete question[1], which is widely applied in solving material cutting, cargo loading and information encryption etc. At present, there are several deterministic algorithms (e.g. Dynamic Programming, Backtracking Algorithm etc.) and approximate algorithms (such as Greedy Algorithm, GA etc.) solved the KP question. Recently, the approximate algorithms solving the KP question become hotter.Liang Ma[3] proposed a method based Ant Colony Algorithm to solve KP question, however, this algorithm introducing ACA is a simple application in solving knapsack problem. Constraint conditions cannot meet and solution quality is difficult to be guaranteed.

2 AICA Algorithm and Basic Principle of ACA Algorithm

2.1 AICA Glorithm

The AICA algorithm is a new Stochastic Optimization method [4]. Antigen, antibody, and the affinity between antigen and antibody correspond to the objective function

and constraints, candidate solution, and the matching degree between solution and objective function. Antibody Clonal Selection theory believes that when antigens invade the body, through Colnal Selection mechanism, the corresponding immune cell (which can identify and eliminate the corresponding antigen) can be selected, then activating, differentiating, proliferating immune response. So, the procedure of immune response has the same mechanism as issue-driven optimization process. But, the process emphasizes adaptive change of antibody which bases on antigenic's characteristics. This algorithm can maintain the diversity of antibody, and prevents the solution from falling into local optimum area.

The immune body clone operator is by the compatible induction immune body stochastic mapping, the immune body group stochastic shift situation may express the following stochastic process:

$$A(k) \xrightarrow{\quad clone \quad} A^{'}(k) \xrightarrow{\quad immune \quad} A^{''}(k) \xrightarrow{\quad select \quad} A(k+1)$$

Where, the antibody population $A(k) = [A_1(k), A_2(k), \cdots, A_m(k)]$ is a m-tuple, which is an element of the space I^m.

Antibody clonal selection operator that splits a element of solution space $A_i(k) = [x_{i1}(k), x_{i2}(k), \cdots, x_{in}(k)]$, $A_i(k) \in A(k)$ (where n is the dimension of problem) into q_i identital points $A_i^{'}(k) \in A^{'}(k)$. New antibody population $A(k+1)$ will be obtained after clonal immunity and clonal selection.

2.2   ACA Glorithm

ACA proposed by M. Dorigo Italian scholar in the early 90s of last century is a new stochastic and optimized algorithm. This algorithm inspired by the observation of real ant colonies searches for optimal solution though the accumulation and updating of pheromone. The results show that ants deposit on the ground a substance called pheromone while walking from food sources to the nest and nice versa. Ants can smell pheromone and, when choosing their way, they tend to choose, in probability, paths marked by strong pheromone concentrations. Also, it can be used by other ants to find the location of the food resources found by their nestmates. The more the pheromone of the way, the more possible the way be changed.

3   Slovves 0-1 the KPAICACA Algorithm of the Knapsack Problem

3.1   0-1 The Description of Knapsack Problem

Knapsack problem can be described as follows: there are n items and one knapsack. The volume of item $i (\mathrm{i} = 1,2,\cdots,\mathrm{n})$ is $v_i (\mathrm{i} = 1,2,\cdots,\mathrm{n})$, and the value is $p_i (\mathrm{i} = 1,2,\cdots,\mathrm{n})$, Knapsack's capacity is V. How can we maximize the total benefit considering that we are subject to an absolute capacity limit V.

Since there are two chances (loaded and not loaded) for an item, the variable x_i indicates whether item is selected. If item is selected, then $x_i = 1$, else $x_i = 0$. Therefore, the goal of Knapsack Problem is as follows:

$$Max \sum_{i=1}^{n} p_i x_i \qquad (1)$$

Subject to,

$$\sum_{i=1}^{n} v_i \times x_i \leq V \qquad (2)$$

$$x_i \ (i = 1,2,\cdots,n) \in \{0,1\} \qquad (3)$$

Where, the solution of 0-1 knapsack question can be regarded as a sequence, $(x_1, x_2, \cdots, x_n)$, $x_i \in \{0,1\}$.

3.2 KPAICACA Algorithm

Attributes Item i includes volume v_i and value p_i. So, this paper presents a new method that ant deposits pheromone on items. The visible of item is : $\eta_i := p_i / v_i$.

Each ant is provided with a knapsack and a taboo table. According to the volume of item, ant decides whether to load into the backpack. When the selected item is loaded into the knapsack, the total volume does not exceed capacity, continue to select the next item. Otherwise, the number will be recorded in its taboo table, and the item will not be selected in the future. Through n times' choice, the solution described by item number is constructed.

For easy to describe, several marks are introduced as follows:

m：the number of ant population；$Tab[k]$：taboo table of ant k，recording the number of selected item by ant k, but not loaded in knapsack. $k = 1,2,\cdots,n$；S：set of all items，$S := \{1,2,\cdots,n\}$；$J[k]$：set of items which is visible to ant k; $E[k]$：set of items belonging to knapsack of ant k.

At any moment, assumes that $S = Tab[k] \cup J[k] \cup E[k]$，and $Tab[k] \cap J[k] \cap E[k] = \varnothing$. When ant k constructs a solution which may be transformed into a knapsack question sequence solution $(x_1, x_2, \cdots, x_n)$ by the formula (4) according to the set $E[k]$:

$$x_i = \begin{cases} 1 & \text{if } i \in E[k] \\ 0 & \text{else} \end{cases} \qquad (4)$$

At t moment, Ant k chooses item j by roulette based on probability $P_j^k(t)$, the formula （5）as follows:

$$P_j^k(t) = \begin{cases} \dfrac{[\tau_j]^\alpha [\eta_j]^\beta}{\sum\limits_{s \in J[k]} [\tau_s]^\alpha [\eta_s]^\beta} & \text{if } j \in J[k] \\ 0 & \text{else} \end{cases} \qquad (5)$$

Where α and β are the relatively importance of pheromone and heuristic factor respectively., $J[k] = S - Tab[k] - E[k]$.

When m ants construct solution, these numbers are transformed into 0-1 sequence solutions by formula (4), each 0-1 sequence can be regarded as an antibody. Therefore, the initial antibody population is obtained. Where the encode of each antibody has length n and the size of antibody population is equal to the num of ants.

The main operations for antibody population include cloning operation, mutation operation and selection operation. Define clonal operator T_c as follows:

$$T_c = T_c(A(k)) = [T_c(A_1(k), T_c(A_2(k)), \cdots, T_c(A_m(k))] \qquad (6)$$

Where $A_i^{'}(k) = T_c(A_i(k)) = I_i \times A_i(k)$, $i = 1,2,\cdots,m$, $i = 1,2,\cdots,m$, I_i is q-dimensional row vector whose element is 1, that means q clone for antibody A_i, Where q is a constant. This method reaches a compromise between efficiency and effectiveness.

After cloning, the population is as follows:

$$A^{'}(k) = [A(k), A_1^{'}(k), A_2^{'}(k), \cdots, A_m^{'}(k)] \qquad (7)$$

Generally, there are two strategies (clonal and immune) for clonal operator. This paper uses mutation strategy. The mutation operator, based on a stochastic array $r[i] \in (0,1)$, $i = 1,2,\cdots,n$, acts on the antibody population, that $A^{''}(k) = T_m(A^{'}(k))$. To preserve the original antibody population information, mutation operator does not effect the original antibody population. Where the threshold of mutation is $r_0 \in (0,1)$. If $r[i] < r_0$, The first i-st bit of corresponding antibody has to take inversion mutation.

Define the affinity between antibody and antigen as follows:

$$f(A_i(k)) = \sum_{j=0}^{m} p_j x_{ij} - Z\left[\min\left\{0, V - \sum_{j=0}^{m} w_j x_{ij}\right\}\right]^2 \qquad (8)$$

Where Z is a positive constant.

Define clonal selection operator as: $\forall i \in \{1,2,\cdots,m\}$, if there is an antibody $A_{ij}^{''}(k) = \max\{f(A_{is}^{''}(k)), s \in \{1,2,\cdots,q_i\}\}$, makes the following formula established.

$$f(A_{ij}^{''}(k)) > f(A_i(k)) \tag{9}$$

Then antibody $A_{ij}^{''}(k)$ replaces antigen $A_i(k)$, that antibody $A_i(k+1)$ is obtained. The new antibody population $A(k+1)$ can be obtained by updating.

Update pheromone of items: After antibody immune algorithm completes the search, calculate the sum value of items in knapsack by formula (1),and record the latest optimal solution $A_{best}(k+1)$. The corresponding value is P_{best}. By formula (10), the optimal solution can be updated pheromone on the items.

$$\tau_i \leftarrow (1-\rho)\tau_i + \Delta\tau_i \quad i = 1,\cdots,\text{n} \tag{10}$$

$$\Delta\tau_i = \begin{cases} Q \cdot p(i)/P_{best} & \text{if } A_{best}(k+1)[i] = 1 \\ 0 & \text{else} \end{cases} \tag{11}$$

Where $\Delta\tau_i$ represents the increment of pheromone, parameter ρ is evaporation coefficient of pheromone, and Q is a positive constant. When completing the update of pheromone, the algorithm completes an iteration.

3.3   KPAICACA Algorithm Description

Step 1. Initialization. Iterative counter, pheromone in the items , max iterative number, scale of clonal, threshold of mutation and scale of ants are respectively initialized to 0, τ_0, $N_{\max}$, NB, r_0 and m.

Step 2. for k=1 to m, do
while $(J[k] \neq \emptyset)$, do

(1) Ant k calculates selection probability by formula (5), and selects next item by roulette method.

(2) If the selected item does not meet the constraints, put the figure into the taboo table, otherwise, continue to select next item.
End while

(3) By formula (4), the solution is transformed into encode of antibody. Then, clonal operator T_c, mutation operator T_m and selection operator T_s act on the encode of antibody.
End for

Step 3. Calculate the best solution for this iteration, and update the global optimal solution.

Step 4. By formula (10), calculate the pheromone on the optimal solution items. $t := t+1$.

Step 5. If the iteration number is smaller than N_{max}, then no stagnation, jump the step 2. Otherwise, the algorithm ends and outputs the result.

In the KPAICACA, the role of taboo table reduces blindness in searching, and the time complexity of the algorithm. To increase the randomness of choice, select item by roulette method. The introduction of antibody clonal operator increase diversity of solution.

4   Smulation Experiment Result

4.1   KPAICACA Algorithm Parameter Impact Analysis

Select the programming tool VC6.0. CPU is 2.8GHz, and OS is Window XP. Through solving 80-item knapsack problem, the influence of clonal scale NB and mutation threshold r_0 to KPAICACA is tested. Parameters have test value as follows: $NB = \{1, 5, 10, 20\}$, $r_0 = \{0.01, 0.05, 0.08\}$. When test one parameter, other parameter takes default value (the default value: $\alpha = 1.0$, $\beta = 2.0$, $m = n$, $\rho = 0.2$, $Q = 5$, $NB = 10$, $r_0 = 0.05$). For any parameter assemble, the algorithm KPAICACA calculates for 20 times.

4.2   In KPAICACA Algorithm Solution Literature 0-1 Knapsack Question

The simulation uses instances in literature [3,8,9]. Each example is calculated for 20 times. Table 1 compares the corresponding results among KPAICACA, Quick Ant Colony algorithm [10] and Ant Colony algorithm [11]. For KPAICACA, the parameter is set to: $\alpha = 1.0$, $\beta = 2.0$, $\rho = 0.2$, $m = n$, $Q = 5$, Clone proportion $NB = 10$, mutation threshold $r_0 = 0.05$. As can be seen from table 1, KPAICACA is effective, and has better convergence performance than KPQACA and KPACA.

Table 1. The experiment comparison of KPAICACA and other algorithms

Goods n	KPAICACA			KPQACA (literature[10])			KPACA			Knapsack capacity
	Alg	Valu	Weig	Alg	Value	Weig	Alg	Valu	Weig	
20[3]	1	1024	871	2	1024	871	2	1024	871	878
50[8]	2	1610	1123	8	1610	1123	15	1610	1123	11258
60[9]	4	1870	3199	22	1870	3199	19	1870	3199	3200

5   Summaries

The algorithm proposed in this paper inspires from Immune Clonal Algorithm and Ant Colony Algorithm. With KP's own characteristics, this paper firstly uses Ant Colony Algorithm to calculate optimal solution, then, Antibody clonal selection algorithm can expands the searching space. This method makes KPAICACA better on the rate of convergence and searching capability. There are many issues about hybrid algorithms worth studying. The future work will focus on finding suitable ways to apply the proposed algorithm to other optimization problems.

References

[1] Garey, M.R., Johnson, D.S.: Computers and intractability: a guide to the theory of NP-Completeness, pp. 261–262. W. H. Freeman and Co., San Francisco (1979)

[2] Fréville, A.: The multidimensional 0-1 knapsack problem: An overview. European Journal of Operational Research 155, 1–21 (2004)

[3] Dorigo, M., Maniezzo, V., Colorni, A.: The ant system: Optimization by a colony of cooperating agents. IEEE Transactions on Systems Man and Cybernetics, Part B 26(1), 29–41 (1996)

[4] Dorigo, M., Di, C.G., Gambardella, L.M.: Ant algorithm for discrete optimization. Artificial Life 5(2), 137–172 (1999)

The Application of Virtual Reality Technology in Mechanized Mining Face

Li-Rong Wan, Long Gao, Zhi-Hai Liu, and Liang Wang

Shandong University of Science and Technology, Qingdao 266510, China

Abstract. In recent years, with the development of virtual reality technology and informatization requirements of the coal industry, virtual reality technology has been initially applied to mine production. According to the relevant virtual reality technology, with the software engineering methods, using software such as MultiGen Creator and Vega, a fully mechanized coal face simulation system which can simulate process of longwall face was established; and the scene database design and processing, and other key technology was completed. Finally the simulation of complex process of fully mechanized coal face was effectuated and a new method and platform for the production of guidance and training of miners was supplied.

Keywords: Virtual Reality, Fully mechanized coal face, Modeling, Simulation.

1 Introduction

Virtual Reality (VR) technology is a highly realistic human-computer interaction technique that simulates human vision, hearing, haptic, tactile and other perceived acts. It is a cross-disciplinary technology on the basis of digital image processing, computer graphics, multimedia technology, human - Machine interface technology, computer simulation technology, sensor technology and many other information technology[1]. Virtual reality technology has a very broad prospect. It begins with the needs of the military field, and has played a very important role in the simulation and training of military and aerospace field. With the maturity and development of related technologies, virtual reality technology has been used in medicine, industry, entertainment, education and many other fields.

With the development of market economy, our growing coal industry tends to be large-scale, centralized, information-based, and fully mechanized coal mining has become an important part of coal mining. VR technology can simulate the work on the underground mechanized mining face and generate a series of real-time simulated scenes, the operator can not only see through the three-dimensional animation of the operational status of each device but also achieve the parameters related to each equipment operating, real-time monitoring of mining face is also achieved. In addition, virtual reality simulation of the production process in fully mechanized coal face will have superiorities of less investment, safety, reliable and repeatable operations in teaching and training miners, it will also play an important role in planning and design, program demonstration, decision support and integrated management of mechanized mining face[2].

2   System Development Tool

Currently, there are virous types of modeling tools in virtual reality environment modeling, such as VRML, WTK (World Tool Kits), MultiGen Creator, Java 3D API, anfy3dAPI and other tools. MultiGen Creator/Vega and C++are used to complete the virtual scene modeling and simulation in this paper.

2.1   Modeling Tool

MultiGen Creator is the leading real-time three-dimensional database generating system now, to generate simulation-oriented, well realistic and large area of virtual scenes meeting the premise of real-time, it can provide twenty different types of image generators with modeling systems and tools, and its OpenFlight format has become the most popular image formats in real-time three-dimensional field and the industry standard for the simulation field.

2.2   Simulation Tool

Programs compiled with Vega software have the cross-platform compatibility as high as 99%, supporting OpenFlight, 3D Studio, VRML2.0 and other database formats. Vega has many additional modules, in which LynX graphical interface has powerful, fast and efficient characteristics. Vega also includes a complete C language application program interface, which can link C++ simulation programs and the Vega environment seamlessly after putting three-dimensional entities, scene-driven special module into C++ environment. The virtual reality simulation environment will be realized with the runing program binding to the Vega module[3].

3   System Design

Virtual reality simulation system is based on the modeling scene, including three-dimensional visual and three-dimensional auditory modeling. This system focuses on three-dimensional visual modeling, including geometric modeling, motion modeling, physical modeling, etc[4]. The system development process is shown in Fig.1.
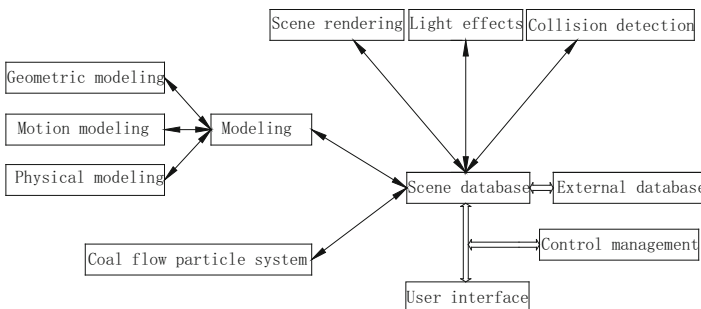
Fig. 1. System flow chart

3.1   Modeling

(1) Geometric modeling. First, to determine the necessary components to establish a solid model and its hierarchy in the scene according to the equipments under the fully mechanized coal face. The main contents of the scene database are entity models (Shearer,etc) and environment (fog, light, material, texture). Solid model hierarchy is shown in Fig.2. The solid modeling is done with MultiGen Creator. Modeling the entities through points, lines, basic shapes and geometric primitives. Models should be simplified as much as possible in the process of modeling, thus reducing the workload and making the system have faster response. Models can be modified to be closer to the actual devices by the material, texture in Creator. The three-dimensional model of hydraulic support is shown in Fig.3.
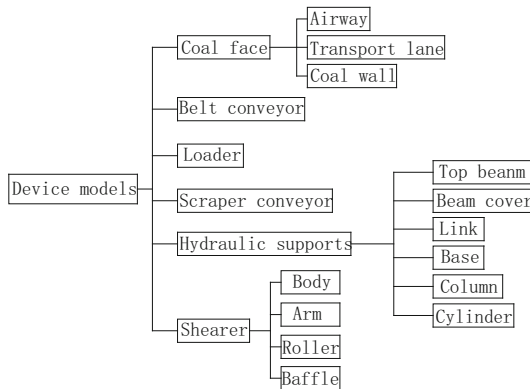
Fig. 2. Solid model hierarchy
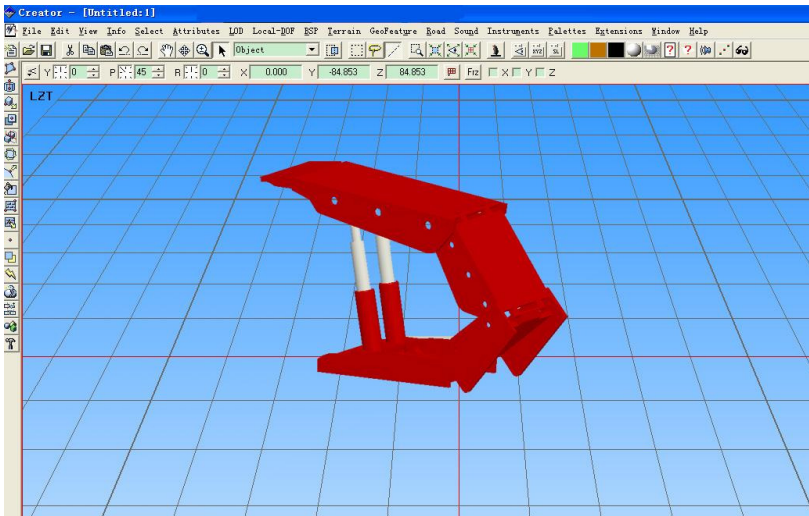
Fig. 3. Hydraulic support

(2) Motion modeling. The main line of the system is based on the movement of shearer, so the process of motion modeling is illustrated by the example of shearer. The shearer body, arms, cylinders and baffles solid models have been completed in the previous step, each part has its particular motion style, but all motions can be seen as the combination of synthesis and rotation. The motion of each parts can be determined according to the type of shearer and it's motion parameters of each component. To realize the movement of models, the DOF (degrees of freedom) of the moving parts should be created in MultiGen, and then a shearer is created in the VC. Finding it in the Vega and determineing the six freedoms of the device model through virtual object coordinate system. Eventually the intended location of the movement of device models will be achieved by modifing and updating the value of freedom. This process is achieved by programming.

3.2   The Establishment of Coal Flow Particle System

In this system, coal flow is the main component, coal flow model is different with other models, it has no fixed shape and size, and it changes as time and space changing. A particle system is constructed to represent the coal flow model in order to meet it's characteristics. The coal particle flow system can be established through a variety of methods in MultiGen and Vega, it is achieved in Vega here. As the following steps:

(1) Create a Vega special effects object-article system, and specify it as the custom type.

(2) The definition of coal particle stream color, shape and size. Coal flow particle's color, shape and other characteristics will not vary as time and space location changeing. In this stream the coal particles is defined as   gray-black color, using polygon particle as particle source shape, then according to shearer and other transport equipment to determine particle size.

(3) The definition of the particles number. Because of the characteristics of the particle system, the system needs to be dynamically updated in real time, but too much particles will cause the pixel processing load, taking up system operation space, and reducing speed. So in this article the particles number of coal flow is limited to just keeping stream of particles continuous visually, expressing the effect of coal flow.

(4) The definition of the velocity distribution of coal flow particles. The speed of coal flow particle is the same on the conveyor, so the velocity distribution of coal flow particles is defined in the way of Planar make particles have the same speed.

(5) The definition of the flow and transform type of the coal flow particles. There are two types of particle flow system: Stream and Burst. Stream shows the particles as continuous particles flow, it is applied in coal flow on the conveyor; Burst releases particles but also removes it, it is applied in coal particles generated in the mining surface in order to achieve a more realistic visual effect.

(6) The definition of the life cycle of coal particle flow. When the particle velocity is defined, the length of the coal flow particle depends on the life cycle of particles. The expected effect is that the coal flow particles should be able to cover all coal transportation equipments in mechanized mining face, so the length of the transportation equipments determines the length of coal flow. Physical effect of coal flow is shown in Fig.4.

3.3 Collision Detection

In the virtual environment, collision, contact or other forms of interaction may occur when the objects move. In order to describe the virtual world better, it is necessary to detect these interactions and make correct response, which is also an important part to treat the physical characteristics.
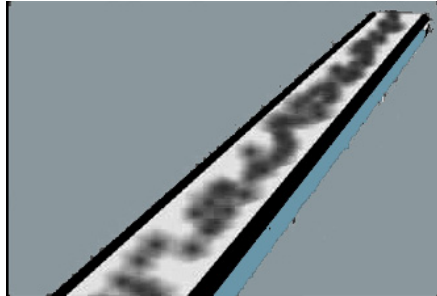

Fig. 4. Coal flow

Currently, there are a variety of collision detection algorithms, which commonly use oriented bounding box method and bounding ball method. Both of these algorithms require only a simple mathematical calculation and fast response, it is much appropriate to use these tow algorithms in the condition excepting easy collision results. In this paper, oriented bounding box (OBB) of the detection algorithm whose most prominent feature is its arbitrary direction, which makes the wrapped object be wrapped as closely as possible according to the characteristic and shape of the object. In three-dimensional space OBB can be represented by the mid-point C, half-length side r1, r2, r3, mutually perpendicular unit vectors v1, v2, v3 and a total of 14 parameters as follows:

$$R = \{C + ar^1v^1 + br^2v^2 + cr^3v^3 \mid a, b, c \in [-1,1]\}$$

Algorithm uses the first and second order statistics of the vertex coordinates object, the division hierarchy of the OBB is based on:

$$\mu = \frac{1}{3n}\sum_{i=0}^{n}(p^i + q^i + r^i)$$

$$C_{jk} = \frac{1}{3n}\sum_{i=0}^{n}(\overline{p}_j^i \overline{p}_k^i + \overline{q}_j^i \overline{q}_k^i + \overline{r}_j^i \overline{r}_k^i), \ 1 \le j,k \le 3$$

In that equation, n is the number of triangles that constitutes the model; μ is the mean; C is the covariance; the vertices of number i triangle is (p^i, q^i, r^i)[5].

3.4   Roaming Mode Setting

In order to meet the user's needs that they can observe the operation of the equipments in the coal face in virtual environment from different perspectives, the roaming should be set up in the system, which allows users to control their own viewing angle and roam in the virtual scene. Vega provides a roaming pattern in itself, which is a good solution to realize roaming. This paper realized roaming directly in the way of Vega.

4   System Function

This system automatically simulates the process of production in fully mechanized coal face under the initial conditions of fully mechanized coal face and allows users to observe the dynamic information of the coal face from different perspectives by setting the roaming; it provides a convenient interface for real-time monitoring system on the mine face; it can take human-machine interaction according to the needs of actual production on coal face; in addition, it can provide pre-job training for miners and multimedia presentations for students related to this industry to understand the work conditions on coal face. Fig.5 shows the interface of the system.
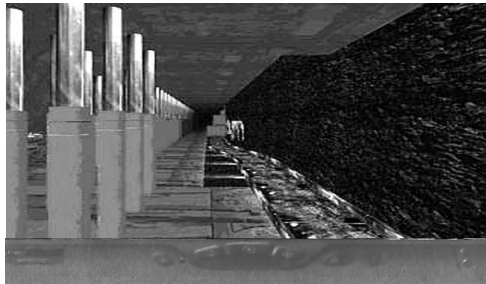
Fig. 5. Simulation of coal face

5   Summaries

This paper constructed a system model based on the research of virtual reality simulation system of fully mechanized coal face, achieved the simulation of the production process and provided a cell roaming capability. Some of the key technologies in the system has been solved, such as the construction of motion modeling and coal flow particle system, etc. The system provides the basement for the construction of the digital mine. It is very useful for the simulation of mine production system, real-time monitoring, accident reconstruction and miners' training. This paper basically realized the simulation of the system, but there are still some flaws in human-machine interaction on the movement of   hydraulic supports and continuity of scraper conveyor, so there's still much work needed to do to perfect the system.

References

[1] Hong, B.-R., Cai, Z.-S., Tang, H.-X.: Virtual reality and its application. National Defence Industry Press (May 2005)

[2] Chang, D.-F., Han, K.-Q., Wei, L.-J.: Application of Virtual Reality Technology in the Simulation of Fully Mechanized Coal Caving Face. Mining R & D 24(1) (February 2004)

[3] Wei, Y.-S., Yang, X.-L., Wang, F.: Virtual reality and system simulation. National Defence Industry Press (January 2004)

[4] Li, J.-Z., Chen, H.-Z.: Study on Virtual Scene's Modeling of Fully Mechanized Mining Face. Journal of System Simulation (August 2006)

[5] Ma, D.-W., Ye, W., Li, Y.: Survey of Box-based Algorithms for Collision Detection. Journal of System Simulation (2006)

Application of Improved Genetic Algorithm in Medical Image Registration

Qiliang Zhu and Qiuhua Shi

North China University of Water Resources and Electric Power

Abstract. The process of Medical image registration is essentially to find a optimal transformation between two medical images, Because of the advantages and disadvantages of various optimization algorithm is relative, this paper proposed by the simplex algorithm and genetic algorithm hybrid optimization algorithm combining for registration parameter optimization, analysed and compared with the traditional genetic algorithm through experiment, verify the simplex genetic algorithm whether in registration precision or in registration efficiency are obviously better than the traditional genetic algorithm, Is a kind of practical optimization algorithm.

Keywords: Medical image registration, Mutual information, Genetic algorithm, Simplex algorithm.

1   Introduction

With the continuous development of visualization technology, information processing of medical image has more and more important in modern medicine. Medical images are playing a very important role in clinical diagnosis, teaching research. Medical image registration is one of the popular research direction in medical image processing area recently. Medical image registration refers to a space change with a picture of a medical image, let it reached space consistent with the corresponding points of another medical image. It is to find a measure of similarity between the two images. Though transforming parameter constantly, it can make the similarity measure to achieve optimal. At last it is transformed into the optimization problem of more parameters and multi-peak value. At present, optimization algorithms used in medical image registration method are: Powell method, the simplex method, genetic algorithm and simulated annealing algorithm, particle swarm algorithm etc[1].The advantages and disadvantages of each optimization algorithm is relative. Hybrid optimization algorithm is effective to avoid the problems of local extremum and slow speed in registration, through the advantage complementary methods, becomes the hot spot of application and research.

   This paper puts forward registration parameters optimization using hybrid optimization algorithm combined by simplex algorithm and genetic algorithm, based on mutual information of medical image registration algorithm. analysed and compared with the traditional genetic algorithm through experiment, verify the simplex genetic

algorithm is a kind of practical optimization algorithm. whether in registration precision or in registration efficiency are obviously better than the traditional genetic algorithm.

2   The Principle of Medical Image Registration Based on Mutual Information

Mutual information is a related measure of two random variables,It is a measure of one variable contains another variable information. If we put two picture prepare registration image gray value as two random variables A and B. Its gray value range is 0 ~ N. Its probability density function is written for $p_r(a)$ and $p_f(b)$ respectively. The joint probability density functions for $p_{rf}(a,b)$, the mutual information of random variables A and B can be described by entropy as formula (1):

$$I(A,B) = H(A) + H(B) - H(A,B) \tag{1}$$

Among them, H (A) and H (B) are the respectively individual entropies of image A and B, H (A, B) is the combined entropy, defined as formula (2):

$$\begin{cases} H(A) = -\sum_a p_r(a) \lg p_r(a) \\ H(B) = -\sum_b p_f(b) \lg p_f(b) \\ H(A,B) = -\sum_a \sum_b p_{rf}(a,b) \lg p_{rf}(a,b) \\ a,b \in [0,N] \end{cases} \tag{2}$$

In the image registration, when the two image based on common anatomic structure achieve the best match . The information that one image expresses about the other image(also is their mutual information), should achieve maximum [2]. this is the principle why biggest mutual information can become registration measure. Registration process is seeking mutual information's largest space transformation through iterative. Often accelerate mutual information optimization by some optimization algorithm. Due to the largest mutual information measure need not do any assumption to Different imaging mode grey of relation, also don't need to do any segmentation or pretreatment to image, so it is widely used in medical image registration, as CT/MR, PET/MR.

Despite the mutual information measure is successfully applied in the medical image registration. Because the overlap size of the two images has a great influence to the mutual information measure, overlap reduce, pixel number that participate in the number of pixels statistical mutual information decreases cause mutual information value decreases. Mutual information is proportional to how much area the two images overlap. Secondly, some scholars Studholme etc[3]say that the increasing of mistakenly number may enlarge the mutual information value. Although the mutual information value achieves maximum, we are not sure we have the right registration

results. To solve this problem, the relationship between the mutual information and the registration parameters can be reflect more accurately by the objective function. Studholme etc puts forward a normalized mutual information measure as formula (3):

$$NMI(A, B) = \frac{H(A) + H(B)}{H(A, B)} \qquad (3)$$

F.maes. etc [4] also put forward a similar standardized information registration criterion. Normalized mutual information was widely used by many scholars recently. It has been proved by practice, that Normalized mutual information is stronger robustness than traditional mutual information in the rigid registration.

3   Genetic Simplex Hybrid Optimization Algorithm

GA is a highly parallel, random and adaptive general optimization algorithm. The coding technology and genetic operation is simple, restrictive conditions cannot constraint optimization. Two of the most significant characteristic is implied parallelism and global solution space search. Although in theory of genetic algorithm with probability 1 convergent features. It always appears "premature" convergence or convergence slow shortcomings in the actual application[5]. Cause the optimal solution quality is not high. Simplex Method, also known as the variable polyhedron search Method. it is deterministic decline method , the local search ability is very strong .the two algorithms which exist such differences in search mechanism are mixed. Benefit to rich search behavior, enhance the search ability and efficiency in global and local sense.

(1) Determine fitness function
Mutual information is the measure of medical image registration. by a floating image geometric transformation, we can find the geometry transform parameter that has the maximum mutual information with floating image. Therefore, in medical image registration of mutual information, fitness function can be identified as the mutual information computation function, taking geometric transformation parameters as the main parameters.

$$\text{MI}_{\text{max}} = f(kx, ky, p) \qquad (4)$$

Among them, kx, ky and p is respectively parameters of x, y axis-translation and rotation transformation.
    This paper is to find each equation parameters that meet fitness function using genetic algorithms.

(2) Coding scheme
An important step of designing genetic algorithms is coded representations of variables solutions. It usually adopts binary code. The main advantage of binary code is the easy-simple operation of encoding and decoding. Crossover and mutation genetic operation is easy to realize. But for many parameters optimization, if we use binary code, it will cause coding too long and searching low efficiency. Real number coding has many advantages, for example: the search space is larger, the global search ability is stronger, the search speed is fast and not easy to fall into the local extreme, and can

avoid notation conversion. Numerical optimization has higher precision and efficiency. It is suitable for a parameter optimization [6]. Therefore, this paper utilizes real number coding.

Establish $x(j)$ for the first j optimization variables, $[a(j),b(j)]$ is the variation of $x(j)$. Use below linear transformation as formula (5):

$$y(j) = \frac{x(j)-a(j)}{b(j)-a(j)} \qquad (j=1,2,\ldots,p) \tag{5}$$

Put the first j optimization variables which initial interval is[a(j),b(j)] correspond to the real interval on the [0, 1], $y(j)$ is gene. Real-number-coding treat direct each variable as gene directly. Also is a kind of no coding encoding. By the real number coding, the scope of all optimized variables are interval [0, 1].

(3) Genetic operation

a. Selection strategy
In the evolutionary process, it always revise three population information: Father generation population, son generation population of father generation population after crossing, son generation population of father generation population after variation. Sorting is according to the fitness of each individual in three populations order. Then we will choose using the rules change geometric distribution selection tactics，and to ensure that each generation final individual number unchanged. This choice strategy not only expands the search space, but also need not worry that crossover operator damage excellent individual and random disturbance caused by variation.
b. Genetic operator
In the genetic algorithm of real number code, we must ensure that gene value is in a given interval limit range. The crossover and variation genetic operator, which used in genetic algorithm, must ensure the new individual genes values that produced by computing within the limits.

Crossover operators which always used in real number coding are discrete crossover and arithmetic crossover. Because arithmetic crossover is not simply swap places like discrete cross. it produces two new generations interpolating around two father solution vectors like mathematical interpolation operation, therefore search range of arithmetic crossover is bigger than discrete crossover. So this paper adopts arithmetic crossover. Set two individual corresponding gene (namely parameters) is a and b. new individual genes produced by crossover operation is a' and b'.

$$a^{'} = \alpha * a + (1-\alpha)*b \tag{6}$$

$$b^{'} = \alpha * b + (1-\alpha)*a \tag{7}$$

α is a random number in (0, 1)

Adopt non-uniform mutation method. For a given individual father generation v .if it's gene (namely parameters) selected by random to variations, the corresponding elements of generating offspring $v^{'}$ change according to the following two possible opportunity randomly:

$$x_k^{'} = x_k + \Delta(t, x_k^{u} - x_k) \tag{8}$$

or

$$x_k^{'} = x_k - \Delta(t, x_k - x_k^l) \qquad (9)$$

In formula (8)and(9) , x_k^u And x_k^l are the bounds of x_k respectively. Function $\Delta(t, y)$ gives a value of interval [0, y]. make $\Delta(t, y)$ tending to zero with the increase of t (t is algebra),if take $\Delta(t, y)$ for :

$$\Delta(t, y) = y \times r \times (1 - t/T)^b \qquad (10)$$

In formula (10), 'r' is the random interval [0, 1],T is the largest evolution algebra. b is the parameters to determine the inhomogeneous degree.

(4) Hybrid algorithm process
20 individuals are Random selected (an array consisted by each image transformation parameters) as the initial population, each individual is a chromosome, each chromosome contains three genes. Because the number of population in the registration process determines the actual search steps in each iteration, therefore populations cannot too big, otherwise it will greatly reduce the efficiency of registration, firstly calculated each chromosome's fitness of initial population when genetic operation began, each group of geometry transform parameter corresponds to the size of the mutual information, and according to the descending order, The first two chromosome as optimal chromosomes directly into the next generation of new species, the third to 10th chromosomes of every two group into 4 groups, combined with the optimal chromosomes again get four simplex, each chromosome as a vertex simplex. The new vertex calculated through the unconstrained simplex method as new chromosome into the next generation populations. The rest of the 10 chromosomes in new populations gets by the original population genomes through classical choice, crossover and mutation. Thus produced a new population's new individuals, repeat the process until meet the requirements, the main processes of calculation shown as figure1.
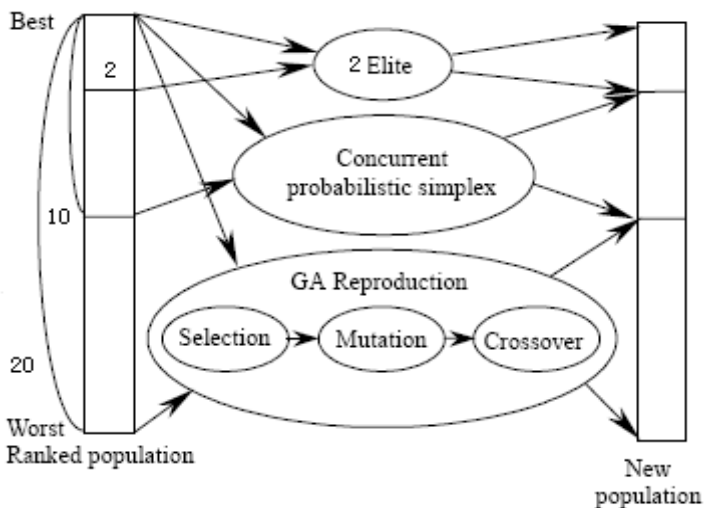
Fig. 1. Genetic mixed Simplex algorithm operators

(5)Simplex coefficient   improvement

Considering the population size is 20 in the hybrid algorithm, individual lack of diversity, easy occurrence premature convergence when genetic operation. We improved the simplex method. Classic simplex method optimize through a series of the transform of reflection and compression , Iterative formula is as follows:

$$X_e = \bar{X} + \alpha(\bar{X} - X_w) \tag{11}$$

$$X_c = \bar{X} - \alpha(\bar{X} - X_w) \tag{12}$$

formula (11) is reflective transform formula , X_w is the simplex vertices transformed by reflect, $\bar{X}$ is the geometric center of all other vertex, X_w is the worst of vertex, $\alpha \in (0,2)$ is reflection coefficient . formula (12)is compressed transform formula, X_c is the vertices transformed by compression. $\beta \in (0,1)$ is compression coefficient. α , β is a fixed value in the original algorithm, in order to make the search space more freedom when simplex iteration , we set it to random value in hybrid algorithm. strengthened the diversity of population in the whole hybrid algorithm, avoid premature convergence.

4   Experimental Results and Analysis

In order to test the performance of the hybrid algorithm, and compare with the traditional genetic algorithm, this paper using hybrid algorithm and the traditional genetic algorithm registration experiment on the same pair images. Registration original image is shown in figure 2, Figure a is a pair of brain CT image, Figure a is clockwised 2 degrees On the plane by Figure a. And respectively shift two pixels in the x, y, coordinate. The environment of Registration experiment is in the PIV 2.0, 256M memory hardware, using Matlab programming for registration experiment. Both two experimental use nearest neighbor interpolation algorithm. Because of certain randomness of Genetic algorithm choice, each experiment   registration 15 times Separately using hybrid algorithm and the traditional genetic algorithm, record registration time、convergence algebra、registration parameters and max mutual information value, experimental results comparison such as table 1 and table2 shows:
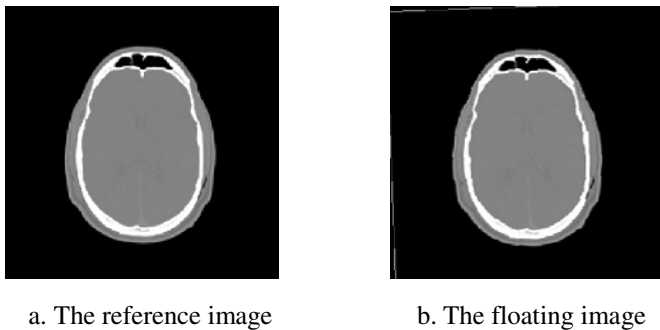

a. The reference image   b. The floating image

Fig. 2. The original image

Table 1. The experiment results of Genetic mixed simplex algorithm

serial number	convergence algebra	registration time	registration parameters	max mutual information
1	7	39. 254	−2, −2, −2	1. 3666
2	10	51. 325	−2, −2, −2	1. 3666
3	5	24. 361	−2, −2, −2	1. 3666
4	6	29. 985	−2, −2, −2	1. 3666
5	8	45. 263	−2, −2, −2	1. 3666
6	6	31. 024	−2, −2, −2	1. 3666
7	9	48. 237	−2, −2, −2	1. 3666
8	5	25. 081	−2, −2, −2	1. 3666
9	6	30. 099	−2, −2, −2	1. 3666
10	7	38. 256	−2, −2, −2	1. 3666
11	7	41. 081	−2, −2, −2	1. 3666
12	8	43. 981	−2, −2, −2	1. 3666
13	6	27. 625	−2, −2, −2	1. 3666
14	9	50. 648	−2, −2, −2	1. 3666
15	11	55. 462	−2, −2, −2	1. 3666

Table 2. The experiment results of Traditional genetic algorithm

serial number	convergence algebra	registration time	registration parameters	max mutualin formation value
1	16	105.263	-2, -2, -2	1.3666
2	18	118.469	-2, -2, -2	1.3666
3	15	98.237	-2, -1, -2	1.3576
4	17	110.263	-2, -2, -2	1.3666
5	17	114.359	-2, -2, -2	1.3666
6	20	148.25	-2, -2, -2	1.3666
7	20	145.896	-2, -2, -2	1.3666

Table 2. (*continued*)

8	17	108.397	-2, -2, -2	1.3666
9	17	109.862	-2, -2, -2	1.3666
10	11	72.49	-1, -2, -1	1.3297
11	16	110.231	-2, -2, -2	1.3666
12	14	89.745	-2, -1, -3	1.3369
13	18	115.389	-2, -2, -2	1.3666
14	15	95.423	-2, -2, -2	1.3666
15	19	123.263	-2, -2, -2	1.3666

In order to more direct observation of the performance of the two kinds of optimization algorithm, we compare the experimental results of table1and table2, get the contrast results such as shown in table 3 .

Table 3. Comparison of the experimental results

optimization algorithm	not convergence times	average convergence algebra	Average registration time
genetic simplex hybrid algorithm	0	7.33	38.78
traditional genetic algorithm	3	16.67	111.03

We can clearly see that from tab.1 data, at 15 times experiments, the hybrid genetic algorithm Not convergence times is 0, while the traditional genetic algorithm for 3 times, registration precision greatly improved, and the average convergence algebra and average registration time also is greatly reduced. So, genetic simplex hybrid genetic algorithm whether in registration precision or registration efficiency are improved significantly, it is a kind of practical optimization algorithm.

References

[1] Chen, H., Varshney, P.K.: A cooperative search algorithm for mutual information based image registration. In: Proc SPIE, vol. 4385, pp. 117–28 (2001)
[2] Maes, F., Collignon, A., Vandermeulen, D., et al.: Multimodality image registration by maximization of mutual information. IEEE Transactions on Medical Imaging 16(2), 187–194 (1997)

[3] Studholme, C.: Measure of 3D medical image alignment. Pattern Recognition 32(1), 71 (1999)

[4] Maes, F., Collignon, A., Vandermeulen, D., Marechal, G., et al.: Multi-modality image registration using mutual information. IEEE Transactions on Medical Imaging 16(2), 187–198 (1997)

[5] Wang, L.: Intelligent optimization algorithms and its application. Tsinghua University Press, Beijing (2001)

[6] Zhou, D.H., et al.: Multiparameter optimization Based on floating-point coding GA. Journal of Computer Simulation 20(9), 98–100 (2003)

Impedance Transformation Trace Generation Algorithm Based on the Smith Chart

Yong Fu, Xiaoshi Zheng, and Qiang Guo

Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center, Jinan, China
`fuy@keylab.net`

Abstract. A circuit schematic for radio frequency (RF) impedance transformation is designed at first; then the elements in the schematic constitute a binary tree. Next the operation expression for impedance transformation is generated from the binary tree. The impedance transformation trace generation algorithm is proposed at last. In the last part of the paper a practical RF amplifier's impedance matching network is simulated and analyzed to verify the algorithm. It can be seen that the algorithm is precise and efficient. This algorithm suggests applications to simulation and design of impedance matching networks for RF amplifiers.

Keywords: Smith chart, Binary tree, Wireless communication, Impedance transformation.

1 Introduction

Radio Frequency (RF) Amplifiers are wildly used in modern wireless communications [1-4]. Power Amplifiers (PA) are used to increase the transmission power and Low Noise Amplifiers (LNA) are used to improve the receiver sensitivity. There are many differences between the low frequency amplifier design and the RF amplifier design. At first, the operation wavelength is comparable with the conductive track's length in RF circuits and the track must be handled as a circuit component; and the impedance matching, it means the input impedance and out impedance should be transformed to some specified value, in most conditions the value is about 50ohm, is critical[2].

Usually RF engineers use a tool called Smith Chart [5, 6] to help them do impedance matching work and study the performances of designed amplifiers. Unfortunately the impedance matching is a boring and time-consuming job, though lots of programs and EDA software based on the Smith chart [7-10] can be used to simulate and analyze impedance transformation circuit, few of them give the impedance transformation trace, which provides some useful information, such as circuit bandwidth and loaded quality factor. The role of the elements in RF circuits could be better understood by analyzing the impedance transformation trace in the Smith chart. So the impedance transformation trace can be used to guide the adjusting of the elements' parameters in RF circuits. In this paper, circuit elements and schematic are designed and realized at first, then a 2-port impedance transformation trace generation algorithm is proposed, at last the algorithm is employed to realize a practical RF amplifier impedance matching network.

2   The Smith Chart and Impedance Transformation Method

The Smith chart, which is invented by P. H. Smith, is a graphical aid designed to assist in solving problems with transmission lines and matching circuits in radio frequency. The Smith chart is plotted on the complex reflection coefficient plane in two dimensions and is scaled in normalized impedance and admittance, which is often known as the ZY Smith chart. The Smith chart can be used to represent many parameters, such as impedances, admittances, reflection coefficients, SWR, quality factor, etc. It's the most basic and popular tool in RF circuit designs. Fig. 1 is a Smith chart developed by us, the size of the chart is chosen as 500*500 points, and it will provide enough precision. Every point in the chart contains following information: coordinate, reflection coefficient, SWR, impendence and admittance. Further more, different colors are used to mark different SWR; meanwhile quality factor lines are displayed.
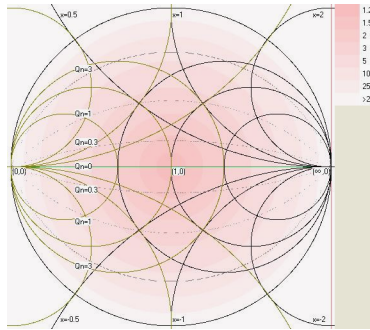
Fig. 1. Smith chart

To achieve maximum power transfer or minimum noise figure, the impedance of the load should be matched to that of the source, which is accomplished by impedance matching networks (IMN). These networks usually contain only passive elements: capacitance (C), inductance (L), transmission line (TL) and resistance (R). A typical LNA with matching networks is shown in Fig. 2[11].
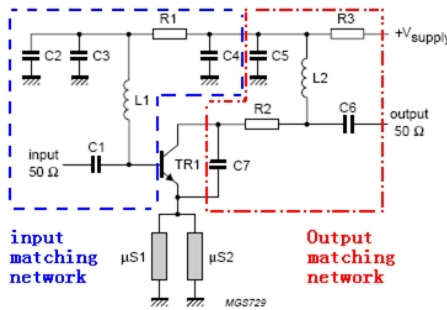
Fig. 2. Typical LNA circuit

The connecting of elements to IMN will transform impedance in the Smith Chart according of connecting type and element's type:

(1)   Series connection of L and C elements. Assuming the original impedance value is $Z_0 = r + jx$. The real part of the impedance r doesn't change, the imaginary part x is added by the impedance value of the elements x_s:

$$Z = r + j(x + x_s)_,$$ (1)

In the Smith chart, the impedance moves around constant resistance circle.

(2)   Series connection of R elements. Assuming the original impedance value is $Z_0 = r + jx$. The imaginary part of the impedance x doesn't change, the real part r is added by the impedance value of the elements r_s:

$$Z = (r + r_s) + jx_,$$ (2)

In the Smith chart, the impedance moves around constant reactance circle.

(3)   Parallel connection of L, C and TL elements. Assuming the original admittance value is $Y_0 = g + jb$. The real part of the admittance g doesn't change, the imaginary part b is added by the admittance value of the elements b_p:

$$Y = g + j(b + b_p)_,$$ (3)

In the Smith chart, the impedance moves around constant conductance circle.

(4)   Parallel connection of R elements. Assuming the original admittance value is $Y_0 = g + jb$. The imaginary part of the admittance b doesn't change, the real part g is added by the admittance value of the elements g_p:

$$Y = (g + g_p) + jb_,$$ (4)

In the Smith chart, the impedance moves around constant susceptance circle.

(5)   Series connection of TL elements. Assuming the original impedance value is Z_0. Transmission line's characteristic impedance is Z_{line}, phase constant is β and line length is l, the impedance is transformed to:

$$Z = Z_{line} \frac{Z_0 + jZ_{line} \tan(\beta d)}{Z_{line} + jZ_0 \tan(\beta d)}_,$$ (5)

The impedance position in the Smith chart is rotated by twice of the electrical length of the transmission line $el = 2 \times \beta d$.

By now, preparations for the impedance transformation are done. In order to simplify calculation, three restrictions are defined:

> (1) The element placement area is divided into square cells, and one clement in one cell;
> (2) There are 3 connectors at most in one cell;
> (3) The impedance is transformed from the RL port to the Zin Port.

Also, three elements are introduced to make up a circuit: connecting line, RL port and Zin Port. Based on these rules, a schematic can be easily realized. Fig. 3 is a demo schematic.
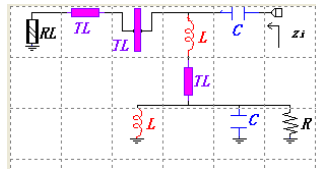
Fig. 3. Circuit schematic

3   Impedance Transformation Trace Generation Algorithm

In order to generate the impedance transformation trace, the operation expression should be extracted from the schematic. The cells in the schematic are numbered in the format of "CCR", "CC" is the column No., which is ranged from 00 to 17; "R" is the row number, which is ranged from 0 to 9. Also a structure is introduced to store the parameters of cell (Fig. 4-a). The structure is defined as follow:

(1) It has three pointers, one point to its father node, one point to the left child node and the last one point to the right child node.

(2) It has a Layer parameter. It is used to identify the position in the expression binary tree.

(3) It stores parameters of the element in the cell.

(4) Zi port is the tree's root node.

(5) RL cell is always an end node on the left side.

Now all cells are connected to a binary tree by connecting type. The Zi port are found out as the tree's root node at first, then check its parameters to find its left child node and the right child node. Then we check its left child node and its right child node; do the same thing till the node has no child node[12]. After this work, a binary tree is generated (Fig. 4-b). The calculation will started from the RL cell, find its father node and check the node list, then decide what operation to do. When the father node is Zi cell, a calculation expression is generated. Fig. 4-c is the final binary tree of the circuit.
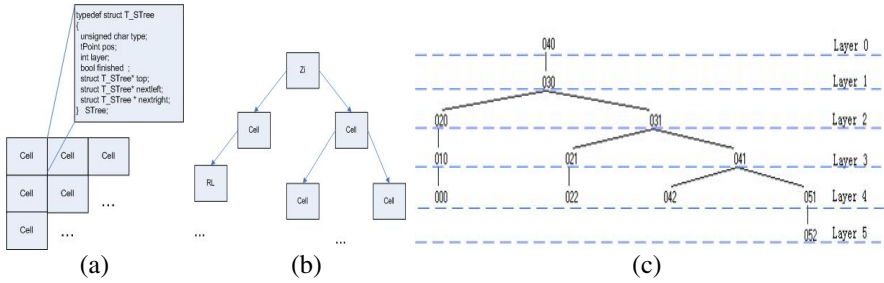
Fig. 4. Binary tree generation from the Schematic. (a) The original schematic; (b) The Binary tree; (c) Binary tree of the circuit in Fig. 3.

Then we generate the operation expression. Every cell generates an expression[13], which have one or two child nodes position information and the connecting information. There are two connecting types: '+' means series, and '/' means Parallel. The priorities of these operators are: '+' > '/'. All expressions are sorted by layer and stored to an expression list; it can be seen in Figure 5. After that the expressions are merged in this way:

(1) Find the position information from the expression list, replace the position information by the new one and delete the expression from the list;
(2) Repeat (1), until there is only one expression left.

After the last expression generated, four things have to do: all connecting lines are removed at first; then all serial transmission line are found out and their operators be changed to '>', which means series transmission line; third unnecessary parentheses in the expression are removed; and at last, the expression is reversed, this is used for impedance transformation trace generation. The finial expression is started with the RL cell and all neighbor cells are connected by '+','/' or '>'.

Finally, the impedance transformation trace can be generated. Position information in the expression is replaced by the elements' parameters in the cell.

Fig. 5. The operation expression generation process

The calculation is started from the RL port, and then its impedance is operated with its next node according to equation (1)-(5). Meanwhile the impedance transformation trace could be plotted in the Smith chart. The operation is repeated until the next node is Zin port. The final result is shown in Fig. 6, and the circuit quality factor can be read directly from the Smith chart, in this demo, the quality factor is 8.691 The algorithm code is wrote by the C Language, the code size is 3.4K bytes and memory cost is 2.8M bytes. The calculation in this paper was performed on a 2.8GHz Pentium® D computer, the calculation time cost is less than 1 second.
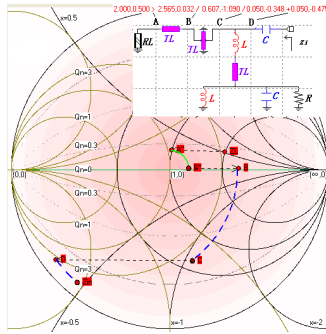
Fig. 6. Impendence match calculation process

4   Results and Discussion

An 2.45Ghz power amplifier circuit from [14] is used to verify our algorithm. We take the most complex part of impedance matching networks: inter-stage matching network, to be our task. The inter-stage matching is to match the load impedance conjugation of the first transistor Z_{Q1load} to the source impedance of the second transistor $Z_{Q2source}$: $Z^*_{Q1load} \xrightarrow{IMN} Z_{Q2source}$, in this circuit,

$$Z_{Q1load} = 45 + j37 ,$$

$$Z_{Q2source} = 8 - j11 .$$

The circuit is shown at the upper part of Fig. 7. The impedance transformation trace is generated and displayed in the Smith chart. The quality factor is 2.356, and the calculation result is 12-j13.5, which is pretty close to the theoretical value 8-j11. Considering that the elements in the schematic are simplified, our algorithm is precise and effective.
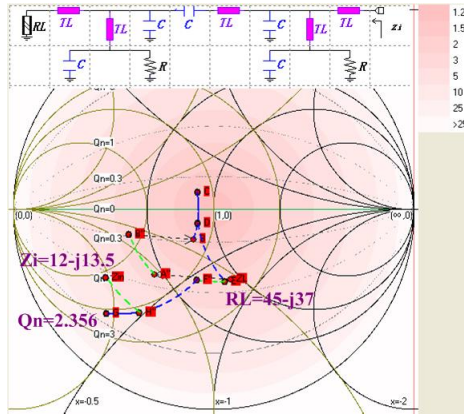
Fig. 7. Inter-stage matching circuit

5   Summaries

In conclusion, we proposed an impedance transformation trace generation algorithm for RF amplifier design. We also have verified the algorithm by simulating a practical RF amplifier impedance matching network. It can be seen that the algorithm is easy to realize and it needs only small memory spaces. It is precise and efficient at the same time. This algorithm can be employed to simulate and design impedance matching networks for RF circuits.

References

[1]   Haykin, S.S., Moher, M.: Modern wireless communication. Prentice Hall (2004)
[2]   Ludwig, R., Bogdanov, G.: RF circuit design: theory and applications. Prentice Hall (2008)
[3]   Sayre, C.W.: Complete wireless design, 2nd edn. McGraw-Hill Professional (2008)
[4]   Cripps, S.C.: RF power amplifiers for wireless communications. Artech House (2006)
[5]   Smith, P.H.: An improved transmission line calculator. Electronics 17, 130 (1944)
[6]   Smith, P.H.: Transmission line calculator. Electronics 12, 29–31 (1939)
[7]   Loo, C.H., Elmahgoub, K., Yang, F., Elsherbeni, A.Z., Kajfez, D., Kishk, A.A., Elsherbeni, T., Ukkonen, L., Sydanheimo, L., Kivikoski, M.: Chip impedance matching for UHF RFID tag antenna design. Progress in Electromagnetics Research 81, 359–370 (2008)
[8]   Shamim, A., Radwan, A.G., Salama, K.N.: Fractional smith chart theory. IEEE Microwave and Wireless Components Letters 21, 117–119 (2011)
[9]   Hoarau, C., Corrao, N., Arnould, J.D., Ferrari, P., Xavier, P.: Complete Design and Measurement Methodology for a Tunable RF Impedance-Matching Network. IEEE Transactions on Microwave Theory and Techniques 56, 2620–2627 (2008)
[10]   Kundert, K.S.: Introduction to RF simulation and its application. IEEE Journal of Solid-State Circuits 34, 1298–1319 (1999)
[11]   Zhu, X., Ying, S., Ling, L.: Multimedia sensor networks design for smart home surveillance. In: Control and Decision Conference, pp. 431–435 (2008)
[12]   Knuth, D.E.: Art of Computer Programming, Volume 1: Fundamental Algorithms. Addison-Wesley Professional (1997)
[13]   Becchi, M., Crowley, P.: An improved algorithm to accelerate regular expression evaluation. In: Proceedings of the 3rd ACM/IEEE Symposium on Architecture for Networking and Communications Systems, pp. 145–154 (2007)
[14]   Semiconductors NXP, 2.4 GHz power amplifier with the BFG425W and the BFG21W (2000), `http://www.geniusnet.sk/om3bc/datasheets/AI_BFG425W_21W_2400M_1.pdf`

A Niche-MAS Framework for Evolving Artificial Neural Networks

Yudong Wang and Gengming Zhu

School of Computer Science and Engineering Hunan
University of Science and Technology, Xiangtan, 411201, China

Abstract. In order to adopt PEAs for evolving neural networks and to implement or simulate them efficiently, we introduce a distributed framework, called NMfENN in this paper. The NMfENN framework is embedded with a general PGA, which guide the evolution of the whole neural network population. But the PGA is not an explicit entity; it carries out by the Niche-MAS, which consists of a number of NAs (Niche Agents) and a CA (Commonweal Agent). By the implementation of NEAT, SANE, and MOBNET, NMfENN is demonstrated to be easily and conveniently employed to realize users' parallel evolutionary algorithms or strategies for evolving neural networks.

Keywords: neural networks, Niche Agents, Commonweal Agent, Niche-MAS.

1 Introduction

The problem of finding a suitable architecture, including its topology and corresponding weights, of the Artificial Neural Network (ANN) for a given problem is a very complex task [1, 2]. A network smaller than needed would be unable to learn, as larger than needed would probably end in over-training. The use of evolutionary algorithms (EAs) [3, 4] has been a popular approach to find a neural network suitable to solve a given problem [1, 2]. EAs are particularly useful for dealing with large complex problems which generate many local optima by traditional gradient-based search algorithms. They can even deal with problems where no explicit or exact objective function is available [1].

The most important phases in EAs are reproduction (crossover), mutation, fitness evaluation and selection (competition). Among these phases, fitness evaluation is very often the most CPU intensive part [5]. Computation times of more than 1 CPU year have been reported for a single run in complex domains [5]. Moreover, when the population is very large (e.g. evolution of ANN applied in some special domains), EAs are very memory-consuming [5]. In some cases this makes it impossible to run an application efficiently using a single machine.

To tackle these problems, one popular approach is to employ parallel EAs (PEAs, e.g. Parallel Genetic Algorithms, PGAs). In addition, EAs can search in parallel different subspaces of the whole search space, thus making it less likely to become trapped by low-quality subspaces [5, 6]. In the research of evolving neural networks, several algorithms similar to PEAs are studied [2, 7, 8, 9, 10]. But these researches only have the training convergence in mind and overlook the potential parallelism of

the algorithms. Further more, there are few effective mechanisms to adopt PEAs and to implement or simulate the algorithms for evolving neural networks.

In this article, based on niching method [11] and MAS (Multi-Agent Systems) theory [12, 13], we introduce a distributed framework, with a PGA embedded, to efficiently evolve neural networks. We call this framework Niche-MAS for Evolving Neural Networks (NMfENN).

In the next section, we briefly review niching method and MAS theory. They are the context and foundation of NMfENN. Section three specifies three major issues of NMfENN: 1) the procedure of the embedded PGA; 2) the implementation of the PGA by MAS; 3) the mechanism of individuals' migration. Then we demonstrate the effectiveness of NMfENN by several experiments in section 4. In the final section, we draw some conclusions.

2   Niches and MAS

In ecology, a niche is a term describing the function or position of an organism or a population within an ecological community. The niche concept was formulated to determine the role of organisms in a community [14]. It is well known that natural evolutionary processes maintain a variety of species, each occupying a separate ecological niche. However, traditional EAs rapidly push an artificial population toward convergence [11].

Niching methods are the extension of EAs into the multimodal domain. They aim to maintain the diversity of certain properties within the population; and this diversity allows parallel convergence within those subpopulations into multiple good solutions for the given problem [11, 15]. So, niching methods are applicable for two sorts of problems: 1) problems need to delay convergence to increase exploration, so that a better, single solution can be obtained; 2) problems demand multiple, final solutions [15].

The field of multi-agent systems (MAS) is a vibrant part of distributed artificial intelligence. It aims to lay down the principles for the construction of complex systems involving multiple agents and to provide mechanisms for coordination of independent agent behaviors.

An agent is a computer system that is capable of independent action on behalf of its user or owner [12]. A multiagent system is one that consists of a number of agents, which interact with one-another [12]. In the most general case, agents will be acting on behalf of users with different goals and motivations. MAS are essentially distributed.

3   The NMfENN Framework

The framework of NMfENN is depicted in Fig. 1. This framework is embedded with a PGA (EPGA), which guide the evolution of the whole neural network population. But the PGA is not an explicit entity. It carries out by the Niche-MAS (NMAS). The MAS consists of a number of NA (Niche Agent) and a CA (Commonweal Agent). As described in the precious section, niches are brought to maintain diversity of the population, and agents in MAS have their own intentions. The NAs with specific objectives are designed to manage the niches with a subpopulation of neural network, and to communicate with the CA in the process of evolution. In NMAS, taking advantage of the Blackboard, the CA acts as a manager and communicator.
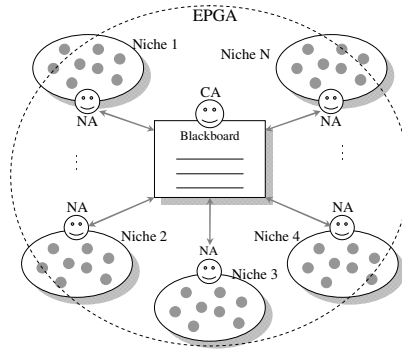
Fig. 1. The framework of NMfENN. The broken line indicates that the PGA is an embedded on but not an explicit entity.

3.1   The Embedded PGA

According to the nature of MAS, the PGA employed in NMfENN, namely Embedded PGA (EPGA), is a coarse-grained PGA [5, 6]. EPGA divide the neural network population into some number of niches (subpopulations). Niches are separated from one another, and individual neural networks compete only within a niche. From time to time, some individual migrate from one niche to another.

The data structures and operators of EPGA are similar to that in [16], which are shown in Fig. 2. The significant difference is the data structures and operators associated with Niche. In EPGA, the Niche has a Local Clock, a local Evaluator and a local Migration Manager. The Local Clock enables a niche to have a local time and to evolve ahead with the parameter Lookahead. The local Evaluator evaluates the individuals in a niche. The Migration Manager migrate individuals in a niche to the whole population according to the migration mechanism.
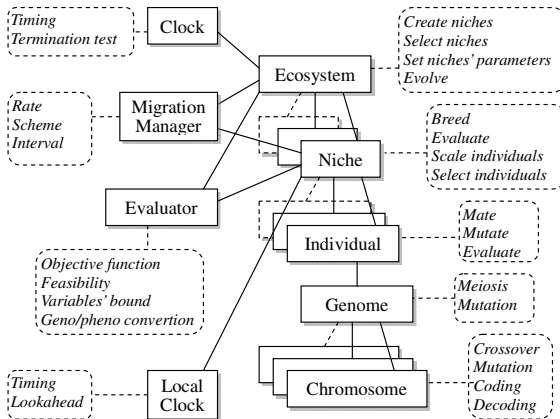
Fig. 2. The data structures and operators of EPGA

Based on the data structures and operators, the general procedure of EPGA can be described by the following pseudo code. It is revised from the NES in [17].

```
t = 0;
InitializeParameters();
CreateNiches(i,t) ∀ i∈N;
SetNichesParameters(i,t) ∀ i∈N;
Evalute(Niche(i,t)) ∀ i∈N;
iterate
    t := t+1;
    NicheEvolution(i,t) ∀ i∈N;
    Evalute(Niche(i,t)) ∀ i∈N;
    At Certain Interval
        NicheCompete();
        GenesSelect();
        GenesMigrate()
until Terminated();
display solution
```

In this procedure, t is the generation number and N denotes the number of niches. The procedure of CreateNiches(i,t) create the desired number of niches for the run. SetNichesParameters(i,t) gives the parameters of each niche, such as crossover rate, mutation rate, subpopulation size, migration scheme, migration rate and migration interval. The procedure of NicheEvolution(i,t) performs the independent evolution of each niche. In NicheCompete() the worst niche dies. GenesSelect() selects the proper individuals in the whole population. GenesMigrate() has two functions: migrating the selected individuals among niches and creating a new niche to replace the extinct one.

3.2   The Niche-MAS

The Niche-MAS is designed to implement EPGA. However, instead of the implementation of a certain PGA, it is a general framework suitable for a group of PGA, strategies or models for NN evolution (refer to section 4). The MAS consist of a number of NA and a CA. The agents in NMAS are runnable entities with their own data.

The Commonweal Agent. Just as its name implies, the Commonweal Agent (CA) provides common facilities for evolution and has the commonweal of the MAS in mind. The typical CA contains three classes: a Evaluator, a Clock, a Parameter and a MigrationConcolor, and two procedures: initialize() and iterate(). Fig. 3 depicts the correlative classes and their relationships with CA.
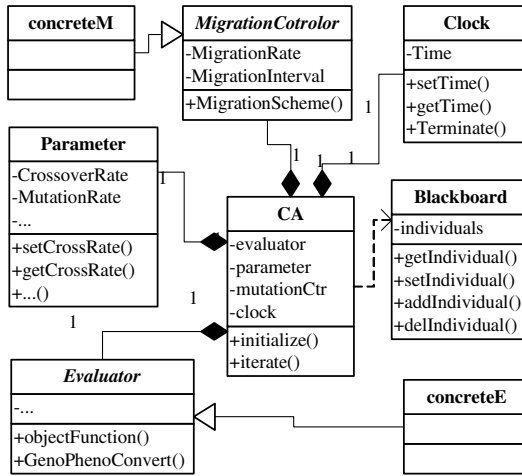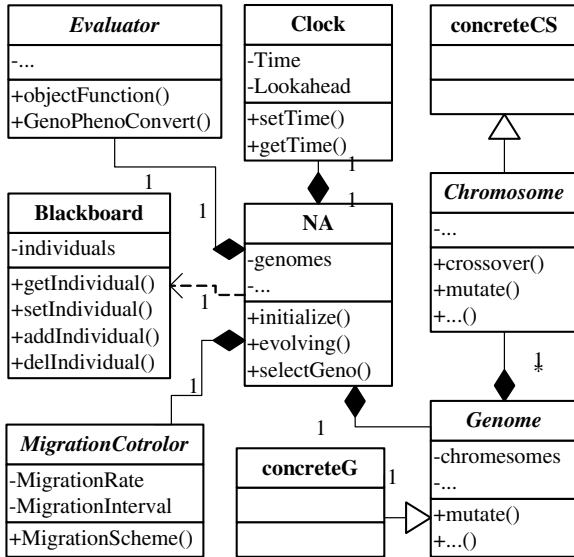
Fig. 3. The correlative classes and their relationships with CA

The concrete E and concrete M denote the classes of concrete Evaluator and concrete Migration Concrolor respectively.

To achieve the generality of NMfENN, the classes of MigrationConcolor and Evaluator are set to be abstract, which need to implement by users. The Blackboard (also in Figure 1) is the information sharing mechanism with a hash table, which contains the elite individuals selected by the NAs. In CA, the procedure iterate() implement the iteration part in EPGA, while initialize() do the initialization part. But In the iteration, CA do not explicitly call NicheEvolution(); and it only supervises niches' evolution by time management.

The Niche Agent. The Niche Agents (NAs) manage the niches created by CA. Each NA is embedded with a traditional GA which performs the evolution of its managed niche. Although the GAs' courses in NAs are supervised by CA, they have their own clock which can go ahead by CA's permission *Lookahead*. In NAs, individuals are regarded as genomes, that is to say, we only employ genotype in the niches' evolution. The use of phenotype only appears in the evaluation.

The correlative classes and their relationship with NA are showed in Fig. 4. Similar to *MigrationConcolor* and *Evaluator*, *Genome* and *Chromosome* are set to be abstract; and they are implemented by users. In each NA, the type of parameter *genomes* is *Genome*; and this parameter contains all the individuals' genomes in the NA's managed niche. In evolution, NA calls procedures in the abstract *Genome* which hook the identical procedures in the concrete one.

Fig. 4. The correlative classes and their relationships with NA. The concrete G and concrete CS denote the classes of concrete Genome and concrete Chromosome respectively.

3.3   Individuals' Migration

In multiple-population coarse-grained GAs, the exchange of individuals among subpopulations is called migration which plays an important role in GAs' behavior [5, 6]. In NMfENN, individuals' migration is carried out by the Blackboard which is mentioned in previous subsections. Besides the hash table of elite individuals, Blackboard provides several RPCs (Remote Procedure Calls) for CA and NAs. Using these RPCs, CA can manage (e.g. select individuals and assign migrants) the migration of individuals; and NAs can offer the selected individuals or get the assigned migrants.

Based on the mechanism provided by Blackboard, the two typical migration model, island model and stepping stone model [5], can easily be realized. Since individuals can migrate to any other niche in island model, migration can simply be carried out by NAs' freewill selection of other NAs' elite. In stepping stone model, individuals can migrate only to neighbouring niche; and by maintaining a topology that defines the connections between the niches, CA can distribute the migrants to corresponding adjacent niches.

4   Empirical Results

As mentioned above, NMfENN is a general framework suitable for the implementation or simulation of a group of algorithms, strategies or models to evolve neural networks. To show this virtue of NMfENN, we conduct the implementation experiments of several representative algorithms or strategies for evolving neural networks. They are NEAT [7], SANE [8], and MOBNET [2].

4.1   NEAT

There are toolboxes for NEAT [18]. However, we completely implement it in NMfENN. TNEAT is a method to evolve neural networks. It consist of a set strategies: 1) employing a principled method of crossover of different topologies, 2) protecting structure innovation using speciation, and 3) incrementally growing from minimal structure [7].

The first strategy defines the genome format, mutation policy and crossover pattern. All of these can be carried out by the concrete classes of Genome and Chromosome depict in Fig. 4. Strategy 2) is coincident with the niching method used in NMfENN; and it can be naturally realized using the compatibility distance define in [7]. The third strategy means that NEAT begins with a uniform population of networks with no hidden nodes, and grow new structure only as necessary. We achieve these goals by the initialization in CA and evolution and evaluation in NAs. Moreover, in NMfENN, MEAT is parallelized.

Also, the pole balancing problem is used to demonstrate the validity and efficacy. We deploy the implemented NMAS over a PC-cluster including 8 PCs. One PC runs the CA which selects the best network to control the pole; other 7 PCs run 7 NAs each with a niche of 143 or 142 individuals. Experiment shows the validity of our implementation; and because of the parallelization, NEAT in NMfENN runs almost 3.7 times faster than the original one.

4.2   SANE

Different from other approaches to neuroevolution in which individual in the population represent a complete neural network, SANE method restrict the scope of each individual to a single neuron; and complete neural networks are built by combining several neurons. In SANE, the notion of blueprints is used to denote the records of neuron combinations (neural networks). So, SANE maintains and evolves two populations: a population of neurons and a population of network blueprints. The neuron evolution searches for effective partial networks, while the blueprint evolution searches for effective combinations of the partial networks [8].

To implement SANE in NMfENN, one of the major issues is how to manage the blueprint population. It is natural to maintain blueprint population in CA, for blueprints is the record of neuron combinations. But this arrangement aggravates the labor of CA. Our solution is to distribute blueprint to niches maintained by NAs. Blueprints get neurons' information by migration mechanism. CA only needs to maintain a virtual topology of niches as that in Fig. 5. It also can allow the migration among neuron niches or blueprint niches.
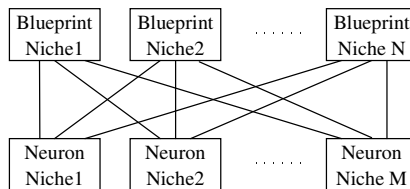

Fig. 5. The virtual topology of niches maintained in CA. However, in some cases, migration among neuron niches or blueprint niches

Another important issue is the evaluation of the neurons because it depends on neurons' participation rate as well as neurons themselves. In NMfENN, this issue can also be settled by migration mechanism. We let the migrant of blueprint niches be the participation rates of each neurons recorded in the blueprints. Neuron niches can get this information to evaluate their individuals.

In addition, other implementations are trivial. So, as discussed above, SANE can easily be realized in NMfENN.

4.3   MOBNET

Similar to SANE, MOBNET uses cooperative coevolutionary algorithms to evolve neural networks and maintains two different populations: population of subnetworks (nodules) and population of networks. As usual, MOBNET specified the strategies for evolving nodule population and for evolving network population. The new idea of MOBNET is the introduction of multi-objective optimization in the process of subcomponent fitness assignment [2]. In MOBNET, the proposed multi-objective evolutionary algorithm for obtaining the fitness of individuals of the population is based on the concept of Pareto optimality; and it is basically an adaptation of NSGA [19] in which the underlying idea is the use of a ranking selection method to emphasize current non-dominated individuals and a niching method to maintain diversity in the population [2].

It is well-known that fitness calculation is one of the steps of evaluation in GAs. So, the subcomponent fitness assignment can naturally be involved in the process of evaluation. Accordingly, to implement MOBNET in NMfENN, the multi-objective evolutionary algorithm and the principles of multi-objective optimization are carried out in the concrete evaluators in CA and NAs. Techniques for other implementation issues are similar to SANE. There are no migrations in MOBNET; and each network is the combination of one nodule of each subpopulation of nodules. So, actually, except the concrete evaluators the implementation of MOBNET is easier than that of SANE.

5   Summaries

For evolving neural networks, PEAs, especially cooperative coevolutionary algorithms (e.g. algorithms in SANE and MOBNET), offer a promising alternative to the standard EA methods. It is important to have some organisms to implant them. Moreover, the implementations or simulations of these PEAs have their practical significance. Taking advantage of the niching methods, the NMfENN framework embeds a general PGA that carried out by MAS. By means of abstraction, NMfENN allows users to specify the genotypes and phenotypes of the objective neural networks, to realize the demanding operators and schemes at their will, and to regulate the parameters for the evolution correspond to their purpose. So NMfENN is a very general solution to implementation of the PGAs for evolving neural network. Moreover, users can easily and conveniently employ NMfENN to realize their notion of parallel evolution of neural networks; and NMfENN is very laboursaving for it. All of these merits of NMfENN are demonstrated in practice.

References

[1] Yao, X.: Evolving artificial neural networks. Proceedings of the IEEE 87(9), 1423–1447 (1999)

[2] García-Pedrajas, N., Hervás-Martínez, C., Munoz-Pérez, J.: Multi-objective cooperative coevolution of artificial neural networks (multi-objective cooperative networks). Neural Networks 15(10), 1259–1278 (2002)

[3] Goldberg, D.E.: Genetic algorithms in search, optimization and machine learning. Addison-Wesley, Boston (1989)

[4] Michalewicz, Z.: Genetic Algorithms + Data Structures = Evolution Programs, 3rd edn. Springer, Heidelberg (1996)

[5] Mariusz, N., Riccardo, P.: Parallel genetic algorithm taxonomy. In: Proceedings of the Third International Conference on Knowledge-based Intelligent Information Engineering Systems (KES 1999), Adelaide, Australia (1999)

[6] Cantú-Paz, E.: A Survey of Parallel Genetic Algorithms. Technical Report IlliGAL 97003, University of Illinois at Urbana-Champaign (1998)

[7] Stanley, K.O., Miikkulainen, R.: Evolving Neural Networks Through Augmenting Topologies. Evolutionary Computation 10(2), 99–127 (2002)

[8] Moriarry, D.E., Miikkulainen, R.: Forming neural networks through efficient and adaptive coevolution. Evolutionary Computation 5(4), 373–399 (1998)

[9] Potter, M., De Jong, K.: Evolving Neural Networks with Collaborative Species. In: Proceedings of the 1995 Summer Computer Simulation Conference, Ottawa, Ontario, Canada, pp. 24–26 (1995)

[10] García-Pedrajas, N., Hervás-Martínez, C., Munoz-Pérez, J.: COVNET: A cooperative coevolutionary model for evolving artificial neural networks. IEEE Transactions on Neural Networks 14(3), 575–596 (2003)

[11] Mahfoud, S.W.: Niching Methods for Genetic Algorithms. PhD Thesis, University of Illinois at Urbana-Champaign, IlliGAL Report No. 95001 (1995)

[12] Wooldridge, M.: An Introduction to Multi-Agent Systems. John Wiley and Sons Limited, Chichester (2002)

[13] Ferber, J.: Multi-Agent Systems: An Introduction to Distributed Artificial Intelligence. Addison- Wesley Longman Inc. (1999)

[14] Hutchinson, G.E.: Concluding remarks. In: Cold Spring Harbour Symposium on Quantitative Biology, vol. 22, pp. 415–427 (1957)

[15] Shir, O.M., Bäck, T.: Technical Report: Niching in Evolution Strategies, TR 2005-01, LIACS, Leiden University (2005)

[16] Pedroso, J.P.: Niche search: an evolutionary algorithm for global optimization. In: Ebeling, W., Rechenberg, I., Voigt, H.-M., Schwefel, H.-P. (eds.) PPSN 1996. LNCS, vol. 1141, pp. 430–440. Springer, Heidelberg (1996)

[17] Porter, F., Xue, B.: Niche Evolution Strategy for Global Optimization. In: Proceedings of the 2001 Congress on Evolutionary Computation, vol. 2, pp. 1086–1092. IEEE Press, New Jersey (2001)

[18] http://nn.cs.utexas.edu/keyword?neat

[19] Srinivas, N., Deb, K.: Multi-objective function optimization using non-dominated sorting genetic algorithms. Evolutionary Computation 2(3), 221–248 (1994)

The Preliminary Exploration of Difference Schemes in Three-Order Partial Differential Equation

Lingyan Han[1], Guanghe Cheng[1], Jiyong Xu[1], and Xiaogang Qu[2]

[1] Shandong Computer Science Center, Shandong Provincial Key
Laboratory of Computer Network
[2] Xi'an University of Architecture and Technology

Abstract. Research on two-order partial differential equation tends to be mature. This paper takes initial-boundary problem of dispersive equation for example. I have followed the way of other scholars and created new methods and got two eight-point difference schemes. I have calculated and contracted their truncation errors and analyzed their stability analysis. Based on the unstable difference scheme, the Matlab program is obtained by the algorithm. Although one of the schemes is unstable, it created a new way and solution for the high-order difference equations theoretically.

Keywords: Constitution of the schemes, Three-order partial differential equation, Stability analysis, Dispersive equation.

1 Introduction

In reality word, the relationship among the majority of object is very complicated. If we find the correlation between the variation regularities among these conditions, these forms are partial differential equations. Most partial differential equations' definite solutions are difficult to express as an analytical form due to the multiple boundary conditions and relations. Although we can get the analytic solutions in particular cases, it also can out of work because of the complicated expression.

The numerical solution of partial differential equation is of great use for the situations which can not get the exact solution. While for the situations which the exact solutions are available, it is meaningful in the constitution method of Math. No matter how large and high-speed the computer developed, it is essential leading in the numerical solutions when meet the high dimension problems. The practical application requires higher effective and accurate solutions for partial differential equations.

Presently, the research on two-order partial differential equation tends to be mature. For example, the Richardson scheme is unstable and it can turn into Crank-Nicolson scheme, which is stable. Ritz method can constitute two-order, four-order difference schemes. It is out of work for odd order partial differential equations. By researching the difference scheme on high order especially odd order partial differential equations, this paper can provide offer to higher order and more complicated partial differential equations' schemes.

2   Materials and Experiments

Take this dispersive equation initial and boundary value problem for example:

$$\frac{\partial u}{\partial t} = a\frac{\partial^3 u}{\partial x^3}, \qquad (0 < x < 1, \quad t > 0)$$

$$u\big|_{t=0} = \varphi(x), \qquad (0 \le x \le 1)$$

$$u\big|_{x=0} = f_1(t), \quad \frac{\partial u}{\partial x}\big|_{x=0} = f_2(t), \quad u\big|_{x=1} = g(t) \qquad (t > 0)$$

a is a known constant, $\varphi(x), f_1(t), f_2(t), g(t)$ are known functions.

We assume this problem's solution to exist and it is unique and sufficiently smooth in the solving region. $h = \Delta x, \tau = \Delta t$ are separately space step length and time step length.

Use $x_m = mh$, $t_n = n\tau$, $(m = 0,1, 2,\cdots,m_0, m_0 h = 1; \ n = 0,1,2,\cdots,n_0, n_0\tau = T)$ these parallel lines constituted rectangular grids which cover the whole solving region $(x,t)\big|\{0 \le x \le 1, 0 \le t \le T\}$. Now, we give the difference scheme H-Y,

$$u_{m+1}^{n+1} + u_m^{n+1} = ru_{m+3}^n - ru_{m+2}^n + (1-2r)u_{m+1}^n + (1+2r)u_m^n + ru_{m-1}^n - ru_{m-2}^n$$

Let, $r = \dfrac{\tau}{2h^3}$. The following sketch map is eight grid points.
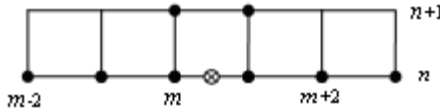
Fig. 1. The eight-point scheme

This scheme uses (x_m, t_{n+1}), (x_{m+1}, t_{n+1}) two points in n+1 layer and $(x_{m-2}, t_n), (x_{m-1}, t_n), (x_m, t_n), (x_{m+1}, t_n), (x_{m+2}, t_n), (x_{m+3}, t_n)$ six points in n layer. This is a new difference scheme and it's unique among all the eight-point difference schemes. The truncation error is $O(\tau + h^2)$ and it can be proved unstable by discrete Fourier transform. As a numerical example, we let

$$\varphi(x) = e^x, \quad f_1(t) = e^t, \quad f_2(t) = e^t, \quad g(t) = e^{1+t}$$

It can be proved easily that the exact solution is $u(x,t) = e^{x+t}$. We divide the interval [0,1] into M portions and [0,T] into N portions equally. Let u_m^n represent the value of m row and n column, $m = 0,1,2,\cdots,M; n = 0,1,2,\cdots,N$.
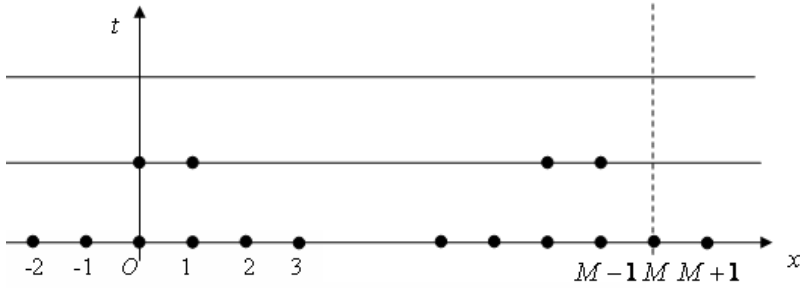
Fig.2 shows the algorithm of this difference scheme.

Fig. 2. The algorithm of H-Y scheme

We utilize initial condition $u(x,0) = e^x$ $(0 \le x \le 1)$ and get $u_m^0 = e^{mh}$; utilize boundary condition $u(0,t) = e^t$, $u(1,t) = e^{1+t}$ $(t > 0)$ and get $u_0^n = e^{n\tau}$, $u_M^n = e^{1+n\tau}$. According to the character of the scheme, before using it, we have to utilize the points outside the grids besides the initial and boundary layers. Based on the initial situation $u(x,0) = e^x$ $(0 \le x \le 1)$,

let $u_{M+1}^0 = e^{(M+1)h}$ and use $\dfrac{\partial u}{\partial x}\big|_{x=0} = e^t$, we can get

$$\frac{u_0^n - u_{-1}^n}{h} = e^{n\tau} \quad u_{-1}^n = u_0^n - h e^{n\tau} = e^{n\tau} - h e^{n\tau} = (1-h)e^{n\tau}$$

According to linear interpolation, we can get

$$u_{-2}^n = 2u_{-1}^n - u_0^n, \quad u_{-2}^n = 2(1-h)e^{n\tau} - e^{n\tau} = (1-2h)e^{n\tau}$$

and $u_{M+1}^n = 2u_M^n - u_{M-1}^n$.

When calculate the value of points from 0 layer to 1 layer, it can utilize this difference scheme from left to right explicitly. Similarly, we can get the value of other layers in sequence.

The constitution of eight-point scheme and the algorithm are based on the research in Xi'an when I am a postgraduate student. Because most of the difference solutions of non-linear partial differential equation is generalization, a series of schemes have appeared on difference solutions of linear dispersive equation recent years and solving its difference schemes have the significance of mathematics theory and practical application. I have followed the way of other scholars and created new methods and got eight-point difference scheme. Some of the references mentioned in this paper are [1] to [9]. The following diagram shows the result of this algorithm by using Matlab.
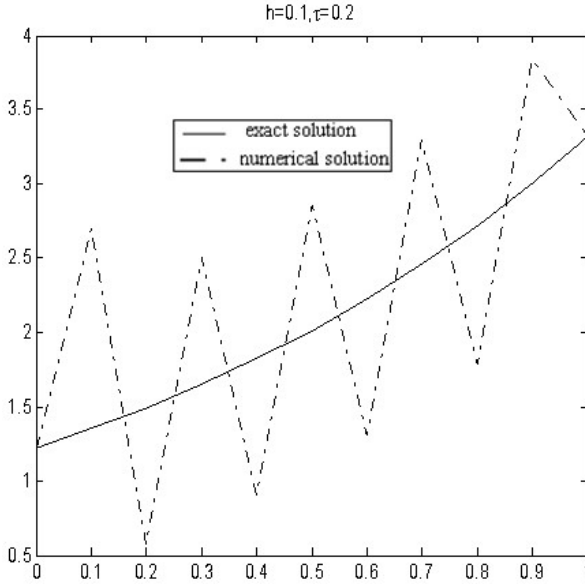
Fig. 3. The calculation of H-Y scheme

This Fig.3 shows $t = 0.2$ the comparison of numerical solution and exact solution. Obviously, the truncation error is hard to control and this match the theoretical analysis.

By further researching on difference schemes, we can get a new eight-point difference scheme H'-Y'.

$$u_{m+1}^n + u_m^n = -ru_{m+3}^{n+1} + ru_{m+2}^{n+1} + (1+2r)u_{m+1}^{n+1} + (1-2r)u_m^{n+1} - ru_{m-1}^{n+1} + ru_{m-2}^{n+1}$$

The new sketch map is equivalent to take x as an axis of symmetry rotate 180 degrees. It can be proved that the new one is absolutely stable. Now, we analysis their stability by using discrete Fourier transform.
Let

$$V_m^n = \lambda^n e^{im\theta}, |\theta| < \pi, i^2 = -1$$

put it in the H-Y scheme,

$$\lambda(\cos\theta + 1 + i\sin\theta) =$$
$$r\cos 3\theta - 2r\cos 2\theta + (1-r)\cos\theta + 1 + 2r + i[r\sin 3\theta + (1-3r)\sin\theta]$$

use trigonometric identity,

$$\cos 3\theta = 4\cos^3\theta - 3\cos\theta$$
$$\sin 3\theta = 3\sin\theta - 4\sin^3\theta$$
$$\cos 2\theta = 1 - 2\sin^2\theta = 2\cos^2\theta - 1$$

get

$$\lambda(\cos\theta+1+i\sin\theta)=4r(1-\cos\theta)\sin^2\theta+1+\cos\theta+i\sin\theta(1-4r\sin^2\theta)$$

$$\cos\theta+1+i\sin\theta\neq0$$

$$\lambda=\frac{4r(1-\cos\theta)\sin^2\theta+1+\cos\theta+i\sin\theta(1-4r\sin^2\theta)}{\cos\theta+1+i\sin\theta}=1-\frac{4ri\sin^3\theta}{\cos\theta+1}$$

while,

$$|\lambda|^2=1+\frac{16r^2\sin^6\theta}{(1+\cos\theta)^2}\geq1$$

So the scheme H-Y is unstable. Similarly, we can analyze the stability of H'-Y'.
    Let

$$V_m^n=\lambda^n e^{im\theta},\ |\theta|<\pi,i^2=-1$$

put it in the H'-Y' scheme,

$$\lambda^{n-1}e^{i(m+1)\theta}+\lambda^{n-1}e^{im\theta}$$

$$=-r\lambda^n e^{i(m+3)\theta}+r\lambda^n e^{i(m+2)\theta}+(1+2r)\lambda^n e^{i(m+1)\theta}$$

$$+(1-2r)\lambda^n e^{im\theta}-r\lambda^n e^{i(m-1)\theta}+r\lambda^n e^{i(m-2)\theta}$$

or

$$\cos\theta+1+i\sin\theta=$$

$$\lambda[-r\cos3\theta+2r\cos2\theta+(1+r)\cos\theta+1-2r+ri(-\sin3\theta+3\sin\theta)+i\sin\theta]$$

Using trigonometric identity, we can get

$$\lambda(\cos\theta+1+i\sin\theta)=\lambda[4r\sin^2\theta(\cos\theta-1+i\sin\theta)+1+\cos\theta+i\sin\theta]$$

$$4r\sin^2\theta(\cos\theta-1+i\sin\theta)+1+\cos\theta+i\sin\theta\neq0$$

$$\lambda=\frac{\cos\theta+1-4ri\sin^3\theta}{\cos\theta+1+16r^2\sin^4\theta(1-\cos\theta)}.$$

$$|\lambda|^2=\frac{1}{1+16r^2(1+\cos\theta)(1-\cos\theta)^3},$$

Because

$$|\theta|<\pi,\quad-1<\cos\theta<1.$$

We can get

$$1+16r^2(1+\cos\theta)(1-\cos\theta)^3>0.$$

While

$$|\lambda|<1$$

The scheme H'-Y' is absolutely stable.

3   Summaries

Every kind of difference schemes have merits and drawbacks: Because of the simple structures and calculations, explicit schemes apparently have calculation superiority, especially for high order equations. Most explicit schemes have strict limitation while the implicit ones have fine stability but can not achieve parallel computation and have a large amount of calculation.

During the process of researching on the difference schemes, we should pay attention to not only the quantitative character but also the qualitative character. For instance, the schemes should be uniform, solvable and convergent. It is better have given order of approximation, maximum ratio of precision and minimum times of operation, etc. Equations of mathematical physics usually use the differential form represent conservation law (energy, heat, quality, charge, etc).

A large number of complex differential equations have practical background. The initial and boundary problem of dispersive equation this paper has mentioned is abstracted from non-linear wave. It is the linear part of particle motion about derivative of time and space. The generalized compound KdV-Burgers equation has different physical meanings according to the different value of parameters. It will be a great helpful when utilize these regularities.

References

[1] Wazwaz, A.-M.: An analytic study on the third-order dispersive partial differential equations. Applied Mathematics and Computation 142, 511–520 (2003)
[2] Djidjeli, K., Twizell, E.H.: Global extrapolations of numerical methods for solving a third-order dispersive partial differential equations. International Journal of Computer Mathematics 41, 81–89 (1991)
[3] Wang, W.: The Alternating segment Crank-Nicolson method for solving convection-diffusion equation with variable coefficient. Appl. Math. Mech. 24(1), 32–42 (2003)
[4] Zhu, S., Zhao, J.: The alternating segment explicit-implicit method for the dispersive equation. Applied Mathematics Letters 14(6), 657–662 (2001)
[5] Mengzhao, Q.: Difference scheme for the dispersive equation. Computing 31(3), 261–267 (1983)
[6] Wazwaz, A.M.: An analytic study on the third-order dispersive partial differential equations. Appl. Math. Comput. 142(2), 511–520 (2003)
[7] Wenping, Z.: Absolute Stable Semi-Explicit Difference Scheme for Solving Dispersive Equation $u_t = au_{xxx}$. Computational Mathematics 3, 248–252 (1988)
[8] Pengcheng, L.: Two Classes of Three-level Explicit Schemes with Higher Stability Properties for a Dispersive Equation $u_t = au_{xxx}$. Journal of Fuzhou University (Natural Science Edition) 1, 6–12 (1990)
[9] Hongen, J., Juhong, G.: Two High Order Accuracy Implicit Difference Scheme for the Dispersive Equation $u_t = au_{xxx}$. China Water Transport 11, 240–241 (2006)

The Design and Implementation of Load Balancing System for Server Cluster

Juanjuan Min, Yi Zeng, Pandong Zhang, and Guangyong Gao

School of Information Science and Technology
Jiujiang University
Jiujiang, China

Abstract. After studying the common load-balancing techniques, a new load balancing way for server cluster is designed in this article. The monitor way of the load information and the estimation method of load factor are used in the design. The design also uses the sliding window skill to solve the problems of the information jam, and the data dispatch is also successful. According to the result of the testing, the system not only makes full use of the equipment, but also provides the better fault tolerance and the high availability for server clusters. In addition, the design improves the performance of the server.

Keywords: Server cluster, Load balancing, Load migration.

1 Introduction

The load-balancing technique uses the advanced multi-layer switching technology, the real-time monitoring of the servers performance and status to distribute the request to the proper server. Different load balancing technologies with different load balancing programs constitute different load balancing systems. A good system should be equipped with flexible and various load-balancing the equalization policy; it can effectively solve the problem of overlarge data flow and network overload; it also can avoid the loss of data flow caused by single point failure of server and it has no need to purchase expensive server with excellent performance.

2 System Design

2.1 System Framework

The load balancing system proposed in this article consists of the following three parts: server cluster, load balancing server, and client side. This design uses a front-end server as a proxy to receive all the client requests[1]. The server cluster is a group of severs respond to the client requests. When the load system has received are beyond more the current capacity of all nodes, you can add new servers in the server pool to meet the growing load requirements. For most Internet services, there is no strong correlation among requests, and requests can be responded in different nodes at the same time, so

the performance of the whole system can basically keep linear growth with the increasing node numbers of server pool, thus it has the good scalability and the fairness.

In order to prevent the failure of the load balancing device causing all servers paralysis, the load balancing system uses a backup load balancing server for mutual supervision. When one of load balancing severs fails, the backup can quickly replace the original one to avoid the loss of a major failure.

2.2   System Functional Modules

Load balancing sever includes three functional modules as show in Fig.1.
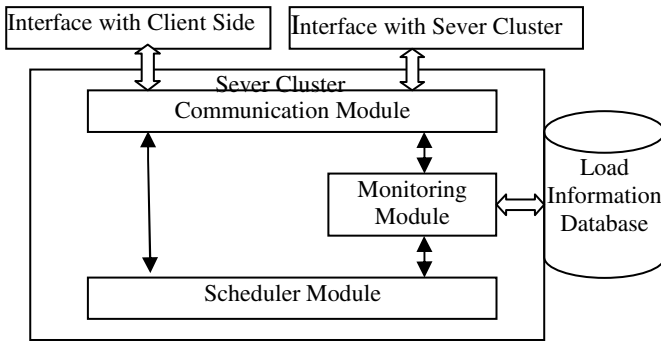
Fig. 1. Structure of system functional modules

Scheduler module communicating with monitoring module, and calculating the latest load information of application server according to scheduling algorithm, and implementing scheduling policy.

Scheduler module is divided into two parts. One is scheduling of request queue, the part is used to select the task in request queue. Usually we use First Input First Output (FIFO) algorithm in which one group of task in request queue is selected for distribution with the help of the sliding window skill. The other part is to select server and distribute requests through load balancing algorithm.

In order to meeting the changing application requirements, the various scheduling algorithm can provide a more diverse range of choices when the server cluster has different application requirements. In the design of this system, we select scheduling algorithms commonly used as load balancing scheduling algorithm, such as weighted round robin algorithm, random algorithm, minimum load algorithm, minimum connection scheduling algorithm and so on.

Monitoring module extracts and stores all load parameters of the application server to prepare for the use when scheduler module implementing scheduling policy.

As to the cycle set of gathering weight of load balancing servers, although short cycle can more accurately reflect the load on each node, frequent collection will be a burden on balancer and may increase unnecessary network load. In addition, since the collector conducts load calculation when collecting, experiment proves that collector shows the intensive jitter of load information on all nodes, causing the balancer can not

accurately capture the real variation trend of load on nodes. Therefore, for solving these problems, on one hand we should adjust the load information collection cycle appropriately, generally 5~10 seconds; on the other hand, we can use moving averages or sliding window to avoid jitter and make the load information collected by balancer display a smooth curve, so that, the adjusting effect of negative feedback mechanism will be better.

Communication module includes receiving module and response module.

Load balancing server interacts with application server cluster and client side through the 3 major functional modules. The client side sends a request to load balancing server through the communication module, load balancing server collects load information of all application server clusters through monitoring module, and then select algorithm and implement scheduling policy by using scheduler module, and finally select one application server to response to client request through communication module.

3   System Implementation

3.1   Load Migration Policy

The load balancing system can not achieve absolute balance whatever scheduling algorithm we use, and there must be overload of some server nodes, so at this time we need to transfer the task of the overloaded node to the light load nodes appropriately. $\varphi(t)$, imbalance factor of estimated time t getting potential benefits through load balancing server, can be defined as difference between the maximum load of server Lmax before balance and the maximum load of server Lbat after balance: $\varphi(t)=$ f (Lmax-Lbat). Suppose $\psi(t)$ is balance cost function, representing the cost of achieving load balancing at the moment t. Balance costs come from three fields, namely, load information transmitting between processors, the decision-making process of selecting migration task and communication delay of task migration. If $\varphi(t) > \psi(t)$, it means load balancing benefits. In the design of the system, the method of sender starting migration has been used when conducting migration task on the overloaded server nodes [2] [3].

With that start method, the sender can randomly or sequentially select the target. Each server has two thresholds, one of which is HWM, the sign showing the server overload, and the other one is LWM showing the server underload. Suppose that the server load can be shown as queue_length of cpu simply. The poll of sender to other processors is random and restricted by poll_limit. When queue_length> HWM, the start algorithm is as follows:

Step1: Set the poll_set empty.
Step2: Select a new server node U randomly, and add it into poll_set.

If queue_length of the node U is smaller than HWM, that is queue_length <HWM, and then transfer current task to U, and the algorithm comes to end.

If node number in poll_set is smaller than poll_limit, that is, node number< poll_limit, and then back to step 2, or back to step 1.
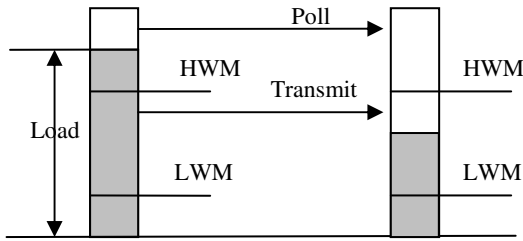
The start algorithm above as shown in Fig.2.

Fig. 2. Sender Start Load Balancing

3.2   Sliding Window Technique

Since many requests may arrive at some point, in order to solve congestion and control problem of flow requesting distribution [4], we apply sliding window technique. According to that method, only tasks within the window can be distributed. The size of the window can be both fixed and changeable. This article chooses the way of change. If there are many arrived requests currently, the overall request execution time is relatively short, and the window value can increase a span; if there are not, the overall execution time is longer, and window value decreases one span.

3.3   MAC Modification Technique

In load migration, the use of MAC modification technique [5] [6] can meet the requirement of circular response, and avoid the complexity process of IP-IP encapsulation. When transferring data packet, server 1 doesn't modify information of IP packet but change MAC address of the head frame. This frame can be transferred to server through LAN 2. After server 2 receiving the frame, if it finds that the IP address is not belong to this machine, the frame should be forwarded under normal circumstances. However, we modify the kernel of server 2 to prevent it from forwarding the packet and to get the IP address of client side from the packet. Then get the MAC address of server 1 from the head frame, and resolve it to IP address of server 1. Load balancers and server cluster are physically connected through unsegmented Local Area Network by using one network card. Server 2 is respond to client requests.

3.4   Estimate Load Index

According to experimental analysis of testing different connection requests, the algorithm defines grade values for each load index at different range including free rate of CPU, free rate of memory, free rate of I / O and so on.Its grade value range is [0,1]. For instance, free rate of CPU is between 50% and 70%, and its grade value is 0.5, but its load index only accounts for 40% of the weight; when the grade value of index value is 0, health values of equipments are all 0, which means that device does not have capacity to provide services for more users.

Formula for calculating the health value of server is:

Health Value = Σ (Grade Value of Index) × Index Weight

By the formula, the better the health status of the equipment, the higher their health value is. In the process of calculating health value, it can be found that if any index

grade value is 0, the health value is 0. If we cannot get one or more health index of the equipment, the device status should be set as -1(unavailable).

Load Value of Server = 1 - Health Value

Specific load information management is to establish a linked list storing load information in load balancing server. We manage the load information through this linked list. The length of linked list is equal to the number of available servers, in which every index load vector contains three load index value including CPU, memory, and I / O. When adding a server, get the latest sample value of this server, then add the element, the new sample value, into the end of linked list of which the length will have an increase of 1.

4   System Tests

The cluster environment is built on the Windows2000 platform. Load balancing server has dual network cards, one of which is used as an external IP address for external request and the other one is used in the communication with servers in cluster. The client side applies multithreading mechanism to simulate connection requests with different intensity.

The operation of the whole system is based on requests. The entire process of one request from starting to getting a response as show in Fig.3.
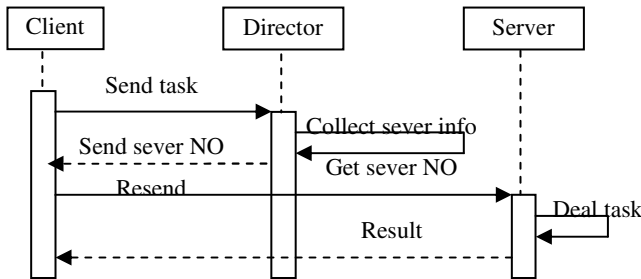
Fig. 3. Sequence Diagram

Above sequence diagram describes the entire process of one request from start to get response. When a request is issued from the client side, it is first send into load balancing server which collect information of servers and give a server number. After getting the server number, the client side relocates to the server; according to their own load circumstances and calculation of balance costs, server decides whether start the load migration policy or not. The selected server or server after migration processes the request and return the results, and then the client side receives the results and displays them.

5   Summaries

The load balancing system designed in this paper can provide better fault tolerance and high availability for server cluster. When the load balancing device fails, there are

good redundancy solutions protecting system from suffering losses of major fault as much as possible. In addition, this design realizes load migration which is helpful to consummate server resources evenly. In that case, it not only improves the resources use rate but also quickly responds to client requests.

References

[1] Herbert Raj, P., Raja Gopalan, S., et al.: Achieving Balanced Traffic Distribution In Mpls Networks. In: Proceedings of 2010 3rd IEEE International Conference on Computer Science and Information Technology, Chengdu, China, pp. 351–354 (July 2010)

[2] Marco, C., Ivan, C.B.: Virtual machines for distributed real-time systems. Computer Standards & Interfaces 31(1), 30–39 (2009)

[3] Moloney, B., Weiskopf, D., Moller, T., et al.: Scalable sort-first parallel direct volum rendering with dynamic load balancing. In: Favre, J.M., Santos, L.P.D., Reiners, D. (eds.) Eurographics Symposium on Parallel Graphics and Visualization, pp. 45–52. A K Perters Ltd., USA (2007)

[4] Sahoo, B., Mohapatra, S., Jena, S.K.: A genetic algorithm based dynamic load balancing scheme for heterogeneous distributed systems. In: PDPTA 2008, United states, pp. 499–505 (2008)

[5] Li, Y.F., Xu, X.H.: Load migration-based resource scheduling mechanism in virtual machine. Huazhong University of Natural Science Edition 37(9), 47–48 (2009)

[6] Renaud, S., Jacques, C., Hervé, K., et al.: On a resource-constrained scheduling problem with application to distributed systems reconfiguration. European Journal of Operational Research 183(2), 546–563 (2007)

A Improved SML Method for Semantic Auto-annotation of Image

Ouyang Jun-Lin, Zhu Geng-Ming, Wen Xing-Zi, and Zhang Shao-Bo

Hunan University of Science and Technology, Xiang Tan, China, 411201
{Yangjunlin0732,coconut168}@163.com, abcdechowxz@sohu.com

Abstract. Semantic auto-annotation of Image becomes focus in image retrieval Based on content. The two step semantic auto-annotation of image method is proposed. Firstly, a supervised multi-class labeling method (SML) is adopted to coarse annotation for image, then an optimal semantic tag annotation based on Ontology method (OOSTIA) is employed to fine annotate. There are four ways to extend coarse annotation result in the OOSTIA method, which can fully mine ample semantic information in images. The proposed method is compared to others, Experiments result show that the proposed method outperforms others.

Keywords: SML, Ontology, Semantic annotation, Image retrieve.

1 Introduction

Image sematic auto-annotation becomes a focus for image retrieve based on content. Automated image annotation can be divided with respect to the deployed machine learning method into co-courrence models[3], machine translation models[4], classification approaches[1,2], grapic models[7], latent space approaches[5,6] and relevance models[8] ect.

A surpervised multiclass labling(SML)[2] methoed based on without segmentation is proposed. The basic idea is that image is decomposed into a set of overlapping 8*8 regions, extracted with a sliding window that moves by two pixels between consecutive sample, images are simply represented as bags of lacalized feature vectors, a mixture density estimated for each image, and the mixtures(associated with all images annotated) with a common semantic label pooled into a density estimate for the corresponding semantic class. Semantic annotation and retieval are then implemented with a minimum probablitity of error rule, based on these class densities. This kind of method is conceptually simple, computationally efficient, and do not mixture density estimated for each image.
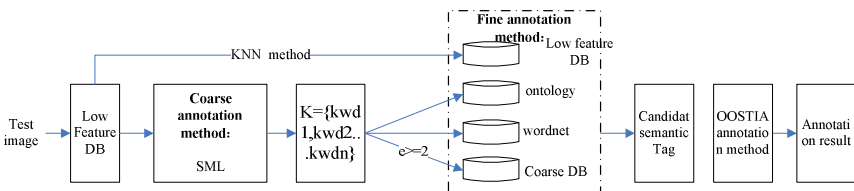

Fig. 1. Two Step Image Semantic Auto-annotation Structure Graphic

By analysing this kind of method, though which get a good efficent in obtaining mostly keywords, a image seems to have thoughts sematic, simple several keywords don't completely describe a image. For example, a image is annotated including {sea, sun, sand}, in fact, mostly semantic isn't annotated such as {ocaan, sunup, sand beach, landscape, scenery, sky, cloud…}. Therefore, a two levels image annotation method is proposed by combining ontology and SML.

2   Surperised Multiclass Labeling (SML) Annotation Methed

2.1   Modling Class without Segmetion

Consider a database T={$I_1,…I_N$} of images I_i and a semantic vocabulary L={$w_1,…w_T$} of semantic labels w_i. It is supposed that a set of class-conditional distributions Pxlw(xli), $i \in \{1, \cdots, T\}$ for the distribution of visual features given the semantic class, it is not difficult to show that both labeling and retrieval can be implemented with a minimum probability of error if the posterior probabilities are available,

$$P_{W|X}(i \mid x) = \frac{P_{X|W}(x \mid i)P_W(i)}{P_X(x)} \tag{1}$$

where Pw(i) is the prior probablity of the ith semantic class. In particular ,given a set of feature vectors X extracted from a (previously unseen) test image I, the minimum probality of an error label for that image is

$$i*(\chi) = \arg \max_i P_{W|X}(i \mid \chi) \tag{2}$$

2.2   Density Estimation

One efficient alternative to the complexity of model averaging is to adopt a heierarchical density estimation method. This method is based on a mixture hierarchy where children densities consist of different combinations of subsets of the parents' components. In the semantic class densities are their parents. it is possible to estimate the parameters of class mixtures directly from those available for the individual image mixtures, using a two stage procedure. The first stage is the naïve averaging. assuming that each image misture has K components, this leads to a class mixture of DiK components with parameters

$$\left\{\pi_j^k, \mu_j^k, \Sigma_j^k\right\}, j = 1,...Di, k = 1,..., K \tag{3}$$

The second is an extension of EM which clusters the Gaussian components into an M-component mixture, where M is the number of components desired at the class level. Denoting by $\left\{\pi_c^m, \mu_c^m, \Sigma_c^m\right\}, m = 1,..., M$ the parameters of the class mixture, the formula can be found in paper [2].

2.3   Algorithm Description

In this section, we describe the two algorithms used in this work, namely, training and annotation. For the training algorithm, we assume a training set D={(I_1,w_1),…, (I_D,w_D)} of image-caption pairs, where $I_i \in T_D$ with $T_{D=}${I_1,…, I_D},and $w_i \in L$,with L={ $w_{1,…}w_T$}.The steps of training algorithm are:

For each semantic class $w \in L$,

a.   Build a training image set $\bar{T}_D \subset T_D$,where $w \in w_i$ for all $I_i \in \bar{T}_D$,

b.   For each image $I_i \in \bar{T}_D$

    i.   Decompose I into a set of overlapping 8*8regions, extracted with a sliding window that moves by two pixels between consecutive samples (note that, in all experiments reported in this work, images were represented in the YBR color space)

    ii.   Compute a feature vector , at each location of the tree YBR color channels, by the application of the discrete cosine transform(DCT).Let the image be represented by

$$B = \left\{ \left[x^Y, x^B, x^R \right]_1, \left[x^Y, x^B, x^R \right]_2, ..., \left[x^Y, x^B, x^R \right]_M \right\}$$

where $\left[x^Y, x^B, x^R \right]_m$ is the concatenation of DCT vectors extracted from each of the YBR color channels at image location $m \in${1,…,M}.Note that the 192-dimensional YBR-DCT vectors are concatenated by interleaving the values of the YBR feature components. This facilitates the application of dimensionality reduction techniques due to the well-known energy compaction properties of the DCT. To simplify notation, we hereafter replace $\left[x^Y, x^B, x^R \right]_m$ with x.

    iii.   Assuming that the feature vectors extracted from the regions of Image I are sampled independently,find the mixture of eight Gaussians that maximizes their likelihood using the EM algorithm.This produces the following class conditional distribution for each image:

$$P_{X|W}(x | I) = \sum_{k=1}^{8} \pi_I^k G(x, \mu_I^k, \sum{}_I^k) \tag{4}$$

where π_I^k 、 μ_I^k 、 $\sum{}_I^k$ are the maximum like-lihood parameters for image I and mixture component k.

c.   Fit Gaussian mixture of 64 components by applying the hierarchical EM algorithm of [17-20]to the image-level mixtures of [21].this leads to a conditional distribution for class w of

$$P_{X|W}(x \mid w) = \sum_{k=1}^{64} \pi_w^k G(x, \mu_w^k, \sum_{w}^{k})$$

(5)

The annotation algorithm processes test images It $I_t \notin T_D$, executing the following steps:

1. step(b-i) of the training algorithm
2. setp(b-ii) of the training algorithm.
3. For each class wi $\in$ L,compute

$$\log P_{W|X}(w_i \mid B) = \log P_{X|W}(B \mid w_i) + \log P_W(w_i) - \log P_X(B)$$

(6)

Where B is the set of DCT features extracted form image I_t,

$$\log P_{X|W}(B \mid w_i) = \sum_{x \in B} \log P_{X|W}(x \mid w_i)$$

(7)

$P_W(w_i)$ is computed from the training set as the proportion of images containing annotation wi ,and $P_X(B)$ is a constant in the computation above across different wi $\in$ L.

4. Annotate the test image with the five classes wi of largest posterior probablility , $\log P_{W|X}(w_i \mid B)$.

3   Semantic Image Annotation Method Based on ontology

By using the result above method to coause annotation for image, in this section,a optimal semantic tag annotation based on Ontology method (OOSTIA) is employed to fine annotate. A ontology dimension D_h, h=1,…,M, is a tree-structured, also called semantic tags. More precisely, a semantic tag st_j is a path in D_h , $st_j = n_0 / n_1 .../ n_k \in D_h$, where each n_i is a node of the ontology. Each image I_i can be concisely represented as I_i ={ P_i , F_i ,K_i ,ST_i }, P_i is represented to image , F_i is represented to image feature, K_i is represented to keyword based on SML annotation. ST_i is represented to semantic tag of image. In this course of fine annotation include two steps: firstly, produces candidate semantic tag, secondly, organizes the candidates into a candidate tree CT, ranks them, and returns the top-m ones.

3.1   Candidate Semantic Tag Predicting Method

For any couse annotation images, which need to fine annotation, defines relevance keyword variable RK and candidate semantic tag variable CST.There are four ways to produce candidate semantic tag, such as figure 2.
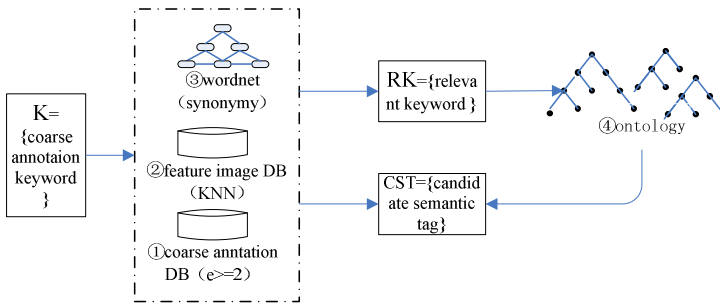
Fig. 2. Four expanded methods of candidate semantic tag

(1) we exploit query keywords K by applying a co-occurrence search on DB image keywords. The search provides a set of images that share at least e terms with K. We rank the iamges on the base of the co-occurrence value and, for the top-p images only, their keywords are added to a set RK of relevant keywords,and all the semantic tags are used to initialize CST. For example, if K={beach , sea , sun}, e=2. and there is an image Ii with K i ={beach, sea, sky}, and STi ={landscape/water/sea}, then sky is added to RK and landscape/water/sea to CST. （2）according to low feature of image F,seach relevant image by using similar degree measure method in the DB , the top-g images which contained at least one keyword is used as candidate image, and the keyword and semantic tag is added to the RK and CST respectively.(3) Expands all relevant keyword RK according to wordnet. For example, if RK has {sea}, synonymous {ocean} is added to RK, and keywords and semantic tag which are contained in the image DB are added to RK and CST respectively.(4) At last, all RK are expanded by domain ontology, and all path contained RK are added to CST.

3.2   Candidate Semantic Tag Annotation Method

The candidate semantic tags CST are bulit into a candidate tree, and then computes the overall top-m as annotation results. Ranking is based on weghts. The weight w_j of st_j is computed as $w_{j=}$ $freq_j * util_j$, where $freq_j$ is the recency of st_j and $util_j$ is so-called utility of st_j wrap all other candidates st_i, defined as:

$$util_j = \sum_{st_i \in ST_h, i \neq j} \frac{len(st_j \cap st_i)}{\max P_h} \tag{8}$$

Where $len(st_j \cap st_i)$ is the length of the common path between st_j and st_i, where $\max P_h$ is the maximum path length within the dimension D_h. Uility measures the amount of overlap between st_j and all other st_i's.

4   Experiment Result

Our experiments is implemented in the environment of Intel-p4 2.8G CPU , XP system. Visual C++6.0 is used as development environment, image DB uses the popular test image DB Corel5K,which contained 5000 images and 50 class, very class contain 100 images. We use 20 image in very class as test example. We compare with four methods: BMRM method [8], information bottleneck cluster annotation method (IBCA)[1], SML[2] and our proposed method (SML-OOSTIA), the annotation result is listed as FIG 3.

Method \ Image				
BMRM	Elephant tree grass	House grass	Flower leaf rose	Sky sea tree
IBCA	Elephant tree grass	House tree grass	Flower leaf rose	Sky sea tree landscape
SML	Elephant tree grass	House tree grass	Flower leaf rose	Sky sea tree hill landscape
SML-OOSTIA	Elephant animal plant tree grass Branch plain forest landscape	House tree animal plant leaf branch grass plain landscape	Flower leaf rose plant tree branch landscape	Sky universe cloud sea ocean water plant tree branch hill mountain landscape

Fig. 3. Annotation result for four different methods

Fig.4 describe   the time performance for the above method. The time complexity can be described by O(DR), where D represent the number of train image, R represent feature of image. the time complexity of SML can be represented by O(TR), where T represent the number of the semantic class of train image. Our method need more time than SML ,but our method is better than MBRM and IBCA.
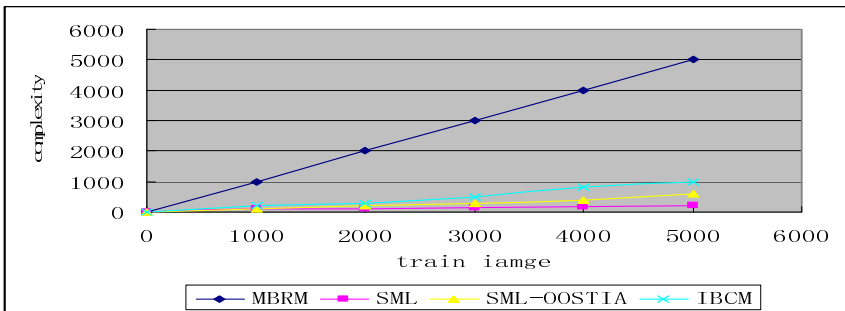

Fig. 4. Performance compared for four different methods

5   Conclusions

A two levels image semantic annotation method is proposed in this paper. Our method compares with other three methods, experiment result show that our method is better than other methods in the whole performance aspect. In the next step, we continue to study the fine annotation methods.

References

[1] Xia, L.-M., Tan, L.-Q., Zhong, H.: Semantic Annotations of Image Based on Information bottleneck method. Pattern Recognition and Artificial Intelligence 21(6), 1199–1205 (2008)

[2] Carneiro, G., Chan, A.B., Moreno, P.J.: Supervised Learning of Semantic Classes for Image Annotation and Retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(3), 394–410 (2007)

[3] Mori, Y., Takahashi, H., Oka, R.: Image-to-word transformation based on dividing and vector quantizing images with word. In: International Workshop on Multimedia Intelligent Storage and Retrieval Management (MISRM) (1999)

[4] Duygulu, P., Barnard, K., de Freitas, J.F.G., Forsyth, D.: Object Recognition as Machine Translation: Learning a Lexicon for a Fixed Image Vocabulary. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) ECCV 2002, Part IV. LNCS, vol. 2353, pp. 97–112. Springer, Heidelberg (2002)

[5] Monay, F., Gatica-perez, D.: On image auto-annotation with latent space models. In: Proceedings of the 11th International ACM Conference on Multimedia, pp. 275–278. ACM, New York (2003)

[6] Monay, F., Gatica-perez, D.: Plsa-based image auto-annotation:constraining with latent space models. In: Proceedings of the 12th International ACM Conference on Multimedia, pp. 348–351. ACM, New York (2004)

[7] Liu, J., Li, M., Liu, Q., Lu, H., Ma, S.: Image annotation via graph learning. Pattern Recognition 42, 218–228 (2009)

[8] Feng, S.L., Manmatha, R., Lavrenko, V.: Multiple Bernoulli relevance models for image and video annotation. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1002–1009 (2004)

Preprocessing Algorithm of Hand Vein Image

Zang Ping[1] and Liu Cai Xia[2]

[1] Sandong Transport Vocational College
Weifang, China
`kikizang@tom.com`
[2] Zaozhuang University
Zaozhuang, China
`cxsqz@126.com`

Abstract. Characteristics of hand vein can be used in identity recognition. The key to carry out this subject is how to effectively extract the skeleton of the hand vein texture. In this paper, we first process the image of the hand vein by limiting its histogram for contrast enhancement, and use the method of wavelet transformation to denoise the image which has been processed. In the basic texture segmentation of the hand vein we present a separate Niblack adaptive threshold method of image and get a good texture segmentation.

Keywords: Vein recognition, Histogram equalization, Image segmentation, wavelet transformation.

1   Introduction

In recent years, biometrics technology is rapidly around the world .It is an important security technology of the computer. This technology is based on face, fingerprint, voice, vein and other human biological characteristics, using image processing and pattern recognition to identify or verify the identity. A number of experiments show that[1] biometrics is the only, and not the same with other people and it is automatically recognized and measurement and verification of physical characteristics or behavior. As biometric fingerprint recognition technology in the field of early start, and with the fingerprint identification system out of the financial, government, airports and other critical industries, more and more appear in the intelligent building, PC computers and other civilian areas of the enterprise, making the fingerprint become the most widely recognized, market share, the largest class of biometric technology. However, there are many drawbacks of fingerprint recognition technology[2], the researchers developed a new vein recognition system, which is currently the world's most advanced use of biometric identity authentication of a technology. The application of the system is broad because it has high recognition rate, low rate of rejection really easy to use, recognition speed, little interference by the environment and a series of advantages. The back of the hand vascular biological characteristics of the human body all the features and characteristics of these creatures in terms of anti-counterfeit and anti-jamming compared to fingerprints, palm prints and other biometric features and advantages are obvious. Typically, the collected image intensity difference is very small veins. Most single-domain and multi-threshold segmentation method are carried

out directly in the acquisition of binary image processing, this will result in some loss of information available background information and even some of the miscarriage of justice. In this regard, this paper proposes a new preprocessing algorithm of Hand Vein Image. It enhances image contrast by histogram equalization while denoising by wavelet denoising techniques. Finally, we use the adaptive threshold method of Niblack in hand vein image segmentation, and have been satisfied with the treatment effect. Ihis has laid the foundation for further feature extraction and vein recognition.

2   Vein Enhancement

Let the gray level transformation s = f (r) is the slope limited by continuous non-differentiable function, that will convert the input image A (x, y) into output image B (x, y). Histogram of the input image is HA (r), and output image histogram is HB (s), then their relationship is shown in the following equation

$$HB\,(s)\,ds\,=\,HA\,(r)\,dr \tag{1}$$

and in the discrete case:

$$f\,(r) = \sum_{i=0}^{k} P\,(r_i) = \sum_{i=0}^{k} \frac{n_i}{n} \tag{2}$$

Where n is the sum of pixels in the image, ni ri is the pixel gray level as the number of pixels, k is to image the total number of possible gray level.

2.1   Global Histogram Equalization

Global histogram equalization algorithm is simple, efficient, and the overall enhancement of the image display better. However, due to the image intensity distribution in the different regions so different from using the same transformation is often difficult to identify the details, and sometimes the noise introduced by the original message to destroy. Fig.1 (a) is the dorsal vein of the original image, Fig.1 (b) the original image histogram. Enhanced by global histogram equalization algorithm processed image in Fig.1 (c).However, intravenous details are still not well enhanced.

2.2   Local Histogram Equalization Limited in Contrast

According to human visual characteristics, on the one hand, the relevant regions of the visual system changes with the change; the other hand, it also affected the surrounding areas related to the environment. R.Cromartie has proposed a constrained local histogram constructor (CLAHE), which is a constrained local histogram considering histograms both in and out of the window. So in this method the histogram is composed of two parts [6,7].

$$h_L(r) = \alpha h_\omega(r) + (1-\alpha)h_B(r) \tag{3}$$

Where $h_\omega(r)$ is the window of the normalized histogram, $h_B(r)$ is the normalized histogram window, and there are $0 \le \alpha \le 1$. Let A_ω and A_B represent the region B,

W and regional area, if $\alpha = A_\omega /(A_\omega + A_B)$, then $h_L(r)=h(r)$. This means that the local histogram is equivalent to a global histogram. If $\alpha > A_\omega /(A_\omega + A_B)$, then the local histogram emphasizes local information. In this case, the local histogram can be adjusted to simulate the environment of the relevant region. If we use this method to construct the local histogram, then all of the local histogram will have the same way as the global histogram of gray level, but each gray level in a different location has different amplitude. After the global histogram equalization enhancement treatment (Fig.1(e)), the veins on the back of the hand was well enhanced.

The various noises reduce the image quality in the image generation and transmission process. Spatial or frequency domain filtering to suppress image noise and improve image SNR. In order to filter the image noise, many methods have been proposed, in which median filtering is a local smoothing technique[8], its pulse interference and impulse noise suppression have a good effect, but also in the noise suppression but can also make the edge of the area to reduce the blurring phenomenon. The results are shown in Fig.1 (g) below.

(a) the original image   (b) histogram of (a)   (c) Histogram equalization

(d) Histogram of (c)   (e) CLAHE   (f) Histogram of (e)

(g) Median filtering result
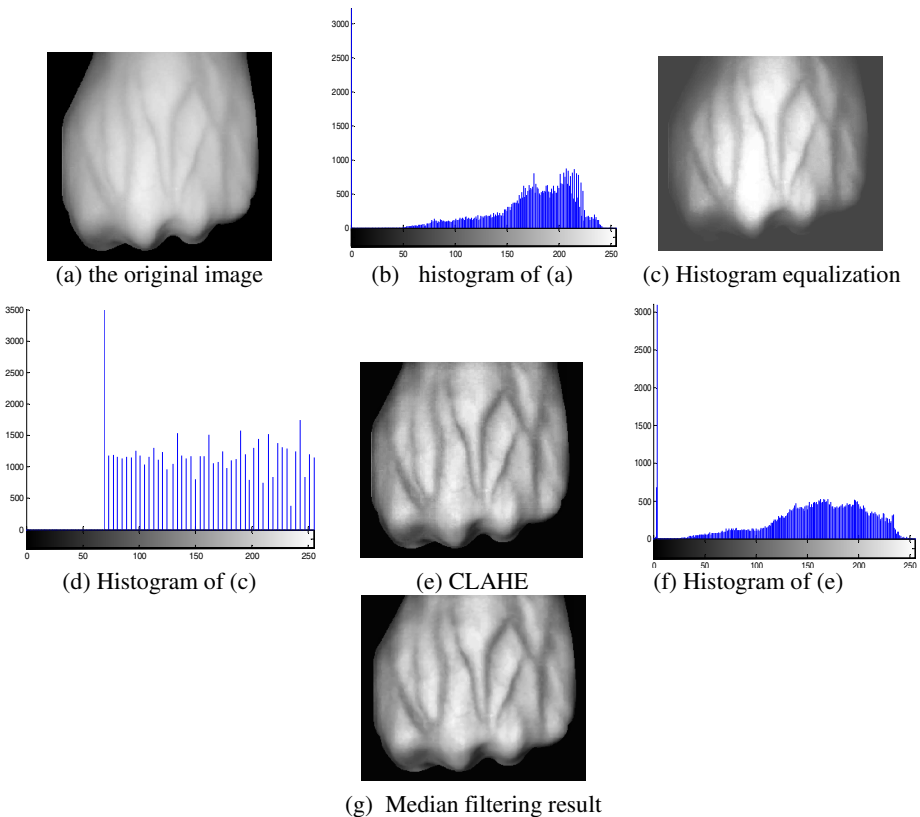
Fig. 1.

3   Wavelet Denoising

Wavelet threshold denoising method is a simple and effective method to achieve better effect in denoising image. It first processed coefficients in layers of wavelet decomposition. The modulus of the coefficients are greater than or smaller than a certain threshold coefficients. Then we used the wavelet coefficients that had been processed to reconstruct the image. The threshold function reflects the wavelet coefficients of different treatment strategies and different valuation methods, the commonly used threshold function is hard or soft threshold function [6].

First , we conducted a three-layer wavelet decomposition of the original image., and in this article we used sym4 as the the wavelet. Because most of the images have noise in their high-frequency, and they have more isolated noise points, thus we can directly use the hard thresholding method to process the image. When the high frequency coefficients are below the set threshold, the coefficient is set to 0, Otherwise, the coefficient is kept. This method can preserve the edge and other local features of the image. In this paper, we use a unified threshold which has been proposed by Donoho in[9]. We adopt the following equation to select the threshold.

$$T = \sqrt{2\ln(N)}\sigma \tag{4}$$

Where N represents the total number of pixels in the image; σ is the noise standard deviation. The result of the image histogram equalization is shown in Fig.2
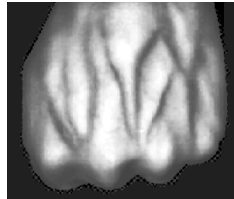
Fig. 2. Result of the image histogram equalization

4   Image Segmentation

The key technology in image segmentation is the threshold selection, because the value of the threshold can directly affect image segmentation results and follow-up treatment. To get a more accurate and reliable threshold, many scholars have conducted extensive and in-depth studies [10]. But they mainly took a single attribute of the image as the standard to determine the threshold, which is difficult to meet the requirement of the image segmentation of multi-attribute. In addition, because of uneven illumination and other confounding ambiguity, resulting in the threshold image segmentation results are usually unsatisfactory.

Niblack is a simple and effective local dynamic threshold algorithm. The basic idea of this algorithm is that in each point of the image in its r × r neighborhood,. the mean and variance of pixels are Calculated. Binarization is operated by the following equation

$$T(x, y) = m(x, y) + k * s(x, y) \tag{5}$$

Where for each pixel (x, y), T (x, y) is the threshold for the point, m (x, y) is the average for points of the r × r neighborhood pixels, s (x, y) is the standard deviation for points of   the r × r neighborhood pixels, k is the correction factor. Assuming that pixel (i, j) at the gray value is f (i, j), then m (x, y) and s (x, y), respectively, by equation (5) and equation (6) are calculated.

$$m(x, y) = \frac{1}{r^2} \sum_{i=x-r/2}^{x+r/2} \sum_{j=y-r/2}^{y+r/2} f(i, j) \tag{6}$$

$$s(x, y) = \sqrt{\frac{1}{r^2} \sum_{i=x-r/2}^{x+r/2} \sum_{j=y-r/2}^{y+r/2} f(i, j)^2} \tag{7}$$

Here the correction factor k generally determined by the empirical value. It can be seen in Figure 3, when the k value is -0.25, the results are good.

| k=0.4 | k=-0.25 | k=-0.1 | k=0.1 |
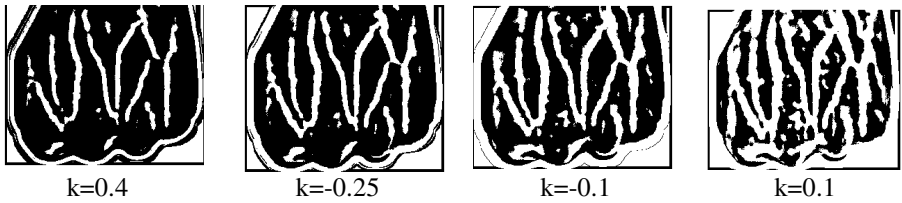
Fig. 3. Results of direct Niblack of Fig.2

Calculated in each window, we make that

$$ni(i, j) = (f(i, j) - \kappa) * (f(i, j) - \min) \tag{8}$$

$$nj(i, j) = (\max - \min + 1)^2 \tag{9}$$

$$kw = m(x, y) / (s(x, y) + 1) \tag{10}$$

$$k = \begin{cases} -\log(kw) * 0.1 & kw > 1 \\ \log(kw) & kw <= 1 \end{cases} \tag{11}$$

| w=11 | w=15 | w=27 |
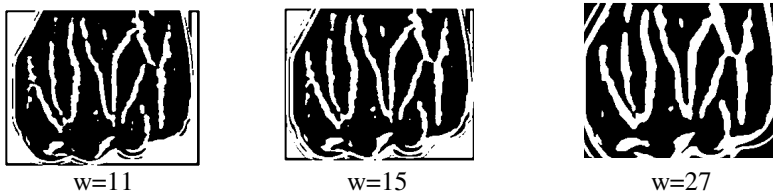
Fig. 4. The use of adaptive methods for image segmentation Niblack

Where κ is the average of the gray value of pixels within the window, min for the gray value of pixels within the window the minimum, max is the maximum value. Figure 4 shows that, whatever the size of the template, our segmentation results generated less noise and better segmentation. In comparison, 15 * 15 template more in line with demand.

5   Conclusions

Because the image contrast of back of the hand is low and there is noise, we first carried out two image enhancement methods of histogram equalization and contrast limited histogram equalization. By comparison, contrast limited histogram equalization method can more effectively highlight the vein detail more in line with our needs. To solve the noise problem, we introduced a denoising based on wavelet transform technology, enhance image contrast while reducing noise, highlighting the image detail. Finally, we proposed based on Niblack's adaptive thresholding method, image segmentation, has been a clear hand vein patterns.

References

[1] Miller, B.: Vital signs of identity. IEEE Spectrum 31(2), 22–33 (1994)
[2] Luo, D.-S.: Authentication in information age. Science and Technology Press, Beijing (2004)
[3] Gonzalez, R.C., Woods, R.E.: Digital image processing, 2nd edn., pp. 70–74, 482–486. Publishing Housing of Electronics Industry (2003)
[4] Lin, X.-R., Zhuang, B., Su, X.-S.: Human dorsal vascular image feature extraction and matching. Journal of Tsinghua University 43(2), 164–167 (2003)
[5] Wang, K.-J., Ding, Y.-H., Zhuang, D.-Y.: Threshold value Segmentation of Hand Vein Image. Control Theory and Applications 24(8), 19–22 (2005)
[6] Jin, C.-W., Guo, Y.-M., Qiang, Y.-Q.: Preliminary study of Image enhancement based on histogram equalization. Imaging Technology 22(4), 466–469 (2006)
[7] Xu, L.-P., Cai, Y.-Y.: Pneumoconiosis x-ray enhancements based on CLAHE. Computer Applications, 388–389 (2007)
[8] Shen, T.-Z., Fang, Z.-W.: Digital image processing and pattern recognition. Beijing Institute of Technology Press (1998)
[9] Hou, J.-H.: Image Denoising based on wavelet and statistical properties. Huazhong University of Science (2007)
[10] Jiang, H.-Y., Pan, D.-L., Xia, D.-S.: Rapid implementation of Two-dimensional Adaptive thresholding algorithm of otsu. Automation Journal 33(9), 968–971 (2007)

A Faster Than Quicksort of Mutation Algorithm

Xiuli Wei and Furong Ji

Shandong Transport Vocational College
Weifang, China
sdjzywxl@foxmail.com,
ejifurong@163.com

Abstract. Sorting is a very important research subject of computer science. Quicksorting is a key word more based on a kind of performance which is good sequencing algorithm, by widespread application in the theoretical teaching and the algorithm design. In this paper the traditional quicksort is analyzed, puts forward a new kind of quicksorting method, and the analysis of the algorithm.

Keywords: Sort, Quicksort, Time complexity, Algorithm.

1 Introduction

In data processing, sorting is a kind of use higher frequency arithmetic. In order to facilitate the search, its function is to a data elements (or record), to form an arbitrary sequence by keyword orderly sequence. Due to the number of records to sort is different, make the sorting process involves memory can be different, sorting method divided into two kinds: one is the internal sorting, refers to stay sorting records stored in computer random access memory for sorting process; Another kind is the external sorting, refers to the number of records to sort that memory a very full record, can hold in sorting process is needed to save the foreign visit sorting process[1]. This article mainly discuss the quicksort internal sort. Quicksort is improved, its bubble sort through a exchange can eliminate multiple reverse, so can reduce the consumption of scanning in reverse and data exchange frequency. Based on the comparison of the algorithm, the performance of the quicksort in large quantities, best the sorting process data. It often hitches Optimal situation, it's time for O (nlogn) complexity. But there are two quicksort insufficient. First,When it is sort of data basic orderly, to row sequence data value equal or existence outlier data circumstances, its efficiency below or far less than O (nlogn), speed, but most slow, the time complexity even to O (n2); the second, it is to sort results not stable. Sorting algorithm is stable, so if from a key word order, and then from another key word order, the first key word the sort of the results can be used for the second key word order, and to reduce the time and space complexity. Radix sort is like this, according to the first order according to the high, low, low successive sort of the same elements the order to high same also will not be changed. In addition, if sort algorithm based on the comparison of the stable, sort algorithm, the number of elements in exchange will be less.

2   Quicksort Algorithm

2.1   Basic Ideas

Quicksort is a kind of bubble sort of improved algorithm. The basic idea is that by a trip to sort will stay row records split into two parts, the independent part of the key words are better than record to another part of the key word, small record can be the two parts respectively on record to continue to achieve the entire sequence sort order[4].

   To sort of sequence S from [1... p], the first freely choose a record (generally choose the first record), and then press the following principles fulcrum rearranged the rest records: will all keyword its smaller records in front of it all keyword positions, bigger than it is in its record of behind position. Thus the "pivot" record the position of the final fall line, I make will sequence S [1... p] split into two disorderly son sequence, S [1... (I - 1)] and S [(I + 1)... p]. This process is called a trip to quicksort. Then the two sons were the sort of the above sequence of process, until finally all elements sequenced so far.

2.2   Algorithm Described

```
int partition(Sqlist &S, int low,int high)
{// S for stay row sequence, ancillary two pointer, high low
// Comparison and exchange low and high Pointers, make
//the element position than fulcrum small element in front of protection
placements, than in element are fulcrum behind the protection placements
pivotkey=S[low].key
// The first record in table do key point
while(low<high)
// From the table to the ends of the alternately among scans
  {
while(low<high&&S[high].key>=pivotkey)
--high;
S[low]<-->S[high];
  // Small record than the pivotkey to low-end exchange
while(low<high&&S[low].key<=pivotkey)
S[low]<-->S[high];
// The record than the pivotkey big switch to high-end
  }
return low;
  // Return to pivot location
  }
void quicksort(Sqlist &S,int low,int high)
{// quicksort of recursion algorithm
if(low<high)
  // Sorting table length is more than 1
  {
pivotloc=partition(S,low,high);
// Divided into two sons sequence
  quicksort(S,low,pivotloc-1);
// The    lower table of recursion sequence
```

```
    quicksort(S,pivotloc+1,high);
    // The higher table of recursion sequence
  }
  }
```

2.3   Algorithm analysis

The so-called stable sorting, is equal before the two Numbers, sort order, after what is also what order sequence. For example a = 1, b = 3, 2 this number, a ranking in the b front, and if originally can ensure sorting, a or b ahead, is stable sorting, otherwise it is not stable sort. Quicksort is an unstable sorting method. Quicksort has two directions, the left has been low subscript right away, when S [low] < = S [pivotloc], which picvotloc is the fulcrum of the array subscript, general elements off for array element 0 a. And right to left was subscript always, when a > a [[high] pivotloc]. If the low and high all can't walk anymore, low < = high [low], exchange a [was], and a repeat this process until low > high. Exchange was] and [a]], [pivotloc a complete a trip to quicksort. In the fulcrum elements and a [was] exchange, it is likely the stability of the elements, such as sorting listed as to disrupt 39,29,56,97,68,08,15) now pivot element (39 and 15 exchange will be the stability of the elements, so disrupted 15 quicksort of an unstable sorting algorithms, instability occurs in the fulcrum elements and a [high] exchange moments.

For example: stay sorting the value of the array S were: (initial fulcrum elements: X = 39) note pivot element X forever, forever is compared with X, no matter what the purpose of the seat, finally the X in the middle, is small put behind the front, big put.

S[0] 、 S[1]、 S[2]、 S[3]、 S[4]、 S[5]、 S[6]：
39   29   56   97   68   08   15
For the first time after exchange: 15 29 56 97 68 08 39
(begin from behind in accordance with the algorithm for)
On the second after exchange: 15 29 39 97 68 08 56
(according to algorithm from the front started looking for > X values, 56 > 39, both exchange, right now: low = 3)
The third after exchange: 15 29 08 39 97 68 56
(according to algorithm and from later started looking for)
For the fourth after exchange: 15 29 08 39 68 97 56
(from front and in accordance with the algorithm of value larger than X started looking > 39, both, 97, when: low = exchange was = 4, 6)
Now and then execute algorithm low = just found was, ending a trip to quicksorting, then after a trip to quicksort the result is: after 15 29 08 39 68 56, namely so more than 97 several all in 39 39 behind the number, so less than 39 in front of all 39.

Quicksorting is a recursive call this process - in 39 for midpoint segmentation this data sequence respectively, the front part of the part and behind the quicksort of similar, thus finish all the data sequence of quicksort, finally put this data sequence into a orderly sequence, according to this for the idea of the above array S the whole process of quicksort shown below:

Initial condition (39 29 56 97 68 08 15)
After a quicksort into (15 29 08 39 (68) 97 56)
In two parts respectively before quicksort:

After the second quicksort(15 29 08)changed into (08 15 29) sorting.
After the second quicksort(68 97 56)changed into (56 68 97) sorting.
Of the above process, we can generally see this algorithm in Eric's.

First, in the second exchange was not found before than 39 big element, need to constantly moved, high --;

Second, when low after exchange referred to ensure elements referred to moment low elements must be less than 39, namely low pointer, and may continue to stay in place, can't move. So, this will cost many stay time.

Quicksorting algorithm for the average time complexity O (nlogn). Because two pointer, low was scanning again array, spending generated is O (n), best case scanning, the worst logn times next scanning n times. Also can use mathematics knowledge proof, quicksort for the average time complexity O (nlogn). Based on the comparison of the sorting, finally are never escape the fate of O (nlogn). We consider quicksorting process: random selection of an element do "pivot element", will be all the more than pivot element move to the left, the rest of the moves to the right. According to this process, quicksort the first more is the one element and fulcrum elements compare, at this time, "it is greater than" and "less than half of each possibility. This is a beautiful comparison.

However, quicksort of second comparison is not so clever: we assume that the fulcrum for pivot, the first element of the comparison results is s1 pivot, then s2 and "pivot relationship has two kinds, s2 < > pivot, pivot and s2 for the latter, very easy to assess the relationship between pivot, namely for s1 < < s2. And pivot, s2 < relationship between and not sure, and also a need for further analysis, and there are two kinds of circumstance s1 < s2 > s2, so with s1 relationship between may be s1 < s2 < pivot or s2 < s1 < pivot, the incidence of relations among the same, so the possibility of pivot less s2 is 2/3. This is quicksort of reason and not so fast, because it does not do each comparison can cut half the possibility of rest.

3   Quicksorting Mutation Algorithm

3.1   Basic Idea

Many times Quicksorting algorithm by means of "distribution" and "collection" two kinds of operations to row a sorting sequence internal sorting method. Combined with the radix sort and quicksort, is the characteristics of the string sorting compared efficient algorithm. This algorithm takes two v1 and v2, hub elements every time division, according to "less than v1", "v1 and v2" and "between the greater than v2" array element is divided into three parts. And then a recursive division of the three parts. Until each part number of elements in the less than a predetermined value recursive over. After a trip to sort and make sure that can be the position of the three elements, and the rapid sort algorithm after a trip to quicksort can be determined after a element position, the former will stay row sequence down into three to row sequence, and the latter only is decomposed into two series, so this point of view, with orders of magnitude more times sequence, many times Quicksort performances in superior to the quicksort, space, both on the same performance [2].

3.2   Algorithm Described

The algorithm firstly to sequence of decomposition, put this sequence interval is decomposed into three stay row sequence, small element than division element in s [l]... S [j], equal division element in s [j + 1]... S [I - 1], greater than the divided elements in s [I]... S [r], then call can be done twice recursion sequence.

To complete the task, program it is equal to the key word division element on the left (l and j between) and to the right (I and r between), in the division cycle in a scanning subscript stop and I and j after the data exchange, it checked each item of data items to see if it equal division element. If an element is equal to the left now divided elements, put it on the left part of the array exchange; If an element of the equal right now divided elements, put it to the array right part of the exchange. When subscript crisscross, equal division from an array of elements in both ends exchange. And then when recursion calling, these keywords can remove from child files. If files have lots of equal keyword, there is an element will eventually be elected division, and those key element of once put in position will not need further treatment.

Adopt many times Quicksorting algorithm scheduling method are as follows:

```
void timesquicksort(ElemType s[], int l, int r)
  { int k;
ElemType v1 = s[l],v2=s[r];
if (r <= l)
  return;
    int i = l-1, j = r, p = l-1, q = r;
    for (i;i<r;i++)
      {
        while (s[++i] < v) ;
  // According to the value of the division series v into three parts
        while (v < s[--j])
if (j == l)
break;
           if (i >= j)
break;
        exchange(s[i],s[j]);
  // Smaller moved to the front of the sequence
        if (s[i] == v)
{
  p++;
exchange(s[p],s[i]);
}
  // Equal moved to the middle of the sequence
        if (v == s[j])
{
q--;
exchange(s[q],s[j]);
}
  // Bigger moved to the back of the sequence
        }
```

```
    exchange(s[i], s[r]);
    j = i-1;
    i = i+1;
    for (k = l   ; k <= p; k++, j--)
    exchange(s[k],s[j]);
    for (k = r-1; k >= q; k--, i++)
                    exchange(s[k],s[i]);
        timesquicksort(s, l, j);
        timesquicksort(s, i, r);
    }
```

3.3   Algorithm Analysis

The standard of quality evaluation algorithm basically has two points: time complexity and space complexity. If the auxiliary space is not dependent on the size of the problem, called the local sorting, the auxiliary space is O (1); The sort of general requirement on auxiliary space is O (n). However, time complexity of the algorithm is involved itself depends on the comparison between the record number and exchange frequency.

How to ensure that the efficiency of the quicksort, its core is that can in any case to make the best data block, ensure the son tell the sequence's length, but the traditional broadly similar quicksort thought can't guarantee it. In recent years, many scholars commission of various improved quicksort algorithm, such as balance quicksort, external quicksort, etc, to a certain extent, enhance the quicksort the stability of the performance, but fundamentally, the improved method can ensure quicksort in any kind of data distribution situation its performance all for O (nlogn).

The proposed algorithm, implementation process similar to our straighten out a deck of CARDS process: if you have a N (N < = 13) pokers, sequence is very untidy, how to straighten out? We assume that the table with 13 position, then we will hand of poker a put out, if is 2, put the position 2, if is Q, on location after 12, put 1 to position from the position of poker (13 to collect all the position without poker on not collect any poker).

One way to understand why speed will be so quick, efficiency are so high: because we are put the first I + 1 CARDS, front I zhang is already set on their corresponding position, in fact is equivalent to "suddenly" is established it and former I zhang poker, the size relations with O (1) operation will this piece of poker correctly inserted before I came to the right position on zhang poker, the effect is equivalent to the insertion sort first I wheel originally need to compare O (I) times, now only need O (1).

When the first poker was putted in position after, placed first I + 1 piece of poker about when I + 1 possibility, because previous I zhang poker 13 position time-have divided I + 1 interval, first I + 1 piece of poker can fall on any interval. So placed the first I + 1 poker is just like is asking this question: "should have this piece of poker in that interval?" The answer is I + 1 possibility, so it will be left to the possibility of partitioning became I + 1 (probability events), suddenly reduced the problem to the original 1 / I + 1. And quicksorting because each comparison only two kinds of results, so most can only add the remaining possibilities reduced by half. So, this algorithm for the time complexity O (n), greatly improving the sort of efficiency.

4   Closing

"Program = data structure + algorithm", data structure is mainly is to learn the thought, algorithm design is the core of the data structure course task, and also one of the program design of the soul. Whatever the quality of the program, which depends on a large part algorithm. Internal order in a data structure is one of the important chapters, is also a difficult point in the learning process, but as long as the idea of sort algorithm, reasonable design program of the process and variable, the problem will be solved. This paper proposed to the quicksort a variation algorithm is more times quicksort algorithm, and analyzes the problems in the same order of magnitude, in time was better than quicksort. When the input sequences are orders of magnitude more times with can be used when quicksorting algorithm to improve the sorting speed. We in the study not only to learn to use of knowledge to solve problems, and to solve the problem of the algorithm is analyzed, and the time and space complexity from to evaluation, and choose the suitable algorithm is presented, and the algorithm is to try to research and improvement.

References

[1]   Yan, W., Wu, W.: The data structure. Tsinghua University Press, Beijing (1997)
[2]   Fan, W.: Quicksorting why so fast liu weipeng. CSDNbolg. Mathematical Beauty (June 2008)
[3]   Ge, J.M.: Quicksort of an improved algorithm. Computer Era (August 2008)
[4]   Dong, P.: improved quicksorting algorithm and recursion. Journal of Anyang Tech. (June 2008)
[5]   Yin, R.: The data structure. Tsinghua University (March 2008)

A Game Theory and Food Chain Algorithm Based QoS Multicast Routing Algorithm

Junwei Wang and Tao Wu

Northeastern University at Qinhuangdao

Abstract. Due to the difficulty on the exact measurement of network parameters and complete expression of the user requirement, a flexible QoS multicast routing scheme based on FCA (Food Chain Algorithm) is presented. In this paper, the corresponding model and its mathematical description are introduced. Under the inaccurate network status information and the flexible user QoS requirement, whether Nash equilibrium between the user utility and the network provider utility can be achieved on the candidate edge is determined by gaming analysis. Combing knowledge of fuzzy mathematics and optimum searching ability of FCA, the proposed algorithm tries to find the multicast tree with the user utility and the network provider utility maximized. Simulation results have shown that the proposed algorithm is both feasible and effective.

Keywords: QoS, multicast routing, Food chain algorithm, Inaccurate network information, Nash equilibrium.

1 Introduction

It is urgent to design reasonable multicast routing algorithm due to the fact that the topology changes frequently and network parameters are inaccurate in NGI (Next-Generation Internet). For multimedia applications involved in real-time communications, they usually require the underlying network to provide a number of QoS (Quality of Service) guarantees to users. It has been proved that multi-constrained QoS multicast routing is NP complete [1]. By now, a lot of solutions have been proposed. [2-5].

A careful analysis of the existing algorithms explored in QoS multicast routing reveals that they often rely on deterministic knowledge of QoS parameters. However, inaccuracies in these parameters are generally unavoidable. Moreover, the existing algorithms are based on vivid QoS parameters, but the requirements of multicast user are usually elastic and an interval of QoS parameter can be accepted. In this paper, a flexible QoS multicast routing algorithm based on FCA (Food Chain Algorithm) [6] is presented. The knowledge of fuzzy mathematics is introduced to address inaccurate parameters. Based on global optimization ability of FCA and under user flexible QoS constraints, the proposed algorithm tries to find the multicast tree with the maximum user utility, network provider utility and RD (reliability degree) of meeting with user QoS requirement.

2   Mathematic Models

2.1   Problem Model

A network can be modeled as a graph $G(V, E)$, where V is node set and E is edge set. an edge e_{ij} has the following parameters: available bandwidth $bw_{ij} \in [bw_{ij_{min}}, bw_{ij_{max}}]$, delay $dl_{ij} \in [dl_{ij_{min}}, dl_{ij_{max}}]$ and error rate $l_{ij} \in [l_{ij_{min}}, l_{ij_{max}}]$. Considering information inaccuracy, *min* and *max* are introduced as subscripts to represent minimum and maximum of QoS parameter that an edge can afford respectively.

In order to support flexible QoS, user requests are defined as $(s, M, \Delta_{Bw}, \Delta_{Dl}, \Delta_{Ls}, \Delta_C)$. Here, $s \in V$ is source node, $M \subseteq V$ is destination nodes set and its number is $|M|$. $\Delta_{Bw} = (\Delta_{Bw_1}, \Delta_{Bw_2}, ..., \Delta_{Bw_{|M|}})$ is the vector of user bandwidth requirement where $\Delta_{Bw_k} = [\Delta_{Bw_k}^{low}, \Delta_{Bw_k}^{high}]$ $(k = 1, 2, …, |M|)$ represents bandwidth constraint interval of the kth user. Similarly, $\Delta_{Dl} = (\Delta_{Dl_1}, \Delta_{Dl_2}, ..., \Delta_{Dl_{|M|}})$ and $\Delta_{Ls} = (\Delta_{Ls_1}, \Delta_{Ls_2}, ..., \Delta_{Ls_{|M|}})$ are the vectors of user delay and error rate requirements. Moreover, $\Delta_{Dl_k} = [\Delta_{Dl_k}^{low}, \Delta_{Dl_k}^{high}]$ and $\Delta_{Ls_k} = [\Delta_{Ls_k}^{low}, \Delta_{Ls_k}^{high}]$ represent delay and error rate constraint interval of the kth user. $\Delta_C = (C_1, C_2, ..., C_{|M|})$ is the vector of user cost requirement and C_k is the upper cost that the kth user is willing to pay. A QoS multicast routing tree $T(s, M)$ supporting communication of s and M should be found.

2.2   Fuzzy Evaluation

Three fuzzy sets $F_l(\sigma)$, $F_m(\sigma)$ and $F_h(\sigma)$ are designed for each QoS parameter σ based on trisection method of fuzzy mathematics, which are related to user satisfaction low L, middle M and high H respectively. Then the membership functions that bandwidth satisfaction degree is low $(I = Bw)$ or delay and error rate satisfaction degrees are high $(I = Dl, Ls)$ of e is:

$$\mu_I(x) = 1 - \Phi\left(\frac{x - I_1}{\beta_I}\right) \tag{1}$$

The membership function of that bandwidth, delay and error rate satisfaction degrees are middle $(I = Bw, Dl, Ls)$ of e is:

$$\mu_I(x) = \Phi\left(\frac{x - I_1}{\beta_I}\right) - \Phi\left(\frac{x - I_2}{\beta_I}\right) \tag{2}$$

The membership function of that bandwidth satisfaction degree is high $(I = Bw)$ or delay and error rate satisfaction degrees are low $(I = Dl, Ls)$ of e is:

$$\mu_I(x) = \Phi(\frac{x - I_2}{\beta_I}) \tag{3}$$

Here, $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ and $\beta_I (I = Bw, Dl, Ls)$ is positive constant.

Given bandwidth interval Δ_{Bw_i} of an edge, for the ith user, its membership degree of satisfaction degree low, middle and high are $T(F(\Delta_{Bw_i}), F_l(Bw))$, $T(F(\Delta_{Bw_i}), F_m(Bw))$ and $T(F(\Delta_{Bw_i}), F_h(Bw))$ respectively. Similarly, the membership degree of the other QoS parameters satisfaction degree low (middle, high) can be obtained.

Table 1. Fuzzy rules

Rule	P_1	P_2	P_3	FR
1	L	L	L	EL
2	L	L	M	VL
3	L	L	H	L
4	L	M	M	M
5	L	M	H	M
6	L	H	H	H
7	M	M	M	M
8	M	M	H	H
9	M	H	H	VH
10	H	H	H	EH

The centre of each fuzzy result is: $el = 0$, $vl = 16$, $l = 33$, $m = 50$, $h = 66$, $vh = 83$, $eh = 100$ ..

The corresponding user QoS satisfaction degree of kth on e is:

$$Q_k(e) = \frac{el * EL + vl * VL + l * L + m * M + h * H + vh * VH + eh * EH}{100 * (EL + VL + L + M + H + VH + EH)} \tag{4}$$

2.3   Game Analysis

Hierarchical payment is adopted in the algorithm. The guarantee QoS of e is divided into m grades and $p_i(e)(i = 1, 2, ..., m)$ is the cost which user should pay when e provides ith QoS.

Game between user and network provider is carried on e . The user game strategies is m QoS grades US_1 , US_2 ;$\cdots$,and US_m . US_{m+1} represents the user abandons the current link e . The network provider has two game strategies: (NS_1, NS_2), denoting whether it is willing to provide the current service or not respectively.

Under strategy (US_i, NS_j), a_{ij}, the kth user utility on e is defined as follow:

$$UU_k^e = \begin{cases} (Q_k(e) - \omega) * \dfrac{C_k}{p_i(e)} & i \neq m+1 \wedge j = 1 \\ \omega & i \neq m+1 \wedge j = 2 \\ -\mu_1(Q_k(e) - \omega) & i = m+1 \wedge j = 1 \\ 0 & i = m+1 \wedge j = 2 \end{cases} \qquad (5)$$

Here, μ_1 is a penalty factor and its value is set bigger than 1, for the user, meaning that rejecting one willing network provider would bring the negative effect. ω is expectation satisfaction degree of user.

Accordingly, b_{ij}, the network provider utility is defined:

$$UN^e = \begin{cases} \dfrac{p_i(e) - C_e(Q_i) - f(i)}{C_e(Q_i)} & i \neq m+1 \wedge j = 1 \\ -\mu_2 \dfrac{p_i(e) - C_e(Q_i)}{C_e(Qi)} & i \neq m+1 \wedge j = 2 \\ 0 & i = m+1 \wedge j = 1 \\ 0 & i = m+1 \wedge j = 2 \end{cases} \qquad (6)$$

Here, $C_e(Q_i)$ is cost price of ith QoS guarantee of e. $f(i) = K * \dfrac{\Delta b_i(e)}{b_{max} - b_{min}}$, represents the impact or blocking on the other users when e provides ith QoS guarantee. $\Delta b_i(e)$ is the bandwidth interval which should be assigned for the user. μ_2 is a penalty factor and its value is set bigger than 1. K is a constant value.

For the utility matrix of e, each value can be obtained. Find the Nash equilibrium strategy pair (US_i^*, NS_j^*) and computing its Pareto value

3   Algorithm Design

3.1   Definitions of Artificial Life

The basic definition of artificial life is shown as follows:

h_i^0: Initial energy of Cr_i. h_i: Current energy of Cr_i, relating to the agent which has maximum energy within Cr_i. $g_i = \log_{1+\eta}(\dfrac{h_i}{h_i^0})$: Current level of Cr_i, η is a constant coefficient. h_{ij}^0: Initial energy of jth agent in Cr_i. h_{ij}: Current energy of jth agent in Cr_i.

3.2   Competing Rule

When Cr_i abandons the waste, assume Cr_j has n approaches to acquire it (there are n connected edges between Cr_j and Cr_i $i, j = 1, 2, \ldots Y$). The probability of Cr_j acquires the food through edge e_r $(r = 1, 2, \ldots, n)$ is defined as follow:

$$gf_{ij}^r = \frac{h_j * pa(e_r)}{\sum\limits_{j \in \delta \wedge j \neq i} \sum\limits_{r=1}^{n} h_j * pa(e_r)} \tag{7}$$

Here, δ is the neighbor colony set of Cr_i.

3.3 Sharing Rule

The energy of jth agent which is obtained by sharing rule within Cr_i is h_{ij}^s and the energy of preying agent in Cr_i is h_{ip}. The relationship between h_{is} and h_{ip} is defined as follow:

$$h_{ij}^s = \frac{h_{ip}}{m} \tag{8}$$

Here, m is number of sharing agents.

3.4 Accumulating Rule

The relationship between energy level of kth growing up of jth agent in Cr_i and initial energy is shown as follow:

$$h_{ij}^k = (1 + \rho)^k h_{ij}^0 \tag{9}$$

The energy which is consumed in a time periodic of jth agent in Cr_i is defined as follow:

$$h_{ij}^c = \lambda_i h_{ij} \tag{10}$$

The dying out energy level of jth agent in Cr_i is defined as follows:

$$h_{ij}^d = (1 - \tau)h_{ij}^0 \tag{11}$$

3.5 Algorithm Descriptions

Step 1. Initialize network topology and partition colony. Set initial energy of each agent. Pareto optimal set is null and the number of destination agents is M.

Step 2. I foods are placed on each destination agent and $i = 0$.

Step 3. $i=i+1$. If $i > I$, go to Step 9, otherwise, $m = 0$ and go to Step 4.

Step 4. $m = m+1$. If $m > M$, go to Step 7, otherwise, the colony which includes mth destination agent is set to be current colony.

Step 5. If current colony includes source agent, then sharing request are send from source agent to preying agent, and preying agent shares the food with source agent according to formula (8) and go to Step 4, otherwise, go to Step 6.

Step 6. Preying agent of current colony shares the food with ejectable agent according to formula (8) and adjacent colonies compete for food according to formula (7), the colony which obtains the food is set to be current colony and go to Step 5.

Step 7. Update Pareto optimal set, the agent on T accumulates energy, and all of the agents consume energy according to formula (10). Determine whether agent grows up or vanishes according to formula (9) and (11), go to Step 3.

Step 8. Normalize the solutions in Pareto optimal set and output the corresponding multicast tree of biggest solution, the algorithm ends. If the Pareto optimal set is empty, consult with user and run the algorithm again.

4   Performance Evaluations

Simulations have been done on NS2 (Network Simulator 2) platforms. The proposed algorithm (called F), fuzzy QoS multicast routing algorithm based on general GA in paper [11] (called G) and multicast routing algorithm based on Prim (called P) have been performed over CERNET (T1), NSFNET (T2) and a virtual topology (T3, generated by Waxman2 with average node degree 3.5).

In topology 2, P algorithm is taken as benchmark, comparison results of network provider utility and user utility of three algorithms are shown in Tab. 2. It can be concluded that the proposed scheme has better performance.

Table 2. Comparison of user utility and network provider utility

The ratio of destination nodes	F:G:P	
	User utility	Network provider utility
16.2%	1.206:1.206:1	1.183:1.183:1
32.4%	1.303:1.225:1	1.211:1.15:1
48.6%	1.215:1.18:1	1.253:1.126:1
64.8%	1.253:1.197:1	1.149:1.117:1
81.0%	1.174:1.129:1	1.24:1.203:1
97.3%	1.28:1.209:1	1.177:1.114:1

5   Conclusions

In this paper, considering the inaccurate parameters of NGI and flexibility QoS requirement of user, a QoS multicast routing algorithm based on FGA is proposed. Gaming between user utility and network provider utility is employed to design routing principle. Combing knowledge of fuzzy mathematics and optimization ability of FCA, the proposed algorithm tries to find the multicast tree with the maximum user utility and network provider utility of meeting with user QoS constraint.

References

[1] Wang, Z., Crowcroft, J.: Quality of Service for Supporting Multimedia application. IEEE Journal on Selected Areas in Communication 17(7), 1228–1234 (1996)

[2] Moses, C., Joseph, N., Baruch, S.: Resource Optimization in QoS Multicast Routing of Real-Time Multimedia. IEEE/ACM Transaction on Networking 12(2), 340–348 (2004)

[3] Turgay, K., Marwan, K.: Bandwidth-Delay Constrained Path Selection Under Inaccurate State Information. IEEE/ACM Transaction on Networking 11(3), 384–398 (2003)

[4] Oh, J.S., Bae, S.-i., Ahn, J.-h., Kang, S.: Route Reinforcement for Efficient QoS Routing Based on Ant Algorithm. In: Kahng, H.-K., Goto, S. (eds.) ICOIN 2004. LNCS, vol. 3090, pp. 342–349. Springer, Heidelberg (2004)

[5] Junwei, W., Xingwei, W., Min, H.: A Hybrid Intelligent QoS Multicast Routing Algorithm in NGI. In: IEEE PDCAT, pp. 723–727 (2005)

[6] Haifei, Y., Dingwei, W.: Food–Chain Algorithm and Application to Supply–chain Planning. Journal of System Simulation 15(5), 1195–1199 (2005)

[7] Ping, C., Tianling, D.: A Fuzzy Genetic Algorithm for QoS Multicast Routing. Computer Communications 26(1), 506–512 (2003)

The Study about Feature Extraction of Analog Circuit Fault Diagnosis Based on Annealing Genetic Hybrid Algorithm

Lin Hai-Jun, Wan Si-Bo, Zhong Hui, and Wang Qi-Gao

Harbin University of Science & Technology, Harbin, China

Abstract. This paper forms the annealing genetic hybrid algorithm, which is about the genetic algorithm and simulated annealing algorithm improved and integrated. For the problem about feature extraction of nonlinear analog circuit fault diagnosis based on Volterra kernel, using annealing genetic hybrid algorithm to research, put forward annealing genetic intelligent extraction method of circuit fault diagnosis feature, and verified by practical examples. Experiments show that this method has realized effective extraction of circuit diagnostic features and improve the prediction accuracy of fault diagnosis system based on BP neural network.

Keywords: Feature extraction, Annealing genetic algorithms, Analog circuits, Fault diagnosis.

1   Introduction

For the nonlinear analog circuits, the intelligent diagnosis based on neural network is one of the strong practical fault diagnosis methods, due to the nature of intelligent diagnosis is pattern recognition, the key of the fault diagnosis is to find the stable parameters which can reflect the circuit failure characteristics.

Description of nonlinear analog circuits has many ways, due to the time domain kernel and frequency domain kernel of Volterra have the intuitive physical meaning, is commonly used the features of the fault diagnosis system, therefore, this paper does research using the diagnosis based on Volterra   kernel as an example.

In fact, in the intelligent fault diagnosis based on the Volterra kernel do not need to use all the information of the kernel, because Volterra kernel of the nonlinear analog circuits contain a wealth of information, if all the original features are used as a diagnostic feature, is not only makes the neural network for diagnosis complicated, the workload of identification and calculation big, but also increases the probability of error. Therefore, it is necessary to study the feature extraction methods, reduce the number of parameters to obtain a set of concentrated characteristics, so as to decrease early warning error probability.

There are many Feature extraction methods, commonly used Karhunen-Loeve transform method, Fisher correlation recognition method, etc. These methods have a common feature, namely dimension reduction mapping based on the linear

transformation. In recent years, with the continuous development of artificial intelligence, feature extraction and selection methods of using artificial intelligence theory gradually appeared, such as feature extraction based on neural network method, feature extraction based on rough set method and feature extraction and selection based on genetic algorithm method. In that regard a variety of intelligent methods were gradually used, but was not enough systematic and impeccable, especially in terms of efficiency need to be further improved.

Aiming at the feature extraction of non-linear analog fault diagnosis, this paper using annealing genetic hybrid algorithm to study, and then   put forward feature extraction and selection based on annealing genetic hybrid algorithm methods.

2   Improved Annealing Genetic Hybrid Algorithm

Genetic Algorithm (abbreviation is GA) is a kind of optimization algorithm imitated biological evolution process. This algorithm codes for possible solutions, constitutes the initial population, and choose, cross, and mutate individual of initial population follow the genetic rules, then choose the individual according to certain probability or the individual's fitness value, thus to hope produce the better individual.

GA algorithm is a parallel algorithm with parallel processing capabilities and global optimizing ability, in theory, is equal to the optimal solution, however, slow evolution or precocious phenomena occasionally occur in actual operation.

Simulated Annealing (abbreviation is GA) algorithm is an optimization algorithm imitated the high-temperature annealing principle. According to thermodynamics, the object at a certain temperature, which atomic energy is distributed by Boltzmann equation, with the temperature of the object gradually decreases, the atomic volume which constitute the object decreased slowly, until it reaches equilibrium under the termination temperature, in this state, atomic energy is the lowest. when using the SA algorithm to optimize, usually regard the optimization objective as atomic energy, and make optimizing calculation to their distribution probability, take temperature as control parameters of Optimization process, and Generation after generation to search by the method of dropping temperature appropriately, in order to avoid local optimal state, while SA searching towards the direction of optimization, it also receives deteriorative states by a certain probability, so as to search the global optimal solution[1].

The sudden jump of the probability of SA makes it get the ability to avoid falling into local optimum, but does not have parallelism, and the conditions for convergence of the algorithm is relatively rigor, it requires a high enough initial temperature, low enough terminate temperature and slow enough cooling rate, so the search with a high quality needs a long time [2] [3].

In order to acquire a better speed and quality of optimization, In this paper, these two above algorithms is improved, and combined the features of the sudden jump of the probability of simulated Annealing algorithm and the parallel search of genetic algorithm, propose an improved genetic annealing hybrid optimization algorithm (Improved SAGA, abbreviation is ISAGA).

In the algorithm, based on diversity determine accept to prevent prematurity of genetic algorithm, adopting the worst initial acceptance probability to determine the initial temperature, using the scale parameter optimization method to speed up the convergence rate of annealing algorithm and the self-adaptive adjustment crossover

and mutation parameters, strive to find the global optimum result in a little time to improve the algorithm efficiency. The detailed treatment method was shown in flow 1.

In improved optimization process of the annealing genetic hybrid, adopt uniformly the fitness function as the optimizing evaluation function. In ISAGA, due to use the GA's fitness function to evaluate part performances of SA, therefore, we must take the increase of fitness value as optimization direction.

Flow 1 introduction:

The first step: Individual code and Initialize the optimization process;

Set the number of individuals of initial population as M, generate initial state of all the individual x(i) randomly;

Decide what is used as fitness function or objective function, and determine the formula;

Determine the initial temperature and temperature range, In order to avoid the affection of getting too high or too low initial temperature to the searching speed and quality, initial temperature is selected as $t_0 = (f_{min} - f_{max})/\ln p_r$, f_{min} is minimum value of objective function in Initial population, f_{max} is maximum value of objective function in Initial population, p_r is the worst initial acceptance probability.

Determine the appropriate annealing strategies. In this paper, index fell strategy was used;

Determine genetic fitness threshold and the maximum number of generations, the initial parameters of cross p_{ch}, p_{cl}, and the initial parameters of the mutation rate p_{mh} and p_{ml}.

The second step: Disturbance and update operations of population;

According to the formula $x'(i) = x(i) + h \cdot \delta$ makes disturbance to Each of the individuals in the population, in formula, h is the scale parameter, δ is a random variable, $x'_{(i)}$ is new state, $x_{(i)}$ is old state. Then we can obtain the neighborhood function by adopting optimizing the scale parameters.

The third step: Decision of individual's acceptance or rejection;

Calculate the fitness function of the original state and new state respectively, and calculate the difference between them $\Delta J(i) = J'(i) - J(i)$, the i represents individual of population. When $\Delta J(i) > 0$, accept the new state, otherwise, according to Metropolis criterion with probability $p(\Delta J(i)) = exp((-\Delta J)/(kT))$ to accept the new state, where, k is the Boltzmann constant. In the actual calculations, randomly generated number α between 0 and 1, if $p(\Delta J(i)) > \alpha$, then determine $x'(i)$ as the new state, otherwise, remains the original $x(i)$ state.

The fourth step: Selection;

Use the roulette wheel selection method to select the individual, the more Fitness large the more individual selected probability big.

The fifth step: Cross [4] [5];

Select individual of population randomly according to the cross rate Pc, two individuals identified as a father generation, then determine the hybrid position randomly, and using two point method to hybridize. In order to avoid falling into local optimum or searching does not converge, need to protect the good individuals, we use self-adaptive adjustment method of P_c to improve the individual cross rate with low fitness and drop the individual cross rate with high fitness.

Where J_{tc} is the larger of the two individual fitness value, J^t_{max} means the largest individual fitness value of   the current population, t is the current population times, p_{ch} and p_{cl} is the set parameters, and meet $1 > p_{ch} > p_{cl} > 0$, J^t_{avg} means the mean of all individual fitness of   the current population.

The cross-rate formula realizing self-adaptive adjustment is:

$$P^t_c = \begin{cases} p_{CL} - (p_{ch} - p_{cl})(J^t_c - J^t_{avg})/(J^t_{max} - J^t_{avg}) & J^t_c \geq J^t_{avg} \\ p_{ch} & J^t_c < J^t_{avg} \end{cases}$$

The sixth step: Variation;

This operation can improve the ability global searching of optimization algorithm and possibilities of all the solution space searched. Mutation rate P_m also used this self-adaptive method and the specific methods similar to P_{tc}'s.

The seventh step: Judge and accept new population according to the diversity;

The specific method is to generate [0, 1] uniform distributed random number α, Accept the new generation of individuals according to the following formula:

$$X(k+1) = \begin{cases} X'(k), & \text{当 } \alpha < C_k, \\ X(k), & \text{当 } \alpha \geq C_k. \end{cases}$$

Where $C_k = \min\left(1, (\lambda(X'(k))/\lambda(X(k)))^{\beta/k} (F_s(X'(k))/F_s(X(k)))^k\right)$, $\lambda(X\ (k))$, $\lambda(X'(k))$, $F_S(X\ (k))$ and $F_S(X'\ (k))$ is varied degrees and overall fitness of new populations and old population, β is a constant.

The eighth step: Optimized termination discrimination;

If one of the following two conditions is met, optimization process is over, otherwise, cooling down according to the determinate cooling strategy, then go to the second step. One of the conditions is that groups have met the fitness thresholds set; the second condition is that optimized times have reached a given upper limit.

3   The Annealing Genetic Intelligent Extracting of Circuit Fault Diagnosis Features

Feature extraction is through the transform or mapping to make characteristic vector reduction, or we can say that is all original features produced better features by extracting, which is more representative and can reflect the essence better.

Feature extraction can be summarized as that through a mathematical transformation to compress n features to less m features, as $A = H^T B$, among them, H is the n×m transformation matrix. Transformed vector A is the column matrix of m dimension; vector B before transformation is the column matrix of n dimension; and they meet n > m.

It is visible that the key link of feature extraction is find out the best transformation matrix, which makes category separability criterion value to be the largest in the m dimension model space after transformation.

In order to improve Volterra nuclear feature extraction efficiency of during the fault diagnosis, this paper adopts ISAGA feature extraction algorithm. The basic idea: Taking feature extraction as optimization problem, determining the characteristic

evaluation function, and taking it as the optimized objective function, adopting ISAGA algorithm to global optimize transfer matrix to get the optimal transformation matrix.

Feature extraction based on ISAGA algorithm mainly to complete the following several jobs:

First of all, go ahead optimization preparation, code transferring matrix, which as individual code when code optimized; Build category space feature vector, using the extracted characteristic value of transformed each fault state corresponding to build characteristic vector according to certain rules which used for the discrimination of classification effect; Determining the characteristic evaluation function which used to measure the effect of a transfer matrix feature extraction; Determine the initial parameters and the strategy of optimization algorithm to make the algorithm   global optimization effect well and speed quickly. Secondly, adopt the improved genetic algorithm to produce new transfer matrix. then, through model transformation to get transform features A. Finally, evaluate the change matrix, and loop optimization until obtaining the optimal transformation matrix.

The flow of Volterra kernel diagnosis feature extraction which based on MSAGA algorithm is as follows:

The first step: Code individual. Each column of the transfer matrix listed as an optimal individual, code transferring matrix parameters for subsequent optimization.

The second step: Determine characteristic vector structure of category spatial. It includes eigenvector dimension and the method calculated transform feature by new transfer matrix.

The third step: Build evaluation function. In order to it can reflect the difference degree of feature extraction, use each line of the target matrix to constitute evaluation vector, and use the lumped Euclidean distance of the evaluation vector as evaluation function, namely the optimized fitness function.

The fourth step: Execute ISAGA arithmetic.

The fifth step: Output the feature transfer matrix optimizing calculated, and through transferring matrix to calculate transformed features.

4   Experimental Study

In order to test feature extraction method based on ISAGA algorithm, take the Volterra kernel data under the condition of non-linear no source filter circuit soft fault as sample to extract fault features, as shown in Fig. 2. In different states of the circuit, data of Volterra kernel is known, as show in Tab. 1.
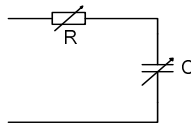
Fig. 2. Nonlinear passive filter circuit

Table 1. Feature data of Volterra kernel under different conditions

Parameter	1	2	3	4	5	6	7	8
normal state 0	1.25+0.0024i	1.25+0.0021i	-1.85-0.0298i	-1.80-0.0268i	-1.95-0.0250i	-1.89-0.0219i	-1.95-0.0125i	-1.92-0.0116i
fault state 1	2.00+0.0060i	2.00+0.0055i	-5.14-0.1273i	-4.94-0.1130i	-5.60-0.1097i	-5.33-0.0947i	-5.60-0.0549i	-5.45-0.0507i
fault state 2	1.25+0.0024i	1.25+0.0021i	-1.94-0.0305i	-1.87-0.0272i	-2.08-0.0260i	-1.99-0.0226i	-2.08-0.0130i	-2.03-0.0120i

The ISAGA feature extractions are shown in tab. 2. In this case, the original parameters of each state have 8 and after extraction turn to 3.

Use the extracted feature data and original data to train the neural network to diagnose fault, Training goal set 0.001, and then, adopt other samples include extracted and unextracted for comparison text. The results show that the sample after ISAGA feature extracted, convergence speed of training neural network and accuracy of diagnosis are all superior to the original data without the optimization.

Table 2. ISAGA feature extraction results

parameter	1	2	3
normal state 0	149.6625+0.2223i	-487.0450-7.5697i	33.6415+6.7848i
fault state 1	202.9765+0.0386i	-1374.0671-32.9121i	1330.6473+33.7958i
fault state 2	145.7227+0.2185i	-513.1592-7.7548i	102.9150+7.2751i

5   Conclusion

During the non-linear analog circuit fault diagnosis, Volterra feature parameters under ISAGA feature extraction, Dimension of the parameter which reflect the characteristics of fault status is reduced, features are concentrated, the diagnostic accuracy were improved obviously.

References

[1] Xu, Z.-B., Gao, Y.: The feature analysis and prevention of phenomenon of premature convergence of genetic algorithm. Science in China (E series) 8, 364–375 (1996)
[2] Wang, Z.-C., Zhang, T., Wang, H.-W.: The simulated annealing optimization algorithm based on chaotic variables. Control and Decision 8(4), 381–384

[3] Lan, H., Wang, X., Wang, L.: A kind of function optimized performance analysis of Genetic-Annealing Algorithm. System Simulation 8, 111–113 (2001)
[4] Yin, S.-R., Chen, G.-Y., Xie, Y.-L.: Analog circuit fault diagnosis incentive optimizing based on genetic algorithm. Control Technology 6, 20–22 (2007)
[5] Lu, J.-G.: Genetic Algorithms and Engineering Applications, 12th edn., pp. 1–174. China Mining University Press (1997)

Finite Stage MDP Method and Algorithm
for Task Allocation in UCAVS

Peng He and De-Yun Zhou

Northwestern Polytechnical University, Electronic Information Institute, Xi'an, China

Abstract. Multi-UCAV (Unmanned Combat Air Vehicle) UAV operations task allocation is an important part of decision-making. It is also for solving a class of difficult combinatorial optimization problem for which there is no mature and high effective solving schemes. The problem of task allocation can be treated as grading sequence decision programming by Markov decision process (MDP), and the single integer coded method is used for current stage. The solution algorithm based on finite phase MDP is also proposed in the paper. The simulation results show that the computing rate and precision achieve well in solving multi-UCAV task allocation problem using this algorithm.

Keywords: unmanned combat aircraft, Markov decision, task allocation, combinatorial optimization.

1   Introduction

UCAV (Unmanned Combat Air Vehicle) appears in modern local wars in an increasing frequency, becoming one of the key technologies in building the future information battleground. Multi-UCAV cooperative action can achieve a higher combat efficiency than a single UCAV, but it requires complex tactical mission planning system with the support, and multi-UCAV cooperative task allocation is such a most important part of such mission planning system that it gets much attention at home and abroad. In addition to the traditional integer programming optimization methods, in recent years, genetic algorithm [1,2], particle swarm algorithm [3] and ant colony algorithm and other heuristics have been tried to apply in this field. Genetic Algorithm iterates the prior values according to some optimal genetic iteration rules until the value reaches the ideal ones. Genetic Algorithm achieves high precision with long iteration time which goes against its utility in programs. Particle Swarm Algorithm computes the posterior distribution density through the sample, and then evaluates the number through the posterior distribution. This algorithm is out of the type of interferefering origin's control, but usually with the problems of weight degeneration and sample shrinking. Ant Colony Algorithm is a kind of evolutionism simulation algorithm, a kind of probability algorithm that finds the optimal path in the graph, which can achieve well in some occasions but not universal. An a finite stage of MDP based algorithm to solve the multi-UCAV task allocation problem and to realize the quick accurately allocations is proposed, the simulation results show that the algorithm has very good speed and accuracy performance.

2   Problem Description

Assume that the number has been given a set of UCAV, it needs a pre-determined series of attacks on ground targets, the core of task allocation issue is how to satisfy the constraint conditions for the optimal target allocation scheme to maximize a performance. Performance depends on the set of principles or policy of engagement which can be expressed as one or more objective function. For example, we want to select a possible target assignment UCAV programs, it has the greatest damage value to enemy targets, making minimal loss of it's own UCAV fleet. This problem can be described as:

$$\max J = \sum_m \sum_n C_{mn} x_{mn} \tag{1}$$

Where, x_{mn} is the decision variables, if the target m is assigned to the n th UCAV, then, $x_{mn} = 1$, otherwise $x_{mn} = 0$; and :

$$C_{mn} = W_D * P_{id} * V_m * PS_{mn} * PKT_{mn} - W_A(1 - PS_{mn}) \tag{2}$$

In formula (2), P_{id} is target recognition probability; V_m is the target value; PS_{mn} is the survival probability of UCAV n to target m; PKT_{mn} is the kill probability of UCAV n to target m; W_D and W_A is the their relative importance of target damage and self-risk which is the choice of the performance of the decision-making personal preference. Assuming a single UCAV can carry k weapons, the kill probability of one weapon on target is PKI, then $PKT_{mn} = 1 - (1 - PKI)^k$.

Problem constraints for each target must be assigned to the uniquely identified UCAV:

$$\sum_{n=1}^{N} x_{mn} = 1 \text{, for } m = 1, 2, \cdots, M \text{, and } x_{mn} \subset \{0,1\} \tag{3}$$

Where, M is the target number, N is the UCAV number. At the same time the number of weapons on-board require to meet the constraints:

$$\sum_{m=1}^{M} x_{mn} \leq k \text{, for } n = 1, 2, \cdots, N \tag{4}$$

It can be seen multi-UCAV task allocation needs to solve the problem which has a huge search space, for targets and UCAVs, there is kinds of possible allocation schemes, and its number relates with target number in exponential growth being proved to be an NP problems.

3   MDP-Based Solution for Multi UCAV Task Allocation Problem

3.1   Finite Phase of MDP

MDP (Markov Decision Process) model is a special kind of sequential decision problems, its character is that its action aggregate, instant reward and transition probability depend on current system state and action only, have nothing to do with the past.

T is the time or stage set for the decision-making, and $T = \{0, 1, \cdots, N-1\}$, $0 < N < \infty$; all possible states of the system are $i \in S$; the set of available actions is $A(i)$; the next time system state is decided by the probability distribution $p(\cdot \mid i, a)$, $r(i, a)$ is the bounded payoff function defined on S and $A(i)$, $\{T, S, A, p(\cdot \mid i, a), r(i, a)\}$ is a limited stage MDP. For $N \geq 0$ and all general strategy $\pi \in \prod$, the N periods expected total return utility function is defined as:

$$V_N(i, \pi) \equiv \sum_{t=0}^{N-1} E_\pi^i[r(Y_t, \Delta_t)] + E_\pi^i[r(Y_N)], \qquad i \in S \qquad (5)$$

Where, π is the using strategy, starting at time 0 under the conditions of state $i \in S$, the system obtained expected total return until the moment of time N, $r(Y_N)$ is the process of termination pay. For Markov strategies class $\pi \in \prod_m^d$, it can prove [4] :

$$\sup_{\pi \in \prod} V_N(i, \pi) = \sup_{\pi \in \prod_m^d} V_N(i, \pi), \qquad i \in S \qquad (6)$$

u_t^π is denoted as the sum of the expected return from time t to time N using strategy $\pi \in \prod_m^d$, and $u_0^\pi(i) = V_N(i, \pi)$. The definition of finite stage MDP optimal equation is:

$$u_t(i_t) = \sup_{a \in A(i_t)} \left\{ r_t(i_t, a) + \sum_{j \in S} p_t(j \mid i_t, a) u_{t+1}(a, j) \right\} \qquad (7)$$

When $t = N$, the up boundary conditions $u_N(i_N) = r_N(i_N)$ is added.

Assumption is that equation (7) of the solution, after resolving optimal equation, those actions combination who meet the upper boundary become the optimal strategies $a_t^*(i_t)$:

$$a_t^*(i_t) \in \arg\max_{a \in A(i_t)} \left\{ r_t(i_t, a) + \sum_{j \in S} p_t(j \mid i_t, a) u_{t+1}^*(a, j) \right\} \qquad (8)$$

3.2   Model Foundation

Algorithm based on finite stages MDP is used to solve the task allocation problem of multi-UCAV described in the first section, battlefield assuming is that UCAV number of our own is N, to number of enemy ground targets to fight is M, and $N \le M$, assuming formular (4) is always satisfied, but adding constraints :

$$\sum_{m=1}^{M} x_{mn} \ge 1, \text{ for } n = 1,2,\cdots,N, \quad x_{mn} = \{0,1\} \tag{9}$$

1. State $T = \{0,1,\cdots,M\}$ is the definition of decision-making stage, the decision-making sequence corresponds to the target sequence in T, $t \in \{0,\cdots,M-1\}$ is decision-making stage, that is, at every moment t, allocation decisions is made to target $m \in \{1,\cdots,M\}$ of the corresponding target sequence.

The possible state i_t is coded by integer, which indicates that the results of task allocation in the current phase. l_n is the coding bit, there $i_t = [l_1 \cdots l_n \cdots l_N]$, $0 \le l_n \le l_{max}$, $l_{max} = M - N + 1$ and $\sum_{n=1}^{N}(l_{max} - l_n) \le M$. For $n = 1,\cdots,N$, the initial state i_0 meets $l_n = l_{max}$; the termination state meets $l_n < l_{max}$.

2. Action Available actions are $a_t = I \in A(i_t), t = 0,\cdots,M-1$, where $I = [l'_1 \cdots l'_n \cdots l'_N]$, and they meet $\sum_{n=1}^{N} l'_n = 1$, $l'_n \in \{0,1\}$.

3. Transition Probability $sum(l_{max})_t$ is denoted as the median sum of $l_n = l_{max}$ in i_t, assuming that $l'_q = 1$ in a_t, in order to ensure that formula (9) holds, when $l_q \ne l_{max}$, action a_t should satisfy the constraints $sum(l_{max})_t \le (M - t - 1)$. Thus the corresponding transition probability is:

$$p_t(j_{t+1} \mid i_t, a_t) = \begin{cases} 1, & j_{t+1} = i_t - a_t \\ 0, & other \end{cases} \quad t = 0,\cdots,M-1 \tag{10}$$

4. Reward The reward function is defined as

$r_t(a_t \mid i_t) = r_{t-1}(a_{t-1} \mid i_{t-1}) + r_\Delta(a_t)$, where $t = 1,\cdots,M-1$, which represent the reward value from initial state to stage t, $r_\Delta(a_t)$ is the incremental returns after choosing action a_t in stage t, and $r_0(a_0 \mid i_0) = r_\Delta(a_0)$. By formula (1) and (2), we know $r_\Delta(a_t) = C_{(t+1)(q)}$, where $l'_q = 1$, $l'_q \in a_t$.

4   A Numerical Example

Reconstructing Based on the algorithm described in Section 2.2, simulation and computation are done through C++ language. Pre-set parameter matrix $V_{1 \times M}$, $PS_{M \times N}$ and $PKT_{M \times N}$, so $P_{id} = 1$, $W_D = W_A = 1$, due to limited space, the specific values of parameter matrix is omitted. To verify the effectiveness of the algorithm, in the same operating environment, GA algorithm [1] and PSO algorithm [3] are compared to these issues in simulations.

Group a: choose $M = N = 10$ and the corresponding parameters. The simulation results are shown in Table 1, the optimal payoff is 4.172.

Table 1. Optimal decision-making programs of set a, $M = N = 10$

Target	UCAV	V_m	PS_{mn}	PKT_{mn}	Target	UCAV	V_m	PS_{mn}	PKT_{mn}
T1	4	0.70	0.94	0.90	T6	8	0.87	0.80	0.95
T2	5	0.89	0.82	0.88	T7	2	0.95	0.90	0.92
T3	3	0.55	0.78	0.68	T8	6	0.90	0.94	0.66
T4	1	0.71	0.90	0.60	T9	9	0.85	0.78	0.89
T5	7	0.36	0.96	0.96	T10	10	0.76	0.95	0.70

Group b: choose $M = N = 20$ and the corresponding parameters. The simulation results are shown in Table 2, optimal payoff is 10.105.

Table 2. Optimal decision-making programs of set b

Target	UCAV	V_m	PS_{mn}	PKT_{mn}	Target	UCAV	V_m	PS_{mn}	PKT_{mn}
T1	9	0.80	0.94	0.86	T11	5	0.84	0.89	0.98
T2	14	0.50	0.92	0.72	T12	18	0.72	0.95	0.90
T3	11	0.61	0.98	0.74	T13	8	0.54	0.90	0.90
T4	4	0.72	0.90	0.77	T14	10	0.21	0.92	0.72
T5	2	0.70	0.96	0.83	T15	19	0.64	0.98	0.91
T6	12	0.98	0.96	0.93	T16	20	0.92	0.50	0.60
T7	17	0.78	0.96	0.80	T17	16	0.85	0.98	0.81
T8	15	0.98	0.53	0.92	T18	6	0.62	0.90	0.77
T9	7	0.61	0.96	0.92	T19	1	0.90	0.99	0.90
T10	3	0.94	0.96	0.92	T20	13	0.32	0.98	0.66
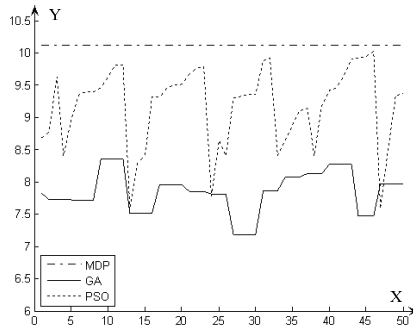
Fig. 1. Simulation results compare of the three algorithms in group a
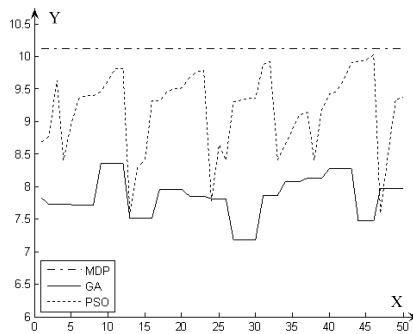
Fig. 2. Simulation results compare of the three algorithms in group b

Fig.1 and Fig.2 are comparison of various algorithms under two sets of experimental conditions, X axis shows us the operation number, Y-axis shows us the optimal objective function value for each run by the algorithm. In experiments, the GA and the PSO population size are both 50, the evolution generations are 500 each time. Simulation results show that the characteristics of traditional GA and the PSO algorithms determines the optimization results to get close to its optimal solution which is difficult to achieve, and solution quality, convergence time and stability depend heavily on the choice of empirical parameters, it may result to a greater volatility of the running results if the parameters is chosen irrelevantly. Infinite stage MDP algorithm can achieve a stable global optimal solution, even in the expanded problem scale, while the solution quality of the other two algorithms have different degrees of decline under the same situation, at the same time that the MDP algorithm has a faster running speed within some complex issues degree.

5   Conclusions

In solving combinatorial optimization problems, the typical optimization algorithm ,such as GA needs to search giant prepare solution space by certain rules under the

guidance of the optimal target, the results and algorithm convergence time largely depend on the relevant parameters selection. a single task allocation decision making is to be broken down into solvable sequence of decision-making process in a finite stage , using the MDP model construction algorithm, the corresponding optimal strategy is found at the time when an optimal value is found in the end of the decision-making process, that strategy is the most optimal task allocation scheme. By comparing the simulation results, we can see that the algorithm can get a better optimal decisions within a reasonable period of running time, providing an effective way for large-scale multi-UCAV task allocation problem.

References

[1] Qi, F.: Projectiles, Rockets. Application of single parent genetic algorithm for task allocation for large-scale UCAVs. Firepower and Command and Control 31(5), 18–21 (2006)

[2] Chen, G., Cruz Jr., J.B.: Genetic Algorithm for Task Allocation in UAV Cooperative Control. In: Proceedings of AIAA Guidance, Austin, Texas, August 11-14 (2003)

[3] Cruz Jr, J.B., Chen, G., Li, D., Wang, X.: Particle Swarm Optimization for Resource Allocation in UAV Cooperative Control. In: Proceedings of AIAA Guidance, Navigation and Control Conference, Providence, Rhode Island, August 16-19 (2004) AIAA-2004-5250

[4] Liu: Practical Markov decision process. Tsinghua University Press, Beijing (2004)

The Split Feasibility Problem in Hilbert Space

Wu Dingping, Duan Qibin, Wang Erli, and Zhao Hang

Department of Mathematic, Chengdu University of Information Technology,
Chengdu, 610225, China
`wdp68@163.com`

Abstract. The purpose of this paper is to introduce and study Ishikawa iterative algorithms for solving the SFP in the setting of infinite-dimensional Hilbert spaces. The main results presented in this paper improve and extend some recent results done by Xu [Iterative methods for the split feasibility problem in infinite-dimensional Hilbert space, Inverse Problems 26 (2010) 105018] .At the end we prove that the accumulation of errors in Ishikawa iterative CQ algorithm is bounded in certain range.

Keywords: Split feasibility problem, CQ algorithm, Ishikawa iterative algorithm.

1 Introduction and Preliminaries

Throughout this paper, we always assumes that H is a real Hilbert space with inner product $< \bullet, \bullet >$ and norm $\|\bullet\|$, respectively. To begin with, let recall that the spilt feasibility problem (SFP) is to find a point x^* with the property that $x^* \in C$ and $Ax^* \in Q$, where C and Q are the nonempty closed convex subsets of the real Hilbert spaces H_1 and H_2 respectively, and A is bounded linear operator from H_1 to H_2.

We use Γ to denote the solution set of SFP

$$\Gamma = \{x \in C : Ax \in Q\} = C \bigcap A^{-1}Q$$

Definition 1.1. Assume H is a real Hilbert space. let $A, T : H \to H$ be nonlinear operators introduced below.

A mapping $T : H \to H$ is said to be nonexpansive if

$$\|Tx - Ty\| \le \|x - y\|, \qquad x, y \in H . \ x, y \in H$$

A mapping $T : H \to H$ is said to be an averaged mapping if it can be written as the average of the identity I and a nonexpansive mapping S ; that is $T = (1 - \alpha)I + \alpha S$,

where $\alpha \in (0,1)$ and $S : H \to H$ is nonexpansive .

A is monotone if $\langle Ax - Ay, x - y \rangle \ge 0$ for all $x, y \in H$.

A is β-strongly monotone, with $\beta > 0$, if

$$\langle Ax - Ay, x - y \rangle \ge \beta \|x - y\|^2 \text{ for all } x, y \in H .$$

A is υ-inverse strongly monotone(υ-ism),with $\upsilon > 0$,if

$$\langle Ax - Ay, x - y \rangle \geq \upsilon \|Ax - Ay\|^2 \quad \text{for all} \quad x, y \in H$$

Proposition 1.1. We have the following assertions:

i. T is nonexpansive if and only if the complement $I - T$ is $\frac{1}{2}$-ism.

ii.   If T is υ-ism and $\gamma > 0$,then γT is $\frac{\upsilon}{\gamma}$-ism.

iii. For $\alpha \in (0,1)$,T is $\alpha-$ averaged if and only if $I - T$ is $\frac{1}{2\alpha}$-ism.

iv.   If T_1 is α_1- averaged and T_2 is α_2- averaged, where $\alpha_1, \alpha_2 \in (0,1)$, then the composite $T_1 T_2$ is $\alpha-$ averaged ,where $\alpha = \alpha_1 + \alpha_2 - \alpha_1 \alpha_2$.

v.   If T_1 and T_2 are averaged and have a common fixed point, then $Fix(T_1 T_2) = Fix(T_1) \cap Fix(T_2)$.

Proposition 1.2. Let K be nonempty closed convex subset of a Hilbert space H . Let $\{x_n\}$ be bounded sequence which satisfies the properties:

- $\omega_\omega(x_n) \subset K$;
- $\lim\limits_{n\to\infty} \|x_n - x\|$ exists for every $x \in K$.

Then $\{x_n\}$ converges weakly to a point in K .

Let $\gamma > 0$ and assume $x^* \in \Gamma$, thus $Ax^* \in Q$ which implies the equation $(I - P_Q)Ax^* = 0$ that in turn implies the equation $\gamma A^*(I - P_Q)Ax^* = 0$, hence the fixed point equation $(I - \gamma A^*(I - P_Q)A)x^* = x^*$,we can see that the solutions of the fixed point equation are exactly solutions of SFP.

Very recently, Xu[3] has viewed the CQ algorithm as a fixed point algorithm for averaged mappings and applied Mann's algorithm to SFP, and he also proved that an averaged CQ algorithm is weakly converge to a solution of SFP.

In this paper, we also regard the CQ algorithm as a fixed point algorithm for averaged mappings, and try to study SFP by the following Ishikawa Iterative process:

$$\begin{cases} x_{n+1} = (1 - \alpha_n)x_n + \alpha_n P_C(I - \gamma A^*(I - P_Q)A)y_n \\ y_n = (1 - \beta_n)x_n + \beta_n P_C(I - \gamma A^*(I - P_Q)A)x_n, \quad n = 0,1,2,\ldots \end{cases} \tag{1.1}$$

where $\{\alpha_n\}$, $\{\beta_n\}$ are assumed to be a real sequence in $[0,1]$.

2   Main Results

Theorem 2.1. Assume the SFP is consistent and $0 < \gamma < 2/\|A\|^2$. Let $\{x_n\}$ be defined by (1.2). Assume that

$$0 < \liminf_{n\to\infty} \alpha_n < \limsup_{n\to\infty} \alpha_n < 1, \quad 0 < \liminf_{n\to\infty} \beta_n < \limsup_{n\to\infty} \beta_n < 1.$$

Then the sequence $\{x_n\}$ converges weakly to a solution of SFP .

Proof: The choice of $0 < \gamma < 2/\|A\|^2$ implies that the defining operator $P_C(I - \gamma A^*(I - P_Q)A)$ is averaged. The $I - P_Q$ is firmly nonexpansive, firmly nonexpansive mapping is $1/2$-averaged, which is 1-ism.

We can see that $A^*(I - P_Q)A$ is $1/\|A\|^2$-ism and $\gamma A^*(I - P_Q)A$ is $1/\gamma\|A\|^2$-ism.

Then $I - \gamma A^*(I - P_Q)A$ is $\gamma\|A\|^2/2$-averaged, therefore noticing that P_C is $1/2$-averaged, so we get $P_C(I - \gamma A^*(I - P_Q)A)$ is α-averaged, where

$$\alpha = \frac{1}{2} + \frac{\gamma\|A\|^2}{2} - \frac{1}{2} \cdot \frac{\gamma\|A\|^2}{2} = \frac{2 + \gamma\|A\|^2}{4} .$$

We denote that $T = P_c(I - \gamma A^*(I - P_Q)A) = (1 - \alpha)I + \alpha S$,

Where $\alpha = (2 + \gamma\|A\|^2)/4 \in (0,1)$ and $S : H \to H$ is nonexpansive .

Then we prove that .

For any $x^* \in \Gamma$, noticing $Tx^* = x^*$, we have

$$\|Ty_n - x^*\| = \|Ty_n - Tx^*\| = \|(1 - \alpha)(y_n - x^*) + \alpha(Sy_n - Sx^*)\|$$
$$\leq (1 - \alpha)\|y_n - x^*\| + \alpha\|Sy_n - Sx^*\|$$
$$\leq (1 - \beta_n)\|x_n - x^*\| + \beta_n\|Tx_n - x^*\| \leq \|x_n - x^*\|. \tag{2.1}$$

By (2.1), we have

$$\|x_{n+1} - x^*\| = \|(1 - \alpha_n)x_n + \alpha_n Ty_n - x^*\| \leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n\|Ty_n - x^*\|$$
$$\leq (1 - \alpha_n)\|x_n - x^*\| + \alpha_n\|x_n - x^*\| \leq \|x_n - x^*\|. \tag{2.2}$$

So the sequence $\{\|x_n - x^*\|\}$ is nonincreasing.

Hence $\lim_{n \to +\infty}\|x_n - x^*\|$ exists for all $x^* \in \Gamma$.

$$\|x_{n+1} - x^*\|^2 = \|(1 - \alpha_n)x_n + \alpha_n Ty_n - x^*\|^2$$
$$= \|(1 - \alpha_n)(x_n - x^*) + \alpha_n(Ty_n - Tx^*)\|^2 \tag{2.3}$$
$$= (1 - \alpha_n)\|x_n - x^*\|^2 + \alpha_n\|Ty_n - Tx^*\|^2 - \alpha_n(1 - \alpha_n)\|x_n - Ty_n\|^2 .$$

From (2.3) and (2.1),we have

$$\|x_{n+1} - x^*\|^2 = (1 - \alpha_n)\|x_n - x^*\|^2 + \alpha_n\|Ty_n - Tx^*\|^2 - \alpha_n(1 - \alpha_n)\|x_n - Ty_n\|^2$$
$$\|x_n - Ty_n\|^2 = \frac{1}{\alpha_n}\|x_n - x^*\|^2 + \frac{1}{1 - \alpha_n}\|Ty_n - Tx^*\|^2 - \frac{1}{\alpha_n(1 - \alpha_n)}\|x_{n+1} - x^*\|^2 \tag{2.4}$$

$$\left\| x_n - Ty_n \right\|^2 \le (\frac{1}{\alpha_n} \left\| x_n - x^* \right\|^2 + \frac{1}{1-\alpha_n} \left\| x_n - x^* \right\|^2 - \frac{1}{\alpha_n(1-\alpha_n)} \left\| x_{n+1} - x^* \right\|^2)$$

$$= \frac{1}{\alpha_n(1-\alpha_n)} (\left\| x_n - x^* \right\|^2 - \left\| x_{n+1} - x^* \right\|^2)$$

(2.5)

Then we get

$$\lim_{n \to \infty} \left\| x_{n+1} - x_n \right\| = \lim_{n \to \infty} \alpha_n^2 \left\| x_n - Ty_n \right\|^2$$

$$= \lim_{n \to \infty} \frac{\alpha_n}{1-\alpha_n} (\left\| x_n - x^* \right\|^2 - \left\| x_{n+1} - x^* \right\|^2) = 0$$

(2.6)

Next we prove that $\omega_\omega(x_n) \subset S$.

Suppose $p \in \omega_\omega(x_n)$ and $\{x_{n_j}\}$ is a subsequence of $\{x_{n_j}\}$ such that $x_{n_j} \xrightarrow{w} p$;

thus, $x_{n_j+1} \xrightarrow{w} p$ by (2.6).

Then we get

$$\left\| x_{n_j} - Tx_{n_j} \right\| \le \left\| x_{n_j+1} - x_{n_j} \right\| + \left\| x_{n_j+1} - Tx_{n_j} \right\|$$

$$\le \left\| x_{n_j+1} - x_{n_j} \right\| + \left\| (1-\alpha_{n_j})x_{n_j} + \alpha_{n_j}Ty_{n_j} - (1-\alpha_{n_j})Tx_{n_j} - \alpha_{n_j}Tx_{n_j} \right\|$$

(2.7)

$$\le \left\| x_{n_j+1} - x_{n_j} \right\| + (1-\alpha_{n_j}) \left\| x_{n_j} - Tx_{n_j} \right\| + \alpha_{n_j} \left\| Ty_{n_j} - Tx_{n_j} \right\|.$$

Simplify the formula above

$$\alpha_{nj} \left\| x_{n_j} - Tx_{n_j} \right\| \le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j} \left\| Ty_{n_j} - Tx_{n_j} \right\|$$

$$= \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j} \left\| (1-\alpha)(y_{n_j} - x_{n_j}) + \alpha(Sy_{n_j} - Sx_{n_j}) \right\|$$

$$\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}(1-\alpha) \left\| y_{n_j} - x_{n_j} \right\| + \alpha_{n_j}\alpha \left\| Sy_{n_j} - Sx_{n_j} \right\|$$

$$\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j} \left\| y_{n_j} - x_{n_j} \right\|$$

$$= \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j} \left\| (1-\beta_{n_j})x_{n_j} + \beta_{n_j}Tx_{n_j} - x_{n_j} \right\|$$

$$\le \left\| x_{n_j+1} - x_{n_j} \right\| + \alpha_{n_j}\beta_{n_j} \left\| x_{n_j} - Tx_{n_j} \right\|.$$

(2.8)

So we obtain that

$$\left\| x_{n_j} - Tx_{n_j} \right\| \le \frac{1}{\alpha_{n_j}(1-\beta_{n_j})} \left\| x_{n_j+1} - x_{n_j} \right\| \to 0 \quad (as \ n_j \to \infty)$$

(2.9)

Then we apply proposition 2.2, we get that the sequence $\{x_n\}$ converges weakly to a solution of SFP.

This is complete the proof.

Remark 2.2. The sequence $\{x_n\}$ converges weakly, but not in norm. Hundal[6] construct a counter-example as follows:

In the space $H = l^2$, C and Q are two closed convex subsets of H satisfying the properties:

  i. $C \cap Q \neq \varphi$;

  ii. The sequence generated by alternating projections

$$x_n = (P_C P_Q)^n x_0, n \geq 0,$$

with $x_0 \in C$, converges weakly, but not in norm.

When we take $A = I$ and $\gamma = 1$, the sequence $\{x_n\}$ generated by the CQ Ishikawa iterative algorithm (1.2):

$$x_{n+1} = (1 - \alpha_n + \alpha_n(1 - \beta_n)P_C P_Q + \alpha_n \beta_n (P_C P_Q)^2)x_n$$

obviously converges weakly, but not in norm.

Corollary 2.3. The accumulation of errors in (1.2) is bounded and it is not more than the maximum of the absolute errors of $\{\|u_n\|\}$ and $\{\|v_n\|\}$.

Proof: Define the error of Sx_n and Sy_n by $u_n = Sx_n - S\overline{x}_n$ $v_n = Sy_n - S\overline{y}_n$ for all $n \geq 0$, where $S\overline{x}_n$ and $S\overline{y}_n$ are the exact value of Sx_n and Sy_n.

Let $M_u = \sup\{\|u_n\| : n \geq 0\}$, and $M_v = \sup\{\|v_n\| : n \geq 0\}$ be the bounds on absolute errors, and we set $M = \max\{M_u, M_v\}$. In the averaged CQ Ishikawa iterative algorithm

$$\begin{cases} \overline{x}_{n+1} = (1 - \alpha_n)\overline{x}_n + \alpha_n \overline{Ty_n} \\ \overline{y}_n = (1 - \beta_n)\overline{x}_n + \beta_n \overline{Tx_n} \end{cases}$$

So we have

$$\|x_{n+1} - \overline{x}_{n+1}\| = \sum_{j=0}^{n} (\alpha(1 - \alpha)\alpha_j \beta_j u_j + \alpha\alpha_j v_j) \prod_{i=j+1}^{n} (1 - 2\alpha\alpha_i + \alpha^2 \alpha_i) \tag{2.10}$$

$$\|y_n - \overline{y}_n\| = \beta_n \|u_n\| + (1 - \beta_n)\|x_{n+1} - \overline{x}_{n+1}\| \tag{2.11}$$

Then it is known that $\sum_{i=0}^{\infty} (2\alpha\alpha_i - \alpha^2 \alpha_i) = (2\alpha - \alpha^2) \sum_{i=0}^{\infty} \alpha_i = +\infty$ implies that

$$\prod_{i=0}^{\infty} (1 - (2\alpha\alpha_i - \alpha^2 \alpha_i)) = 0$$

Thus

$$\|x_{n+1} - \overline{x}_{n+1}\| = \sum_{j=0}^{n} (\alpha(1 - \alpha)\alpha_j \beta_j u_j + \alpha\alpha_j v_j) \prod_{i=j+1}^{n} (1 - 2\alpha\alpha_i + \alpha^2 \alpha_i)$$

$$\leq M \sum_{j=0}^{n} (\alpha(1-\alpha)\alpha_j \beta_j + \alpha\alpha_j) \prod_{i=j+1}^{n} (1 - 2\alpha\alpha_i + \alpha^2\alpha_i)$$

$$\leq M \sum_{j=0}^{n} (2\alpha\alpha_j - \alpha^2\alpha_j) \prod_{i=j+1}^{n} (1 - (2\alpha\alpha_i - \alpha^2\alpha_i)) \tag{2.12}$$

$$= M \left[1 - \prod_{i=0}^{n} (1 - (2\alpha\alpha_i - \alpha^2\alpha_i)) \right]$$

$$\leq M \left[1 - \prod_{i=0}^{\infty} (1 - (2\alpha\alpha_i - \alpha^2\alpha_i)) \right] = M$$

$$\|y_n - \overline{y}_n\| \leq \beta_n M + (1 - \beta_n)M \leq M . \tag{2.13}$$

From (2.12) and (2.13),we can get that The accumulation of errors is not more than the maximum of the absolute errors of $\{\|u_n\|\}$ and $\{\|v_n\|\}$.

References

[1] Ishikawa, S.: Fixed point and iteration of a nonexpansive mapping in a Banach spaces. Proc. Amer. Math. Soc. 73, 7365–7371 (1976)

[2] Qin, X., Cho, Y.J., Kang, S.M.: Viscosity approximation methods for generalized equilibrium problems and fixed point with applications. Nonlinear Anal. 72, 99–12 (2010)

[3] Xu, H.-K.: Iterative methods for the split feasibility problem in infinite-dimensional Hilbert space. Inverse Problems 26, 105018 (2010)

[4] Xu, H.-K.: Averaged Mapping and the Gradient-Projection Algorithm (in press)

[5] Xu, Y., Liu, Z., Shin Min, K.: Accumulation and conreol of random errors in the Ishikawa iterative process in arbitrary Banach space. Comput. Math. Appl.

[6] Hundal, H.: An alternating projection that does not converge in norm. Nonlinear Anal. 57, 35–61 (2004)

A Discrete PSO Algorithm for QoS Routing Optimization in Wireless Mesh Network

Muhua Zhong and Yuzhong Chen[*]

Department of Computer Science, Fuzhou University, China
mhzhong1988@163.com, yzchen@fzu.edu.cn

Abstract. The performance of wireless mesh network is influenced significantly by the link state and interference between links. In this paper we seek routes for QoS connection requests with bandwidth requirements in consideration of the link performance parameters of data link layer and QoS requirements. In order to get the solution of QoS routing problem, the problem is converted to multi-objective combinatorial optimization problem and a discrete PSO algorithm is proposed for multi-objective optimization to realize the QoS routing problem. Experiment results show that compared with enumeration method and GA algorithm, our scheme for QoS routing is feasible and efficient.

Keywords: Wireless mesh network, Interference, PSO, QoS.

1 Introduction

Wireless mesh network [1] is a new type of wireless broadband network structure, namely a distributed wireless network with high capacity and high speed. It is similar with Ad hoc network in network protocol, but the mobility of WMN is weaker and the rate of topology change is smaller.

Currently there are several common routing algorithms, such as DSDV, DSR, ABR and AODV. AODV and DSR are based on the shortest path. Without considering the influence of interference on the multi-hop wireless network, communication distance and other factors, the performance of delay, throughput and packet success rate from source node to destination node will get worse [2]. And several typical routing criteria (such as ETX [3], ETT [4], WCETT [4]) use a single QoS as the routing criterion and neglect other effects, which results in finding a not ideal routing.

Literature [5] proposed the concept of intra-flow interference and inter-flow interference and interference clique (a collection of links interfering with each other). In [6] the author considered the interference problem as the independent sets' scheduling problem, and just computed a small number of independent sets to obtain the upper and lower available bandwidth of the route. Literature [7] proposed a method of computing a route satisfying the QoS requirements considering the interference in wireless networks, but it doesn't consider mutual interference of the links in the route, which led the route to waste a lot of network resources although the route satisfying the

---

[*] Corresponding author.

QoS requirements. Literature [8] uses two-hop as interference distance in the process of searching for all the route's interference cliques, which is not realistic. In real world interference still exists in most of the links those distances more than two-hop distance.

This paper focuses on the QoS routing problem under constraints in the wireless mesh network, which proves to be a NP-complete problem. Particle Swarm Optimization (PSO) has many advantages, such as fast convergence, simple model and easy to be implemented, it has been successfully applied to many fields, thus we solve this problem by using PSO, in addition this paper adopts actual Euler distance as the link interference distance to when computing the optimal routes.

The rest of this paper is organized as follows. We analyze the problem in Section 2. Section 3 utilizes the discrete PSO to discover optimal QoS routes. Section 4 compares our algorithm with the existing algorithms. Section 5 concludes this paper.

2   Problem Description

It's supposed that the location of Mesh network's nodes is fixed, besides each nodes have fixed transmission range and interference range. QoS routing problems can be described as in the precondition of satisfying given bandwidth seeking a series of optimal paths which meet specific QoS requirements [9].

We use an undirected graph $G(V,E)$ to model the wireless mesh network where V is the set of n vertices and E is the set of m edges. $(i, j) \in E$ corresponds to a link in the network; d_{ij} represents the delay between the link(i,j); b_{ij} represents the bandwidth betwwen the link(i,j); per_{ij} represents the packet error rate of the link(i,j).

The object of Many QoS routing problem is that searching for a series of optimal paths P_{st} which meet the requirements, besides P_{st} should meet a QoS constraint: the available bandwidth of P_{st} is not less than a given bandwidth requirement. The problem can be described as follow:

$$\min \quad Delay = \sum_{(i,j)\in Pst} d_{ij} \tag{1}$$

$$\min \quad Psr = 1 - \prod_{(i,j)\in Pst} (1 - per_{ij}) \tag{2}$$

$$\min \quad B = \min\{each\ clique's\ F_j\} \tag{3}$$

$$\text{s.t.} \quad B \geq B_{req} \tag{4}$$

In the above formulation, *Delay* denotes the delay of the path. *Psr* denotes the packet success rate of the path .*B* denotes the available bandwidth of the path and F_j is defined in section 2.1. B_{req} denotes the given bandwidth requirement.

The available bandwidth of the path is the maximum throughput it can achieve subject to some conditions. In order to get the solution of the above formulation, we need to compute the available path bandwidth, and then search for the routes satisfying the above formulation by utilizing PSO algorithm.

2.1   The Method of Calculating Available Bandwidth

As to a given path, we can construct a contention graph of the path, and find out all the clique j from the contention graph. Due to the reason that each link in the clique interferes with each other, the links in the clique can not communication with each other at the same time, therefore the maximum achievable throughput F over the path in the form $\sum_{l \in j} F / B_l \leq 1$, where B_l denotes the bandwidth of the link. Then we can draw a conclusion $F \leq F_j$, where $F_j = \sum_{l \in j} (1 / B_j)^{-1}$, namely the value of the available bandwidth of the clique is equal to F_j, B_j denotes the bandwidth of the link l. Fig.1 shows a example of the computation of the available bandwidth of the path.
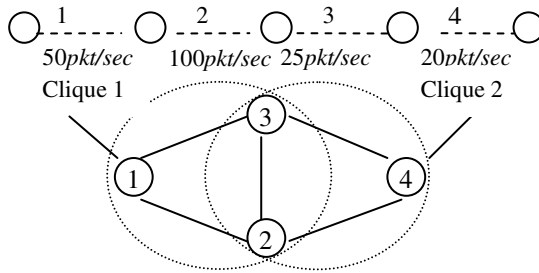
Fig. 1. Contention Graph

According to the definition mentioned above, the available bandwidth of the path in Fig.1 is min{ $(1/50+1/100+1/25)^{-1}$, $(1/100+1/25+1/20)^{-1}$ }=10$pkts/sec$.

2.2   Algorithm for Seeking Available Bandwidth

Algorithm 1. Search for all the cliques containing the link m

Step_1. Encode the links of the path P_{st} (1,2…n), construct the corresponding contention graph. Then we define a array $Clique[n][n]$ to store all the clique, initialize $Clique$ to be null, $currentCliqueNum$=0;

Step_2. Set $Clique[0][m]$=1, which means storing link m in the first line of $Clique$.
    Traverse all the links of P_{st}.
    if the link of P_{st} interferes with a part of links of $Clique[i]$(i from 0 to $currentCliqueNum$), construct a new clique consisting of a part of links of $Clique[i]$ and the link of P_{st}, $currentCliqueNum$++; then traverse $Clique$, delete the repeated clique, $currentCliqueNum$--;

Algorithm 2. Calculate the available bandwidth of the path

Step_1. Initialize $availBandWidth$= ∞ ($availBandWidth$ denotes the available bandwidth of the path P_{st})
Step_2. Traverse all the links of the path P_{st} (i from 1 to n)

Call Algorithm 1, obtain all the cliques containing the link i, then calculate F_j of all the cliques; if (F_j <availBandWidth), availBandWidth= F_j .

3   PSO Algorithm for QoS Routing Optimization

PSO is initialized with a population of random solutions. The updates of the particles are accomplished according to the following equations. Equation(5) calculates a new velocity for each particle(potential solution)based on its previous velocity (v_{id}^{t-1}) and the personal best location which the particle has achieved so far and the global best location of the population in history. Equation (6) updates i-th particle's position in solution hyperspace.

$$v_{id}^t = w.v_{id}^{t-1} + c_1.r_1.(p_{id}^t - x_{id}^t) + c_2.r_2.(p_{gd}^t - x_{id}^t) \tag{5}$$

$$x_{id}^t = x_{id}^{t-1} + v_{id}^t \tag{6}$$

where t is the iteration index, d is the number of dimensions, w is the inertia weight, c_1 and c_2 are two positive constants, called acceleration constants, r_1 and r_2 are two random numbers within the range [0,1]. A constant called Vmax is used to limit the velocities of the particles.

3.1   Encoding

In our routing problem, the method of integer encoding is used to encoding implementation of the particle, each particle represents the sequence of nodes, namely a path from source to destination. For example, a particle $P_{st}=[v_1, v_2,\dots, v_t]$, where v_1 denotes source node, v_t denotes destination node.

3.2   Particle Initialization

In this paper, in order to initialize the particle swarm, we utilize depth-first search based on the recursive method to obtain the sequence of nodes from the source to the destination.

3.3   Improved Discrete PSO

The discrete PSO algorithm incorporates the crossover and mutation operators in Genetic Algorithm (GA) [10].

At first, the notion of mutation operator is incorporated into the inertance keeping part of (5).

$$A_i^t = F_1(X_i^{t-1}, w) = \begin{cases} M(X_i^{t-1}), r1 < w \\ X_i^{t-1}, otherwise \end{cases} \tag{7}$$

Where F_1 denotes the mutation operator with the probability w, and r_1 is a random number.

Secondly, the cognitive component and social component of (5) adopt the notion of crossover operator in GA.

$$B_i^t = F_2(A_i^t, c_1) = \begin{cases} C_p(A_i^t), r_2 < c_1 \\ A_i^t, otherwise \end{cases} \tag{8}$$

$$X_i^t = F_3(B_i^t, c_2) = \begin{cases} C_g(B_i^t), r_3 < c_2 \\ B_i^t, otherwise \end{cases} \tag{9}$$

Where F_2, F_3 denotes the crossover operator with the probability c_1, c_2 respectively. r_2 and r_3 are two random numbers.

Then, we can get the equation

$$X_i^t = F_3(F_2(F_1(X_i^{t-1}, w), c_1), c_2) \tag{10}$$

Where X_i^t represents the result of i-*th* particle at the *t-th* generation.

3.4 Fitness Function

For multi-objective optimization, the personal best position can be used to store non-dominated solutions generated in the past. While iterating, a set of non-dominated solutions is maintained, and a global best should be selected from this set.

The fitness function proposed in [10] is applied, in which a particle is evaluated by both pareto dominance and neighborhood density. For a global best, a non-dominated solution with lower fitness value is selected. In particular, if several solutions have the same fitness value, we choose a random one. The detail of the fitness function is given below.

Definition 1: The fitness function of *i*-th particle is

$$F(i) = (1 + D(i)) \times (1 + N(i)) \tag{11}$$

Where $D(i)$ indicates the number of the particles which are better than *i-th* particle in the current population, $N(i)$ is defined as follow.

Definition 2: The neighborhood density of *i-th* is

$$N(i) = \sum_{j=1}^{p} sh(fd_{ij}) \tag{12}$$

where $sh(fd_{ij}) = \begin{cases} 1, if \ fd_{ij} \leq \sigma_s \\ 0, otherwise \end{cases}$, σ_s denotes the sharing parameter, fd_{ij} denotes the distance between *i-th* particle and *j-th* particle.

3.5 Algorithm Overview

Step_1. Load the network topology and initialize the parameters of the PSO algorithm.

Step_2. Initialize the population, calculate the available bandwidth, delay, packet success rate and fitness of each particle.

Step_3. Initialize the leading particle, store it in the external document.

Step_4. Do corresponding operation on each particle according to equation (7)-(9), update the location of each particle, and calculate the available bandwidth, delay, packet success rate and fitness of current particle.

Step_5. Update the leading particle in the external document.

Step_6. If termination condition is satisfied, the algorithm stops; otherwise, go to Step_4.

4   Experiment Results

Our algorithm was implemented in C++ language. The simulation experiment is carried out in the grid network topology. We set the parameters of links：$0 \leq$bandwidth≤ 54, $0 \leq$delay≤ 30, $0 \leq$packet error rate≤ 0.1; the value of interference distance of the node is 50; the parameters for PSO algorithm were set to the following values：w decreased linearly from 0.9 to 0.4, c_1 decreased linearly from 0.82 to 0.5, and c_2 increased from 0.4 to 0.83.

Our algorithm is compared with enumeration algorithm and traditional GA algorithm. Enumeration algorithm adopts enumeration method of add pruned, where delay, available bandwidth and packet error rate of the path are used to as the pruned condition, in order to all the optimal accurate routes; and GA algorithm adopts weighted sum of delay, available bandwidth and packet error rate of the path as fitness function to search for an optimal route.
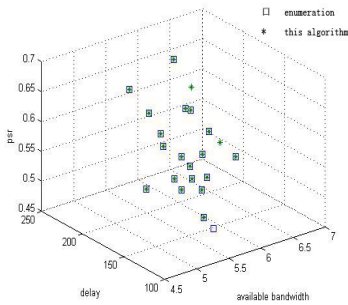
Fig. 2. Comparison of this algorithm with enumeration in 7×7 grid network
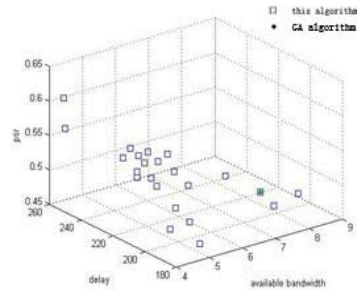
Fig. 3. Comparison of this algorithm with traditional GA in 8×8 grid network

Fig.2 shows the comparison result of the proposed algorithm with enumeration. We can see the proposed algorithm can find out almost all the solutions, furthermore, the unaccurate solutions can be avoided and unfound solutions can be found by increasing the iteration number. So the accuracy of the proposed algorithm can be guaranteed.

GA algorithm is used for getting a single optimal route. As shown in Fig.3, compared with traditional GA algorithm, the optimal routes found by our algorithm include the solution of GA algorithm. Thus we can get a appropriate route according to the user's different QoS requirements.

5   Summaries

Mutual interference of the links is a significant factor influencing the performance in wireless mesh network. Therefore, we need to get an appropriate route with interference as little as possible for the sake of decreasing the bandwidth consumption. In this paper we consider the influence of the intra-flow interference in wireless mesh network, and the experiment shows that our algorithm is feasible and efficient.

References

[1]   Akyildiz, I.F., Xudong, W., Weilin, W.: Wireless mesh networks: a survey. Computer Networks 47, 445–487 (2005)

[2]   Couto, D.S.J.D., Aguayo, D., Bicket, J., et al.: A high-throughput path metric for multi-hopwireless routing. In: Proceedings of ACMMOBICOM, San Diego, CA, pp. 134–146 (2003)

[3]   De Couto, D.S.J., Aguayo, D., Bicket, J., Morris, R.: A High-Throughput Path Metric for Multi-Hop Wireless Routing. In: ACM Mobicom (2003)

[4]   Draves, R., Padhye, J., Zill, B.: Routing in Multi-Radio, Multi-Hop Wireless Mesh Networks. In: ACM Mobicom (2004)

[5]   Salonidis, T., Garetto, M., Saha, A., Knightly, E.: identifying high throughput paths in 802, pp. 21–30 (October 2007)

[6]   Jain, K., Padhye, J., Padmanabhan, V., Qiu, L.L.: Impact of Interference on Multi-hop Wireless Network Performance (2003)

[7]   Yang, Y., Kravets, R.: Contention-aware admission control for ad hoc networks (2005)

[8]   Hou, R., Lui, K., Chiu, H., Yeung, K.L., Baker, F.: Routing in multi-hop wireless mesh networks with bandwidth guarantees. Technical Report, February (2009)

[9]   Tang, J., Xue, G., Zhang, W.: Interference-aware topology control and QoS routing in multi-channel wireless mesh networks. In: Proceedings of the 6th ACM International Symposium on Mobile ad Hoc Networking and Computing (May 2005)

[10] Guo, W.Z., Chen, G.L., Huan, M., Chen, S.L.: A discrete particle swarm optimization algorithm for the multi-objective permutation flowshop sequencing problem. In: Proceedings of the Second International Conference of Fuzzy Information and Engineering, pp. 323–331 (2007)

A Local Routing Strategy on Weighted Scale-Free Networks

Xiao-Xi Wang and Guo-Long Chen

Department of Computer Science, Fuzhou University, China
`axp@live.com, cgl@fzu.edu.cn`

Abstract. In this paper, the packet traffic flow on weighted scale-free networks is guided based on local link information and a single adaptive preferential factor α which characterizes the preferential probability in choosing links to forward packets: $P_{l \to i} = \dfrac{w_{li}^{\alpha_{li}}}{\sum_{j} w_{lj}^{\alpha_{lj}}}$ .The adaptive preferential factor α increases when the queue length of the link is smaller than the capacity, and decreases otherwise. The capacity of links is set to be the weight of the link. It is shown by simulations that using adaptive preferential factor α could both reduce the packets delivering time and increase the capacity of huge communication networks compared with using fixed value of α.

Keywords: Weighted scale-free networks, Routing strategy, Average packets delivering time, Adaptive preferential factor.

1 Introduction

Since the small world (SW) effect discovered by Watts and Strogatz [1]and the scale-free (SF) property with degree distribution following Power-law by Baraba′si and Albert(BA model) [2], complex networks have become a very hot activity on physical and computer field, mainly because a wide range of large communication networks such as the Internet [3] and WWW [4] can be described by complex networks. The increasing importance of large communication networks calls for the need for high efficiency in handling and delivering information. Therefore, a variety of studies have been focused on developing better packet routing strategies to enhance traffic flow and to avoid traffic congestion on complex networks.

In previous studies, there are mainly two types of routing methods. One is based on global information, the other is based on local information. Global routing methods include the shortest path[5], the efficient path[6], and Optimal betweenness[7].Such research selects the best path by minimizing a cost function (for example, the total number of hops, the sum of node degrees and the maximum betweenness centrality, etc.) These routing rules allow each node to have the whole network's topological information, which may be practical for small or medium size networks but not for very large networks in modern society such as the Internet or WWW. local routing methods include the random walking[8-10], the nearest-neighbor search strategy[11], the next-nearest-neighbor search strategy[12] and the integration of local static and

dynamic information[13]. These routing rules allow each nodeonly knows the information of its neighbors.

At present, the traffic dynamics have been investigated mainly in un-weighted networks. However, most real networks such as the Internet and WWW are weighted networks [14, 15]. Weighted networks can carefully portray the node connections and mutual interactions that not only reflect the topology of real networks, but also reveal physical and dynamic characteristics of real-world networks. Therefore, study of traffic dynamics in weighted networks is needed. In this paper, a local routing strategy with adaptive preferential factor for scale-free networks is proposed.

Packets are guided based on local link information and a single adaptive preferential factor α which characterizes the preferential probability in choosing links to forward packets. The adaptive preferential factor α increases when the queue length of the link is smaller than the capacity, and decreases otherwise.

The rest of this paper is organized as follows.In the following section,the network model is introduced, in Section 3 traffic model is described in detail, in Section 4 simulation results of traffic dynamics are provided,and Section 5 gives the conclusion.

2   Network Model

To generate the underlying network infrastructure, this paper uses a weighted scale-free network model proposed by Wang et al. [14],in which power-law distributions of degrees, weights and strengths are all in good accordance with real observations of weighted technological networks. The model rules of [14] can be described as follows.

Starting from m_0 nodes fully connected by links with assigned weight $w_0=1$, the system is driven by two mechanics:

(1) the strength dynamics: the weight of each link connecting i and j is updated as

$$w_{ij} \rightarrow w_{ij} + 1 \tag{1}$$

with probability

$$P_{ij} = W \times P_{ij} = W \times \frac{s_i s_j}{\sum_{a<b} s_a s_b} \tag{2}$$

where $s_i = \sum_{j \in \Gamma_i} w_{ij}$ is the strength of node i and Γ_i is the neighboring set of node i.

(2) the topological growth: a new node n is added with m links that are randomly attached to a node i according to the strength preferential probability:

$$\prod_{n \rightarrow i} = \frac{s_i}{\sum_j s_j} \tag{3}$$

where j runs over all existing nodes.

3   Traffic Model

Once the network is generated, it remains fixed, and our model is described as follows:

a. At each time step, there are R packets generated in the system, with randomly chosen sources and destinations.

b. For simplicity, we treat all nodes as both hosts and routers for generating and delivering packets. The capacity of the link that is the number of data packets the link can handle at most each time step is set to be the weight of the link.

c. To navigate packets, each node performs a local search among its neighbors. If the packet's destination is found within the searched area, it is delivered directly to its target; otherwise, it is forwarded to a neighbor node according to the preferential probability of each node:

$$P_{l \to i} = \frac{w_{li}^{\alpha_{li}}}{\sum_j w_{lj}^{\alpha_{lj}}} , \qquad (4)$$

where w_{li} is the weight of the link connecting nodes l and i, the sum runs over the neighbors of the node l, and α_{li} is an introduced adaptive preferential factor characterizing the preferential probability in choosing links to forward packets. Once the packet arrives at its destination, it will be removed from the system. We should also note that the queue length of each link is assumed to be unlimited and the FIFO(first in first out)discipline is applied at each queue.

d. Update all adaptive preferential factors. The rules can be described as follows. When the queue length of the link is smaller than the capacity,the adaptive preferential factor of the link increases by a step λ; otherwise, decreases by a step λ. Simultaneously, the upper and lower threshold of the adaptive preferential factors is set to be $\alpha_{max} (1.4 < \alpha_{max} < 1.6)$ and $\alpha_{min} (\alpha_{min} < 1)$. The adaptive preferential factors can take the value of step λ between one percent and one twentieth of α_{max} (in this paper λ=0.1 and the initial value of all adaptive preferential factors is set to be 1.0).

4   Simulation Results

The network overall capacity is measured by the critical generating rate R_c at which a continuous phase transition occurs from free state to congestion. The free state refers to the balance between created and removed packets at the same time. If the system enters the jammed state, packets will continuously accumulate in the system and only a few packets can reach their destinations. In order to describe the critical point accurately, we investigate the order parameter [16]:

$$H(R) = \lim_{t \to \infty} \frac{1}{R} \frac{\langle \Delta Q_p \rangle}{\Delta t} \qquad (5)$$

Here $\Delta Q_p = Q_p(t + \Delta t) - Q_p(t)$, with $\langle \ldots \rangle$ takes the average over a time window of width Δt,and $Q_p(t)$ is the total number of packets in the network at time t. For $R < R_c, \langle \Delta Q_p \rangle = 0$ and $H(R) = 0$, indicating that the system is in the free state in which the numbers of added and removed packets are balanced without traffic congestion; Otherwise for $R > R_c$, $H(R)$ increases suddenly from zero ,the system will collapse ultimately because congestion emerges and spreads in the system and

packets will continuously accumulate in the system.. Therefore a phase transition occurs at $R = R_c$ and R_c is the maximal generating rate under which the system can maintain its normal and efficient functioning. Thus the maximal handling and delivering capacity of the system is measured by R_c. We note that there can be different ways to define the network capacity, but we will mainly consider the critical value of R_c in the following discussion.
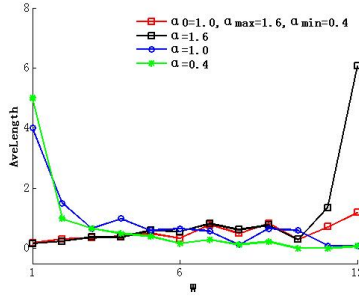
Fig. 1. Average queue length of links with different weights divided by the link weight versus W(link weight) for weighted scale-free networks. The other parameters are $N = 1000$, $m_0 = m = 5$, $W = 2$.

In Ref. [15], Hu proposed a local routing strategy based on link weight information for weighted scale-free networks. He found in most cases, the optimal value of the parameter α characterizing the preferential probability in choosing links to forward packets appears at α>1.0, which means taking advantage of links with large bandwidth. However, in Fig.1 which reflects the traffic load distribution among the links, one can see when α >1.0, with α increasing, packets will continuously accumulate on links with large weights more. That implies only a few links with large weights entering the jammed state, instead of most of the links, leads to the whole system entering the jammed state. With adaptive preferential factor, the value is roughly constant for all the link weights. Thus the capacities of all weights are more fully used, and all the links enter the jammed state at the same time.
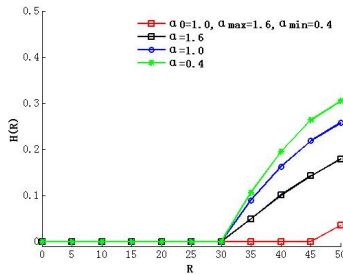
Fig. 2. (Color online) The order parameter $H(R)$ versus R for weighted scale-free networks with different routing strategy. The other parameters are $N = 1000$, $m_0 = m = 5$, $W = 2$.
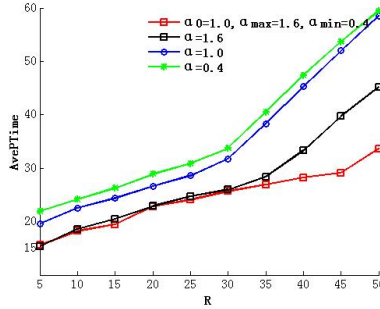
Fig. 3. (Color online) Average packets delivering time versus R. with different routing strategy .The other parameters are $N = 1000$, $m_0 = m = 5$, $W = 2$.

In Fig.2, compared with fixing $\alpha = 1.6, 1.0, 0.4$, the critical generating rate R_c with adaptive α is bigger, in fact up to 45. This indicates that the routing strategy with adaptive α maximized the overall capacity of the system by fully using all links.

Fig.3 illustrates average packets delivering time increases with decreasing α. However, by using adaptive α, average packets delivering time is not larger than fixing α=1.6. Especially, when $R > 30$ with adaptive α, average packets delivering time is much smaller than fixing α=1.6 .That's the result of bigger critical generating rate R_c .

5   Summaries

In this paper, a local routing strategy with a single adaptive preferential factor which characterizes the preferential probability in choosing links to forward packets on scale-free networks is proposed. The adaptive preferential factor α increases when the queue length of the link is smaller than the capacity, and decreases otherwise. Our simulation results show that the capacities of all links are more fully used and compared with fixed value of α, using adaptive preferential factor α could both reduce the packets delivering time and increase the capacity of huge communication networks.

References

[1]   Watts, D.J., Strogatz, S.H.: Collective dynamics of small-world networks. Nature 393, 440–442 (1998)
[2]   Barabasi, A.L., Albert, R.: Emergance of scaling in random networks. Science 286, 509–512 (1999)

[3] Pastor-Satorras, R., Vázquez, A., Vespignani, A.: Dynamical and correlation properties of the Internet. Physical Review Letter 87, 258701 (2001)

[4] Albert, R., Jeong, H., Barabasi, A.L.: Diameter of the World Wide Web. Nature 401, 130–131 (1999)

[5] Zhou, T.: Mixing navigation networks. Physica A 387, 3025–3032 (2008)

[6] Yan, G., Zhou, T., Hu, B., Fu, Z.-J., Wang, B.-H.: Efficient Routing on Complex Networks. Physical Review E 73, 046108 (2006)

[7] Danila, B., Yu, Y., Marsh, J.A., Bassler, K.E.: Optimal Routing on Complex Networks. Physical Review E 74, 046106 (2006)

[8] Noh, J.D., Rieger, H.: Random walks on complex networks. Physical Review Letter 92, 118701 (2004)

[9] Wang, S.-P., Pei, W.-J., He, Z.-Y.: Random Walks On the Neural Network of C.elegans. Physica A 387, 4699–4708 (2008)

[10] Yang, S.-J.: Exploring complex networks by walking on them. Physical Review E 71, 016107 (2005)

[11] Wang, W.-X., Wang, B.-H., Yin, C.-Y., Xie, Y.-B., Zhou, T.: Traffic dynamics based on local routing protocol on a scale-free network. Physical Review E 73, 026111 (2006)

[12] Yin, C.-Y., Wang, B.-H., Wang, W.-X., Yan, G., Yang, H.-J.: Traffic dynamics based on an efficient routing strategy on scale free networks. Euro. Phys. Journal B49(2), 205–211 (2006)

[13] Wang, W.-X., Yin, C.-Y., Yan, G., Wang, B.-H.: Integrating local static and dynamic information for routing traffic. Physical Review E 74, 016101 (2006)

[14] Wang, W.-X., Wang, B.-H., Hu, B., Yan, G., Ou, Q.: General dynamics of topology and traffic on weighted technological networks. Physical Review Letter 94, 188702 (2005)

[15] Hu, M.-B., Jiang, R., Wu, Y.-H., Wang, W.-X., Wu, Q.-S.: Routing on a weighted scale-free network. Physica A 387, 4967–4972 (2008)

[16] Arenas, A., Díaz-Guilera, A., Guimerá, R.: Communication in networks with hierarchical branching. Physical Review Letter 86, 3196–3199 (2001)

A Virus Spread Model Based on Cellular Automata in Weighted Scale-Free Networks

Ning-Ning Wang and Guo-Long Chen[*]

Department of Computer Science, Fuzhou University, China
ningwang1361@yahoo.cn, cgl@fzu.edu.cn

Abstract. In this paper, based on cellular automata, a virus spread model, is proposed to investigate the virus spreading in weighted scale-free networks. In this model, each node of the network can be in one of four states: susceptible, exposed, infective and recovered. The transition rate from S to E of each susceptible node is related to its neighbors' states and its connected weights with neighbors. Through some useful results obtained by the analysis and simulations, we can understand virus spreading in weighted scale-free networks better.

Keywords: Weighted scale-free networks, Virus spreading, Cellular Automata.

1 Introduction

Complex networks can properly describe many biological, social and technological networks[1-3] with vertices standing for individuals and links denoting the interactions between them. As a kind of complex networks, Scale-free networks(SF network) have attracted much attention due to the ubiquity of a power-law degree distribution in real-life networks. Epidemic spreading, one of the typical dynamical processes built on Scale-free networks, has been investigated intensively.

Most recent studies on epidemic models are focused on unweighted scale-free networks. for example, in 2005, Barthelemy et al. [4] studied the susceptible-infected(SI) model in Barabasi-Albert (BA) networks. Yang et al. [5] proposed a modified susceptible-infected-recovered(SIR) model. In [6], Zhou et al. studied behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. Fu et al. examined epidemic thresholds for disease spread using susceptible-infected-susceptible models with variable infectivity in [7]. However, in lots of real networks one of the most important properties is the weight. For instance, on the internet the weight can stand for the knowledge of its traffic flow or the bandwidths of routers [8]. Especially, In epidemic spreading studies, the weight can indicate the extent of the contacting between nodes, the larger the weight the more possibly a susceptible node will be infected through the edge. So we intend to provide a virus spread model in weighted SF networks in this paper.

Most epidemic models in recent studies use differential equations. Nevertheless, they have serious drawbacks. Firstly, what they describe is just the average tendency

---

[*] Corresponding author.

of virus spreading in which the probabilistic event is disregarded. Furthermore, they neglect the local character of the spreading process, and do not include variable susceptibility of individuals. In advance, due to the lack of spatial concept, it is difficult to describe the spatial-temporal process of virus spreading. To overcome those drawbacks, we propose a virus spread model based on Cellular automata (CA) in weighted Scale-free networks. CA is a discrete dynamical system whose individual components are rather simple, yet that can exhibit highly complex and unpredictable behavior due to these simple components' mutual interaction and synergy[9,10].

The rest of the paper is organized as follows: Cellular Automata on weighted graph is briefly introduced in Section 2. The virus spread model based on CA is formulated in Section 3. In Section 4, the simulations of virus spreading based on the proposed model are given. Section 5 concludes this paper.

2   Cellular Automata on Weighted Graph

A weighted graph G^w is a pair (V,E,W) where V is an non-empty finite set of elements called nodes (or vertices), E is a finite family of pairs of elements of V called edges, and W is the set of weights of edges. Two nodes of the graph, $i, j \in V$, are said to be adjacent (or neighbors) if there exists an edge in E of the form (i, j), and the weight of the edge is w_{ij}. We consider undirected graphs, that is, $(i, j)=(j, i) \in E$.

A cellular automata on an undirected weighted graph $G^w=(V,E,W)$ is a 4-uple $A =(V,S,N,r)$ where: The set V is the cellular space, in which each cell(node) can be in one of a finite number of distinct states at each discrete time. S is the finite state set whose elements are all the possible states of the cells. $S_i(t)$, the state of the cell i at t-th time is defined as in formula (1). It is denoted by $S_i(t-1)$ and the state of its neighborhood by $S_{Ni}(t-1)$, where N_i denotes the neighborhood of cell i. N_i is defined as the set of cells which are adjacent to i, that is: $N_i=\{j \in V$ such that $(i, j) \in E\}$. At last, r denotes the local transition rules, which specifies how the dynamics run for the CA. According to the transition rules, at the next time each cell transforms from its current state to a new state based on its current state and the states of its neighbors.

$$S_i(t) = r(S_i(t-1), S_{N_i}(t-1)) \qquad (1)$$

3   The Proposed Model

The model is constructed as follows:

Cells and Cellular Space: All nodes in the network are cells. The set consisting of all cells is the cellular space (that is V).

State set: Our model is based on the traditional SEIR model [11]. The SEIR model has four states: S (susceptible), E (exposed), I (infective), and R (recovered). Susceptible nodes have not been infected by the virus. Exposed nodes are those that have been infected but not yet infective. On infective nodes, the virus is activated and it is able to propagate to another node. Recovered nodes, those having acquired immunity to virus, will not be infected.

Neighborhood: N_i is the neighborhood of cell i. s_i, the strength of a node v_i, is defined as the sum of the weights of the edges between v_i and its neighbors in formula (2).

$$s_i = \sum_{j \in N_i} w_{ij} \qquad (2)$$

Transition rules: The state transition chain in classic SEIR is S→E→I→R. We add the follow state transition paths to it. (1) S→R: it represents that some susceptible nodes take countermeasure of immunization before being infected; (2) E→R: it is to describe the fact that some nodes in E remove virus and gain immunization without activating the virus; (3)E→S: it is added because some nodes in E may succeed to clean the virus before becoming infectious but fail to acquire immunity; (4)I→S: given that some infectious nodes may get rid of the running virus after obtaining a part of technical support (antivirus), however they do not take immune measures so that they reenter the state S. The state transition diagram of cell i is shown in Fig.1.
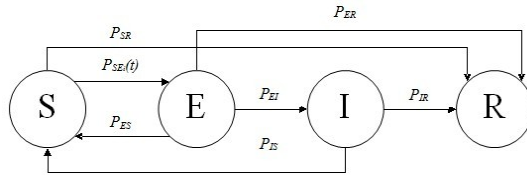
Fig. 1. The state transition diagram of cell i

Suppose cell i is at discrete time $t-1$, the transition rules at t-th time is as follows:

Rule 1: if i is in state S, next time it becomes recovered with probability p_{SR}; or it turns to be exposed with infection rate $p_{SE_i}(t)$ if it has neighbors in state I, and $p_{SE_i}(t)$ is defined in formula (3),

$$p_{SE_i}(t) = 1 - \prod_{j \in N_i^{\text{infectious}}} \left(1 - \lambda_{ij}\right) \qquad (3)$$

where $N_i^{\text{infectious}}$ is the set of all i's infected neighbors at time t, and λ_{ij}, the spreading rate for i being infected by j($j \in N_i^{\text{infectious}}$),is defined in formula (4),

$$\lambda_{ij} = \alpha \left(\frac{w_{ij}}{w_{i\max}}\right)^{\beta} \quad (\alpha > 0; \beta \geq 0, w_{i\max} = \max_{k \in N_i} w_{ik}) \qquad (4)$$

where β and α (α is the spreading rate for I being infected by k having the max weight with i, when k is in state I.) are two modulating parameters, and $w_{i\max}$ is the largest value of w_{ik} ($k \in N_i$). When $\beta = 0$, our definition in formula (4) corresponds to the constant infection rate in the traditional epidemic models.

Rule 2: if i is in state E, then it transforms to be recovered with probability $p_{ER;}$ or reenters state S with p_{ES}; or it becomes infectious with probability p_{EI}.

Rule 3: if i is in state I, next time it is cured and transits to state R with probability p_{IR} ; or it becomes susceptible again with probability p_{IS}.

4   Simulation

We perform the simulations of our virus spread model on a weighted scale-free network model proposed by Wang et al. [12], in which power-law distributions of degrees, weights, and strengths are all in good accordance with real observations of weighted technological networks. The parameters N(network size, that is, the total number of nodes), $m0, m, W$ of   network are set as 1000,5,5,2 respectively.

In simulation, $I(t)$ denotes the number of infected nodes(the infected nodes contain exposed and infectious nodes) at t-th discrete time. Spreading velocity $v(t)$ is defined in formula (5). Let $I(0)=1/N$,we assume that the initially infective node is randomly chosen from all nodes on the given network. Each result is the average of 100 simulations.

$$v(t) = \frac{I(t+1)-I(t)}{N} \tag{5}$$

(1) Effect of infection parameter α
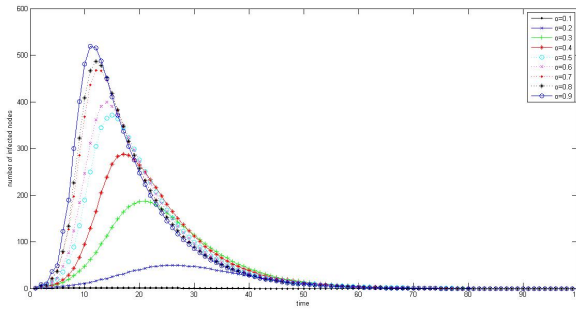
Fig. 2. $I(t)$ vs time with increasing α from 0.1 to 0.9(0<α≤(1-p_{SR}=0.95)). Other parameters are β=1.0, p_{SR}=0.05, p_{IS}=p_{ES}=0.01, p_{ER}=0.02, p_{EI}=0.5, p_{IR}=0.1.

Fig.2 shows simulation results for different spreading rates ranging from 0. 1 to 0.9. the larger α is, the steeper the curve is, that is, the more rapidly number of infected nodes increases. The susceptible nodes are infected faster with α increasing. As α raises, the peak of the curve is greater, which represents that more susceptible nodes are infected.

(2) Effect of infection parameter β
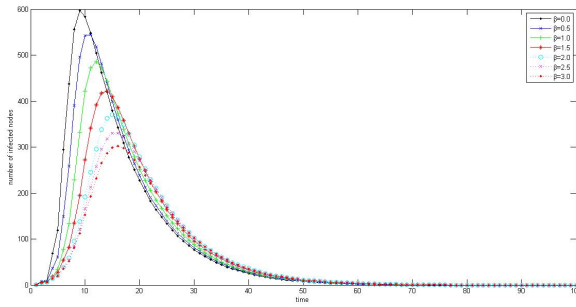
(a)
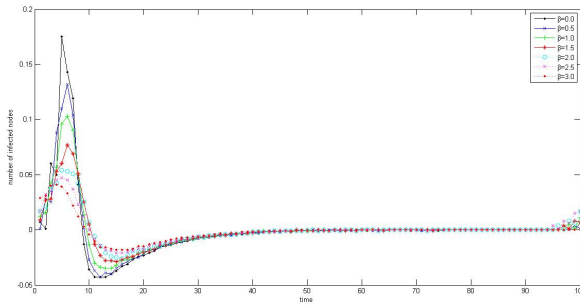
(b)

Fig. 3. *I(t)* and spreading velocity *v(t)* vs time with β=0.0, 0.5, 1.0, 1.5, 2.0, 2.5 and 3.0 respectively. Let α=0.8, p_{SR}=0.05, p_{IS}=p_{ES}=0.01, p_{ER}=0.02, p_{EI}=0.5, p_{IR}=0.1.

Fig.3(a) shows that the smaller β is, the more quickly virus spreads and the larger the total number of infected nodes is. In Fig.3(b), obviously, the spreading velocity goes up to a peak quickly that similar to the situation in BA network, leaving us with very short response time for the deployment of control measures. *I(t)* increases at the fastest rate in the peak of spreading velocity at the *t-th* time, which is defined as the moment of the outbreak [13], showing a transition from the small incidence of the virus to a macroscopic outbreak. The larger β the greater the peak of *v(t)* is, therefore the virus outbreaks faster.

(3) Effect of p_{EI}
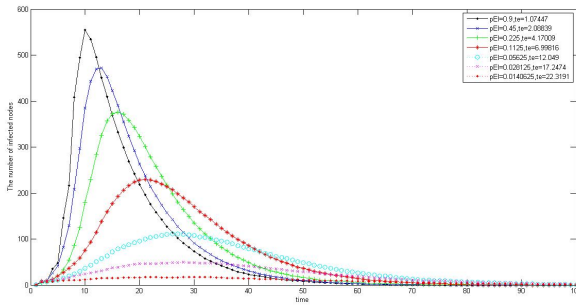
Fig. 4. $I(t)$ vs time with p_{EI} =0.9, 0.45, 0.225, 0.1125, 0.05625, 0.028125 and 0.0140625 respectively. Other parameters are α=0.8, β=1.0, p_{SR}=0.05, p_{IS}=p_{ES}=0.01, p_{ER}=0.02, p_{IR}=0.1.

In Fig.4, average incubation period t_e is the average time for exposed nodes staying in state E. With p_{EI} decreasing, t_e increases and the peak of $I(t)$ declines, thus virus breaks out more slowly, and fewer susceptible nodes are infected.

(4) Effect of the strengh s of initially infective node
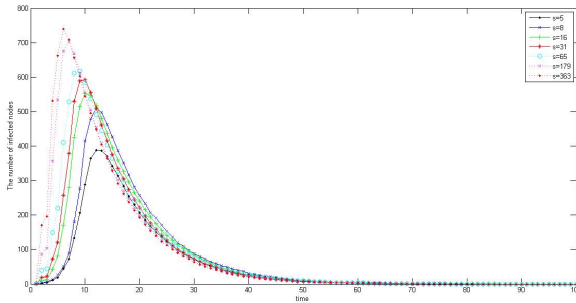
Fig. 5. $I(t)$ vs time when s=5, 8, 16, 31, 65, 179 and 363 respectively. Other parameters are α=0.8, β=1.0, p_{SR}=0.05, p_{IS}=p_{ES}=0.01, p_{ER}=0.02, p_{IR}=0.1, p_{EI}=0.5.

In Fig.5, the total number of infected nodes increases and the outbreak of virus becomes quicker with s increasing. Therefore, in immunization strategy, we should give priority to the immunization of nodes with larger s to weaken the spreading of virus.

5   Summaries

In this paper, we propose a virus spread model based on cellular automata in weighted scale-free networks. The model add some possible state transition paths neglected by classic SEIR. Differing from being a constant in tranditional model, the rate for

different susceptible nodes being infected varys in our model. With analysis and simulations of the proposed model in a weighted scale-free network, some useful results are acquired, such as the effect of some parameters on the number of infected nodes and spreading velocity. It contributes to a better understanding on virus spreading in weighted scale-free networks.

References

[1] Albert, R., Barabasi, A.-L.: Statistical mechanics of complex networks. Reviews of Modern Physics 74, 47–97 (2002)

[2] Newman, M.E.J.: Ego-centered networks and the ripple effect. Social Networks 25, 83–95 (2003)

[3] Dorogovtsev, S.N., Mendes, J.F.F.: Evolution of Networks: From Biological nets to the Internet and WWW. Oxford University Press, Oxford (2003)

[4] Marc, B., Alain, B., Romualdo, P.-S., Alessandro, V.: Dynamical patterns of epidemic outbreaks in complex heterogeneous networks. Theoretical Biology 235, 275–288 (2005)

[5] Yang, R., Wang, B.-H., Ren, J., Bai, W.-J., Shi, Z.-W., Wang, W.-X., Zhou, T.: Epidemic spreading on heterogeneous networks with identical infectivity. Physics Letters A 364, 189–193 (2007)

[6] Zhou, T., Liu, J.-G., Bai, W.-J., Chen, G., Wang, B.-H.: Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. Physical Review E 74, 056109 (2006)

[7] Fu, X., Small, M., Walker, D.M., Walker, D.M.: Epidemic dynamics on scale-free networks with piecewise linear infectivity and immunization. Physical Review E 77, 036113 (2008)

[8] Pastor-Satorras, R., Vespignani, A.: Evolution and Structures of the Internet: A Statistical Physics Approach. Cambridge University Press, Cambridge (2004)

[9] Fuentes, M.A., Kuperman, M.N.: Cellular automata and epidemiological models with spatial dependence. Physica A 267, 471–486 (1999)

[10] White, S.H., Rey, A.M.d., Sanchez, G.R.: Modeling epidemics using cellular automata. Applied Mathematics and Computation 186, 193–202 (2007)

[11] Anderson, R.M., May, R.M.: Infectious Diseases of Humans: Dynamics and Control. Oxford University Press, Oxford (1992)

[12] Wang, W.-X., Wang, B.-H., Hu, B., Yan, G., Ou, Q.: General dynamics of topology and traffic on weighted technological networks. Physical Review Letters 94(18), 188702 (2005)

[13] Barthelemy, M., Barrat, A., Pastor-Satorras, R., Vespignani, A.: Velocity and hierarchical spread of epidemic outbreaks in complex networks. Physical Review Letters 92, 178701 (2004)

An Improved Channel Allocation Scheme in IEEE 802.11s Mesh Network

Linhan Feng, Zhihong Qian, and Dongcheng Jin

College of Communication Engineering
JiLin University
Changchun, China
{Fenglinhan,jindongcheng}@jl.chinamobile.com, dr.qzh@163.com

Abstract. With the increasing demand on high quality wireless broadband access service, the traditional access technologies become inadequate. As a new and promising next generation network, wireless mesh network draws researchers' attention by its advantages of large coverage, low deployment cost, convenient access and so on. In this paper, we research on the performance of channel allocation scheme in IEEE 802.11s WMN. The property of the traditional fixed assignment algorithm and the mesh interference model are discussed. Then an improved channel allocation scheme based on these will be introduced and implemented. The simulated performance of the new improved algorithm will be shown as the result.

Keywords: Wireless mesh network, Channel allocation scheme, Greedy algorithm, Ripple effect, Channel interference.

1   Introduction

Wireless mesh network (WMN) is a promising wireless technology for several emerging and commercially interesting applications. Different from the traditional network, WMN is dynamic self-configured, self-organized and self-healing. This characteristic makes it more attractive with lower up-front cost, easier network maintenance work, more robust network architecture and more stable business coverage. Because of wireless mesh network's traffic merge by employing the multi-hop technique, the performance of data stream drops rapidly after the multi-hop transmissions. The traditional wireless access technology cannot satisfy the requirement of larger range coverage because only supporting on the single hop and being set near the wired network. Thus the focal point of the research works on wireless mesh network is to how to reduce the competition and interference among links and improve the network capability.

The major problem faced with WMN is the scarce bandwidth. Wireless channel is characterized by broadcast nature, when   a link between two ends are occupied, other nodes within its coverage range cannot transmit which cause the low throughput of single channel WMN. Thus the users' requirements on wide bandwidth applications cannot be satisfied[5,6]. Hence, utilizing multiple orthogonal channels to process the datagram transmission will improve the throughput evidently. So establishing

effective way of resource assignment has great significance on the performance improvement of WMN. Resource assignment for WMN contains two aspects: one is channel allocation mechanism; the other one is multi-channel routing protocol. The former aspect is our main issue of this paper-the channel allocation algorithm.

This paper firstly introduces the theme of the traditional fixed channel assignment algorithm and the model of the interference for IEEE 802.11s WMN. Then based on these discussions, an improved channel allocation scheme will be introduced which contains evaluation and allocation two stages. The new algorithm will be implemented and the simulated. The ripple effect probability and the interference performance comparing with the original allocation scheme will be shown as the result.

2   Related Work on the Channel Allocation Scheme in IEEE 802.11s WMN

IEEE 802.11s task group (TG) is founded to solve the multi-hop transmission problems in wireless local area network (WLAN). In IEEE 802.11s wireless network, mesh is defined as a set of nodes that are connected by radio link. These points are characterized by the ability of automatic topology and dynamic path selection[2]. The typical architecture of the IEEE 802.11s wireless mesh network includes two types of nodes: mesh point (MP) and non-mesh point. The mesh point means the nodes that support the wireless mesh service like path selection and packet forwarding; the non-mesh point is defined as the terminal stations (STA). Except for the mesh service, mesh access point (MAP) can also provide the radio access service. STA cannot be directly added to mesh network whereas it can enter to mesh network.

In IEEE 802.11s draft standard, the MAC layer behavior is introduced and the PHY layer is designed with extensibility to other IEEE 802.11 standard, like IEEE 802.11a, IEEE 802.11b and IEEE 802.11g. As mentioned above, channel allocation is the issue in multi-channel system which is one scheme utilized to improve the throughput. Utilizing multiple orthogonal channels to process the datagram transmission will decrease the interference and improve the throughput evidently. The multiple channels on PHY layer are available in IEEE 802.11 standards, where IEEE 802.11b/g and IEEE 802.11a provide 3 and 12 non-overlapping channels respectively.

However, a pair of STAs must share a common channel to set up a virtual link. On another hand, in order to reduce the interference, mesh STAs should reduce the number of neighbor STAs who share the common channel. Thus, our goal of the channel allocation scheme is to finding a balance point between network connectivity and reduction of interference. To give the improved algorithm, we must consider two related basic concept:

2.1   Fixed Channel Assignment Scheme

We assume the mesh STAs of WMN is on the high configuration so that the multi-interface with multi-channel mode is suitable. One mesh STA can utilize multiple

channels to send and receive packet data, so the end-to-end delay drops and end-to-end bandwidth improves markedly. Under this condition, we often take the fix channel assignment.

Fixed allocation schemes pre-assign each channel to each interface, allocation scheme stays the same until the algorithm is reactivated. Many schemes of this kind have been put forward. And Greedy algorithm is involved in most of them. The advantage is that no overhead of channel coordination and no need to modify the current MAC protocols. The disadvantage is low flexibility to adapt to dynamic business. Ripple effect is unwanted and algorithms should try to reduce and avoid it.

2.2 Interference Model

Link interference model is used to analyze the interference c when using a certain channel. we define the set of interference is *Intf(i)*. The measurement of the interference for Link *i* can be simply calculated by the following equation:

$$I(i) = \sum\nolimits_{j \in Intf(i) \text{ and } j \neq i} T(j) \tag{1}$$

3 The Improved Channel Allocation Algorithm

In this part, the improved channel allocation algorithm is introduced. Because of the currently existing hardware's incapability of fast switch and huge amount of overhead, our work focused on the fixed allocation scheme. The actual throughput available to application is halved when all the overheads are accounted, not to mention more overheads imported by dynamic coordination and switch.The algorithm has two stages: evaluation stage and allocation stage. The first stage is responsible for collecting the necessary inputs for the second stage. At the same time, it also determines if the allocation stage should be activated. Next I will introduce them in details.

3.1 Evaluation Stage

In the process of the establishment of end-to-end model, every mesh STA (mesh portal gateway is also included) gathers statistics regularly, and then sends the traffic to the mesh gateway node. The gateway node establishes the traffic model on the basis of the traffic information just received. After that, comparison between the current traffic situation and the last traffic situation is made by gateway node to determine if there is a need to reallocate the channels, namely activating the channel allocation algorithm. If the traffic does not have a significant change, the channel allocation scheme is stayed the same; otherwise, channel allocation algorithm is activated to reallocation the channels. We proposed a weighted average index process aiming at obtain the long term stable traffic information that should be conducted by portal mesh node.

$$T(s,d) = \alpha T_{old}(s,d) + (1-\alpha)T(s,d) \tag{2}$$

After the model established, a traffic matrix stores the expected link load of each link should be computed to provide basic information for the next stage. The expected link load of Link i can be found by:

$$\Phi(i) = \sum_{s,\,d} P_i(s,d) B(s,d) / P(s,d) \tag{3}$$

Where s and d are the source and destination STAs, $\Phi(i)$ is the expected link load of Link i, $P(s, d)$ is the total number of acceptable paths between STA pair s and d, $Pi(s, d)$ is the number of acceptable paths which go through Link i between STA pair s and d.

3.2   Allocation Stage

Most of the fixed channel allocation schemes utilize greedy algorithm at the allocation stage. By taking this idea, interference reduction and traffic awareness are implemented. However, ripple effect may exist to increasing the algorithm complexity by revisiting some STAs that have already visited. This is a common issue in C-Hya algorithm. The improved scheme is to force each STA could be visited only once. So the STAs are arranged to stand in line waiting to be visited before the links stand in line, which can be considered as a more greedy method. After this process, the independency of STA is eliminated. Thus the ripple effect is ameliorated a lot. However, it cannot be avoided completely, the problem of link revisiting still may happen. But via this new method, the problem is faced only when the radio and channel resources are extremely scarce compared to the scale of network, e.g. 25 nodes network with each of them equipped with 2 two radios (Further observation and discussion on this will be made in next part). We also utilized greedy algorithm when dealing with the issue of STA queuing. We consider the mesh STA undertaking more traffic and have the worst working conditions should be visited first which means it has the largest priority. The working conditions are the minimum hops from the mesh portal point and the number of radios. As mentioned above, wireless mesh network is characterized by traffic convergence, which means the nearer to the gateway, the more traffic may undertake. And more number of radios will augment node's capability. So referring to literature [9], the node priority can be found by:

$$Node(i).Priority = \sum Link(j).LinkLoad / Node.HopsCount * NumOfRadio \tag{4}$$

where $Link(j)$ belongs to $Node(i)$. Then we come to the secondary greedy algorithm. According to the number of unassigned radios of two STAs, there will be three scenarios: Two STAs have unassigned radios; One STA has unassigned radio but the other does not; Two STAs have no unassigned radio. For the first scenario, we come to an improved point of mine. Other algorithms choose the channel with the lowest traffic as the best choice. However if there existed channels allocated out of the interference range, one of them can be reused as the best choice. Generally, researchers consider the interference-to-communication ratio is between 2 to 3 for a typical value [3]. Fig.1 is the flow chart showing the complete algorithm.
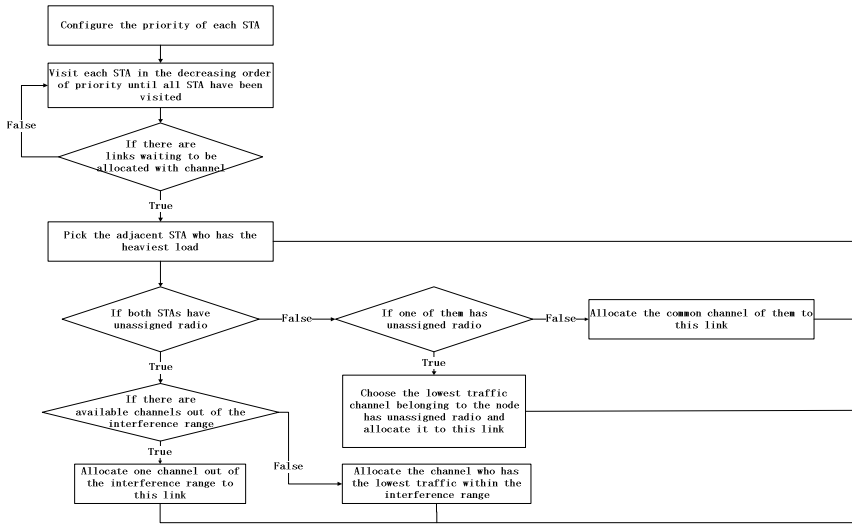
Fig. 1. Improved allocation algorithm

4   The Improved Channel Allocation Algorithm

We will discuss the improvement algorithm performance on two aspects: ripple effect and interference reduction. We create a 5x5 grid network topology. Assume the center STA is the mesh portal gateway node. All the necessary network parameters are configured. For the ease of experiment, the load of each link is randomly generated between the interval (1, 54), thus the first stage is simplified. We conduct the experiment with the number of radios equipped with each STA set to 2 or 3 and number of non-overlapping channels set to 3 or 12(corresponding to 802.11b/g and 802.11a, respectively).

4.1   Ripple Effect

Firstly, we explore the approximate occurrence condition of ripple effect to validate the effectiveness of eliminating ripple effect by involving node queuing. Four situations should be conducted to get a comprehensive result which is shown in Tab.1.

Table 1. Cutting parameters of simulation milling

	2 radios, 3 channels	2 radios 12 channels	3 radios 3 channels	3 radios 12 channels
Probability of ripple effect happening	9%	35%	0%	0%

For the first configuration, 2 radios with 3 channels, the probability is not a significant value, almost can be ignored. But when it comes to the second configuration, 2 radios with 12 channels, the probability gets an obvious increment. This is because the mechanism I used in the allocation process. The mechanism would allocate the common channel to the link preceding other channels. When the number of non-overlapping channels is increased to 12, the common channel shared by the pair of STAs may not exist due to the more diversity of prior allocation compared with the scenario of 3 non-overlapping channels. The rest situations both have a probability of 0%. Hence, we can deduce that the sufficient physical resources can reduce the probability of ripple effect happening and the number of radios has a greater influence than the number of non-overlapping channels. In order to reduce it adequately, the scale of network topology, the number of radios and the number of non-overlapping channels should have a certain matching relationship.
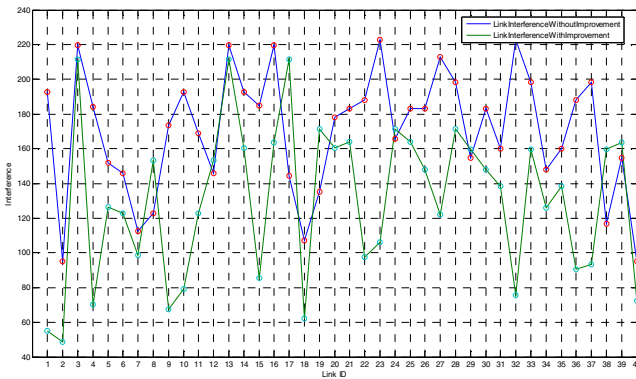
Fig. 2. Traditionally making the car model

4.2   Reduction on Interference

Then validate if the link interference can be reduced with interference range is considered. Fig.2 shows the contrast result of the interference between improved and unimproved channel allocation scheme. From it, we can clearly see that most of the links get a smaller interference except link 8, 12, 17, 19, 24, 29, 38 and 39. The number of inoperative nodes only account for 20%. For 80% of them, the involvement of interference does serve the intended purpose. It reduces about 23% interference for each link, which is a rather considerable improvement on link interference. Hence, the effectiveness of the involvement of interference range has been verified.

6   Summaries

This paper lays the stress on the research of fixed channel allocation in IEEE 802.11s wireless mesh network. In the major factors of consideration, network connectivity is usually the unyielding pursuit of target and the interference is the unwanted adversity.

Algorithm research is made under the principle of assuring connectivity and reducing interference as much as possible, extensive study is made to get a better idea for improvement. With the foreseeable future maturity of advanced technologies in PHY layer, the current bottle-necks will be freed to broaden the design angle of channel allocation. So they are the future plans that should be started to give key considerations.

References

[1] IEEE P802.11s[TM]/D5.0. Draft STANDARD for Information Technology—Telecommunications and information exchange between systems—Local and metropolitan area networks—Specific requirements—Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications—Amendment 10: Mesh Networking (April 2010)

[2] Hossain, E., Leung, K.: Wireless Mesh Network Architectures and Protocols. Springer Science + Business Media (2008)

[3] Chiueh, A.R.: Architecture and algorithms for an IEEE 802.11-based multi-channel wireless mesh network. In: Proc. IEEE Infocom, pp. 2223–2234 (2007)

[4] Skalli, H., Ghosh, S., Das, S.K., Lenzini, L., Conti, M.: Traffic and interference aware channel assignment for multi radio Wireless Mesh Networks. Technical Report, IMT Lucca (2006)

[5] Jain, K., Padhye, J., Padmanabhan, V., Qiu, L.: Impact of Interference on multi-hop wireless network performance, In: Proceedings of ACM 9th Annual International Conference on Mobile Computing and Networking, MobiCom 2003, pp. 66-80 (2003)

[6] Xu, S., Safadawi, T.: Does the IEEE 802.11 MAC Protocol Work Well in Multi-hop Wireless Ad Hoc Networks? IEEE Communications Magazine, 130–137 (2001)

[7] So, J., Nitin, Vaidya, H.: Multi-channel MAC for Ad Hoc Networks: Handling Multi-Channel Hidden Terminals using a Single Transceiver. In: Proceedings of the ACM International Symposium on Mobile Ad Hoc Networking and Computing, MobiHoc (2004)

[8] Bahl, P., Chandra, R., Dunagan, J.: SSCH: Slotted Seeded Channel Hopping for Capacity Improvement in IEEE 802.11 Ad-Hoc Wireless Networks. In: Proceedings of ACM 10th Annual International Conference on Mobile Computing and Networking, MobiCom (2004)

[9] Skalli, H., Ghosh, S., Das, S.K., Lenzini, L., Conti, M.: Channel Assignment Strategies for Multi-radio Wireless Mesh Networks: Issues and Solutions. IEEE Communications Magazine, Special Issue on "Wireless Mesh Networks" (2007)

Weighted Networks of Object-Oriented Software Systems: The Distribution of Vertex Strength and Correlation

Gao Yang[1], Liu Jia[1], Shao Shuai[1], Xu Guoai[1], and Cheng Gong[2]

[1] Information Security Center, Beijing University of Posts and Telecommunications
Beijing, P.R. China
[2] National Computer Network and Information Security Administration Center
Beijing, P.R. China

Abstract. In this paper, empirical investigation results on the distribution of vertex strength and correlation of four open software systems have been presented with the view of reasonable description of software structure. We make use of the software coupling weighted networks to characterize and describe the macroscopic properties of four open software systems. Implement examining of the distribution of vertex strength for software coupling weighted networks and find that the in- strength distribution have the scale-free property but the out- strength distribution can be well described by stretched exponential distribution. And then we calculate correlation of the software coupling weighted networks, and draw a conclusion that the network is disassortative mixing in unweighted and undirected versions, and negative correlation between in-strength and out- strength of software coupling weighted networks has existed.

Keywords: software coupling weighted networks, vertex strength, SED, correlation.

1 Introduction

Software architecture affects function and performance of the software systems, the internal attributes of software is essential for quality control. A reasonable description and effective quantification of software structure is the basis for an in-depth study of software system complexity [1-2].

The recent progress in complex network theory motivates many researchers to pay attention to the complexity of software structure as complex networks and have made many achievements [3-5]. Such as, it is found that software network follows the power-law distribution and small world property to varying degrees. Myers unearthed the scale free property in software network after analyzing a significant number of C and C++ software system and shown that the differences between in-degree and out-degree have existed [6]. Valverde made a discovery of the scale-free and small-world property of object-oriented (OO) software after investigating the inheritance and interface relationship in some OO systems, including JDK etc. using the undirected graph module [7-8]. Afterwards, according to granularity of software entities,

researchers examined quantities of open source applications at the level of packages, classes and methods, thus further validated the conclusions above [9-11]. Also some researchers proposed software metrics to measure the structural complexity of software systems based on complex networks [12-13].

In our work, we examine the distributions of in- and out-vertex strength based on software coupling weighted networks (SCWN), and found the differences between them. We use stretched exponential distribution (SED) which was proposed by Laherrete and Sornette in 1998 [14] to describe the out- strength distributions of SCWN and find that it is better than power law distribution. Thereafter, we empirically investigate the correlation of in- and out- strength and summarize the general features.

The organization of the paper is as follow. In section 2, we introduce the vertex strength of SCWN. In section 3, we study the distributions of in- and out- strength of four open sources OO software systems to get the conclusion that the differences between in- and out-strength distributions have existed. In section 4, empirical investigation results on correlation of in- and out- strength in SCWN have been presented.

2   Vertex Strength of Software Coupling Weighted Networks

In our early work, a new light on understanding complex coupling interactions among software components has been brought, and then constructing software coupling weighted network(SCWN)[15].

A significant measure of the weighted network properties in terms of the actual weights is obtained by extending the definition of vertex degree $k_i = a_{ij}$ in terms of the vertex strength s_i, defined as [16]

$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij}$$

This quantity measures the strength of vertices in terms of the total weight of their connections.

Barrat's work has focused on undirected network, In the directed SCWN, we employ $s_{in}(i)$ and $s_{out}(i)$ to denote the in-strength which measures the strength of a given vertex i in terms of the total weight of its incoming links and out- strength which counts the strength of a given vertex i in terms of the total weight of its outgoing links, and $s_{tot}(i)$ represents the total weight of all links of vertex i in SCWN. Hence, we have

$$s_{in}(i) = \sum_{j=1}^{N} a_{ij} w_{ij}, \quad s_{out}(i) = \sum_{j=1}^{N} a_{ji} w_{ji}, \quad s_{tot}(i) = \sum_{j=1}^{N} (a_{ji} w_{ji} + a_{ij} w_{ij})$$

The vertex strength count the number of the statements processed. The in-strength of a vertex is the number of statements sent out from the class, which is equal to the number of method invocations. Adversely, the out-strength of a vertex is the count of the number of received statements sent out from other classes to this class, which is the number of methods invoked.

In the following sections, we have empirically analyzed the distribution of vertex strength and correlation of SCWNs for four different open-source software systems including Apache Ant 1.5.2、Apache Tomcat 6.0.20、J2SDK 5.0 、Eclipse 3.5.1, and report on some basic statistical results about SCWN.

3   The Distribution of Vertex Strength

In our previous work, we have found that the topology of SCNs exhibit small-world property, and reconsider the degree distributions of the in-degree, out-degree and total-degree of four SCNs.

For the degree distributions, in our previous experiments of SCN, we confirm that the asymptotic existence of power laws of in-degree is well validated in the log-log plots, whereas the relevance of power laws of out-degree is less clear-cut, and thus SCN is not strictly scale free for the distribution of out-degree. However, the out-distributions can be better described using stretched exponential distribution (SED) which is given by $P(x)dx = \mu(x^{\mu-1} / x_0^{\mu})\exp(-(x/x_0)^{\mu})dx$ and its accumulative distribution is $P(x) = \exp(-(x/x_0)^{\mu})$, where μ is between 0 and 1, the degree distribution is between a power law and an exponential function.

And then, by fitting the distribution of in- and out-vertex strength in the log-log plots, we have found the same properties of the distribution of vertex strength and degree. The in-strength distribution have the scale-free property and the out-strength distribution have an exponential decay. We have fitted the distribution of out-strength by SED again and found that fitting curve is linear (Fig. 1).

(a)	(b)	(c)	(d)
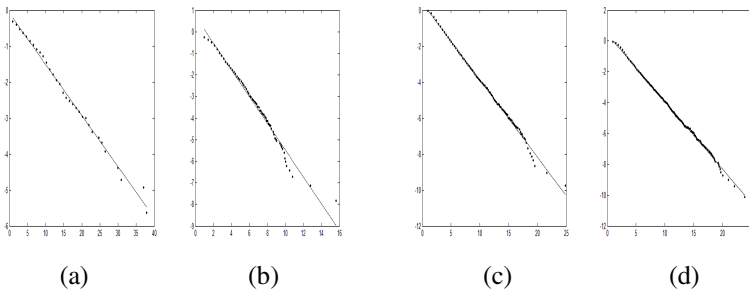
Fig. 1. Cumulative distribution with SED of out-strength for the four systems: (a) Ant, (b) Tomcat, (c) JDK and (d) Eclipse

We find the distributions of out-strength appears a significantly curvature compare with the distributions of in-strength in the log-log plots. Tab.1 shows the exponents μ of SED scaling of the distributions of in-strength and out-strength for four SCNs, which reveal the distributions of in-strength tend to extend to 0 (that is, they fit power law distribution) and the distributions of out-strength cross over to an exponential decay.

Table 1. The exponents of SED of the distributions of in- and out-strength

Exponents	Ant	Tomcat	JDK	Eclipse
μ_{out}	0.5	0.4	0.4	0.4
μ_{in}	0.07	0.09	0.06	0.07

The in-strength may indicate how dependent the implementation of the local methods is on the methods in other classes. The smaller in-strength a class has, the more independent the class is, the easier it is to reuse in another classes. So there are a few of vertexes with large in-strength and the others which is a vast part with small in-strength. On the other hand, a class with large out-strength may be difficult to reuse in another context, so the exponential decay has existed, which suggests that the out-strength distributions tend to be homogeneous.

4   Correlation

We examine the relationship between in- and out-strength in the SCWNs, and Tab.2 shows the correlation coefficient. From Tab.2, the relationship between in- and out-strength is weak correlation.

As shown in Fig.2, we can see that, normally, there are few classes with both large in-strength and large out-strength in the SCWNs; that is, if the out-strength of a class is large, its in-strength is generally small, and vice versa. These classes which both have large in- and out-strength have remarkable complexity and could be problematic.

Table 2. The correlation coefficient between in- and out-strength

Exponents	Ant	Tomcat	JDK	Eclipse
μ_{out}	0.5	0.4	0.4	0.4
μ_{in}	0.07	0.09	0.06	0.07

| (a) | (b) | (c) | (d) |
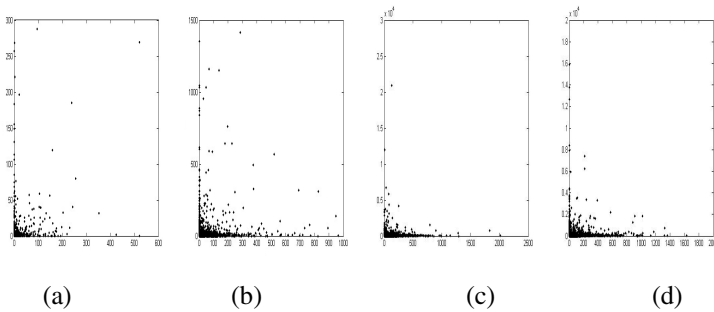
Fig. 2. Scatter plot of the number of in- and out-strength for the four systems: (a) Ant, (b) Tomcat, (c) JDK and (d) Eclipse

5   Summaries

In this work, we employ a directed software coupling weighted network to characterize and describe the macroscopic properties of software systems. By empirical analysis of software coupling weighted networks of four open source software systems, we survey the in-vertex strength and out-vertex strength in a different perspective. We confirm that the property of scale-free has existed in in-strength distributions and fit the curve of the out-strength distribution with stretched exponential distribution (SED), that is, an exponential decay has been possessed in out-strength distributions. In addition, we examine the correlation of in- and out-strength of software coupling weighted network, there is a remarkable weak correlation in software coupling weighted network. Furthermore, we analyzed the causes of above behaviors from the perspective of software develop technology, develop process and developer's habit respectively.

References

[1]   Briand, L.C., Daly, J.W., Wüst, J.K.: A Unified Framework for Coupling Measurement in Object-Oriented Systems. IEEE Transactions on Software Engineering 25, 91–121 (1999)

[2]   Dijkstra, E.W.: The structure of the "T. H. E." multiprogramming system. Communications of ACM 11, 453–457 (1968)

[3]   Valverde, S., Ferrer, R., Solé, R.V.: Scale Free Networks from Optimal Design. Europhys. Lett. 60, 512–517 (2002)

[4]   Concas, G., Marchesi, M., Pinna, S.: Power-laws in a large object-oriented software system. IEEE Transactions on Software Engineering 33, 687–708 (2007)

[5]   Moura, A., Lai, Y.C., Motter, A.E.: Signatures of small-world and scale-free properties in large computer programs. Physical Review E 68, 017102 (2003)

[6]   Myers, C.R.: Software systems as complex networks: Structure, function, and evolvability of software collaboration graphs. Physical Review E 68, 046116 (2003)

[7]   Valverde, S., Solé, R.V.: Hierarchical Small Worlds in Software Architecture, Working paper of Santa Fe Institute, SFI/03-07-44 (2003)

[8]   Valverde, S., Solé, R.V.: Logarithmic growth dynamics in software networks. Europhys. Lett. 72, 858–864 (2005)

[9]   Potanin, A., Noble, J., Frean, M., et al.: Scale-free geometry in OO programs. Communications of the ACM 48, 99–103 (2005)

[10]   Baxter, G., Frean, M., Noble, J., et al.: Understanding the shape of Java Software. ACM SIGPLAN Notices 41, 397–412 (2006)

[11]   Concas, G., Marchesi, M., Pinna, S., et al.: Power-laws in a large object-oriented software system. IEEE Transactions on Software Engineering 33, 687–708 (2007)

[12]   Liu, J., He, K., Ma, Y.T., Peng, R.: Scale Free in Software Metrics. In: Proc. The 30th Annual International Computer Software and Applications Conference, Chicago, pp. 229–235 (2006)

[13] Ma, Y.T., He, K.Q., Du, D.H.: A Qualitative Method for Measuring the Structural Complexity of Software Systems Based on Complex Networks. In: Proc. of 12th Asia-Pacific Soft. Eng. Conf., Taiwan, pp. 257–263 (2005)

[14] Laherrere, J., Sornette, D.: Stretched exponential distributions in nature and economy: "fat tails" with characteristic scales. Eur. Phys. J. B 2, 525–539 (1998)

[15] Xu, G.A., Gao, Y., Liu, F.F., Chen, A.G., Zhang, M.: Statistical Analysis of Software Coupling Measurement Based on Complex Networks. In: 2008 International Seminar on Future Information Technology and Management Engineering, pp. 577–581 (2008)

[16] Barrat, A., Barthelemy, M., Pastor-Satorras, R., Vespignani, A.: The architecture of complex weighted networks. PNAS 101, 3747–3775 (2004)

Security Evaluation Methods of Computer Networks Based on BP Neural Network

Liping Li

Computer Center
Laoning University of Technology
Jinzhou, Liaoning Province, China
lg_llp@sina.com

Abstract. An index system of security evaluation of computer network system is constructed based on the tree-shaped security evaluation model of computer network system in terms of the problem of security evaluation of computer network system. According to the principle of BP neural network, the BP security evaluation model of the neural network information system is constructed and the specific steps are discussed. For the purpose of solving the problem of the relatively larger security indexes in computer network system, the paper proposes the fuzzy comprehensive evaluation method to preprocess evaluation indexes in order to streamline the complexity of the evaluation system.

Keywords: BP neural network, security evaluation, fuzzy comprehensive evaluation.

1 Introduction

At present the computer network systems have been widely used in various fields. The computer network systems are becoming bigger and bigger and their structures are more and more complex. Therefore, the security problems of computer network systems have become increasingly striking, which has become the biggest obstacle to the applications of computer networks. Solving security problems of the computer networks has become the most important and basic work in developing computer network application.

At present, many scholars have applied the asset evaluation theory to information security evaluation and have obtained some results [1, 2]. But the methods of the risk assessment need some prior knowledge, such as the probability of threats, the assignment of the intangible assets, etc. It is difficult to get these accurate data in practice. Some studies use such methods as fuzzy logic [3], the rough set theory [4] and the gray theory [5] to establish evaluation models of information security, but most of them focus on the evaluation of local systems. Applying the BP neural network theories to security evaluation of the network systems can present unified metrics to evaluate the risk indexes of various factors. Then according to certain evaluation methods, the risk ranks of the whole system are given to obtain more practical solutions to the security evaluations. The evaluation indexes have been processed beforehand in order to implement the security evaluation of the corresponding methods proposed in the paper so that they can be applied to this method.

2   The Basic Model of the Safety Evaluation of the Computer Network Systems

As fig.1 shows, this paper discusses a basic model of the security evaluation of the computer network systems based on the security model of the security tree-shaped structure of the information systems. According to the effect of the various security factors in the computer network system on the overall security of the system and the degrees of the effect, the security of the whole system is measured.
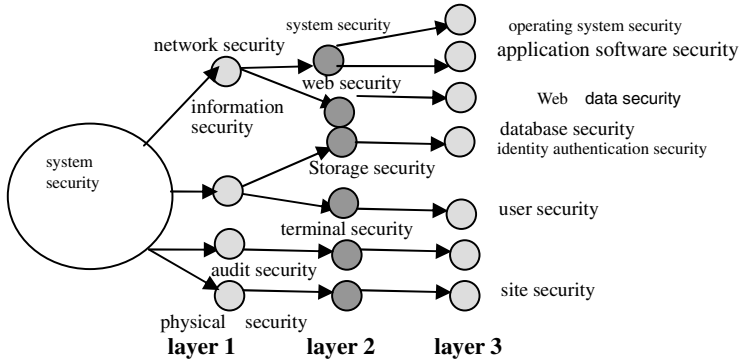
Fig. 1. A Tree-shaped Security Evaluation Model of the Computer Network System

Because the index system of the security evaluation of the computer network systems contains many factors, fig.1 briefly presents some typical factors influencing the security of the computer network system. To begin with, the original security evaluation factors (abbreviated for factor analysis) are analyzed and classified and the appropriate nodes in the tree-shaped model are chosen to construct the security evaluation model. The security factors in layer 2 in fig.1 include more than 30 factors, such as "data encryption", "undeniable signature" and "identity authentication security", etc. The factors in layer 1 can be reduced to eight factors, including "web security", "network operation security", "information storage security". Each factor in layer 1 generally contains a number of factors in layer 2.

Table 1. Security Evaluation Indexes of the Computer Network

root	security factors in layer 1	security factors in layer 2
security of computer network system	system security	operating system security application software security
	network operation security	data backup and recovery firewall security LAN security
	Web security	security of web application servers web database security
	information storage security	database security terminal security
	information audit security	audit security of the information content verification security of the main information
	security of the computer rooms environment	security of the security management system site safety
	hardware protection security	protection devices of the equipment information input/output control measures

3   The Evaluation Methods Based on BP Neural Network

3.1   BP Neural Network

The group of scientists headed by Rumelhart and McCelland put forward BP (Back Propagation) network in 1986. BP neural network, which is widely used at present, is a multilayer feedforward network trained by an error anti-transmission algorithm. BP network can learn and store a lot of mapping relationships of input-output models without prior revealing the mathematical equations to describe the mapping relation. The topology structure of BP neural network model includes an input layer, a hidden layer and an output layer.
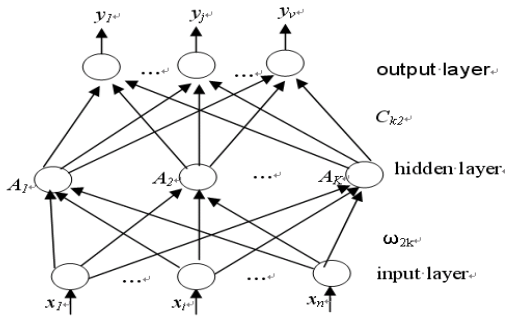
Fig. 2. Three layer feedback structure of BP neural network BP network

3.2   Security Evaluation Model of a Computer Network Based on BP Neural Network

As is shown in Fig.2 in which security evaluation model of a computer network is constructed based on BP Neural Network. The model is composed of four modules: the module of parameter pretreatment, BP network, sample input and evaluation results. The module of parameter pretreatment normalizes the processing of the input data. BP network is the core of the assessment model which forms a reasoning mechanism through a great deal of learning, realizing the nonlinear reflection of the input evaluation index. The sample input module inputs the learning samples to the knowledge-base and finally outputs evaluation results through the evaluation result module.
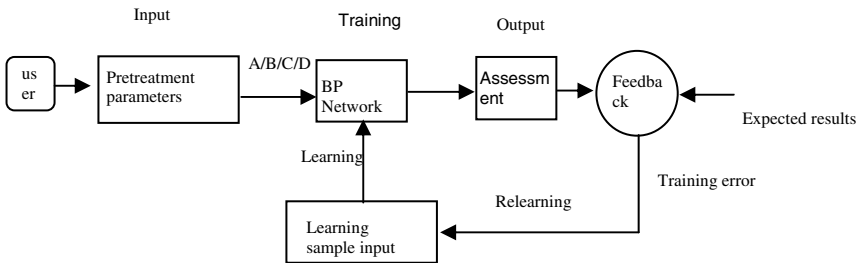
Fig. 3. A system security assessment model based on BP neural network

Evaluation procedures are as follows:

(1)Training the Neural Network
The input, which takes the security index of the computer network as BP neural network, obtains the evaluation result by the calculation of the neural network. Compared with the result of the experts' evaluation, the result in question gives the error feedback to the hidden layer of the neural network, adjusts the proportions of the weights in the parameters of each layer and evaluates the input indexes again until the result of the evaluation and the result of the experts' evaluation meet the demand. BP neural network that meets the demand is obtained after multi-group training.
(2) Evaluation of the Computer Network
The security index of the computer network being the input, the evaluation result is obtained after the calculation of the neural network.
(3) Parameter Pretreatment
Because there are more security evaluation factors in the computer network system, both the complexity of the neural network and the calculation time will be greatly increased if the second-class evaluating index is adopted as the input of the neural network. The index needs retreating in order to reduce the input amount of the index. This paper adopts the fuzzy comprehensive evaluation to streamline the assessment index.

4   Pretreatment Evaluation Indicators

4.1   Theory of Fuzzy Set

Theory of fuzzy set is proposed by Zadeh who is a cybernetics expert in the United States in 1965 , it uses fuzzy set to describe the   fuzzy numbers. Given the universe U, in the U, a fuzzy number can be described by the fuzzy set F on U [90]. Membership function: μF: $U \rightarrow [0, 1]$, for any u ∈ U,   μF (u) said that u belongs to membership grade of fuzzy set F , fuzzy set F is :

F = {μF (u1)/u1, μF (u2)/u2, ..., μF (un)/un}

When U is an infinite set, fuzzy set F is: $F = \int_{u \in U} \mu_F (u)/u$

4.2   Fuzzy Evaluation Function

Definition 1: A function of n variables f: $[0, 1]^n \rightarrow [0, 1]$ satisfied the following conditions:

(1) Boundary conditions: $f(0, 0,..., 0) = 0, f(1, 1,..., 1) = 1$;

(2) Monotonicity: if $x_i \le x'_i$, $f(x_1, x_2,,..., x_n) \le f(x'_1, x'_2,,..., x'_n)$

(3) Continuity: $\lim_{x_i \to x_{i_0}} f(x_1, x_2,..., x_n) = f(x_{1_0}, x_{2_0},..., x_{n_0})$

(4)Additivity: $f(x_1 + x'_1,..., x_n + x'_n) = f(x_1,..., x_n) + g(x'_1,..., x'_n)$

So f is called evaluation function. g: $[0, 1]^n \rightarrow [0, 1]$

4.3   A Integrated Evaluation Model

To evaluate a factor or a part of a certain object according to the criteria established is called single evaluation. To get the overall evaluation from a certain object is called integrated evaluation. According to the evaluation function mentioned above can draw the following model of integrated evaluation: $A \circ R = B (b_1, b_2, …, b_m)$ among which:

$$A = (a_1,...,a_n) , \quad \sum_{i=1}^{n} a_i = 1 , \quad a_i \geq 0 \quad R = (r_{ij})_{n \times m} , \quad r_{ij} \in [0,1] \qquad b_j = \sum_{i=1}^{n} a_i r_{ij} ,$$

$j = 1,…, m$

b_j is a function belongs to $r_{1j}, r_{2j}, …, r_{nj}$, is evaluation function

For the integrated evaluation, there are three elements:

(1) factor set $U = \{U_1, …, U_n\}$, a collection of various factors which are judged object ;
(2) determination set $V = \{ V_1, …, V_n \}$, a collection of reviews;
(3) single factor determination, also is a judgment for sigle factor U_i (i = 1, …, n),can draw the fuzzy set $(r_{i1}, r_{i2}, …, r_{im})$, it is a fuzzy mapping from U to V, f: $U \rightarrow \mu (V), \quad U_i \longmapsto (r_{i1}, r_{i2}, …, r_{im})$

According to Theorem 1, fuzzy mapping f can determine a fuzzy relation $R \in \mu_{n \times m}$, which is called evaluation matrix.

$$R = \begin{pmatrix} r_{11} & r_{12} & … & r_{1m} \\ r_{21} & r_{22} & … & r_{2m} \\ … & … & & … \\ r_{n1} & r_{n2} & … & r_{nm} \end{pmatrix}$$

It is composed of F set which consists of the single factor evaluation. As the status of each factor may not equal, the factors must be weighted. Using the F set on U A = (a1, a2, …, an) expresses that the weight distribution of each factor, synthesized the evaluation matrix R is a integrated evaluation of various factors, so we can draw the evaluation model mentioned above.

5   Summaries

According to the characteristics of the security of the computer network and the techniques of BP neural network, the paper constructed security evaluation model of a computer network based on BP Neural Network. In terms of the unique advantages of the neural network in numerical parallel computing of the complex systems and the power of the fuzzy system in dealing with the expert's knowledge and experience, the neural network is evaluated by combining the fuzzy comprehensive evaluation methods with BP neural network.

References

[1] Zhao, D., Liu, H., Liu, C.: Risk Evaluation of Information Safety Based on BP neural network. Computer Engineering and Application 43(1), 139–142 (2007)

[2] Liu, H., Wang, W., Cai, G.: A Safety Evaluation Model of Information System Based on Neural Network. Computer Engineering and Science (11), 16–18 (2008)

[3] Zhu, L., Dai, G.: An Evaluation Model of Information Security of Fuzzy Comprehensive Evaluation. Journal of Information Security and Communications Confidence 8(2), 13–15 (2006)

[4] Qing, Y., Zuo, C., Liu, D.: A Safety Evaluation Model Based on the WSNs Routing of the Optimized BP Neural Network of. Journal of Software 6(9), 11–13 (2008)

[5] Liu, P., Liu, M.: The Application of Gray Theory in the Evaluation Model of Information Security. Journal of Micro Computer Information 22(12), 95–97 (2006)

[6] Li, X., Liu, J.: Security Evaluation of the Electronic Commerce system Based on Fuzzy Comprehensive Evaluation. Computer Engineering and Design 30(17), 4002–4006 (2009)

[7] Skowron, A., Rauszer, C.: The Discernibility Mat Rices and Functions in Information Systems. In: Intelligent Decision Support-Handbook of Applications and Advances of the Rough Set Theory, pp. 331–362. Kluwer Academic Publishers, Netherlands (1992)

Modularized Real-Time Communication Modem Design Based on Software Defined Radio of Underwater Acoustic Network

Zhi Shaolong, Fang Dong, Liu Xun, Li Yu, and Huang Haining

Institute of Acoustics, CAS, Beijing, 100190

Abstract. Demand of underwater acoustic network communications is growing faster since the development of Interest of Thing. The definition and model of underwater communication modem is presented in this paper. This paper also discusses a design method of real-time underwater acoustic communication modem based on Software-Defined-Radio and Modularization. This method includes the ideas of SDR to design universal signal processor as the core of system and modularized transmitters and receivers. The core is in all-digital mode. Transmitters, receivers, transducers with different performances and types connect to the core as blocks through the system bus. This approach has portability and compatibility to adapt to the rapid development and continuous updating of the communication system and technology. A example of this method based on dual-core DSP was implemented in the paper. Lake experiment proved the feasibility of design method.

Keywords: Underwater network, Modularization, Software-Defined-Radio, dual-core DSP.

1   Introduction

As one of the five national emerging strategic industries, Internet of Things promotes development of a series of new technologies and plays an important role in industrialization and information technology. 70% of earth's surface is ocean, underwater wireless network is therefore an indispensable part of Internet of Things [1][2]. An underwater wireless network as shown in Fig.1.
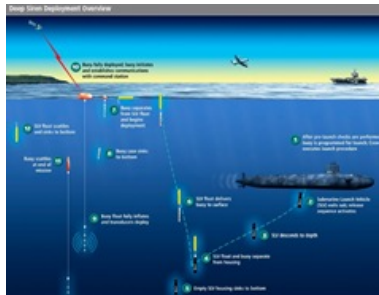
Fig. 1. Underwater wireless network

Sound wave is by far the most available carrier of information transmission underwater. Underwater acoustic wireless communication technology is the foundation of underwater wireless network. The original hardware-based method for underwater acoustic communication modem design cannot meet application demands which are more and more rapidly developing [3][4][5][6]. For example, Orthogonal Frequency Division Multiplexing (OFDM) technology is an effective technology against frequency selective fading for short-range high-speed transfer of large amounts of data while the direct spread spectrum (DSS) technology has a good anti-jamming capability for remote low-speed of text communications. However, the existence of high Peak-to-Average Power Rate (PAPR) is the main difficulty of OFDM and there is an obstacle of long time reliability of high-power transmit in spread spectrum system [7] [8]. Therefore different communication systems require different designs; hardware-based design methodology brings a lot of work.

Underwater communication modem is an underwater acoustic communication system that is made up of signal processor, transmitter, receiver, transducer, power source, system bus and certain mechanical structure with stable, reliable, effective information transfer capability.

Some key technologies of underwater communication modem:

1) universal signal processor, real-time processing and power consume
2) stable high power and high efficiency transmitter
3) superior performance receiver
4) stable, credible, available system bus
5) transducer technology
6) reliable and efficient power technology
7) the mechanical structure of solid and convenient

The diversity of performances such as signal frequency, system bandwidth and communication distance, etc. leads to the different design projects of transmitter, receiver and transducer. Modularization design concept is that the universal signal processor is the core of the system, transmitter, receiver, transducer with different performances and types in the form of module are connected to the core by system bus to integrate the modem unite. Reasonable division in accordance with structure and function improves the expansibility and compatibility of system.

Characteristics of software Defined Radio (SDR) method are compatible and open [9]. Universal signal processor design is based on idea of SDR. Specifically, SDR is to load different software to the universal signal processing platform to achieve different communication functions. To achieve interaction of information under different communication system by using different software which implements different communication standards and protocols to constitutes highly flexible acoustic communication equipment.

System bus connects every parts of the modem as link. Bus protocol must meet speed and bandwidth requirements of data transmission inside system. Different bus distributions constitute different system topologies. Relative to the RF signal, underwater acoustic signals of low frequency, narrow bandwidth, common bus (e.g. SPI bus) or the special high-speed data interface between the same family of digital signal processing devices (e.g. Linkport of SHARC series DSP) could satisfy the data transmission requirements.

System power is the source of energy for communication modem. Communication modem requires stable power supply. Transmitter requires a large transient current, rated voltage and rated current must meet the transmitter's maximum emission power.

In addition, most of underwater communication modems work in the closed environment. The working mechanism of the system must therefore be optimized to reduce system power consumption as much as possible.

Through analysis of characteristics of underwater acoustic signal and working mechanism of underwater communication modem above, it is feasible that the method of SDR and modularization used in real-time underwater acoustic communication modem design.

This paper describes the system model of underwater communication modem. Design method of SDR and modularization in real-time underwater acoustic communication modem, including universal signal processor, transmitter, receiver and transducer is put forward. A realization of this method to constitute communication modem of dual-core DSP based on the QPSK modulation is approached. Lake experiment results are given in the end.

2   System Description and Design Principles

Underwater acoustic communication modem is not only a terminal in point-to-point underwater communication, but also the network node which could acquire ocean information, detect underwater target in applications based on network. Underwater acoustic communication system model as shown in Fig.2.
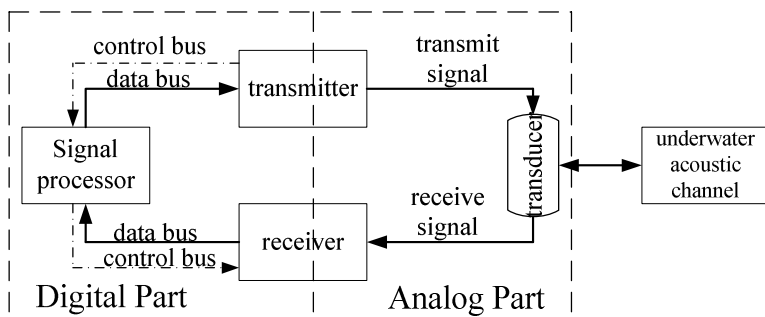
Fig. 2. Underwater acoustic communication system model

(1) Signal Processor

As the core of communication modem, signal processor is the computing platform for all underwater acoustic communications algorithms, including signal modulation and demodulation, channel coding and decoding, synchronization, equalization, error correction and other functions. Underwater target detection, location, tracking and other algorithms based on network, data transmission and relay, network organization and control are also implemented in signal-processor.

According to the idea of SDR, all-digital mode is introduced in universal signal processor design. Signal processor processes digital signal from the receiver and

sends modulated digital signal to the transmitter as shown in Fig.2. Hence, compared to the signal processor, transmitter is equivalent to a Digital-to-Analog Converter (DAC) and receiver is equivalent to an Analog-to-Digital Converter (ADC). Besides, the signal processor is also responsible for the businesslike affairs and control tasks of communication modem, such as power management, underwater environmental information monitoring, system status self-test, etc.

Signal processor's computing capability should meet real-time, accuracy, speed and other requirements of algorithms. But high capability usually leads to increased power consumption. Therefore, the optimization of the algorithms and improvement of the processor work mechanism are needed to prepare before system design. In addition, the signal processor is also required to equip reliable, flexible and compatible peripheral interfaces that take control tasks and businesslike affairs of communication modem.

For characteristics of underwater acoustic channel and working mechanism of underwater acoustic communication modem, "nDSPs + FPGA" mode is proposed in signal processor as shown in Fig.3.
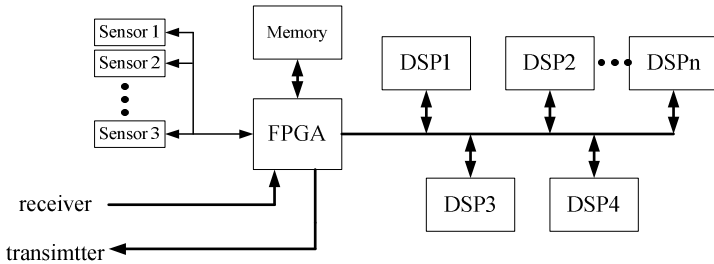
Fig. 3. "nDSPs + FPGA" mode of signal processor

(2) Receiver

Receiver obtains communication signals from transducer. Receiver manages transducer signals and converts them to digital ones. The receiver with superior performance is the premise of effective communication. Transducer is usually broadband, the role of the receiver is before the signal processing, to extract the useful signals form the system bandwidth and suppress noise in order to increase Signal to Noise Ratio (SNR).

Signal first passes through the preamplifier, filter and voltage gain control amplifier (VGA). Then it is converted to digital bits by the high-precision ADC. Digital signal transfers to signal processor through the system bus.

In underwater acoustic communication, it is necessary to minimize the power consumption of the receiver because the receiver has been working all the time.

(3) Transmitter

Transmitter sends communication signals to transducer. Modulated signal from processor is power amplified by transmitter to drive the transducer. High-power, small size, light weight, high efficiency and high reliability of the transmitter are prerequisites for achieving the underwater acoustic communication.

The types of power amplifiers are A, B, AB, D, etc. In underwater field, AB and D class power amplifiers are common used. Before loading into the transducer, signal has to be optimized according to the parameters of transmitter and transducer. Adjustment of the parameters of output matching networks for matching impedance and resonant frequency improves the efficiency of electro-acoustic energy conversion.

(4) Transducer

The role of the transducer is to convert electronic and acoustic signal into each other. Functionally, transducers can be divided into the transmit transducer, receive transducer. In order to save cost and size, the combined set of transmit and receive transducer is selected. Transducer converts electrical signal into acoustic signal while transmitting and converted acoustic signal into electrical signals while receiving.

Underwater acoustic communication modem with this kind of transducer is of half-duplex mode. The manufacture process of transducer is very complex. Some important parameters of it are the voltage response, power response and sensitivity. As transmitter's load and receiver's front, transducer plays a fatal role in communication. It is necessary to design transducer in cooperate with both of transmitter and receiver.

Transducer is the interface between underwater acoustic communication modem and underwater channel. Match of transmitter and transducer consists of two aspects:

   a) resonant frequency matching
   b) impedance matching

The underwater acoustic transducer is generally capacitive. In order to achieve resonant frequency and impedance matching, there should be a match circuit between transmitter and transducer which is made up of matching inductor, matching capacitors and other components.

Match of receiver and transducer is mainly impedance matching. Preamplifier input impedance must be much larger than the transducer output impedance to drive signals.

(5) System bus

The system bus has two meanings:

   a)   Transfer protocols of signal processor, receiver and transmitter
   b)   Hardware and circuit design of them

Optimized system bus protocol and improved circuit interface design could promote the transmission capacity. System bus design should consider transfer rate, reliability, versatility, etc. Because underwater acoustic communication signal bandwidth is quite narrow and digital signal processing algorithms don't have so much data to transfer, common bus, such as Data Address Bus, SPI protocol, PPI protocol could satisfy data transmission of the system.

3   Modem Design and Experiment Based on Dual-Core DSP

The ADSP-BF561 processor is a high performance member of the Blackfin family of products. At the heart of this device are two independent Analog Devices Blackfin processors：CoreA and CoreB. In another words, BF561 could be seen as two DSPs.

BF561 processor has 328K bytes of on-chip memory. Each Blackfin core includes its own memory [10].   Additionally, there is on-chip memory L2 SRAM. CoreA and CoreB have access to L2 memory, L2 is used to store shared data.

In underwater acoustic networks, the system is divided into the physical layer, data link layer, network layer and application layer. The Key technology of dual-core DSP is how to schedule internal system resources, including interruptions, memory space, DMA controllers, etc. and access to external devices. Such resource is limited.

In the dual-core DSP, tasks of coreA and coreB are divided as follows:

CoreA: synchronization, equalization

CoreB: decoding, demodulation

Signal processor working flow as shown in Fig.4. Universal signal processor is shown in Fig.5 and modularized receiver is shown in Fig.6.
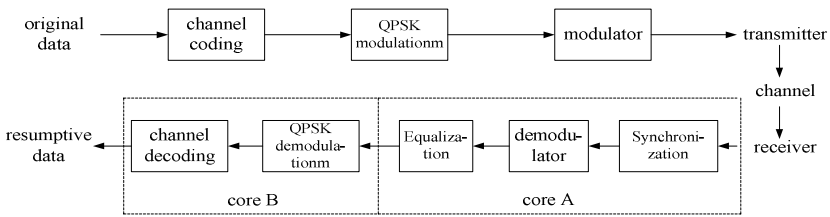
Fig. 4. Signal processor working flow
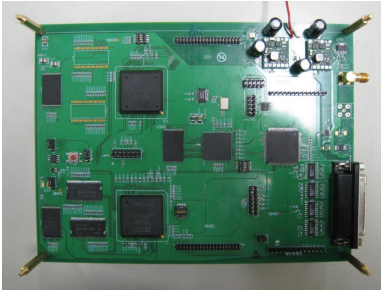
Fig. 5. Signal processor
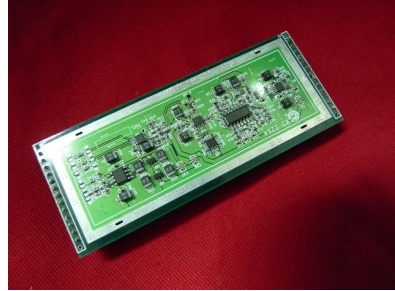
Fig. 6. Modularized receiver

Experiment in Qiandaohu Lake， Zhejiang Province in the April 2011 tested the system. The received waveform, constellation as shown in Fig.4 and Fig.5 ; Table 1 shows the experimental conditions

Table 1. Experimental conditions

communication system	depth(m)	transmit acoustic source level(dB)	data rate (kbps)	distance(km)
QPSK	10-12	165	8	8
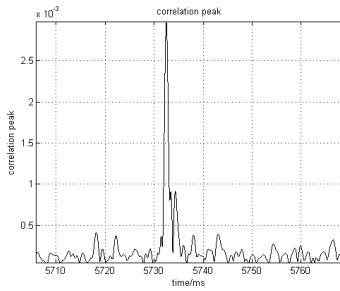
Fig. 7. Result of synchronization
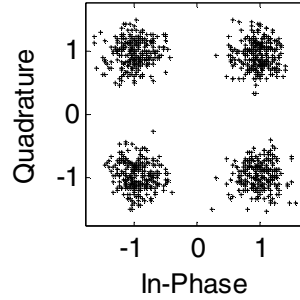
Fig. 8. Result of demodulation

Result of synchronization is shown in Fig.7 and result of demodulation is shown in Fig.8.

4 Summaries

Requirements of underwater acoustic communication modem design are more advanced as the rapid development of the underwater network. The original design method based on hardware cannot keep up with the update rate of communication system and technology. A modularized modem is advanced in this paper. Universal signal processor is designed by method of SDR with strong computing and businesslike affairs processing capability; transmitter, receiver, transducer, etc. for different communication bandwidth are connected to the core through a certain bus to form a flexible communication modem. This approach is of portability and compatibility. This paper implemented an underwater acoustic communication platform based on dual-core DSP. Design of algorithm and hardware was completed. Lake experiment proves that the system has practice value and the design method is feasible.

References

[1] Tan, L., Wang, N.: Future Internet: The Internet of Things. In: 2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), vol. 5, pp. 376–380 (2010)

[2] Dohr, A., Modre-Opsrian, R., Drobics, M., Hayn, D., Schreier, G.: The Internet of Things for Ambient Assisted Living. In: 2010 Seventh International Conference on Information Technology: New Generations (ITNG), pp. 804–809 (2010)

[3] Jeon, J.-H., Hong, C.-G., Park, S.-J., Kim, C., Kim, S.: Robot Control Using an Underwater Acoustic Modem. In: 2010 IEEE/IFIP 8th International Conference on Embedded and Ubiquitous Computing (EUC), pp. 331–336 (2010)

[4] Won, T.-H., Jeon, J.-H., Park, S.-J.: An Underwater Communication Modem Using TMS320C6416 DSK. In: International Conference on Computational Science and Engineering, CSE 2009, pp. 978–981 (2009)

[5] Benson, B., Li, Y., Faunce, B., Domond, K., Kimball, D., Schurgers, C., Kastner, R.: Design of a Low-Cost Underwater Acoustic Modem. IEEE Embedded Systems Letters, 58–61

[6] Nowsheen, N., Benson, C., Frater, M.: Design of a high frequency FPGA acoustic modem for underwater communication. In: OCEANS 2010, pp. 1–6. IEEE, Sydney (2010)

[7] Park, M., Jun, H., Cho, J., Cho, N., Hong, D., Kang, C.: PAPR reduction in OFDM transmission using Hadamard transform. In: 2000 IEEE International Conference on Communications, ICC 2000, pp. 430–433 (2000)

[8] Flikkema, P.G.: Spread-spectrum techniques for wireless communication. IEEE Signal Processing Magazine, 26–36

[9] Marojevic, V., Reves, X., Gelonch, A.: Computing Resource Management for SDR Platforms. In: IEEE 16th International Symposium on Personal, Indoor and Mobile Radio Communications, PIMRC 2005, pp. 685–689 (2005)

[10] ADSP-BF561 datasheet, Rev. E, Analog Devices, Inc. (2009), http://www.analog.com

Routing Optimizing Algorithm of Mobile Ad-Hoc Network Based on Genetic Algorithm

Wu Xiao-Yan and Liu Yang

Polytechnic School of Shenyang Ligong University
Fushun 113122

Abstract. A routing optimizing arithmetic of Mobile Ad-Hoc Network Based on genetic algorithm is designed. The routing calculated is correct and convergence velocity is quicker, which shows the performance of the arithmetic is favorable.

Keywords: Mobile Ad-Hoc network, Routing optimizing, Genetic algorithm.

1   Introduction

Mobile Ad hoc network is a multi-hop temporary autonomous system which composed of a mobile terminal including a group of wireless transmitters and receivers. Different from traditional wireless networks, mobile ad-hoc network needs neither a fixed network structure nor a special fixed site or router as the network center. In the network, each node has a router function, which is responsible for discovering and maintaining the router to the other nodes, and transmitting packets to the neighbor nodes. In the mobile ad-hoc networks, because of the limited transmission range of terminal wireless, the two end nodes which can not directly communicate often transmit through multiple intermediate nodes, and the route between nodes usually compose of multi-hop.

QoS routing technique of mobile ad-hoc network is one of the difficult and direction in the current studying of routing technology [2]. The important index of optimizing network resources is to reduce the cost of network resources. The cost is relevant to the hops if the delay of network we can ignore. The more hops, the greater cost of the network resources, so an important goal of reducing the cost of network resources is to reduce the hops which the route goes through. Therefore, the paper uses genetic algorithm to realize routing optimizing between the source node and destination node, the number of hops as a measure of router, between the source node and destination node to select a relatively short route from numbers of accessible routes.

2   Introduction of Genetic Algorithms

Genetic algorithm (GA) was put forward by Professor J. Holland from the United States in 1975 in his book 《Adaptation in Nature and artificial systems》. Genetic algorithm is an efficiently random searching and optimizing algorithm based on the evolution. Since the whole searching strategy and optimizing calculate of genetic algorithm does not depend on the gradient information, so it has been successful

applied in many fields because of its widely used. [3][4][5]. Genetic algorithm may not get the optimal solution, but it can get a better solution quickly. When the iteration time is big enough, it may get the optimal solution or nearest optimal solution. So the genetic algorithm can make good compromise between searching optimal solution and searching efficiency. The main advantage of genetic algorithm for optimizing problem as follow: It can solve several optimization problems only through using the value of adaptive function but not functional derivative or other assistant information; It also has a global optimization, robustness and implicit parallelism. It can be real-time choosing the best path in real-time communication environment, high dynamic topology and the QoS requirements of mobile Ad Hoc network.

Genetic algorithm is a searching algorithm through the iterative process of generate-and-test. Macroscopically, stepping for the genetic algorithm is basically fixed. Genetic algorithm includes coding, population initialization, calculating of adaptive function, choosing, crossing, variation operation and other key content. A typical diagram of the genetic algorithm implementation is shown in Fig.1. Fig.1 shows the evolution ideology which genetic algorithms relies on: the initial population generate a better next-generation groups through continuous crossover and mutation, evolution from junior to senior. [6]
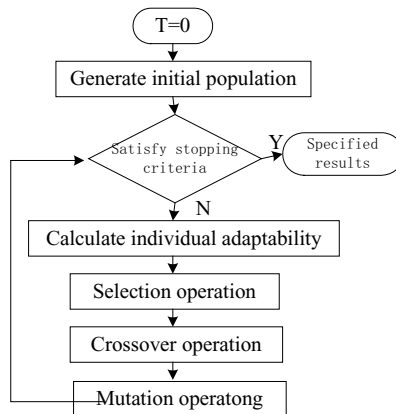
Fig. 1. Process of Genetic Algorithm

3   Routing Optimizing Algorithm Based on Genetic Algorithm

(1)Coding Scheme

By using the binary encoding, when there is a larger number of network nodes, it will be bound to cause redundant of character "0", which will be a waste of storage space, and reduce efficiency. A coding scheme for variable length nodes sequence is proposed in the paper. Each individual means a routing. If the number of network nodes is 512, the nodes number is from 0 to 511, the individual length is between 2 and 32.

(2) Adaptive Function

In the genetic algorithm, the probability of the individual is genetic to the next generation according to the size of the individual adaptability. The bigger adaptability means the better key and the greater probability.

In this paper, the number of hops is a measure of mobile Ad Hoc network routing optimization, the adaptability of each path only related to the number of hops. The calculation formula of adaptability as:

$$\begin{cases} f = 1, hop < \min \\ f = k(\max - hop), hop > \min \end{cases} \tag{1}$$

In the above formula: f is the adaptability of the path.

> hop is the hops of the path.
> min is the minimum-hop of the path.
> max is the maximum-hop of the path.
> k is the Calculation of coefficient.

(3) Selection Operation

This paper adopt random selection with non replay of remainder, using the following procedure：

1) Calculate the expected survival numbers of each router in the next populations:

$$Ni = 32 \times Fi / \sum_{i=1}^{32} Fi \quad (i=1, 2, \ldots\ldots, 32) \tag{2}$$

Where: Fi is the adaptive of the i-th router. Select Ni as the survival numbers in the next generation, so it can be proved there was $\sum_{i=1}^{32}[Ni]$ routers in the next M populations.

2) Calculate the new adaptability of each route as:

$$FFi = Fi - [Ni] \cdot \sum_{i=1}^{32} Fi / 32 \tag{3}$$

Calculate the probability it was selected as:

$$Pi = \frac{FFi}{\sum_{i=1}^{32} FFi} \tag{4}$$

Calculate the accumulation probability $Qi = \sum_{j=1}^{i} Pj$, to do $32 - \sum_{i=1}^{32}[Ni]$ round of selection. Each round of selection generates a uniform random number between 0 and 1 named rand. Let Fsum for the sum of adaptability, according to the selection pointer

to select as following: X=rand*Fsum as selection pointer, if $X \leq Q1$, then select route 1; if $Qi-1 < X \leq Qi$, then select route i, so it can generate $32 - \sum_{i=1}^{32}[Ni]$ new routes which have not yet identified.

(4) Crossover operation

This paper adopt adaptability crossover which relative to the evolution generation and individual adaptability. In the earlier evolution, in order to accelerate the evolution process to avoiding local optima, the higher adaptability the bigger crossover rate; and in the later evolution, each route of populations possess higher performance. In order to keep the superiority of the populations, it should not to be a big variation. To those routes whose adaptability is higher than the average adaptability of the populations, correspond to the lower crossover rate in order to protect the solution to the next generation; and the lower adaptability correspond to the higher crossover rate, the solution should be gone out of use.

Crossover operation as follow: from the crossover route of P1、 P2, find all the same nodes except source node and destination node, and establish common nodes set Ψ. There is no crossover operation between P1 and P2 if $\Psi = \Phi$. Otherwise, random select a node as a crossover from Ψ, then crossover path use the part after P1 and P2. For example, if source node is 0, destination node is 8; the crossover operation is shown as Fig.2. Crossover operation is as follows:
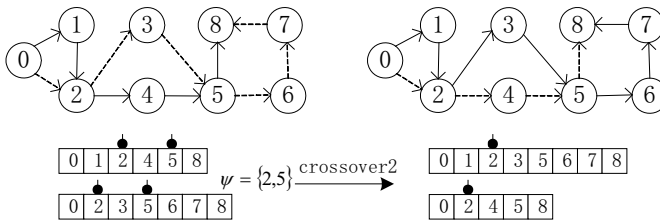
Fig. 2. The process of crossover operation

(5) Mutation Operation

The general trend of mutation operation should decrease gradually and quickly concentrate the populations. From the evolution of populations, mutation rate should be a process which is reduced and to be a steady value finally. In the earlier evolution, we should use a relatively bigger mutation rate in order to avoid local optima; and in the later evolution, because the general performance of the populations is nearer to optimal solution, it should not use a bigger mutation rate in order to keep the superiority of the populations. What's more, the poor individual the bigger mutation rate, and the excellent individual the lower mutation rate. Therefore, we design the adaptability mutation rate as follow: which relative to the evolution generation of genetic algorithm and adaptability of route.

$$m_{tmp} = \frac{pm_{max}}{1+t/T_{gen}} \times e^{\frac{f_{max}-f(x_i)}{f_{max}}}$$ (5)

$$pm_i = \begin{cases} m_{tmp}, & m_{tmp} > pm_{min} \\ pm_{min}, & m_{tmp} \leq pm_{min} \end{cases}$$ (6)

Where:

pm_{max} is the max mutation rate which is preset, let pm_{max} is 0.95.

pm_{min} is the min mutation rate which is preset, let pm_{min} is 0.05.

$f(xi)$ is the t-th generation populations adaptability of route xi.

$fmax$ is the max adaptability of populations.

t is the current generation.

$Tgen$ is the total evolution generation of populations

pm_i is the mutation rate of current route.

4   Simulation Results Analysis

In the paper, the operator is used to optimize the route in the mobile Ad Hoc network. Let the node 0 be the source node, and the destination node is 511, min is 2, max is 32,

individuality	source	interior joint	initial population	information receiver	adaptability
0	0	458, 478, 247, 9, 83, 101, 313, 251, 438, 412, 353, 270, 99, 413, 11, 446, 211, 490, 453, 276, 485, 449, 398, 300		511	0.18
1	0	163, 233, 70, 436, 122, 376, 170, 360, 15, 51, 429, 244, 80, 324, 1, 25, 432, 52, 169, 248, 232, 125, 75, 69, 297, 185, 351, 32, 206		511	0.03
2	0	385, 358, 337, 16, 151, 156, 170, 50, 124, 492, 245, 84, 158, 457, 222, 206, 307, 475, 343, 110, 39, 175, 509, 311, 301, 120		511	0.12
3	0	284, 100, 496, 398, 58, 197, 330, 51, 400, 420		511	0.6
4	0	454, 240, 128, 500, 12, 69		511	0.72
5	0	254, 396, 218, 6, 98, 56, 451, 68, 333		511	0.63
6	0	14, 128, 59, 173, 205, 186, 145, 400, 82		511	0.63
7	0	481, 471, 428, 217, 308, 500, 327, 218, 400, 290,		511	0.63
8	0	191, 178, 323, 55, 415, 468, 219, 100, 321, 507, 181, 86, 351, 48		511	0.48
9	0	169, 40, 406, 447, 221, 443, 21, 133, 209, 390, 419, 459, 412, 486, 237, 107, 333, 326, 53, 170, 89, 289, 341, 451, 245		511	0.15
10	0	26, 381, 175, 82, 312, 258, 501, 76, 302		511	0.63
11	0	293, 318, 201, 121, 145, 25, 90, 34, 125, 269, 469, 503, 347, 111, 95, 66, 7, 210, 400, 409		511	0.3
12	0	32, 169, 63, 128, 378, 337, 349, 489, 343, 217, 212, 176, 138, 371, 229, 231, 21, 188		511	0.36
13	0	153, 192, 498, 61, 70, 35, 99, 64, 10, 88, 469, 400, 387, 298, 4, 12, 29, 67, 97, 57, 288, 370		511	0.24
14	0	11, 319, 135, 424, 508, 387, 122, 98, 159, 29, 379, 497, 60		511	0.51
15	0	291, 59, 492, 192, 359, 274, 443, 258, 79, 369, 460, 432, 397, 9, 498		511	0.45
16	0	459, 161, 306, 317, 40, 94, 60, 149, 243, 238, 312, 340, 432, 272, 302, 64, 319, 219, 433, 39, 163, 157, 442, 370		511	0.18
17	0	501, 36, 412, 138, 434, 405, 295, 246, 5, 70		511	0.6
18	0	381, 8, 23, 90, 75, 510, 321		511	0.69
19	0	187, 386, 291, 73, 167, 176, 61, 253, 460, 395, 198		511	0.57
20	0	7, 268, 228, 149, 344, 181, 476, 107, 400, 44, 363, 185, 196, 15, 447, 115, 309, 99, 435, 417, 277, 440, 145, 404		511	0.18
21	0	144, 189, 271, 179, 80, 246, 380, 145, 332, 272, 347, 356, 217, 315, 51, 308, 389, 62		511	0.36
22	0	281, 201, 469, 10, 446, 245, 392, 322, 346, 253, 53, 382, 12, 454, 485, 464, 302, 501, 173, 380, 117		511	0.27
23	0	207, 100, 445, 130, 88, 2, 495, 253, 354, 54		511	0.6
24	0	85, 316, 187, 65, 23, 419, 359, 191, 203, 408, 372, 430, 496, 235, 89, 338, 273, 123, 439, 182, 18, 367, 383, 247, 103, 300, 107		511	0.09
25	0	453, 389, 291, 493, 135, 326, 373, 337, 278, 422, 366, 264, 245, 33, 372, 277		511	0.42
26	0	273, 413, 251, 276, 236, 160, 263, 281, 433, 302, 185, 377, 60, 93, 228, 220, 394, 217, 278, 84, 27, 76		511	0.24
27	0	73, 13, 456, 432, 62, 167, 418, 46, 97, 390, 308, 398, 368, 455, 438, 299, 266		511	0.39
28	0	473, 24, 421, 326, 144, 108, 5, 398, 275, 224, 434		511	0.57
29	0	206, 118, 131, 191, 497, 287, 49, 258, 231, 21, 408, 22, 95		511	0.51
30	0	292, 510, 367, 198, 291, 107, 71, 287, 494, 380, 289, 278, 134, 398, 256, 15, 480		511	0.39
31	0	217, 357, 17, 488, 458, 379, 260, 87		511	0.66

Fig. 3. Initial routing and adaptability

k is 0.03. The routing algorithm's evaluation index: the accuracy of route and the convergent evolution generation, the function of each operator is verified according to the results of the implementation and the adaptability. Fig.3 shows the initial 32 routes (the initial population) and the corresponding adaptability.

After the evolution of several generations, the initial routing converge, namely to calculate the shortest path (the global optimal solution). Fig.4 shows the route and adaptability that evolved 18 generations. It can be seen, after the evolution of 18 generations, there are 31 optimal paths. The average adaptability of the populations is 0.807187. The populations basically reach convergence. The route after convergence is 0-501-76-302-511. There is no loop in the route. And there is no loop in the every iterative process. That is known, the routing optimization algorithm designed by this paper is correct.

individuality	source	interior joint	replace population	information receiver	adaptability
0	0	501, 76, 302		511	0.810000
1	0	501, 76, 302		511	0.810000
2	0	501, 76, 302		511	0.810000
3	0	501, 76, 302		511	0.810000
4	0	501, 76, 302		511	0.810000
5	0	501, 76, 302		511	0.810000
6	0	501, 76, 302		511	0.810000
7	0	501, 76, 302		511	0.810000
8	0	501, 76, 302		511	0.810000
9	0	501, 76, 302		511	0.810000
10	0	501, 76, 302		511	0.810000
11	0	501, 76, 302		511	0.810000
12	0	501, 76, 302		511	0.810000
13	0	501, 76, 302		511	0.810000
14	0	501, 76, 302		511	0.810000
15	0	501, 76, 302		511	0.810000
16	0	501, 76, 302		511	0.810000
17	0	501, 76, 302		511	0.810000
18	0	501, 76, 302		511	0.810000
19	0	501, 76, 302		511	0.810000
20	0	501, 76, 302		511	0.810000
21	0	501, 76, 302		511	0.810000
22	0	501, 76, 302		511	0.810000
23	0	501, 76, 302		511	0.810000
24	0	501, 76, 302		511	0.810000
25	0	501, 76, 302		511	0.810000
26	0	501, 76, 302		511	0.810000
27	0	501, 76, 302		511	0.810000
28	0	501, 76, 302		511	0.810000
29	0	501, 76, 302		511	0.810000
30	0	501, 76, 302		511	0.810000
31	0	501, 76, 456, 237, 317, 302		511	0.720000

Fig. 4. The route and adaptability after convergent

Fig.5 shows the curves variation of the population average adaptability with the change of the evolution generation. We can see that, as the evolution generations increase, the average adaptability of population increases. When the evolution generation reaches the 18th generation, the populations basically reach convergence.
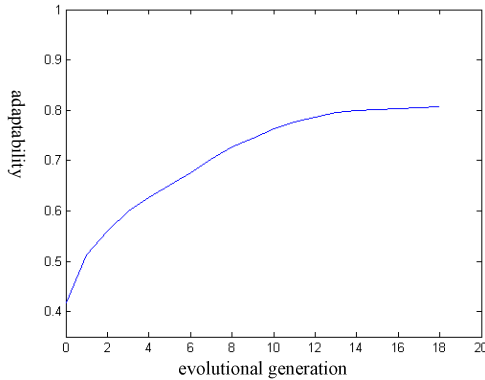
Fig. 5. The adaptability curves of variation of evolutional generation

From the analysis result we can see there is no loop in the route during the iterative process of algorithm, so the routing algorithm is correct. It can reach convergence after the average evolution generation of 18, then it have the better convergence speed relative to the other genetic algorithm. The selection, crossover and mutation operator is rationally designed in the paper, which keep the locality, global and variety of the populations, in order to make the good total performance of evolution result.

5   Summaries

The routing optimizing arithmetic based on genetic algorithm is good performance. The routing calculated is correct and convergence velocity is quicker, which can provide foundation for the routing optimizing of mobile ad hoc network.

References

[1] Ke, M.-Y., Xu, X.-J.: A New Wireless Mobile Ad Hoc Network. Software Guide (2006)
[2] Shi-Jian: QoS Routing Technologies in Modern Wareless Networks. Doctorial Dissertations of Huazhong University of Science and Technology (August 2001)
[3] Partridge, C., Mendez, T.: Host Anycasting Service. RFC 1546 (November 1993)
[4] Hinden, R., Deering, S.: IPv6 Addressing Architecture. RFC 2373 (July 1998)
[5] Johnson, D., Deering, S.: Reserved IPv6 Subnet Anycast Addresses. RFC 2526 (November 1993)
[6] Liang-Rong: A QoS Multicast Routing Algorithms Based on Genetic Algorithm Master thesis of East China Normal University (2004)

Contrast Study on Two Kinds of SIP Trunking Route Scheme Based IMS Network

Zhao Yuan[1], Gu Su[2], and Wang Xiaoyun[2]

[1] Network Technology Research Institute, Beijing University of Posts and
Telecommunications, Beijing 100876
[2] China Mobile Group Research Institute Co., Ltd, Beijing 100053

Abstract. These years, the enterprises more and more have deployed their own call center to provide quality and efficient services for customers, while the traditional calls are transmitted by TDM trunking among the Switches. However, with the advancement of NGN study, IMS network as a pipe can provide efficient SIP trunking interconnection services for large call center and phone users. This paper introduced through the SIP trunking based on IMS network to analyze the routing problem between the call center of enterprises and network applications of IMS network operators.

Keywords: IP Multimedia Subsystem, SIP trunking, Route scheme, Call center.

1   Introduction

Currently, the world's telecommunications industry is facing the most severe challenges since the revolution of information technology. In the traditional circuit-switched PSTN networks and GSM mobile telephone communication system, we mostly use the way of TDM trunking to implement the traffic transmission of data, voice and video. Time Division Multiplex (TDM) is a synchronous transfer mode, the biggest problem which exists in the TDM technology is that the resource scheduling is not flexible to adapt to variable rate business, Business and management concentration is resulting in the insufficient of service network and the room for development of value-added service. Therefore, TDM can't ideally meet the demand of developing the Telecom Industry, especially for the multi-media. the 3GPP R5's IP Multimedia Subsystem[1] (IMS) technology is an open system architecture based on SIP. IMS learns and absorbs the Internet's open mode, borning with an open and fusion feature of end to end full IP system fusion architecture, as a result, the business based IP can be transmitted via the IMS network. Because the session is transmitted and controlled by SIP in the IMS network, we have to face to a SIP trunking routing problem.

2   TDM Trunking and SIP Trunking

2.1   TDM Trunking[2]

TDM is the abbreviation of Time Division Multiplexing, it refers to a technology which transmits multiple digital data, voice and video signals in the same

communications media via the cross-pulse of different channels or time slots. It is the basic characteristics of TDM for any call to assign a fixed rate of channel resources, and dedicate throughout the duration of the call, further, regardless of whether the data need to be transmitted or not, TDM will allocate bandwidth for each channel. Therefore, based on TDM technology in the fixed telephone network PSTN and mobile communication network PLMN, when we need a long distances across the network data, voice and video call, we can transfer call routing by way of TDM trunking in different network domains.

2.2 SIP Trunking

Session Initiation Protocol [3] (SIP) is an application layer signaling control protocol developed by IETF. it's a light-shaped, multi-purpose text-based protocol, which can be used to create, modify and terminate a session, we can use it to implement different services, In a word, SIP is simple, flexible, good scalability, especially for IP Network. It is based on the SIP characteristics IMS network choose it to transmit and control a session. When to use IMS as a pipe, the session needs to be routed among the various network elements, the whole process can have some different SIP trunking[4] route ways, so it is important to study the routing situation of SIP trunking in specific call scenes.

2.3 The Cost Analysis between TDM Trunking and SIP Trunking

With the telecom network evolution to all-IP network, it is more and more difficult for the original TDM trunking to meet the speedily increasing Call traffic, it will be gradually replaced by SIP trunking during the IP application. it has been proven that SIP trunking provides cost savings potential and chance for new applications. Such as Return on investment - the call via IP network can reduce 30% -70% of the cost of telecom service providers among companies. Detailed cost analysis in different segments as show in Fig.1 [5]:

Trunk Type	Access	Trunks	Usage	Usage	Features	Sub Total
	T1/PRI		Inter-State	Intra-State	Basic	
Toll Free	33	750	4250000	750000	750	
Default	$300	0	0.02	0.05	$50	
Sub Total	$9,783	$0	$85,000	$37,500	$37,500	$169,783
Long Distance	22	500	3750000	1250000	500	
Default	$300	0	0.02	0.05	$3	
Sub Total	$6,522	$0	$75,000	$62,500	$1,500	$145,522
Local	43	1000	0	2500000	1000	
Default	$300	$35	0	0.01	$5	
Sub Total	$13,043	$35,000	$0	$25,000	$5,000	$78,043
Support					2.5 people	$20,000
Taxes					10%	$39,335
Monthly Total		2,250	$160,000		$44,000	$452,683
Yearly Total						$5,432,196
Trunk Type	Access	Trunks	Usage	Usage	Features	Sub Total
	Ethernet		Inter-State	Intra-State	Basic	
SIP Trunks	4	1500	0	0	0	
Default	$5,000	$25	0	0	$5	
Sub Total	$20,000	$37,500	$0	$0	$7,500	$65,000
Support					1.25 people	$10,000
Taxes					5%	$3,750
Monthly Total						$78,750
Yearly Total						$945,000

(TDM trunks — top section; SIP trunks — bottom section)

Fig. 1. The cost analysis between TDM trunking and SIP trunking

Through the above analysis we can find that the call routing cost using SIP trunks abviously less than the cost of the use of TDM trunks.

3 The SIP Trunking Applications during the Enterprise Call Center Demanding for Direct Connection

When the fixed or mobile phone users dial the large-scale call center which is deployed in the enterprise side, the method of TDM trunking is considered as the traditional call route scheme, but with the rapid development of country's economic, the number of user calls and the total traffic which each large-scale call center have received have an increasing trend from month to month, and will have great growth prospects, whereas they originally have in a state of higher level. The route scheme based TDM trunking gradually unfolds many shortcomings in telecommunications network, with the start of the commercial deployment of IMS network , we obviously can use all IP network capacity of IMS network to complete the call center directly connected, then directly transfer the call to the corporate call center via the SIP trunking of IMS network.

3.1 The Cost Analysis between TDM Trunking and SIP Trunking

First of all, we have to understand the relationship of attribution operators between users and call center, if a business hires an operator's network line and accesses IP private network, when the operator network which customers attribute and the call centers access belong to the same operator or not, it is different for the schemes in directly connected call center, The details will be described in the following circumstances:(Note: HOME NETWORK refers to the same relationship of attribution operators between users and call center, obviously, Not HOME NETWORK is in contrast to HOME NETWORK)

Scene: The user dials a large call center number (such as a Shanghai users dial to a Beijing enterprise call center 955XX)

About scheme A: we can learn from the standard IMS network route scheme Transit and innovate to achieve.
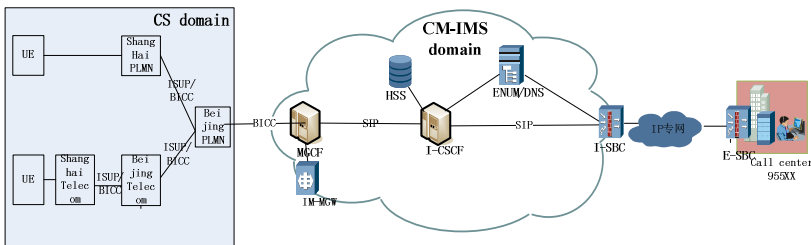
Fig. 2. The call routing of A scheme

As shown in the Fig.2, the SIP trunking is implemented by the fusion of international standard Transit route [6] and china mobile IMS network [7]. The IMS network which the call center lease is deployed by China Mobile, the call center and

the CS domain users aren't registered in IMS network at the same time. The detailed course is described as following:

(1) The call firstly delivers in the CS domain, according to the regulations: Home Network users choose the nearest way into the CS domain and the farthest way to exit the CS domain, Not Home Network users select the farthest way to get out of his network, finally the calls via Home Network PLMN of the call center attribution CS domain route into the IMS domain;

(2) When the call pass the CS domain into the IMS domain, the MGCF network element, which completes controlling protocol conversion from the ISUP / BICC of the CS domain to the SIP of the IMS domain, and routes call to the I-CSCF network element.

(3) I-CSCF is designed with the Transit function. When calls arrive , I-CSCF firstly queries HSS based on the called number, finds that the called user doesn't register in IMS network and enables the Transit functions, then using the called number to query ENUM and get the corresponding SIP URI address, the next, querying the DNS to get IP address of I-SBC which connects to the E-SBC closing to the enterprise call center, finally, routing the call to I-SBC; Obviously, ENUM network elements store the responding information between E.164 number and SIP URI address, DNS network elements store the responding information between SIP URI address and IP address.

(4) I-SBC, which locate in the edge of IMS network, queries the DNS network element to get the next routing IP address according to the called number, then routes the call to E-SBC via IP Private Network, what's more, it also provide Qos controlling, the safe of Signaling and bearer and so on.

(5) The E-SBC which is close to the call center have a simple protocol conversion function, it can complete the conversion between non-IP protocols of call centers and IP protocol running in the IMS network, the E-SBC also provides the security features, interoperability, routing and control and so on.

About scheme B: we can adopt the way of transparent transmission to complete call routing in IMS network.
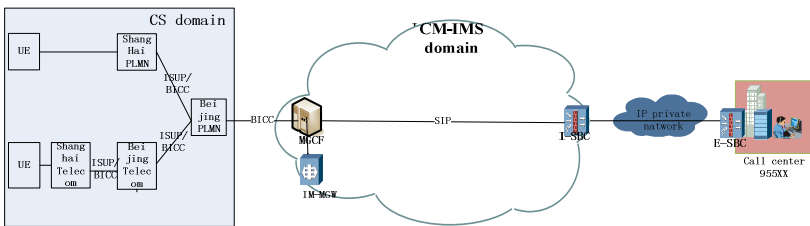
Fig. 3. The call routing of B scheme

As shown in the Fig.3, we learn from Internet network thought to configure static routing table about the call at the network element MGCF. The call routing process is specifically described as following:

(1) The call firstly is delivered in CS domain; the regulation is the same as the described in A scheme.

(2) When call pass the CS domain into the IMS domain, the MGCF network element, which completes the controlling protocol conversion from the ISUP / BICC of the CS domain to the SIP of the IMS domain and routes the call to the I -CSCF, At the same time, MGCF queries the static routing table about the especial call, and directly routes the session to the edge I-SBC of IMS network;

(3) I-SBC statically configure the next route to the E-SBC which is close to the call center, when the session arrive ,it queries the static routing table based the called number and finds out the IP address of E-SBC in the side of enterprise, then pass session to E-SBC via IP Private Network.

(4) The functions of E-SBC are the same as description in Scheme A.

3.2   Comparison of Two Route Scheme

Through the describing in the two route schemes, we can see that they both can route the call correctly handle such routing problems when considering the IMS network as an IP pipe. Therefore, we tested the feasibility of both routing solutions in laboratory, and analyzed the merits and shortcomings of these two schemes from various aspects, the specific analysis described in the Tab.1:

Table 1. Comparing A scheme with B

Comparison dimension	Scheme one: SIP trunking-transit route	Scheme two: SIP trunking-transparent transmission route
The degree of standardization	Standard scheme	Private scheme
Deployment	Need to deploy E-SBC in the side of enterprise	Need to deploy E-SBC in the side of enterprise
Configuring Core network device	Configuring the call center routing in ENUM / DNS	Configuring the call center routing in MGCF and I-SBC
Route	Meet the CM-IMS routing principles, via more network element , three second delay	Meet the CM-IMS routing principles, via less network element , one second delay
Maintenance costs	Lower cost	Higher cost
Other	Meet the evolving future direction of IP call center	Single vendor can implement, there are some private natures.

From the above Tab.1 we can clearly discover, the A scheme will involve more the IMS network elements and interfaces, so it may cause a little larger routing delay, but it is less to change the existing network, only need to simply open the relevant numbers to ENUM / DNS server, therefore, it is a more viable way of using the route in long-term; the B scheme transparent transmission route is larger revelation the function of a pipe provider of IMS network, especially when the number of call center is not bigger, the efficiency is larger in static configuration routing table, so the scheme also has a certain practicality.

4   Summaries

This paper describes these two SIP trunking route schemes based IMS network, which are adapt to the current deployment of Large-size and medium-sized enterprises directly connected to the call center business, it is sure that both options have their advantages and disadvantages, they can effectively reduce the cost of enterprise call center and improve the quality of service. The next, with the development of IMS technology and call center IP-oriented, we can consider the call center as an IMS application platform and directly deploy in the converged IMS network architecture, it will be more economic and efficient to meet the enterprise call Centre directly connected business needs.

References

[1] Poikselkä, M., Mayer, G., Khartabil, H., Niemi, A.: THE IMS:IP Multimedia Concepts and Services, 2nd edn. John Wiley & Sons, Lte (2005)
[2] Zhou, G., Li, M.: TDM / TDMA system trunk network design. Journal of Air Force Engineering University (Natural Science) (1), 33–36 (2001)
[3] Rosenberg, J., Sehulzrinne, H., Camarillo, G.: RFC3261, Session Initiation Protocol. IETF (2002)
[4] Fei, B.: The Research and Implementation of SIP Trunking Gateway in IMS. Beijing University of Posts and Telecommunications, Beijing (2006)
[5] Su, G.: Acme-Packet Deploying SIP Services Successfully –IMS based SIP Trunking (February 2011)
[6] 3GPP TS 23228 v9.1.0-2009. IP Multimedia Subsystem (IMS); Stage 2
[7] Liu, J., Zhi, Y., Wang, Y., Cai, Y.: China Mobile CM-IMS full-service network description. Mobile Communication Technology (6), 3–9 (2009)

Epidemic Analyses on Small Worlds of Tree Topologies of Wireless Sensor Networks

Qiao Li[1], Baihai Zhang[1], Lingguo Cui[1], and Zhaoyao Tao[2]

[1] School of Automation,
Beijing Institute of Technology,
Beijing 100081, China
[2] The Third Department,
Artillery Command Academy,
Hebei 065000, China

Abstract. Tree topologies, which construct spatial graphs with large characteristic path lengths and small clustering coefficients, are ubiquitous in deployments of wireless sensor networks. Due to link additions, small world phenomena exist in tree topologies. Epidemics on small worlds of tree topologies are studied, and we calculate the percolation threshold at which the outbreak of the epidemic takes place. Compared with Cayley tree, the small world has a smaller percolation threshold suffering from the epidemic. In the further study, Epidemiological processes are observed by simulations. The infection extends exponentially with time for the existence of shortcuts.

Keywords: Epidemic, Small world, Tree topology, Wireless sensor networks.

1   Introduction

Recent years have seen the deployments of wireless sensor networks (WSNs) in a variety of applications including habitat and environmental monitoring [1], precision agriculture, security surveillance [2], etc. More and more efficient ways of deploying sensor nodes rise into view recently. The tree topology is a kind of architecture used frequently, which is ubiquitous in the deployment of wireless sensor nodes. Some routing protocols, topology control algorithms and aggregation schedules of wireless sensor networks are helpful to construct tree topologies [3-5].

Most tree topology protocols are multi-hop protocols, which are famous for energy saving in data gathering and transferring. During the work of tree topologies, random link additions among nodes take place inevitably for the use of omnidirectional antennae. Obstacles, adjustments of radio energy, joins of new members and errors of the location precision all incur link additions. Small world phenomena [6] existing in tree topologies of wireless sensor networks are considered [7]. The wireless sensor network is vulnerable to computer viruses. For the existence of shortcuts in small worlds, the epidemic propagation becomes much drastic in the network [8-10].

In this paper, the percolation threshold [11], at which the outbreak of the epidemic takes place in small worlds of tree topologies, is calculated. Epidemiological processes are observed by simulations.

2   Network Model and Basic Idea

In this paper, there are N nodes in the network and we assume that the following properties of the WSN:

(1) Nodes are not moveable after they are randomly distributed in the network.

(2) All nodes are symmetric. They have similar characteristics (e.g., range of radio coverage, energy of batteries, etc.).

(3) There is only one root, which constructs one tree in the network.

An abstract tree model for tree topologies is proposed with K children nodes for each parent node, $K \geq 2$. After the root node is selected out randomly, it broadcasts its location information to other nodes in the field. K neighbor nodes, which are nearest to the root node, are connected to the root node. Then the K nodes broadcast their location information to other nodes which are not connected and each is connected with K-1 nearest neighbor nodes. This process continues until all nodes are included in the tree. If $N \rightarrow \infty$, This tree model is Cayley tree in fact. Random link additions are conducted on it. This abstract simplifies extremely the problem and it represents a spatial graph model. Fig. 1 shows Cayley tree.
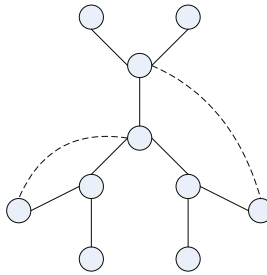
Fig. 1. Cayley tree

3   Epidemic Thresholds on Small Worlds

In our small world model, we start with Cayley tree and add shortcuts between pairs of nodes chosen uniformly at random. More than one bond (link) between any two nodes, or a bond which connects a node to itself, is allowed. A certain fraction h of nodes of tree topologies is assumed to be susceptible to the disease, and the bonds between them represent the physical contacts by which a disease can spread. If the distribution of occupied nodes is random, the problem when an epidemic takes place becomes equivalent to a standard percolation problem. The node is denoted by a site. The percolation probability h_c, at which the outbreak of the epidemic takes place, can be calculated. In this paper, the site percolation is only considered.

In Cayley tree, each node has K children nodes. Some characteristics of Cayley tree have been analyzed in [11], the percolation probability h_c can be calculated by

$$h_c = \frac{1}{K-1} .$$

(1)

The parameter p is defined as the average number of shortcuts per bond on the underlying topology. The probability that two randomly chosen sites have a shortcut between them is calculated as

$$\theta = 1 - (1 - \frac{2}{N^2})^{p(N-1)}$$

$$\approx \frac{2p}{N} .$$

(2)

Occupied sites connected together will construct the local clusters in the small world, and they can be connected together by shortcuts. The average number of local clusters of size i in the network can be calculated by

$$X_i = h^i (1-h)^{2+(K-2)i} N .$$

(3)

In order to construct a so-called "giant component" as in the random graph [12], we start with one particular local cluster, and add all other local clusters to it, which can be reached by traveling along a single shortcut. Then all other local clusters, which can be reached by traveling along a single shortcut, are added to the new ones. This process continues until the connected cluster, the giant component, is constructed.

In order to calculate the percolation probability h_c, a vector $\mathbf{v}$ is defined at each step in this process, whose component v_i is the probability that a local cluster of size i is added to the overall connected cluster. We define another vector $\mathbf{v}'$, whose component v_i' can be gotten in terms of the value of $\mathbf{v}$ at the previous step. At or below the percolation threshold the component v_i is small and we can calculate the vector $\mathbf{v}'$ using a transition matrix $\mathbf{M}$. The following formula reflects the relationship between $\mathbf{v}$ and $\mathbf{v}'$

$$v_i' = \sum_j M_{ij} v_j ,$$

(4)

where

$$M_{ij} = X_i [1 - (1-\theta)^{ij}] .$$

(5)

X_i is the number of local clusters of size i as before. $[1-(1-\theta)^{ij}]$ is the probability of a shortcut, which connects a local cluster of size i with one of size j, and there are ij possible pairs of sites by which these can be connected.

The largest eigenvalue λ of the transition matrix $\mathbf{M}$ is considered. If $\lambda < 1$, the vector $\mathbf{v}$ tends to 0 according to (4). The rate at which new local clusters are added falls off exponentially, and the connected clusters are finite with an exponential size distribution. Conversely, if $\lambda > 1$, $\mathbf{v}$ keeps growing until the size of the connected cluster becomes limited by the size of the whole system. The percolation threshold occurs at the point $\lambda = 1$.

It is difficult to find the largest eigenvalue of the transition matrix $\mathbf{M}$ for finite N. If p is a constant, when $N \rightarrow \infty$ θ tends to 0. (5) can be simplified by

$$M_{ij} = ij\theta X_i .\tag{6}$$

If we set that $v_i^{'} = \lambda v_i$, (4) is rewrote as

$$\lambda v_i = i\theta X_i \sum_j jv_j .\tag{7}$$

v_i can be calculated by

$$v_i = C\lambda^{-1}i\theta X_i ,\tag{8}$$

where $C = \sum_j jv_j$ is a constant. From (8) we know

$$\sum_i v_i = C\lambda^{-1}\theta \sum_i iX_i ,$$

$$\sum_i iv_i = C\lambda^{-1}\theta \sum_i i^2 X_i ,$$

$$\lambda = \theta \sum_i i^2 X_i .\tag{9}$$

When $K=2$,

$$X_i = h^i(1-h)^2 N .\tag{10}$$

This formula is the same as that of the ring understratum proposed by DUNCAN J. Watts and STEVEN H. Strogatz [13]. In this state, λ can be derived by

$$\lambda = \theta Nh\frac{1+h}{1-h} = 2ph\frac{1+h}{1-h} .\tag{11}$$

We set that $\lambda=1$ to get the value of p at the percolation threshold h_c,

$$p = \frac{1-h_c}{2h_c(1+h_c)} .\tag{12}$$

For general K, the average number of local clusters of size i in the network can be given by

$$X_i = h^i(1-h)^{2+(K-2)i} N = (1-h)^2[h(1-h)^{K-2}]^i N .\tag{13}$$

λ can be calculated by

$$\lambda = \theta N(1-h)^2 \frac{g(1+g)}{(1-g)^3} ,\tag{14}$$

where $g = h(1-h)^{K-2}$. Then λ can be calculated as

$$\lambda = \theta N(1-h)^2 \frac{h(1-h)^{K-2}[1+h(1-h)^{K-2}]}{[1-h(1-h)^{K-2}]^3} .\tag{15}$$

We set that $\lambda=1$ to get the value of p at the percolation threshold h_c,

$$p = \frac{[1-h_c(1-h_c)^{K-2}]^3}{2h_c(1-h_c)^K[1+h_c(1-h_c)^{K-2}]}.$$

(16)

p is a constant in our study, and the percolation threshold h_c for general K can be calculated using (16).

Analysis results are used to calculate the percolation threshold. The epidemic should have a single humped form with time, which is similar to the development of fads.

4 Simulation

The percolation probability, at which the outbreak of the epidemic takes place, is observed in this part. When $K=2$, $h_c = \frac{\sqrt{(2p+1)^2+8p}-(2p+1)}{4p}$. While $h_c=1$ in Cayley tree. If p increases from 0.01 to 1, the variety of h_c is observed.

When $K=2$, the tree topology transforms into the chain structure. Fig. 2 shows that the percolation threshold h_c keeps decreasing when the random link addition probability p increases. This result coincides with the reality. In a regular tree topology, the epidemic propagates along the regular bonds with a large percolation threshold. For the existence of shortcuts in small worlds, the epidemic propagation becomes much drastic in sensor networks when the random link addition probability p increases. In an entire random network, the percolation threshold is small. The viruses easily attack the network from one side to another. Without shortcuts, the percolation threshold in Cayley tree is 1, which is shown in Fig. 2. Compared with Cayley tree, the small world has a smaller percolation threshold suffering from the epidemic.
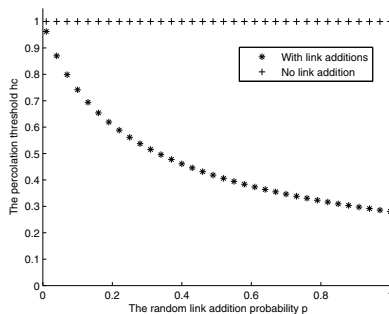
Fig. 2. The percolation threshold h_c with $K=2$

When $K>2$, it is difficult to solve (16), but the situation is similar to that when $K=2$. When the random link addition probability p increases, the percolation threshold h_c keeps decreasing. When $p=1$, there exists the least percolation threshold in the wireless sensor network.

In the further simulations, epidemiological processes are observed with $h > h_c$.

In the following simulation, a certain number of nodes, C_K, are infected on the tree topology in each time unit. Certain parameters are set with $p = 0.3$, $h = 0.8$ and $C_K = 3$. Fig. 3 (a) shows the time evolution of the infected number in the epidemiological process with $N = 500$, 2000 and infinity. Each time unit includes 10 seconds.

If the infected parent node can infect almost all its children nodes in each time unit, Fig 3 (b) shows the time evolution of the infected number in the epidemiological process with $N = 100$, 2000 and infinity. Certain parameters are set with $p = 0.3$, $h = 0.8$, and $K = 3$ in the simulation. Each time unit includes 10 seconds.

(a)   (b)
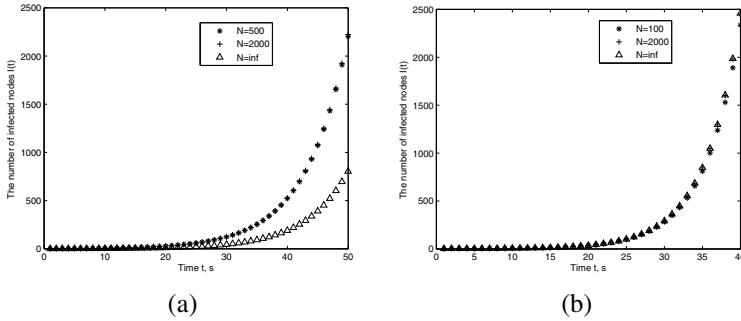
Fig. 3. The number of infected nodes

The simulations show that the infection extends exponentially with time t for the existence of shortcuts in small worlds of tree topologies.

5   Summaries and the Future Work

The tree topology is a kind of architecture used frequently, which is ubiquitous in the deployment of wireless sensor nodes. Due to random link additions, small world phenomena take place and the epidemic propagation becomes much drastic in sensor networks. Based on the abstract tree model, Cayley tree, the epidemic is analyzed, and random link additions are conducted on it. The epidemic on small worlds of tree topologies on large scales is studied and the percolation threshold, at which the outbreak of the epidemic takes place, is calculated. Compared with Cayley tree, the small world has a smaller percolation threshold suffering from the epidemic. The infection extends exponentially with time for the existence of shortcuts by simulations. In the future, we will pay attention to the immunization of tree topologies of wireless sensor networks.

References

[1] Alippi, C., Camplani, R., Galperti, C., et al.: A Robust, Adaptive, Solar-powered WSN Framework for Aquatic Environmental Monitoring. IEEE Sensors Journal 11(1), 45–55 (2011)

[2]   Wang, X., Wang, S., Bi, D.-W.: Distributed Visual-target-surveillance System in Wireless Sensor Networks. IEEE Transaction on Systems, Man, and Cybernetics, Part B: Cybernetics 39(5), 1134–1146 (2009)

[3]   Demirbas, M., Lu, X.-M., Sinqla, P.: An In-network Querying Framework for Wireless Sensor Networks. IEEE Transaction on Parallel and Distributed Systems 20(8), 1202–1215 (2009)

[4]   Wong, Y.-C., Wang, J.-T., Chang, N.-H., et al.: Hybrid Address Configuration for Tree-based Wireless Sensor Networks. IEEE Communications Letters 12(6), 414–416 (2008)

[5]   Zhu, Y.-J., Vedantham, R., Park, S.-J., et al.: A Scalable Correlation Aware Aggregation Strategy for Wireless Sensor Networks. Information Fusion 9(3), 354–369 (2008)

[6]   Stanley, M.: The Small World Problem. Psychology Today 2, 60–67 (1967)

[7]   Li, Q., Cui, L.-G., Zhang, B.-H., et al.: Small Worlds in the Tree Topologies of Wireless Sensor Networks. In: Proceedings of the 29th Chinese Control Conference, Beijing, pp. 4677–4683 (2010)

[8]   Moore, C., Newman, M.E.J.: Epidemics and Percolation in Small-world Networks. Physical Review E 61(5), 5678–5682 (2000)

[9]   Telo De Gama, M.M., Nunes, A.: Epidemics in Small World Networks. The European Physical Journal B 50, 205–208 (2006)

[10] Thomas, E.S., Matthew, M.J., Susan, R.M.: Comparative Effects of Avoidance and Vaccination in Disease Spread on a Dynamic Small-world Network. Physica A: Statistical Mechanics and Its Applications 389(23), 5515–5520 (2010)

[11] Dietrich, S., Ammon, A.: Introduction to Percolation Theory. CRC Press, Boca Raton (1992)

[12] Alon, N., Spencer, J.: The Probabilistic Method. John Wiley, Hoboken (2000)

[13] Duncan, J.W., Steven, H.S.: Collective Dynamics of 'Small-world' Networks. Nature 393, 440–442 (1998)

Slope Stability Evaluation in Open Pit Based on Bayesian Networks

Wen Sheng Liu, Shi Bo Li, and Rui Tang

College of Mining Engineering, Hebei United University, 063009, Tangshan, China

Abstract. Slope stability evaluation is one of the most effective countermeasures to improve the safety in open-pit mine. Bayesian network is a popular tool for modeling many kinds of statistical problems. The method is applied to the stability evaluation of slope system in one open-pit mine of Sanyou Group Corporation Ltd in Hebei province, China. Through the result of the analysis we pointed out the most important factors influencing the stability state. The corresponding countermeasures are put forward.

Keywords: Slope stability evaluation, Open pit, Bayesian Networks.

1 Introduction

Over the last decade, Bayesian networks (BNs) have become a popular tool for modeling many kinds of statistical problems[1]. We have also seen a growing interest for using BNs in many application areas such as the reliability analysis, forecasting, classifying, causal analysis, diagnostic analysis, etc[2,3]. However, this method has not been applied to slope system stability evaluation ever, especially in field of slope stability evaluation. In this paper we will discuss the properties of the modeling framework that make BNs particularly well suited for safety assessment applications, and put forward the countermeasures to prevent slope accident.

2 The Traditional Method and Its Limitations

The safe coefficient is a traditional slope stability analysis method to evaluate slope stability, but this method is only applicable to metal materials approximately because the safety coefficient is defined as a ratio of destruction stress and maximum stress. However, the following disadvantages limit its promotion application in geotechnical problems: (1) Usually, the selection of safety coefficient is arbitrary ; (2)The state of slope stability always different while the safety coefficient are identical; (3)The method ignore uncertainty of slope medium shear strength parameters; (4) The method consider that the factors that influence slope stability are just the mechanical properties of rock itself, ignoring other factors[4,5].

3 The New Method: Bayesian Networks

A Bayesian Networks consists of a series of nodes, representing variables which interact with each other. These interactions are expressed as links between variables; the links, however, are not permitted to form a closed loop. A node representing variable will be linked to a number of 'parent' nodes on which it is dependent. The

links or 'edges' are expressed as probabilistic dependencies, which are quantified through a set of conditional probability tables (CPTs). For each variable the tables express the probability of that variable being in a particular state, given the states of its parents. As more data or knowledge become available it may be necessary to revise these tables to reflect the improved data set. For variables without parents, an unconditional distribution is defined.

A simple Bayesian Networks is shown in Fig 1. The circles are network variables, which represent the most important factors relating to a particular decision or action. They are linked together so that a change in one will result in a chain reaction of impacts on all the linked variables in the direction of the links.
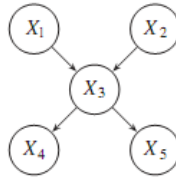
Fig. 1. An example of BN

As shown in Fig.1, The nodes representing variables which interact with each other. These interactions are expressed as links between variables; the links, however, are not permitted to form a closed loop. A node representing variable x_3 is linked to two parent nodes, x_1, x_2 and two child nodes(x_4, x_5), on which it is dependent. The links or 'edges' are expressed as probabilistic dependencies, which are quantified through a set of conditional probability tables (CPTs).For each variable the tables express the probability of that variable being in a particular state, given the states of its parents. For variables without parents, an unconditional distribution is defined [6].

4   The Factors Analysis of Slope Failure

Due to factors of landform, natural environment, rock structures in process of stripping , the rock's stability in the slope depends on the state of the rock's stress balance, when the sliding force larger than sliding resistance, slide is easily appeared. Large area of rock slide will not only hurt people, but also be a serious failure for open stope. Factors influencing stability of open-pit slope are complex. Rock composition and rock structure and groundwater etc are the main factors.

In addition, blasting and earthquake, slope shape also play a part. Through on-the-spot investigation, we have learned that the influence of groundwater and production blasting to the slope can be ignored. In addition, through a preliminary analysis, the slope angle is bigger than the normal. The working platform is relatively flat, but in the back area of the slope top many cracks appear. Crack width is about 1 ~ 4 cm, apart from the slope front 2 ~ 6 m. It can be seen that the possibility of local traffic accident damage is rather high. According to the geologic condition, the open-pit slope is of granular structure, so the most likely failure type of the slope is circular sliding.

In the process of slope management, there still exists some potential unsafe factors that can cause mine casualty accidents, they are all hazards in the open-pit. A few

hazards has greater energy, once cause accidents, often causing irreparable damage for production management.

5   The Slope Failure Analysis by Bayesian Networks

Not only the slope failure would cause group injury events, but also it is likely to cause the whole failure of the open pit, So it is the major hazard in the open-pit. Through investigating and analyzing the accidents reason of slope failure. According to the logic relation of the factors, the Bayesian Networks is constructed, as shown in Fig.1. The number, type ID, name of events and probabilities are shown in table 1. The prior probabilities and CPTs are from the history statistical data of San You Group Corporation Ltd located in He Bei province, China. The prior probabilities of the root node in Bayesian Networks are given by X1~X10 listed in table 1.

The probabilities of non-root nodes are computed by root nodes and the CPTs. We list the CPT of event T as example (Table 2). The other CPTs is not listed one bye one in this paper in order to save space.

Through the prior probabilities of the root node in Bayesian Networks is given by X1~X10, we got the prior probability of top event(T): P(T)=0.37 which means the slope failure accidents is easy to happen. We should take measures to reduce probability. Although the probability of top event will reduces when all the root variable prior probabilities reduce, we should find the some most important factors influencing the top event. To other variables that do not sensitive to the top event we may take less effort to reduce the probabilities.

Bayesian Networks provide a method to find the sensitive variables: backward reasoning. If the top event occurred at a time, we can use the Bayesian Network's backward reasoning function to computed the posterior probability of every basic event, find some important factors which influence the top event significantly. The important factors make us to pay special attention in the process of controlling slope stability.

For example, Supposing there happened a accidents in the slope. We use the backward reasoning function to compute the posterior probability of every basic event X1~X10 which are listed in table 3.
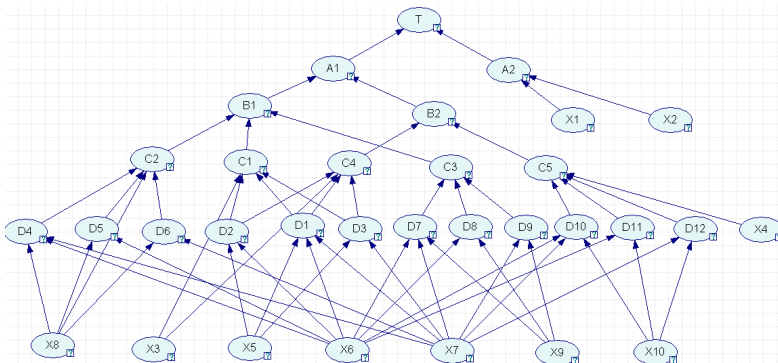
Fig. 2. Bayesian networks of related factors

Table 1. The number, type ID, event name and probability of the slope failure

type ID	Event name	probability	type ID	Event name	probability
T	slope failure	CPT	D7	Compound sliding I	CPT
A1	whole slope failure	CPT	D8	Compound sliding II	CPT
A2	Slope step failure	CPT	D9	Compound sliding III	CPT
B1	whole slope slip	CPT	D10	wedge sliding I	CPT
B2	Muti-steps slip	CPT	D11	wedge sliding II	CPT
C1	plane sliding	CPT	D12	wedge sliding III	CPT
C2	circular sliding	CPT	X1	steep rock dumping failure	0.002
C3	Compound sliding	CPT	X2	caving	0.001
C5	wedge sliding	CPT	X3	Structure surface angle less than slope angle and greater than friction angle	0.003
D1	plane sliding I	CPT	X4	intersection angle less than slope angle and greater than friction angle	0.004
D2	plane sliding II	CPT	X5	structure surface angle great er than friction angle	0.03
D3	plane sliding III	CPT	X6	amount of explosive for one blasting is too much	0.005
D4	circular sliding I	CPT	X7	groundwater effects	0.01
D5	circular sliding II	CPT	X8	granular structure slope	0.007
D6	circular sliding III	CPT	X9	granular structure and along slope structure surface	0.002

Table 2. The CPT of event T

A1		happened		Not happened	
A2	happened	Not happened	happened	Not happened	
happened	0.9	0.8	0.9	0.1	
Not happened	0.1	0.2	0.1	0.9	

Table 3. The prior probabilities and posterior probabilities of basic events of slope stability assessment

event	prior probabilities	posteriori probability	event	prior probabilities	posteriori probability
X1	0.002	0.1132	X6	0.005	0.2425
X2	0.001	0.4103	X7	0.01	0.2568
X3	0.003	0.0148	X8	0.007	0.2103
X4	0.004	0.0042	X9	0.002	0.1858
X5	0.03	0.1304	X10	0.005	0.1645

For example, Supposing there happened a accidents in the slope. We use the backward reasoning function to compute the posterior probability of every basic event X1~X10 which are listed in table 3.

From table 3 we can see that the posteriori probability can be used as the important degree which the basic event affect the top event. The posteriori probability of basic event X 2（caving）are 0.4103 which is the largest of the whole numbers. So, we concluded that X 2 is the most sensitive factors affecting the top event. By reducing the incidence rate of X 2 we can reduce the incidence rate of top event. The other two larger probabilities are x6(amount of explosive for one blasting is too much) and x7(groundwater effects). Reducing these two probabilities, we can also reduce the top event significantly.

At present, amount of explosive for one blasting(X6) is too little to destruct the slope, groundwater effects(X7), are closely related with hydrogeological conditions and belongs to the objective category which make it cannot be influenced by human. The hydrogeological condition in this mining area is relatively simple, so it is not the key factor for the slope failure and can be negligible. So, as long as we control the steps caving(X2), the possibility of slope failure can be reduced.

6   Summaries

This paper analyzed the stability of open pit slope by Bayesian networks, found the main factors causing slope failure, which are for caving(X2), amount of explosive for one blasting(X6) and groundwater effects which illustrate that the possibility of slope failure in this open-pit is high. The slope failure is the main hidden trouble in open pit. Therefore, a stable slope come from many factors such as the design, mining methods, slope treatment etc. Bayesian network can compute the probability of the top event and find the possible factors that influence the top event. The new method can deal with the uncertainty and complex relationship of each factor which is very happened often in analyzing the slope stability.

References

[1] Jesen, F.V.: Bayesian Networks and Decision Graphs, pp. 126–135. Springer, New York (2001)
[2] Jensen, F.V.: An introduction to Bayesian networks, pp. 126–135. Springer, New York (1996)
[3] Dorner, S., Shi, J., Swayne, D.: Multi-objective modeling and decision support using a Bayesian network approximation to a non-point source pollution model. Environmental Modeling & Software, 211–222 (2007)
[4] Zhang, S., Jia, S., Guo, H.: Reliability Analysis of the Rock Slope Stability Rock and Soil Mechanics, pp. 57–66 (1999)
[5] Fang, J.-R., Zhu, H.-H., Cai, Y.-C.: Advancement and Methods of the Slope Stability Research. Chinese Journal of Underground Space and Engineering, 6–7 (2007)
[6] Van der Gaag, L.C.: Bayesian belief networks: odds and ends. The Computer Journal, 97–113 (1996)

Using Bayesian Networks to Measure Web Service QoS

Chang Guofeng

Department of Computer and Information Engineering
Xinxiang College

Abstract. Quality of Service (QoS) has been widely used to support Web Service selection and composition. Thus, QoS is very important for users to select the most suitable Web services. In this paper, we proposed a method to measure QoS of Web services. The key of the approach is to use Bayesian Network to measure each QoS value of Web services. The experimental results show that our proposed method can measure QoS of Web services effectively.

Keywords: QoS, Web service, QoS measure, Bayesian Network.

1   Introduction

Due to the increasing of Web services, it becomes very   important to identify the best Web service to satisfy users' QoS requirements. As it is well known that Web services recommendation for users has become an important research topic. In service recommendation, there are many Web services with the same functional attributes, but different non-functional   attributes (i.e., QoS) such as response time and availability. Thus, QoS is a very important factor, and it should be taken into more consideration during service recommendation.

In an open Web service environment, Web service QoS values are usually delivered by service providers. When different users invoke the same service in different environment, they won't have the same QoS experience on the same service. For instance, if a web service resident in China, a China user may can obtain better QoS experience than a non-China user. Thus, accurate QoS measure for service users is very important, which can help users to find the most suitable services in selecting services [1].

However, it is very difficult to obtain the accurate QoS measure. There are many factors to affect the accuracy of QoS measure. For example, different two users with different user context used a same service, but the QoS values of the service may be different for them. the assumption above is that the customer ratings are considered non-malicious and fairly accurate.   Hence, it is very important to build up reputation model to reduce the affect of malicious ratings. Some researchers propose other simpler models and approaches for reputation of Web services.

The authors of [2] firstly introduced some conceptions of web service and QoS, analyzed the insufficiencies of the current research of web service QoS. At the basis of this, a management framework for web service QoS based on multi-dimension QoS was presented. The applications of multi-dimension QoS in the framework were described and an application example was used to explain how the framework

manages QoS of web service   interaction in the practical application. The authors of
[3] presented a QoS prediction method, taking into account the collaborative filtering
techniques to find the smallest neighbours and the above analysis of the historical
experience of consumers have largely similar, therefore, collaborative filtering
technology was introduced and made some improvements, and thus it was suitable for
the QoS   prediction.

The authors of [4] proposed an approach to predicting QoS for web services, taking
the similarity among consumers' experiences into consideration. The basic
assumption of this approach is that the consumers, who have similar historical
experiences on some services, would have similar experiences on other services. Our
approach consists of two steps. First, The authors figured out the similarity between
each two consumers with their historical QoS data. Second, they predicted the quality
of unused service for consumers based on the data published by other consumers
regarding the similarity. The authors of [5] proposed an adaptive web service QoS
prediction approach based on the technique of forecast combination. Moreover, based
on the idea of forecast combination, a model selection algorithm was proposed for
selecting the potential forecasting models.

Although the studies stated above have obtained many affection, it is difficult to
accurately measure QoS values of web service. To address the problem, this paper
proposes a novel approach to measure QoS of Web services in dynamic and open
Web service environments. The approach firstly employ Bayesian Network to
establish the measure model of QoS, and then use it to measure each QoS values of
web services. In order to evaluate our proposed QoS measure approach, we conducted
a experiment. Experimental results show that our proposed approach is very effective
for improving QoS measure performance of web service in open and dynamic web
service environments.

The rest of the paper is organized as follows. Section 2 describes our proposed
approach. Section 3 gives experiments to evaluate our proposed approach. Finally,
conclusions are given in Section 4.

2   QoS Measure Approach

In order to facilitate understand our proposed approach, we firstly introduce Bayesian
Networks, and then give our proposed approach based on the application of Bayesian
Networks for Web services QoS measure.

Bayesian Networks is built upon the Bayesian approach. The Bayesian approach is
built upon the well known Bayes' Rule, which is itself derived from the fundamental
rule for probability calculus as follows

$$P(a,b) = P(a \mid b) * P(b) \,. \tag{1}$$

In Equation 1, P(a,b) is the joint probability of both events a and b occurring, P(a|b) is
the conditional probability of event a occurring given that event b occurred, and P(b)
is the probability of event b occurring.

Although not included here, further derivation produces Bayes' rule   as follows

$$P(b \mid a) = \frac{P(a \mid b) * P(b)}{P(a)} \,. \tag{2}$$

Bayes' rule not only opens the door to systems that evolve probabilities as new evidence is acquired, but also, as will be seen in the next section, provides the underpinning for the inferential mechanisms used in Bayesian belief networks [9].

Despite its benefits, the Bayesian approach also has drawbacks. One drawback is the difficulty of obtaining accurate conditional probabilities. When adequate data is unavailable, sometimes experts must estimate the missing probabilities subjectively. Another drawback is that the approach can be computationally intensive, especially when the variables being studied are not conditionally independent of one another.

A Bayesian network is a directed acyclic graph that provides a compact representation or factorization of the joint probability distribution for a group of variables. Graphically, a Bayesian network contains nodes and directed edges between those nodes. A simple illustration is provided in Figure 1.
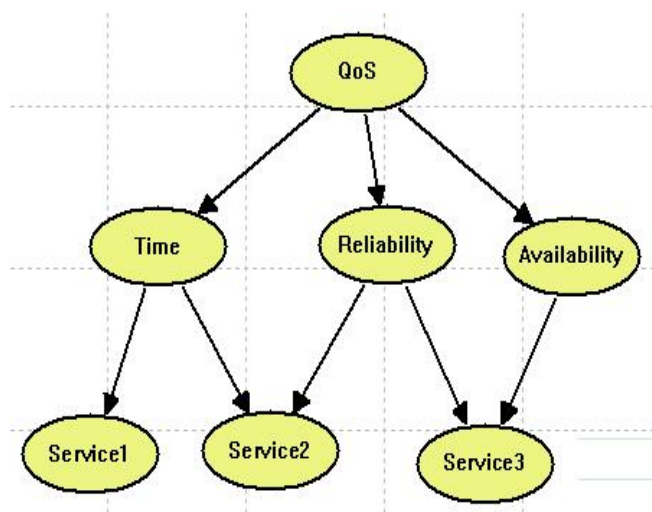
Fig. 1. A Bayesian network

Each node is a variable that can be in one of a finite number of states.   The links or arrows between the nodes represent causal relationships between those nodes. All of the variables in Figure 1 are Boolean variables, but there is no restriction on the number of states that a variable can have. Because the absence of an edge between two nodes implies conditional independence, the probability distribution of a node can be determined by considering the distributions of its parents. In this way, the joint probability distribution for the entire network can be specified [10]. This relationship can be captured mathematically using the chain rule by using

$$p(x) = \prod_{i=1}^{n} p(x_i \mid parents(x_i)) \tag{3}$$

In general terms, this equation states that the joint probability distribution for node x is equal to the product of the probability of each component xi of x given the parents of xi.   Each node has an associated conditional probability table that provides the

probability of it being in a particular state, given any combination of parent states. When evidence is entered for a node in the network, the fundamental rule for probability calculus and Bayes' rule can be used to propagate this evidence through the network, updating affected probability distributions. Evidence can be propagated from parents to children as well as from children to parents, making this method very effective for both prediction and diagnosis [9-12].

By acquiring the information of QoS using probability mechanism, the approach gets the QoS attributes of Web services. The function of the approach is to build up clear expressions and complete understanding of the QoS information, including Response time, Availability, Reliability, Throughput, Compliance, and Latency. By recoding the web service QoS information, the approach give the appropriate content to the each web service and then web services can receive effective measure.

In our approach, it is to build up the network topology between web service publishers and service users. We list the knowledge structure as show in Figure 2. For Figure 2, based on the established Bayesian Network, our approach use the information of service publishers and service users to measure QoS.
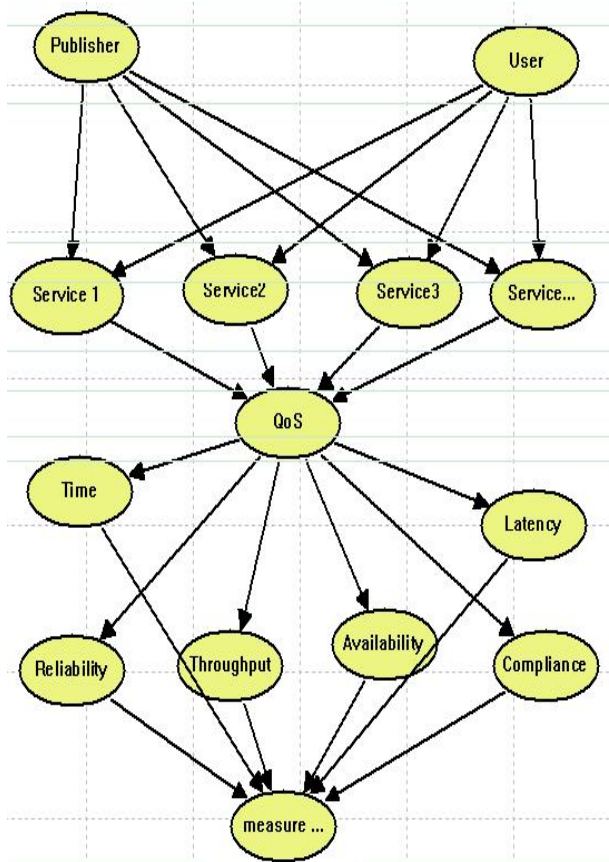
Fig. 2. Bayesian network for measure QoS

3   Experiments

In order to evaluate our proposed approach, we conduct a experiment to verify the effectiveness of our approach. In this experiment, we employ a real web service QoS dataset to establish QoS measure Bayesian Network model. The experimental results demonstrate show that our proposed approach can measure web service QoS effectively.

In our evaluation we experimented with a real QoS dataset. The dataset is the QWS real dataset from [6-7] This dataset includes measurements of 9 QoS attributes for 2500 real web services. The main goal of this dataset is to offer a basis for Web Service researchers. For the dataset, Eyhab Al-Masri have collected 5,000 web services and performed various measurements on this dataset. The services were collected using our Web Service Crawler Engine (WSCE). The majority of Web services were obtained from public sources on the Web including Universal Description, Discovery, and Integration (UDDI) registries, search engines, and service portals. The public dataset consists of 365 Web services each with a set of nine Quality of Web Service (QWS) attributes that he have measured using commercial benchmark tools. Each service was tested over a ten-minute period for three consecutive days. For more details about this dataset we refer the reader to [6,8].

We conducted our experimental results from a PC with Intel Core2 2.0GHz processor, 2.0GB of RAM. The machine is running Windows XP SP3 and RESS Suite 3.0 for establishing Bayesian Network. The experimental results can be obtained on average when the experiments runs for 5 times.
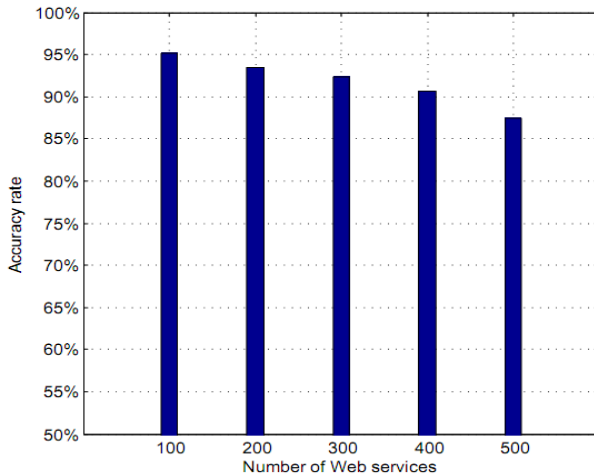
Fig. 3. Accuracy rate

As shown in Fig. 3, by using to the real web service QoS dataset, the experiment results was collected.   The number of web services is from 100 to 500. The results above show that the accuracy rate of our proposed approach is larger than 90% in this experiment. Especially,   when the number of web services is 100, the accuracy rate of our approach is higher than 95%. Thus, according to the experimental

results on accuracy rate, our approach is very effective in QoS measure for web service, i.e., service users can accurately find the most appropriate service by using our proposed approach.

4   Summaries

With the increasing of Web services, it becomes more important to identify the best web service to satisfy an user's requirement. Consequently, the accuracy of web service QoS for users has become an important research topic. Since web services are in an open environment, functional, but also non-functional properties, such as response time and availability, should be taken into consideration during QoS measure. However, it is difficult to measure web service QoS because of some factors. For example, the web service environment is dynamic, which   means that service providers are non-persistent and may become unavailable unpredictably. Moreover, services are owned by various stakeholders with different aims and objectives and there is no central authority that can control all the service providers and consumers.

In this study, we present a web service QoS measure approach. The approach employ   Bayesian Network to establish the measure model of QoS, and use it to measure each QoS values of web services. Experimental results show that our approach can provide accurate QoS measure for web services.

References

[1] Xu, Z., Martin, P., Powley, W., Zulkernine, F.: Reputation-enhanced QoS-based web services discovery. In: Proceedings of 2007 IEEE International Conference on Web Services (ICWS 2007), Salt Lake City, UT, United states, July 9-July 13, pp. 249–256 (2007)
[2] Lu, L., Guangquan, Z.: A management framework for web service QoS based on multi-dimension QoS. In: Proceedings of 2010 5th International Conference on Computer Science and Education (ICCSE 2010), pp. 907–911 (2010)
[3] Li, C.-C., Cui, L.-Q., Deng, Y., Feng, W.-X.: A QoS prediction approach based on improved collaborative filtering. In: Proceedings of 2010 2nd International Conference on Advanced Computer Control (ICACC 2010), pp. 519–522 (2010)
[4] Lingshuang, S., Jing, Z., Yong, W., Junfeng, Z., Bing, X., Hong, M.: Personalized QoS Prediction forWeb Services via Collaborative Filtering. In: Proceedings of 2007 IEEE International Conference on Web Services (ICWS 2007), pp. 439–446 (2007)
[5] Jing, L., Yongwang, Z., Jiawen, R., Dianfu, M.: Towards adaptive web services QoS prediction. In: Proceedings of 2010 IEEE International Conference on Service-Oriented Computing and Applications (SOCA 2010), pp. 1–8 (2010)
[6] Alrifai, M., Risse, T.: Combining global optimization with local selection for efficient QoS-aware service composition. In: Proceedings of the 18th International Conference on World Wide Web (WWW 2009), New York, United States, pp. 881–890 (2009)
[7] Alrifai, M., Skoutas, D., Risse, T.: Selecting skyline services for QoS-based web service composition. In: Proceedings of the 19th International Conference on World Wide Web (WWW 2010), Raleigh, North Carolina, USA (2010)
[8] Al-Masri, E., Mahmoud, Q.H.: Investigating web services on the world wide web. In: Proceedings of the 17th International Conference on World Wide Web (WWW 2008), Beijing, China, April 21-April 25, pp. 795–804 (2008)

Fiber Image Transmission System Base on FPGA

Ke-Lin Sun[1,2,3], Wei-Chao Zhou, and Qin-Zhang Wu

[1] Institute of Optics and Electronics Chinese Academic of Sciences,
Chengdu, Sichuan, China
[2] Univ. of Electronic Science and Technology of China, Chengdu 610054, China
[3] Graduate University of Chinese Academy of Sciences
Beijing, China
colin7612@126.com

Abstract. This design used FPGA and its embedded Rocket IO IP core, to accomplish interfaces conversion between the camera link parallel and the Rocket IO serial, using a custom image transmission protocol, to build a platform which interconnect the sender with the receiver, system realize high-speed real-time image transmission on optical fiber. This paper describes the thought of system collectivity design, the software and hardware design of transmitter, the software and hardware design of receiver, and the key technologies among the system, such as attributes of Rocket IO, transceiver state machines of data-packet and data transmission protocol are introduced in detail. The experimental results indicate that system work rate up 2.5Gb/s, the transmission link is stable, reliable and no error code, the advantages of this system is high bandwidth, high anti-jamming, and reduces the amount of wiring and so on.

Keywords: FPGA, Rocket IO, Camera Link Interface, Fiber Image Transmission, Image Transmission Protocol.

1   Introduction

Aim at optical tracking device, which requires a stability and reliability high-speed real-time image transmission system, the quality of image transmission will directly affect the result of image processing. It is one importance factor of the equipment performance, but traditional image transmission by standard camera link cable, so the cable is interface complex, and susceptible to electromagnetic interference, and transmission distance is limited. Over more, systems need to transmit image through slip ring, such system unable to transmit high-speed real-time image data. In recent years, as high-speed serial transmission interconnect technology has become sophisticated, a high-speed, wide bandwidth, and strong anti-interference serial transmission system is more and more favor of designers, as Xilinx introduces its FPGA embedded Rocket IO IP core solution, building a bridge between FPGA and optical fiber communication, implement high-speed serial transmission from chip-level inter-connected to board-level inter-connected and system-level inter-connected. As optical fiber technology has more and more popular, optical fiber slip ring using optical fiber technology, many advantages of optical fiber slip ring are no contact, no friction, no electromagnetic interference, long life, high speed, and wide bandwidth, so widely used in industrial, military and other fields.

After combined with domestic and abroad research conditions, this design achieve experimental data transmission by optical fiber,   above this basis, this design defining a custom transmission protocol, and used the protocol to achieve high-speed real-time image transmission. The experimental results indicate that system work rate up 2.5Gb/s, the transmission link is stable, reliable and no error code, and also have low cost of development, transplantable and other advantages.

2   System Collectivity Design

Block diagram of as shown in Figure 1, the system compose camera link transform to optical fiber of transmitter and optical fiber transform to camera link of receiver two parts, the transmitter implement camera link data signal input and optical fiber data signal output, the receiver implement optical fiber signal input and camera link signal output, transmission between the transmitter and the receiver by optical fiber. At the transmitter, real-time image from camera, the camera output camera link standard data, and directly connect to transmitter, because of theodolite work at turning, the real time image have to transmission through the optical fiber slip ring. At the receiver, reconstruct image from optical fiber data and output camera link interface data.
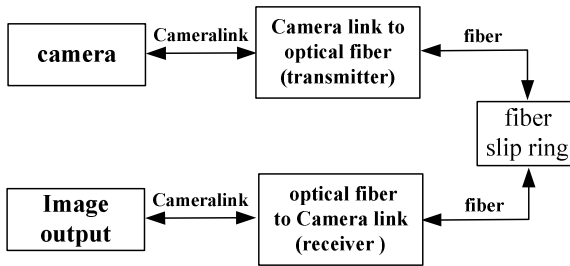
Fig. 1. Block diagram of system

3   Hardware Design

3.1   Hardware Design of Transmitter

Block diagram of camera link transform to optical fiber of transmitter as shown in Figure 2, the transmitter compose camera link demodulation module, data processing module, clock module, program storage module, optical transceiver modules, in order to meet Base-type Medium-type, and Full-type Camera link interface configuration, using DS90CR288A for camera link demodulator chip, to complete camera link data demodulation. Because of the standard Camera link input level is LVDS, at hardware design, have to used camera link demodulator to conversion the LVDS to TTL level, select XC2VP20FG676 series of FPGA for data processing module, there are 8 ROCKTIO core integrates on this chip, each ROCKTIO channel is available from 600 Mb/s to 3.125Gb / s transmission speed, select xcf32p PROM for program storage module, and the clock module is IDT5V9885 chip, select high-precision crystal
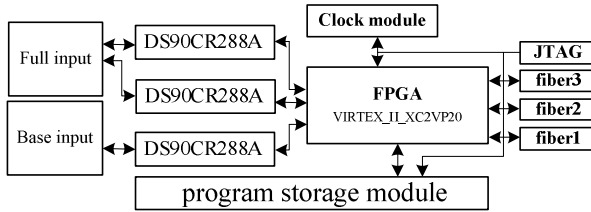
Fig. 2. Block diagram of transmitter system

JFVNY-1007-X075NAGRC of 25.00000MHz for the clock module's input clock source, the clock module can output requires flexible configuration clock frequency, this design uses a clock frequency of 125MHz, Virtex- II of ROCKTIO frequency within 20 times up to 2.5Gb/s transmission speed, using OPT0STAR's CWDM optical transceiver module SFP-2.5G, and the wavelength of CWDM optical transceiver is 1470nm-1610nm, total of eight wavelengths.

3.2   Receiver Hardware Design

Block diagram of fiber optic to camera link transform to receiver as shown in Figure 3, the receiver hardware design is mostly the same as transmitter, only one difference is that the camera link modulation module parts, using DS90CR287 chip to complete the Camera link data modulation, while achieving TTL level to the Camera link LVDS level translator.
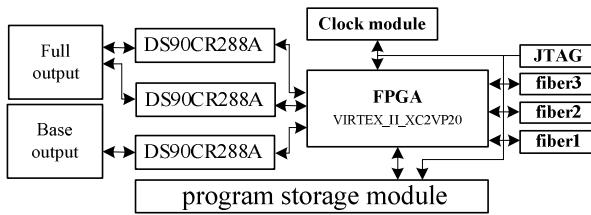
Fig. 3. Block diagram of receiver system

4   Software Design

4.1   Rocket IO Attributes Design

Because of high-speed serial transmission system requires a strict accuracy of clock design, this design used IDT5V9885 chip's output clock as the Rocket IO reference clock,   when Rocket IO transfer rate up to 2.5Gb / s, the reference clock is 125MHz, and clock accuracy is 50ppm, used differential input for the clock input, select BREFCLK as a reference clock input of the Rocket IO, and select static control mode of BREFCLK input, when select REFCLK for reference clock input, Rocket IO only reach the maximum transmission speed is 1.25G/s. select 50 Ohm for impedance terminal, select 8B/10B encoding static control mechanisms, pre-emphasis select 10%,

the output voltage swing is 500mV, both the output voltage swing and the pre-emphasis are not a choice the higher the better, because too high pre-emphasis will overshoot output voltage swing, will lead to increased transmission link error rate, therefore, need to experimental tests to adjust the value of the output voltage swing, making the transmission link to achieve stability requirements. Select comm characters and character alignment check sequence, Align comm MSB option must be chosen so that the character can be properly aligned comm in the high byte, if not the character may be aligned comm in the high byte, or may be aligned in the low Bytes, so the serial data stream occur errors. clock correction, when the write speed of input data rate and the read speed of output data rate is inconsistent in Rocket IO receiver 's FIFO, so Rocket IO need clock correction, the FIFO's input data rate is determined by RXRECCLK recovered clock, and the FIFO's output data rate is determined by RXUSRCLK. In order to eliminate the differences between the two clock is insert or remove idle time slot of the clock correction model, the design of the clock correction sequence length is four bytes, the clock correction and alignment of sequences of characters are as 00111111100, 00000000000, special characters FC00 is K28.7, selected second clock correction sequence template, channel bonding is not used.

4.2 Data Packet Protocol

Image data transmit is sent by data packets, block diagram of a custom definition of data packet transmission protocol format as shown in Figure 4, each packet of data Length is 1K bytes, the data width is16bit, so the length of each packet is 512 clock cycles, it is necessary to consider the maximum time interval through clock correction length, because during the sending images times, only send valid data but not send clock correction sequence, have to calculate the frequency of clock correction, each time during send a packet of data using $512 + 8 = 520$ clock cycles, that is the maximum interval time between the two clock correction is 520 clock cycles, while the reference clock for the stability of 50ppm, in accordance with the requirements of each Rocket IO need at least 10,000 characters to send a clock correction sequence, as is $10000 \div 16 = 625$ clock cycles, this interval time is greater than the designed requirements of Rocket IO indicators. Diagram of state machine transfer as shown in Figure 5, S0 state is clock correction, at state S0 sending alignment, the clock correction sequence, state S1 is character at S1 state sending character alignment sequence, and character alignment sequence need to be sent after send the clock correction sequence, when there is no request for sending data packet, at that times continue to send the clock correction, to maintain the link stability, if request to send valid image data packet, then the state machine turn to S2 state to send data packets, start from the S3 state to sent packet head, packet head is the beginning of the packet, used for the receiver to receive test flag, the flag with the number of K characters, state S4 is packet length, used to indicate the length of valid data to send, in this design, packet length is 1K-byte, state S5 is valid data, sending valid data immediately after sending packet length, when the valid data transmitted after 1K-byte length, sending valid data is end in this packet, state S6 is packet serial number, at state S6 to a corresponding number of each packet, the number will used for the receiver to receive an effective check if the transmission loss packet or no, state S7 is packet end, used to indicate the end of sending of the current package.

Packet head (two bytes)	Packet length (two bytes)	valid data	Packet serial number (two bytes)	Packet end (two bytes)

Fig. 4. Block diagram of data packet format
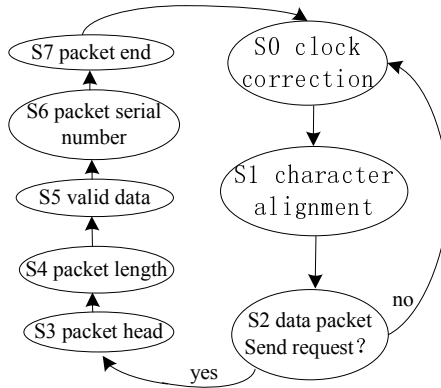
Fig. 5. Diagram of state machine transfer

5   Experiment Test

According to the description of Virtex-II PRO's Rocket IO attribution, at the design, when the transmission link has an error, RXNOTINTABLE signal output an indicates error flag, and the RXNOTINTABLE has two bits, each bits represent the two bytes of data, when transmission link is stability an error-free, RXNOTINTABLE signal has been maintained at the "00"value, when the signal is 1, it said that transmission link error had been occur. Testing method as follows, the transmitter sends the image data in real time, at the receiver, receiving the image data too, and used the Chip-Scope-Pro tools to real-time sampling and monitoring RXNOTINTABLE signal in trigger conditions, experimental results show that the continuous monitoring of 26 hours without an error. The diagram of experimental results are shown in Figure6 and Figure7.
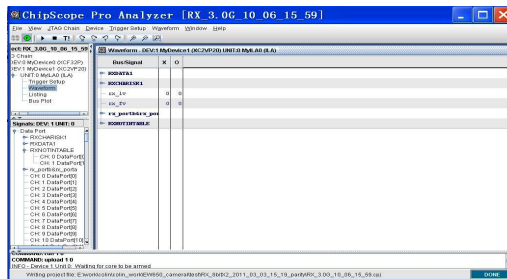
Fig. 6. Diagram of monitoring no error
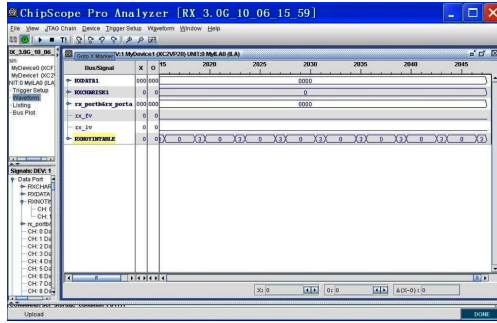
Fig. 7. Diagram of monitoring has error

6   Conclusion and Future Work

This design solved the issues of bandwidth narrow, transmission rate low, and electromagnetic interference in projects. After this system design, including the hardware design, and the software design can provided a system platform, on this basis, used a custom transmission protocol, to achieve the system interconnect stability, to provide a total solution for the project, the performance with this system is reliable and stable. In future, to further enhance the transmission rate and transmission bandwidth, after experimental tests, the transmission rate can up to 3.0Gb/s, and the transmission link can be stable, the   rate will increases up to 3.125Gb/s in future, or to use more powerful Virtex-5 FPGA chip can achieve higher bandwidth. Thus, a Rocket IO channels can be time division and multiplexing to transmit multiple images, making the system direction to the highly integrated, small and micro.

References

[1] Wang, P.K.: Study of High Speed Interconnect Technology in Image Processing. Graduate University of the Chinese Academy of Sciences, Beijing (2005) (in Chinese)
[2] Xilinx. Rocket IO Transceiver User Guide (June 2004)
[3] Xilinx. Virtex-II Pro and Virtex-II Pro X Platform FPGA: Complete Data Sheet (October 2005)
[4] Xilinx. Virtex-II Pro and Virtex-II Pro XFPGA User Guide (November 2007)
[5] Camera Link committee. Camera Link Specifications (October 2000)

The Computer Simulation Design of Error-Correcting Code in Communication Channel

Chu Lili

School of Electronic and Information Engineering, LiaoNing University of Technology, LiaoningJinzhou China 121001

Abstract. Channel coding is an error control measure that widely used in modern communication system. It uses the relationship between the supervise code and the information code to detect the fault codes that may be existed. It has great capability of error detection and easy to realize. This system is the CRC and convolutional coding system., which using the VHDL hardware description language programming, and realize the waveform simulation of the entire encoder on the platform of MAX+plusII. It is certified that the result of simulation is consistent with the theoretical coding.

Keywords: error control, CRC code, convolutional code, VHDL language.

1 The Significance of Error-Correcting Code Research in Communication Channel

With the rapid development of digital communications, data processing and computer communication networks, users need a higher requirement of the reliability of data transfer and storage, so it must be addressed to improve the reliability and validity of information transmission in a digital communication. It is proved that improves processing technology, circuit technology, mechanical technology and improve the reliability of components to improve the reliability of the system in the past, though is a way to improve reliability, while this method is very difficult to improve reliability because that it not only takes high price, but also is difficult to obtain satisfactory efficiency. Nowadays the most popular method to improve system reliability is using error-correcting code technique. As a result of error-correcting code can improve the reliability of the communication, so more attention has been paid. In recent years, the application of algebraic theory provides a theoretical basis for error-correcting code, The development of large scale integrated circuit for error-correcting code provides the material foundation, particular encoder and decode can be simulated by computer software, which promotes the application and development of error-correcting code technology. CRC checksum has become a essential part of transmission lines of various types of CCT proposal in the communication field. In mobile communications, error-correcting code is widely used in analog satellite communication, computer storage and computing system, also widely used in VLSI design.

2   VHDL Hardware Description Language

One of the key technologies of electronic design automation (EDA) is required to use formal method to describe the hardware of digital circuit system, which uses so-called hardware language to describe the hardware. The development of hardware description language has been a few decades, the more famous of which have VHDL language, AHDL language and VerilogHDL language. While VHDL language (VHSIC Hardware Description Language) developed by the U.S. Department of Defense in the late 1980s, became the IEEE standard (designated as IEEE std l076-1987 standard) in December 1987, which laid a solid foundation for popularization and promotion of electronic design automation. Then IEEE modified the version of the 1987, and launched a more comprehensive old version (designated as ANSI / IEEE std l076-1987 standard) in 1993, that made the VHDL language programming more flexible. Since then, more and more people start using VHDL to design digital circuit system.

A complete VHDL language program often has 5 parts including Entity, Architecture, Configuration, Package and Library. The first four is the source design unit that can be compiled respectively. Entity is used to describe the external interface signal of the designed circuit system, while the attribute, system input and output ports are all defined in the entity; Architecture is used to describe the structure and behavior within the system, which uses language to describe functions of the system to be achieved; Configuration is used to select the desired unit from the Library to form different versions of the system design; Package storages data types, constants, subroutines and etc that various design module can be shared ; While Library storages the compiled Entity, Architecture, Package and Configuration. Library generated by the user or by the ASIC chip manufacturers is designed for everyone sharing. This function can be achieved by specific statement. The most important thing In VHDL language program is Entity and Architecture, that are the core of a specific design. If Entity is described a bridge between system and external communication, then Architecture is the signal processing plant to deal with internal and external communication. By VHDL statement and reasonable design, Architecture accepts external input signal. Then various signals in Architecture are coordination. Finally the desired output is obtained, which is sent to the output port and external part to exchange.

3   Error Control Coding in Communication

To add authotity code into infornation code is called error control coding, sometimes called error correction coding.For different coding methods, There are different error detection or error correction, and some coding can only detect error, not correct error. In general, the more cost takes, the more powerful detection(correction) is. The cost mentioned here is that the increased number of authotity code, which is usually

measured by redundancy. For example, if an authotity code is in an average of two information codes in coding sequence, then the redundancy of this coding is 1 / 3. To put it another way, it can be said that the coding efficiency of the coding is 2 / 3. It is Visibled that error control coding in principle reduces information transfer rate (because it sends effective information and error-correcting code that is not effective information at the same time) to improve transmission reliability.In principle, it is various to increase the man-made edundancy and method, which can be divided into two types: If the rule is linear, meaning the relationship between codes is Linear, the pass code is called linear code, otherwise nonlinear code. Although it is likely that nonlinear code have better performance to linear code, but for the lack of depth on the theoretical and practical research, is not discussed here. In addition, the coding manner that the source (signal sender) information corresponds can also be further divided into two types: It is called block code if information sequence of the source is processed and coded according to independently group, otherwise called non-block code, such as convolutional code is the most important category.

The linear of linear block codes is meaning that constraint relation among small code of block codes is linear, while the block is in terms of coding method. Namely each k codes information bits are divided into a block for independent processing in coding process,transformed the binary code block of length n (n> k), and denoted by (n, k), where n is the length of output code block,k represents the input information block.while redundant codes are added according to certain linear rule.

4   Design of CRC Plus Convolutional Coding System

The system in this paper is the CRC plus convolution coding system, whose channel coding process is that the information sequence of each burst contains 25-bit binary information, is conversed into a information sequence of 35 bits after CRC coding. The front 25 bits is original information sequence, the back 10 bits is CRC checksum bit. This result is as the input of convolution coding, then the output of information code became a total of 78 bits after convolutional coding.

VHDL hardware description language is divided into two parts to write programs. First it is composed of the CRC coding part, then is used its output as the input of convolutional encoder to compose convolutional coding part. Finally, the simulation waveform of the encoder is implemented in the MAX + plusII work platform. CRC plus convolutional encoder is a channel encoder. Judging from the whole, its input ports are: asynchronous set input port reset, CRC coding and convolutional coding clock signal input port clk, information sequence input port data-in, ready input port rdy, the local clock (designated as the code output clock after the coding completion). The output ports of channel encoder are: synchronization input displaying output port syn_in, coding result output port data_out, synchronization output displaying output port syn_out.

Pin map of the whole encoder showed in Fig.1.
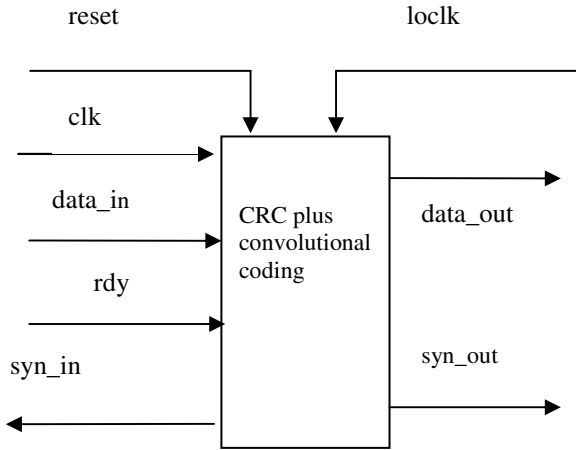
Fig. 1. Pin map of the channel encoder

Simulation result of CRC encoder showed in Fig.2:
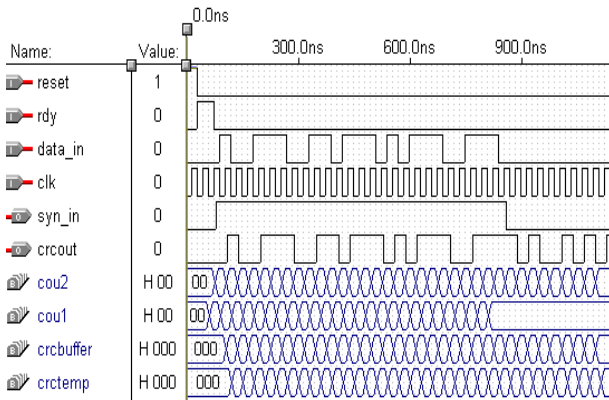
Fig. 2. The waveform simulation of CRC code after the input of information code

The output of CRC encoder will be as the input of convolutional coding after the design of the encoder is completed. Entity design is done according to the pin map of the whole encoding in Fig.1. Designed as follows:

```
library ieee;
use ieee.std_logic_1164.all;
use ieee.std_logic_arith.all;

entity crcjj is
port(clk:in std_logic;--CRC encoder and the convolutional code clock time
localclk:in std_logic;-- local clock of convolution coding output
reset:in std_logic;-- asynchronous set signal input port
```

data_in:in std_logic;-- information sequence input port
syn_in:out std_logic;--synchronous input display output port
rdy:buffer std_logic;-- ready signal port
syn_out:buffer std_logic;--synchronous output display output port
data_out:out std_logic -- coding output port
     );
end crcjj;
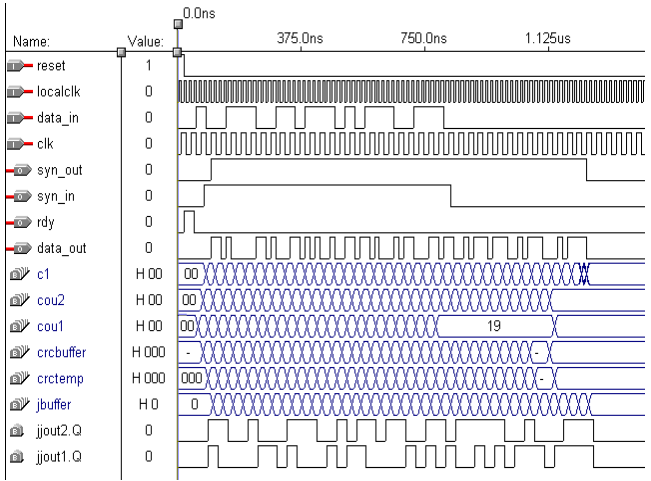
Fig. 3. The waveform simulation of the whole encoder

5   Summaries

The wave simulation result in the design in the MAX + PLUS II working platform is perfectly consistent with theoretical value. And the CRC plus convolutional coding system can achieve the channel coding process. The simulation exports the original information code and check sequence. By detecting any redundant item in channel decoding, the appearance of error code will be judged, while the error control function can be implemented in communication system.

References

[1] Zhao, X.: Design of VHDL and Digital Circuit. Machinery Industry Press, Beijing (2005)
[2] Jiang, Q., Lu, X.: Convolutional encoder fast implementation in communication system. Information Technology (5), 35–37 (2005)
[3] Xu, K.: Technology and application of error correction channel coding. Chongqing Technology and Business University 22, 481–484 (2005); Study
[4] Gu, F., Lai, L., Li, W.: Information classified code for Information sharing and study of management system. Information Science 27(11), 74–77 (2008)
[5] Smarandaehe, R.: Maximum distance separable convolutional codes, construction and decoding. PhD thesis, University of Notre Dame (August 2001)

P-Sets and $\bar{F}$-Memory Information Characteristic

Wang Yang[1], Zheng Xin[1], and Zhang Caijun[2]

[1] Department of Computer Science Huanghuai University, Zhumadian Henan, China
[2] Department of Information Technology
Anyang Power Supply Company
Anyang Henan, China

Abstract. Packet sets is a new mathematical model and mathematical method. By using the structure and dynamic characteristic of packet sets, the concept of $\bar{F}$ - memory informatio, $\bar{F}$ - memory information generation and $\bar{F}$ - memory circle are given, the $\bar{F}$ - memory information existing theorem, the $\bar{F}$ - memory information recovery theorem and the $\bar{F}$ - memory information characteristic theorem are proposed. By using the results, the application of $\bar{F}$ - memory information is given.

Keywords: P-sets, F-memory information, memory measurement, memory characteristic theorem.

1 Introduction

A finite Cantor set X possesses accuracy, boundary certainty and static characteristics. In 1965, L.A.Zadeh proposed fuzzy sets by replacing the boundary certainty with boundary uncertainty [1]; in 1982, Z.pawlak proposed rough sets by replacing accuracy with approximation [2]; in 2008, the Refs. [3] proposed P-sets by replacing static with dynamic characteristic. Dynamic P-sets are obtained by improving Cantor set.

If $X = \{x_1, x_2, x_3, x_4, x_5\}$ is one of information sets of information system W, $\alpha = \{\alpha_1, \alpha_2, \alpha_3\}$ is attribute set of X .If attribute is replenished in α , α separately become $\alpha_1^F = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4'\}$, $\alpha_2^F = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4', \alpha_5''\}$, $\alpha_3^F = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4',$

$\alpha_5'', \alpha_6''', \alpha_7'''\}$, $\alpha \subseteq \alpha_1^F \subseteq \alpha_2^F \subseteq \alpha_3^F$; then X separately become $X_1^{\bar{F}} = \{x_1, x_3, x_5\}$,

$X_2^{\bar{F}} = \{x_3, x_5\}$, $X_3^{\bar{F}} = \{x_5\}$, $X_3^{\bar{F}} \subseteq X_2^{\bar{F}} \subseteq X_1^{\bar{F}}$. This is interesting and common phenomenon, it is often seen about risk investment and benefit analysis by people. memory character of creature is taken into the phenomenon and realize it again: because α become $\alpha_1^F, \alpha_2^F, \alpha_3^F$, $\alpha_1^F \subseteq \alpha_2^F \subseteq \alpha_3^F$; $X = \{x_1, x_2, x_3, x_4, x_5\}$ become $X_1^{\bar{F}} = \{x_1, x_3, x_5\}$, $X_2^{\bar{F}} = \{x_3, x_5\}$, $X_3^{\bar{F}} = \{x_5\}$; $X_3^{\bar{F}} \subseteq X_2^{\bar{F}} \subseteq X_1^{\bar{F}} \subseteq X$, obviously, $X_1^{\bar{F}}$ is from x_1, x_3, x_5 of X remembered , $X_2^{\bar{F}}$ is from x_3, x_5 of X remembered, $X_3^{\bar{F}}$ is from x_5 of X remembered .which is character of P-sets.

The paper provides $\bar{F}$ - memory information and memory measure concept of $\bar{F}$ - memory information, character theorem of $\bar{F}$ -memory information is given

through these concepts, provide identification method of $\overline{F}$-memory information and $\overline{F}$-memory information is used on economical application. More character and application of P-set is seen in [3-15].

2 $\overline{F}$-Memory Information and Its Generation Dependency

Assumption set X, X^F in $\amalg$, is separately regarded $(x),(x)^F$,or $(x) = X,(x)^F = X^F$,and can not be misunderstood .

Definition 1. (x) is one of information U ,what is more $(x) = \{x_1, x_2, \cdots, x_q\}$. $x_i \in (x)$ is called one of information element of (x) , $i = 1,2,\cdots,q$,if (x) possess character set α ,furthermore $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_k\}$.

Definition 2. $(x)^F$ is one of $\overline{F}$-memory information that is generated from $(x) \subset U$,if character set α of (x) and character set α^F of and $(x)^F$ meet $\alpha^F = \alpha \cup \{\alpha' \mid f(\beta) = \alpha' \in \alpha,$ $f \in F\}$.

Definition 3. $\nabla(x)^{\overline{F}}$ is memory lost of $\overline{F}$-memory information, in short $\nabla(x)^{\overline{F}}$ is $\overline{F}$-memory lost, furthermore $\nabla(x)^{\overline{F}} = (x) - (x)^{\overline{F}}$.

Definition 4. information (x) single depend on $\overline{F}$-memory information $(x)^{\overline{F}}$,is signed $(x)^{\overline{F}} \Rightarrow (x)$ .Information (x) double depend on $\overline{F}$-memory information $(x)^{\overline{F}}$,is signed $(x)^{\overline{F}} \Leftrightarrow (x)$. Here : " $\Rightarrow$ ", " $\Leftrightarrow$ "come from symbolic logic ; " $\Rightarrow$ "equal" $\subseteq$ "; " $\Leftrightarrow$ " equal " $=$ ".

   We can get from definition 1-4:

Proposition 1. Character set α^* of information $(x)^*$ and character set α of information (x) meet $\alpha^* - \alpha \neq \phi$. $(x)^*$ must be one of $\overline{F}$-memory information of (x) , $(x)^* = (x)^{\overline{F}}$;otherwise it is true .

Proposition 2. $\overline{F}$-memory information $(x)^{\overline{F}}$ and information (x) meet IDE$((x)^{\overline{F}},(x))$ .Otherwise it is true. Here: IDE=identification.

Proposition 3. $\overline{F}$-memory lose $\nabla(x)^{\overline{F}} \neq \phi$; otherwise it is true.

Proposition 4. Information (x) double depend on $\overline{F}$-memory information $(x)^{\overline{F}}$,t is character set α, α^F meet UNI(α, α^F) .Otherwise it is true. UNI=un identification.

Theorem 1. ($\overline{F}$-memory information exist theorem) has given information (x) , $\alpha = \{\alpha_1, \alpha_2, \cdots, \alpha_k\}$ is character set of (x) , if exist character set α^* meet $\mathrm{card}(\alpha^*) \geq k$.
   Then information $(x)^*$ of character of α^* is $\overline{F}$-memory information of (x) , $(x)^* = (x)^{\overline{F}}$. where: card=cardinal number, $k \in N^+$.

Corollary 1. If character set α^F of $\overline{F}$-memory information $(x)^{\overline{F}}$ and character set α of information (x) meet $\mathrm{card}(\alpha^*) - \mathrm{card}(\alpha) > 0$. Then $(x)^{\overline{F}} \Rightarrow (x)$.

Theorem 2. ($\overline{F}$-memory information restoration theorem) $\overline{F}$-memory information $(x)^{\overline{F}}$ is restored information (x) , or UNI$((x),(x)^{\overline{F}})$ .Its sufficient condition is $(\alpha^F - \{\beta_i \mid \alpha_i \in \alpha^F, \overline{f}(\alpha_i) = \beta_i \overline{\in} \alpha^F, \overline{f} \in \overline{F}\}) - \alpha = \phi$.

   Here: α^F is character set of information $(x)^{\overline{F}}$, α is character set of information (x) .

Corollary 2. If $\bar{F}$-memory information $(x)^{\bar{F}}$ is restored into information (x), then $\bar{F}$-memory lost $\nabla(x)^{\bar{F}}$ meet $\nabla(x)^{\bar{F}} = \phi$.

3 $\bar{F}$-Memory Measure and $\bar{F}$-Memory Property Theorem

Definition 5. γ^F is $(x)^F$ about $\bar{F}$-memory measure of (x), in short γ^F is $\bar{F}$-memory measure of $(x)^{\bar{F}}$, if $\gamma^{\bar{F}} = \mathrm{card}((x)^{\bar{F}}) / \mathrm{card}((x))$.

Definition 6. $\mathcal{O}^{\bar{F}}$ is $\bar{F}$-memory circle is generated from $\bar{F}$-memory information $(x)^{\bar{F}}$, if coordinate origin O is centre of $\mathcal{O}^{\bar{F}}$, $\rho^{\bar{F}}$ is radius of $\mathcal{O}^{\bar{F}}$.

Here : $\rho^{\bar{F}} = \| Y^{\bar{F}} \| / \| Y \|$, $\| Y^{\bar{F}} \| = (Y_1^2 + Y_2^2 + \cdots + Y_p^2)^{1/2}$ is vector 2-normal number of $Y^{\bar{F}} = (Y_1, Y_2, \cdots Y_p)^T$, $Y^{\bar{F}}$ is vector is generated by $Y^{\bar{F}} = \{Y_1, Y_2, \cdots Y_p\}$, $Y^{\bar{F}}$ is information result y_i of information element $x_i \in (x)^{\bar{F}}$ of $(x)^{\bar{F}} = \{x_1, x_2, \cdots, x_P\}$, $i = 1$,

$2, \cdots, p$, $\| Y \| = (Y_1^2 + Y_2^2 + \cdots + Y_q^2)^{1/2}$ is 2-normal number of vector $Y = (Y_1, Y_2, \cdots Y_q)^T$,

Y is vector which is generated by $Y = \{Y_1, Y_2, \cdots Y_q\}$, $Y = \{Y_1, Y_2, \cdots Y_q\}$ is information result y_j of information element $x_j \in (x)$ of $(x) = \{x_1, x_2, \cdots, x_q\}$, $j = 1, 2, \cdots, q$;

$p \leq q$, $p, q \in N^+$; $x_i, x_i \in R$.

Definition 7. $\mathcal{O}$ is information unit circle which is generated from information (x), if coordinate origin O is centre of $\mathcal{O}$, ρ is radius of $\mathcal{O}$.

Theorem 3. ($\bar{F}$-memory information measure theorem) if $(x)_i^{\bar{F}}, (x)_j^{\bar{F}}$ is $\bar{F}$-memory information of (x) separately ,if $(x)_i^{\bar{F}} \Rightarrow (x)_i^{\bar{F}}$ .Then $\gamma_i^{\bar{F}} \leq \gamma_i^{\bar{F}}$.

Proof: If $(x)_i^{\bar{F}} = \{x_1, x_2, \cdots, x_m\}$, $(x)_j^{\bar{F}} = \{x_1, x_2, \cdots, x_n\}$, $n \leq m$, $i < j$,is $\bar{F}$-memory information of (x) ; $(x)_j^{\bar{F}} \Rightarrow (x)_i^{\bar{F}}$ is got from $(x)^{\bar{F}} \Rightarrow (x)$. From $\gamma^{\bar{F}} = \mathrm{card}((x)^{\bar{F}}) / \mathrm{card}((x))$ we can get : $\gamma_i^{\bar{F}} = \mathrm{card}((x)_i^{\bar{F}}) / \mathrm{card}((x)) \leq \mathrm{card}((x)_i^{\bar{F}}) / \mathrm{card}((x)) = \gamma_i^{\bar{F}}$;or , $\gamma_i^{\bar{F}} \leq \gamma_i^{\bar{F}}$,we can get $\gamma_i^{\bar{F}} \leq \gamma_i^{\bar{F}}$.

Corollary 3. If $(x)_i^{\bar{F}}, (x)_j^{\bar{F}}, (x)_k^{\bar{F}}$ is $\bar{F}$-memory information of (x) , further more $(x)_k^{\bar{F}} \Rightarrow (x)_i^{\bar{F}} \Rightarrow (x)_i^{\bar{F}}$ .Then $\gamma_k^{\bar{F}} \leq \gamma_i^{\bar{F}} \leq \gamma_i^{\bar{F}}$.

Corollary 4. If $(x)_i^{\bar{F}}, (x)_j^{\bar{F}}, (x)_k^{\bar{F}}$ is $\bar{F}$-memory information of (x) , further more $(x)_k^{\bar{F}} \Rightarrow (x)_i^{\bar{F}} \Rightarrow (x)_i^{\bar{F}}$ .Then $\eta_i^{\bar{F}} \leq \eta_i^{\bar{F}} \leq \eta_k^{\bar{F}}$.
Here: $\eta_i^F = \mathrm{card}(\alpha_i^F) / \mathrm{card}(\alpha)$, $\eta_i^{\bar{F}}$ is called about $\bar{F}$-memory identification coefficient of (x) of $(x)^{\bar{F}}$ memory; α_i^F is character set of $(x)_i^{\bar{F}}$, α is character set of (x) ; $\eta_i^{\bar{F}} \geq 1$.

Theorem 4. ($\bar{F}$-memory circle theorem) if $(x)_k^{\bar{F}}$ is $\bar{F}$-memory information which is generated from (x), then $\bar{F}$-memory circle $\mathcal{O}_k^{\bar{F}}$ is generated from $(x)_k^{\bar{F}}$ is an internal –the same centre circle of $\mathcal{O}$ which is generated from (x) ,further $\mathcal{O}_k^{\bar{F}} \subset \mathcal{O}$.

Here: symbol " $\subset$ " represents that $\mathcal{O}_k^{\bar{F}}$ is surrounded in $\mathcal{O}$.

Theorem 4 tell us an important fact: research of $\overline{F}$ -memory information property cam be abstracted into simple geometric drawing and is discussed, which make $\overline{F}$ -memory information research visualizing and graphical .

Theorem 5. ($\overline{F}$ -memory information character theorem) if $(x)_\lambda^{\overline{F}}$ is a $\overline{F}$ -memory information is generated from (x) , character set α of (x) and character set α_λ^F of $(x)_\lambda^{\overline{F}}$ mee $\alpha \bigcup \{\alpha' | \beta \in V, \beta \overline{\in} \alpha, f(\beta) = \alpha' \in \alpha, f \in F\} = \alpha_\lambda^F$.

Theorem 6. ($\overline{F}$ -memory information lost theorem) if $(x)_\lambda^{\overline{F}}$ is $\overline{F}$ -memory information of (x) ,then exist information $\nabla(x)_\lambda \neq \phi$,furthermore meet $((x)_\lambda^{\overline{F}} \bigcup \nabla(x)_\lambda) - (x) = \phi$.

Theorem 7. ($\overline{F}$ -memory information character dependency theorem) if $(x)_\lambda^{\overline{F}}$ is $\overline{F}$ -memory information of (x) , α_λ^F , α is character set of $(x)_\lambda^{\overline{F}}$, (x) separately, then α_λ^F , α meet $\alpha \Rightarrow \alpha_\lambda^F$.

In fact , if $(x)_\lambda^{\overline{F}}$ is $\overline{F}$ -memory information of (x) ,or $(x)_\lambda^{\overline{F}} \subseteq (x)$;we can get from P-set structure :character set α_λ^F of $(x)_\lambda^{\overline{F}}$ and character set α of (x) meet $\alpha \subseteq \alpha_\lambda^F$,or $\alpha \Rightarrow \alpha_\lambda^F$,then $\alpha \Rightarrow \alpha_\lambda^F$ can be got .

$\overline{F}$ -memory Information Identification Standard

Given information (x) , α is character set of (x) ,exist information $(x)^*$ and α^* is character set of $(x)^*$,if character of α^* is deleted, which make $\alpha^* \Leftrightarrow \alpha$ .Then $(x) \Leftrightarrow (x)^*$. $(x)^*$ is $\overline{F}$ -memory information of (x) , $(x)^* = (x)^{\overline{F}}$.

4 $\overline{F}$ -Memory Information Application in Enterprise Survival Testament

The data of this section of examples from the profit of a group company data during 2008 to 2009. In the second half of 2009, global-economic crisis outbreak. Assuming $(x) = \{x_1, x_2, x_3, x_4, x_5\}$ is a group of five subsidiaries set. $Y = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$, $Y_1 \sim Y_5$ is annual profit subsidiary. And $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9\}$ is the attribute Set of (x) .For example, the Properties of $\alpha_1 \sim \alpha_9$ means the Characteristics of Market. As commercial secret reason, the Specific names of properties $\alpha_1 \sim \alpha_9$ is omitted. The example of Y data is a technical processed data. It does not affect the results of analysis.

In 2008, the annual profits of five subsidiaries $(x) = \{x_1, x_2, x_3, x_4, x_5\}$ in group company $Y_1 \sim Y_5$ listed in table 1.

Table 1. 2008 annual profit values $Y_i, i = 1,2,3,4,5$

	Y_1	Y_2	Y_3	Y_4	Y_5
x_i	1.83	2.02	1.98	3.05	2.61

In 2009, the annual profits of five subsidiaries $(x) = \{x_1, x_2, x_3, x_4, x_5\}$ in group company $Y_1 \sim Y_5$ listed in table 2.

Table 2. 2009 annual profit values after the economic crisis $Y_i, i = 1,2,3,4,5$

	Y_1	Y_2	Y_3	Y_4	Y_5
x_i	1.83	-	-	-	2.61

In table 2, "-"means" Empty data"(zero data), x_2, x_3, x_4 because the economic crisis is closed.

Instance Data Analysis
In 2008, before the economic crisis, the subsidiary company is in the profits state. In 2009, because the world outbreak the economic crisis, the investment environment become quite bad, because the invade of the risk property $\alpha_{10}{}', \alpha_{11}', \alpha_{12}', \alpha_{13}', \alpha_{14}', \alpha_{15}', \alpha_{16}'$, for example, $\alpha_k', k = 10, \cdots, 16$ means transport cost influence the profits factors. In the (x) of the attribute set, α become $\alpha^F = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5, \alpha_6, \alpha_7, \alpha_8, \alpha_9, \alpha_{10}{}', \alpha_{11}', \alpha_{12}', \alpha_{13}', \alpha_{14}', \alpha_{15}', \alpha_{16}'\}$, (x) become $(x)^{\overline{F}} = \{x_1, x_5\}$, because exsit $\overline{F}$ - memory information.

Through the table 1and table 2:the information (x) corresponding the information character set $Y = \{Y_1, Y_2, Y_3, Y_4, Y_5\}$ =$\{1.83, 2.02, 1.98, 3.05, 2.61\}$ become $\overline{F}$ -memory information .The $(x)^{\overline{F}} = \{x_1, x_5\}$ (x) corresponding the information character set $Y = \{Y_1, Y_5\} = \{1.83, 2.61\}$. Through the definition 6,7 means $\|Y\| = 5.24$, $\|Y^{\overline{F}}\| = 3.19$. Through the definition 3 means $\overline{F}$ is memory loss $\nabla(x)^{\overline{F}} = \{2.02, 1.98, 3.05\}$ .The information (x) corresponding the information character constitute the unit circles radius $\rho = \|Y\| / \|Y\| = 5.36/5.36 = 1$,the memory information $\overline{F}$ constitute the Unit circles radius $\rho^{\overline{F}} = \|Y^{\overline{F}}\| / \|Y\| = 3.19 / 5.24 < 1$ .Due to external risk of the radius attribute, x_2, x_3, x_4 company is closed. The company use the Theorem 4, the $\overline{F}$ - memory information may identify rules, adjust the investment strategy and avoid the risk factors. It become $(x)^{\overline{F}}$ and (x) meet $\text{UNI}((x), (x)^{\overline{F}})$.

5   Discussion

The dynamic features introduced into the common set of limited X and Improve local collection X. The table one and two mentioned a new mathematic model. The P collection $(X^{\overline{F}}, X^F)$ own the dynamic characteristics. The table three to fourteen is given P-set application in many fields. P-set memory feature is one of the characteristics important applications. This paper, introduced the memory characteristics of P-set into the information system. We ask the concept of memory information and given - the multiple features and applications of the memory

information. P-set is a new mathematical methods and tools to study the dynamic information system (dynamic database system). P-set of this model has good prospects, especially in the dynamic information system.

References

[1]   Zadeh, L.A.: Fuzzy sets. Information and Control (8), 338–353 (1965)
[2]   Pawlak, Z.: Rough sets. International Journal of Computer and Information Sciences 32(11), 341–356 (1982)
[3]   Shi, K.-Q.: P-sets. Journal of Shandong University (Natural Science) 43(11), 77–78 (2008)
[4]   Shi, K.-Q.: P-set and its application character. Computer Science 37(8), 1–8 (2010) (in Chinese)
[5]   Wang, Y., Zhang, G.-Y., Shi, K.-Q.: P-set and F - memory Information Characteristic application. Computer Science 37(8), 246–249 (2010) (in Chinese)
[6]   Geng, H.-I., Zhang, G.-Y., Shi, K.-Q.: F - information Camoufladge and Camoufladge-Reduction Identification. Computer Science 38(2), 240–245 (2011) (in Chinese)
[7]   Wang, Y., Zhang, G.-Y., Zhang, L.: P-sets and dynamic separation characteristic of F - information. Journal of Shandong University (Natural Science) 46(3), 35–40 (2009) (in Chinese)
[8]   Shi, K.-Q., Zhang, L.: Internal P-sets and data outer-recovery. Journal of Shandong University (Natural Science) 44(4), 8–14 (2009) (in Chinese)
[9]   Tang, J.-H., Chen, B.-H., Shi, K.-Q.: P-sets and (F, F)-data generation-identification. Journal of Shandong University (Natural Science) 44(9), 19–25 (2009) (in Chinese)
[10] Wang, Y., Geng, H.-I., Shi, K.-Q.: The mining of dynamic informa- tion based on P-sets and its applications. An International Journal Advances in Systems Science and Applications 10(2), 234–240 (2010)
[11] Zhang, G.-Y., Li, E.-Z.: Information gene and identification of its information knock-out/knok-in. An International Journal Advances in Systems Science and Applications 10(2), 308–315 (2010)
[12] Shi, K.-Q., Zhao, J.: Function S-rough sets and Security authentication of hiding law. Science in China F:Information Science 51(7), 924–935 (2008)
[13] Zhang, F., Chen, P., Zhang, L.: P-set separation and application. Shan Dong University Journal 45(3), 18–22 (2010)
[14] Tang, J.-H., Chen, B.-H., Shi, K.-Q.: P-set and (F, F)-date production –identification. Shan Dong University 44(11), 83–92 (2009)
[15] Zhang, L., Xu, M., Shi, K.-Q.: P-sets and applications of internal outer data circle. Quantitative Logic and Soft Computing 2, 581–591 (2010)

Anomaly Detection System of Wireless Communication Network Based on Data Mining

Chen Ningjun and Gao Zhinian

[1] Dept. of Operation Research Center, Dept. of Postgraduate Management, ACC, Nanjing 210045, China
[2] Dept. of Operation Research Center, ACC, Nanjing 210045, China

Abstract. Aiming at problems of wireless communication network that may seriously threaten its security, such as secret leakage or intrusion, which is caused by lost or falsification of portable remote terminals, the concept of anomaly detection of remote terminal behavior is put forward; Then an anomaly detection method of remote terminal by clustering its normal behavior is put forward; While an anomaly detection system of wireless communication network by means of combining spectral clustering and pattern matching is designed, an improved SC algorithm is brought forward; Finally, the article testify the method is feasible and effective by doing experiment on the real log-data of the wireless communication network.

Keywords: Wireless communication network, Anomaly detection, Spectral clustering, SSC.

1   Introduction

Fixed and mobile remote terminals often used for sending and receiving communication signals, which is an important part in wireless communication network. In order to maintain the stability and security, remote terminal should retain normal work. Remote terminal user behavior is the direct embodiment of the wireless communication network anomaly detection. As the miniaturization of remote terminals, the management problems become more and more complex. Unauthorized users can even use few terminals to attack wireless communication network. In non-secure environment, once the terminal was stolen or counterfeited by unauthorized access, false call request may be distributed, even the limited band resources may be engrossed. And fake users can steal secrets, sent false information to convey false orders, which may result in incalculable losses. At present, the wireless network management has not made an effective response to this anomaly detection mechanisms and solutions. This abnormal behavior of attacks using traditional methods is difficult to detect the test. Take advantage of the current maturity of pattern recognition technology, the user through the remote terminal timely detect abnormal behavior, and to take measures to prevent abnormal behavior affect the normal operation of systems or steal confidential information is to ensure that the wireless communication network is safe, reliable and stable operation of the important technical support. Therefore, common research and development on

wireless communication network remote terminal behavior anomaly detection system has important theoretical and practical value.

Anomaly detection aims at generalizing normal behavior, summing up the gap with the normal conduct over the threshold of abnormality [1]. Anomaly detection is to detect the characteristics of a new abnormal behavior; for short of that the false positive rate is relatively high, so the threshold set for anomaly detection is very important. At present, anomaly detection technology is divided into four categories: (1)The Statistical Based anomaly detection. (2)The Prediction models generating Based anomaly detection. (3)The data mining based anomaly detection. (4) The immune Based intrusion detection.

2   Design of Wireless Communication Network Anomaly Detection System

Taking into account each remote terminal of the long run have a certain regularity, such as the time to call it a day, with different communication terminal length of time, the transmission business, and so on, have to meet a certain model, appropriate for the use of fuzzy clustering methods to detect and more difficult to obtain because of unusual samples, the system uses the remote terminal's normal method of clustering the establishment of a normal class acts, by calculating the current deviation from the normal conduct of the unusual degree of judgment at the same time, some of remote terminal Attribute greater certainty, such as an remote terminal where the latitude and longitude, and its scope of the usual communication remote terminal groups, which are relatively stable, more suitable for use pattern-matching approach to detection. The above two points, learn from other areas of the current anomaly detection software experience, we used cluster analysis and pattern matching the combined approach to anomaly detection. The relatively fixed in nature, we use pattern-matching method, and to meet certain long-term behavior of the attributes we use cluster analysis methods. The overall structure is as show in Figure 1.
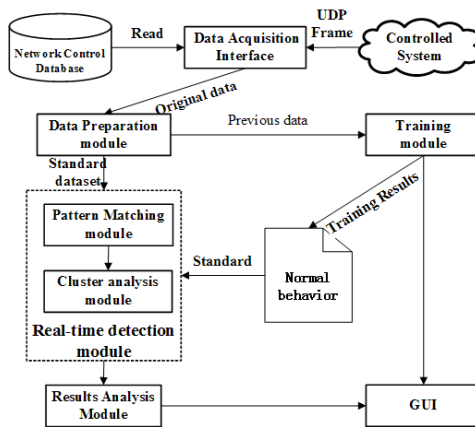
Fig. 1. The total architecture of the system

The course of their work are as follows: data acquisition interface were read by the database and receive UDP data from the database at the way the system and the raw data collected and stored; data pre-processing module is responsible for the original data abstract summarized, in the screening and a major Attributes, such as processing, the standard form, can be learning modules and training modules dealing with real-time detection of data; learning stage, training learning modules from the data pre-processing module access to standard data, and then through the Fuzzy clustering algorithm to deal with these data, each of the Remote terminal and the outline of the normal conduct of its accession to the normal conduct knowledge base, extracted from each remote terminal by joining the model to meet the model. Testing stage, real-time detection module from the data pre-processing module direct access to real-time data, and then first into the pattern matching module, matching failure is unusual warning sent by the cluster model into the match after the detection module statistics, the major input of new terms Detection of data points with the recent cluster centre distance, if the distance is greater than pre-defined threshold, the detection module that detected the anomaly, and sent unusual warning. The anomaly detection system is the threshold of sensitivity to the important parameters, but also to measure an important basis for abnormal levels. Analysis of the results based on test results in the module carrying the message determine the time and place extremely abnormal remote terminal and the basic information, but also in accordance with its counterparts to determine the threshold of serious abnormal levels, such information will be exported through the user interface module, for the next step to quickly and accurately provide the basis for the decision-making.

3   Data Preparation

This is an innovative idea of the article. There are many features concern the behavior of remote terminals, and their importance is different. Some are necessary, the others are not. Use compactness and dispersing degree as the standard of clustering quality. Let n_i denote the member numbers of a cluster c_i, the purity of c_i is represented by

$$P(c_i) = \frac{1}{n_i} \max(n_i'),$$

Let n_i' denote the size of intersection of c_i and set j, so the clustering purity Pur is defined as [7]:

$$Pur = \sum_{i=1}^{k} \frac{n_i}{n} P(c_i),$$

k is the number of clusters. Doing experiment on the wireless communication Network data set under different numbers of features, figure2 shows the clustering purity affected by feature numbers.

Interestingly, clustering purity is not monotonously increasing with the increasing of feature numbers, differently, when feature numbers are too many, it becomes decreasing. That's because the unimportant features affect the accuracy of clustering result.

Therefore, it is better to choose important features and get rid of unimportant ones, which will decrease the clustering quality. Through this way, we can still compress the data, and shorten the processing time. After comparing, we choose 15 important features from the 23 features, and decrease the features of the WCN dataset to 15.
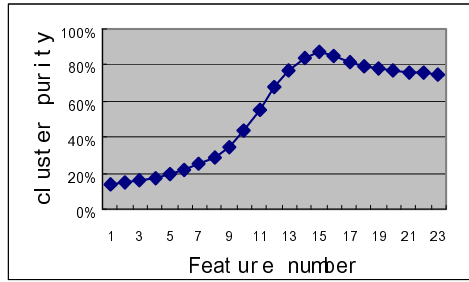
Fig. 2. Cluster purity affected by clustering features

4   Design of Anomaly Detection Algorithm of Wireless Communication Network

Considering the anomaly data is difficult to fetch, we put forward the idea of clustering normal behavior dataset of remote terminals, and define behavior data that deviate from normal behavior cluster over a threshold as anomaly. In this paper, we bring forward an improved spectral clustering algorithm as the anomaly detection algorithm.

One of the SC algorithm's disadvantage is its initialization sensitivity[6]. The improved algorithm Conquers this shotcomings in the way of generating initial cluster center by SOM network. The SOM model of neural network can cluster datasets and find cluster centers automatically, in this way, the disadvantage of choosing initial cluster centers can be fetched up. But the constringing speed for SOM is much too slow(at least 5000 times cycle), as well as its clustering quality is worse than SC algorithm, therefore, we bring forward an new algorithm which combine SC and SOM algorithm and take advantage of both. First, cluster dataset by SOM, and in order to speed up, we do not have to wait until it constringing completely, about 400 times cycle is enough, now the output neurons' weighting of SOM network is near to real cluster centers; Then, calculate clustering result by the k-means algorithm, with output neurons' weighting of SOM network as the initial input cluster centers. This method optimizes the initial input cluster centers of SC algorithm, and the result is more stable and with better quality, compared to choosing initial clustering centers randomly.

The improved algorithm is denoted by SSC, and the steps are as below:

Input: Similarity matrix $S \in \mathbb{R}^{n \times n}$, number k of clusters to construct.
Step1: Construct a similarity graph. Let W be its weighted adjacency matrix.
Step2: Compute the unnormalized Laplacian L.
Step3: Compute the first k generalized eigenvectors $u_1, \ldots, u_k$ of the generalized eigenproblem $L_u = \lambda D_u$.
Step4: Let $U \in \mathbb{R}^{n \times k}$ be the matrix containing the vectors $u_1, \ldots, u_k$ as columns.

Step5: For $i = 1,\ldots, n$, let $y_i \in \mathbb{R}^k$ be the vector corresponding to the i-th row of U.

Step6: Let I denote cycle number, the number of output neurons equal to c, set initial study rate η_0, maximum cycle number of SOM I_{SOM} (only several hundred times);

Step7: Do Input data points $(y_i)_{i=1,\ldots,n}$ to SOM network and train; Until $I \geq I_{som}$;

Step8: Let w_i denote output weighting of SOM network, input w_i as initial cluster centers, and cluster the points $(y_i)_{i=1,\ldots,n}$ in $\mathbb{R}^k$ with the k-means algorithm into clusters $C_1,\ldots, C_k$.

Output: Clusters $A_1,\ldots, A_k$ with $A_i = \{ j \mid y_j \in C_i \}$.

5   Experimental Analysis

In order to verify the effect of SSC, we do comparing experiment on the real datasets with the FCM, SC and SSC algorithm. Set initial study rate to 0.9, final study rate to 0.1, and set the maximum training cycle number of SOM network in the SSC algorithm to 500, fuzzy factor m=2.

Cluster the WCN dataset by the algorithm of SSC, SC and FCM for 10 times and take the average values. Figure 3 shows the effect to cluster purity caused by cluster number.
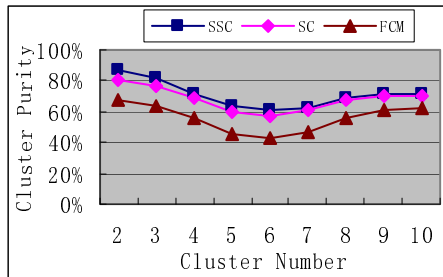
Fig. 3. Cluster purity affected by clustering number
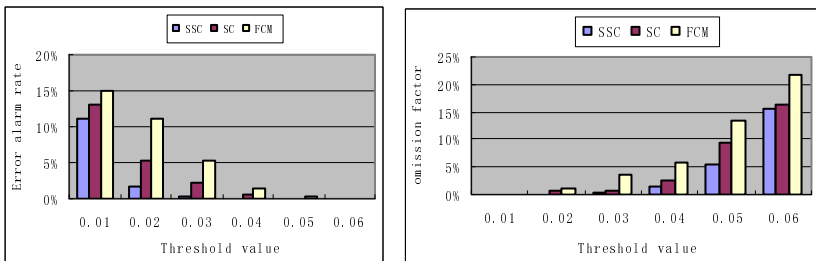
Fig. 4. Error alarm rate and omission rate affected by threshold

Obviously, cluster purity reaches the maximum value when the cluster number is 2, and most time the cluster purity of SSC algorithm is better than others. Although sometimes the result of SC algorithm is better, it's unstable most times.

Carry out simulative experiment on the detection system, after training by each algorithm finished, use testing dataset to test the system, because the anomaly data of the wireless communication network is difficult to fetch, we generate some anomaly data records by doing anomaly operation on the remote terminals, figure 4 shows the error alarm rate and the omission rate caused by threshold.

Experiment result shows that with the reducing of threshold, the error alarm rate of SSC algorithm increases more slowly than the other algorithms, and the omission rate decreases more quickly. When the threshold is equal to 0.02, 0.03, SSC algorithm reaches almost 0 omission rate and a low error alarm rate.

Through the experiments above, we can prove that the system has high detecting rate and low error alarm rate with the SSC algorithm.

6   Summaries

With the more and more widely used of wireless communication network, users will pay more attention to its security. The security of wireless communication network will still be a hotspot in the future research, and anomaly detection of wireless communication network will be an important part of the research.

References

[1]   Winkler, J.R., Page, W.J.: Intrusion and Anomaly Detection In Trusted Systems. In: Proceeding of the Fifth Annual Computer Security Applications Conference, Tucson, AZ (December 1989)

[2]   Anderson, D., Frivold, T., Valdes, A.: Next-Generation Intrusion Detection ExpertSystem (NIDES) A Summary. Technical Report, SRI-CSI-95-07, SRI International, Computer Science Lab (May 1995)

[3]   Teng, H.S., Chen, K., Lu, S.C.: Adaptive real-time anomaly detection using inductively generated sequential patterns. In: Proceedings of the IEEE Symposium on Research in Security and Privacy, Oakland CA, pp. 278–272 (May 1990)

[4]   Lee, W.: A data mining for constructing features and models for intrusion detection system. Ph.D. Dissertation, Columbia University (1999)

[5]   William, P.D., et al.: CDIS: Towards a Computer Immune System for Detecting Network Intrusions (2001)

[6]   von Luxburg, U.: A Tutorial on Spectral Clustering. Statistics and Computing 17(4) (2007)

[7]   Steinbach, M., Karypis, G., Kumar, V.: A comparison of Document Clustering Techniques. Department of Comp. Sci. & Eng. University of Minnesota, pp. 1–20 (2000)

[8]   Han, J., Kamber, M.: Data Mining Concepts and Techniques, 2nd edn., Beijing (2007)

[9]   Li, X.F., Li, J.: Data mining and knowledge discovery. Publisher of High Education (2003)

[10]   Pan, Z.S.: Research on Intrusion Detection Based on Neural Network. Ph.D. Dissertation, NUAA (June 2003)

Study for System Bootloader Based on DSP and FPGA Shared Flash

Zhang Baofeng[1], Zhang Suhao[1], Zhu Junchao[1], and Li Xinzhi[2]

[1] Tianjin University of Technology
[2] Tianjin Design and Research Institute of Electric Drive

Abstract. This paper proposes a method to realize the on-screen display (OSD) that based on TMS320DM643 DSP and FPGA hardware platform system shared by one external FLASH memory. The normal work of DSP system need configure the FPGA in advance. The FPGA configuration is realized by bootloader of DSP. The approach is proved to be accurate and reliable. Compared with conventional methods that using additional EEPROM or FLASH as external memory of the DSP and FPGA, this approach can fully use system resources and can be easily realized. This paper introduces DSP bootloader theory and design scheme for hardware and software in detail.

Keywords: DM643, FPGA, FLASH, Bootloader.

1 Introduction

In order to satisfy the requirements of high-speed of processing and transmission in image applications, many people used to adopt the system which combines use of DSP(Digital Signal Processor) and FPGA (Field Programmable Gate Array) chip[1][2]. Because of the advantages of DSP + FPGA hardware system, the system is getting more attention. With the development of large-scale integrate circuit technology and computer technology, it will become an important subject to simplify peripheral circuit. Conventional methods by using additional EEPROM or FLASH as external memory for the DSP and FPGA increase the size of the hardware and power consumption，which also lead to extra expense.

This paper proposes a method to realize OSD function that base on DSP and FPGA system shared by one external memory FLASH. When the system is reset, the configuration of the FPGA is bootloaded using a secondary bootloader of DSP. The experiment results show that the system is reliable, stable and easy.

2 Introduction for DSP + FPGA System and Function

In this paper the chip of the TMS320DM643 DSP and XC2S300E - 6PQ208C FPGA are adopted. TMS320DM643 (DM643) is a high performance fixed-point DSP device which dedicate to digital media application. With performance of up to 4800 million instructions per second (MIPS) at a clock rate of 600 MHz, the DM643 device offers cost-effective solutions to high-performance DSP programming challenges. The

FPGA is used to realize the function of OSD in the system. Am29LV033C chip is selected for external FLASH memory in this sytem. Figure 1 shows the system block diagram.
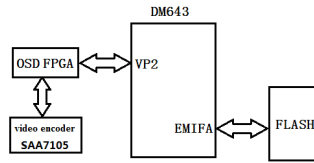
Fig. 1. DSP+FPGA System Block Diagram

The function of FPGA is to provide an interface between the DM643 video port and the SAA7105 video encoder. The interface can assist to realize the function of High-definition video images process, VGA display output and OSD adjustment. This interface works as the following modes:

1)Transparent mode where the 8-bit video output from the DM643 video port is passed directly to the SAA7105 video encoder.

2)Performs 2:1 multiplex operation to support SVGA and HD. The DM643 video port outputs 16-bit video data on the rising edge of the clock to the OSD FPGA. The OSD FPGA converts this to dual clock edge 8-bit video data where the upper 8 bits are clocked out on the rising edge of the clock and the lower 8 bits on the falling edge.

3) One-bit alpha blending with 7-bit color look-up table (CLUT) for 8-bit video output from the DM643. The alpha blended video data is then passed on to the SAA7105 video encoder.

4)One-bit Alpha blending with 7-bit CLUT for 16-bit video output from the DM643 to support HD mode of operation. The alpha blended video data is converted to dual clock edge 8-bit data and passed on to the SAA7105 video encoder.[3]

3   DSP Bootloading Principle

DM643 offers three types of boot configurations: no boot process, ROM boot process, and host boot process. The boot process selected is determined by the configuration of the BOOTMODE pins-AEA[22:21]. Table 1 shows the corresponding relation between the types of boot configurations and the pins-AEA[22:21].

Table 1. DM643 bootloading method

AEA[22:21]	00	01	10	11
bootloading method	no boot process	host boot process	reserves	8-bits ROM boot process

Boot configuration selected for the system is the ROM boot process (also referred to as FLASH boot process). When selected, the FLASH boot process copies a fixed amount of memory located at the beginning of the external ROM to address 0 using

the DMA/EDMA controller automatically. The ROM boot process differs between specific C6000 devices. DM643 belong to 64x series. 64x series EDMA copy 1K bytes from beginning of CE1 to address 0. There is a need for a secondary bootloader because the application size is greater than the 1K bytes of memory copied by the on-chip bootloader. In the applications that require a secondary bootloader, this custom boot code usually resides at the beginning of ROM memory so that it can be automatically transferred by the on-chip bootloader to internal memory, address 0. Once the transfer is complete, the CPU begins executing from address 0 and runs the custom boot code at the same time. Secondary bootloader copies the application program to its runtime memory location. After completion, the secondary bootloader code branches to _c_int00 and Initialize BIOS_init (DSP/BIOS scheduler initialization). Then CPU begins executing from main function (Main ()).[4]

3.1   Design for DM643 and Flash Connection

In the system 4M – 8 bits Am29LV033C is selected for FLASH. Figure 2 shows the connection diagram of DM643 and FLASH.
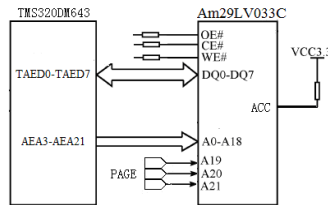
Fig. 2. The connection diagram of DM643 and FLASH

Due to external memory interface (EMIF) having 20 address lines, the largest addressing ability of the CE1 space is 1 Mbyte. To solve the 4 Mbyte addressing ability of FLASH, FLASH paging technique is adopt. The FLASH space is divided to eight pages, and each page is 512k bits, and all are 4M bits. DM643 controls the high three address line of FLASH to choose corresponding eight pages. When A21-A19 is 000, page 0 is selected. When A21-A19 is 001, page 1 is selected, and so on.[5]

3.2   Design for The Connection of FPGA/DSP/SAA7105

The OSD FPGA is connectted to the DM642 EMIF and video port 2. The FPGA also interfaces to the video encoder SAA7105, clock PLL, dual UART, GPIOs, and LEDs. The video port can be configured as either an 8-bit or 16-bit display interface. The Video IF module registers all the data and control signals received from the video port. The Video IF module provides data to the OSD MUX module and the control signals to the OSD Control Logic as well as to the SAA7105 encoder external to the FPGA. The address decoder interfaces to the DM642 EMIF interface. The address decoder module registers all the incoming data and address signals and does first level of address decoding. It divides the CE1 address space into space for the Flash, space for dual UART, and space for asynchronous registers inside the OSD FPGA. It also

divides the CE3 address space into space for external synchronous logic, space for FIFOs inside the FPGA, and space for syn-chronous registers inside the OSD FPGA. Figure 3 shows the connection diagram of video encoder SAA7105/FPGA/DSP.[3]
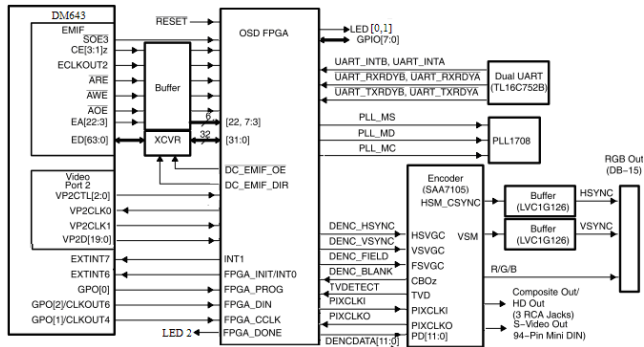
Fig. 3. The connection diagram of DSP/FPGA/SAA7105

4   Software Configuration

The normal work of the system need configures the FPGA in advance. When DSP is reset, EDMA copies 1K bytes from beginning of CE1 to address 0 for the configurations of DSP. Once the transfer is complete, the CPU begins executing from address 0 and therefore runs the custom boot code. The secondary bootloader then copies the FPGA configuration into RAM. The system applys three type of software: Code Composer Studio development environment (CCS) for DSP/BIOS application developed, a hex conversion utility (hex6x.exe), and flash burning utilities (Flash burn.exe).

The steps for Flash Burn are as following:

1) Build the project to generate COFF object (.out) file in Code Composer Studio

    Write application program and Linker command file on Code Composer Studio development environment (CCS) to generate the .out file. The file include FBTC.OUT for Flashburn required and the two file (flash.out and fpga.out) for the hex conversion utility.

    2) Use the hex conversion utility to Convert COFF format (.out) to hex file

    Flash programmers work only with the hex format hence the COFF format(flash.out and fpga.out) obtained from Code Composer Studio must be converted to .hex(flash.hex and fpga.hex) through the hex conversion utility.

    3) Program the flash using the Flash Burn Utility

Once the .hex file is generated, use a flash programming utility to write the hex image to the FLASH. A GUI based Flash Burn Utility is available and is present under the Tools menu in Code Composer Studio.

5   Experimental Verification

To sum up, the success of OSD FPGA configuration can be verified by the following two ways:

1) The connection of the FPGA and three LEDS can be abtained from the connection diagram of FPGA. Three LEDs command can be found in the program of configuration for FPGA. After flash burning, the success of Flash burning can be proved if three lights are turn on, and the FPGA begins to work.Three burning light can be shown in the Fig.4.
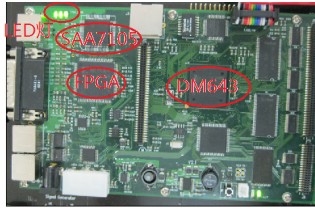
Fig. 4. Actual hardware chart of system

2) First, the image is processed by DSP and FPGA, and then it is passed on the SAA7105 video encoder to be encoded and image output on the screen. The FPGA configuration is successful if the image can be output. The frame image display in CCS and screen can be shown in the Fig.5 and the Fig.6.
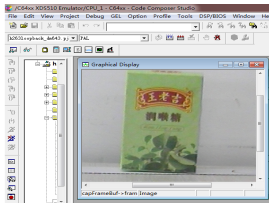
Fig. 5. Frame images on CCS   **Fig. 6.** Display images
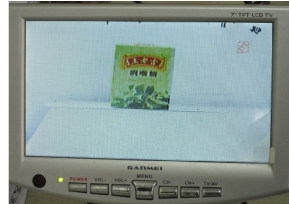
Shown as in Fig5 and the Fig6, video output is success in the system, which proves the FPGA configuration is successful and it is running normally.

6   Summaries

This method to realize OSD of the image based on DSP and FPGA system shared one FLASH can make full use of system's resources, what's more, the function of VGA high-definition video image and OSD adjustment can be easily realized. The method proposed in this paper is simple, convenient, feasible and available.

References

[1] Su, W.-X., Cheng, L.-Y.: Design of Real-Time Video Signal Processing System Based on DSP+FPGA. Chinese Journal of Liquid Crystals and Displays 25(1) (February 2010)

[2] Li, L.: High-Speed Data Communication Between DSP and FPGA in Embedded Panoramic Video Processing System. Journal of Electronics & Information Technology 32(3) (March 2010)

[3] TMS320DM642 EVM OSD FPGA User's Guide. Texas Instruments (July 2003)

[4] Creating a Second-Level Bootloader for FLASH Bootloading on TMS320C6000 Platform With Code Composer Studio,SPRA999A. Texas Instruments (August 2004)

[5] Data sheet for Am29LV033C. AMD (2004)

Research on Time and Frequency Synchronization for 3GPP LTE TDD System over Multipath Fading Channel

Fa-Tang Chen and Lei Ma

Chongqing Key Lab of Mobile Communication, Chongqing University of Posts and
Telecommunications,
Chongqing 400065, China
chenft@cqupt.edu.cn, malei611339@163.com

Abstract. The method of time and frequency synchronization for 3GPP long term evolution (LTE) time division duplexing (TDD) system over multipath fading channels is presented in this paper. The energy differential function of received signals and correlation of primary synchronization signal (PSS) are used to obtain accurate symbol timing synchronization and fractional carrier frequency offset in time domain, respectively. By computer simulation, the performance of our approach is evaluated and compared to other existing algorithms over multipath fading channel.

Keywords: 3GPP LTE TDD system, Time and frequency synchronization, Energy differential function, PSS.

1   Introduction

In order to support long-term competitiveness of its technology, the 3rd Generation Partnership Project (3GPP) is considering LTE for both radio interface and network architecture [1]. LTE TDD is LTE system in time division duplexing mode. Orthogonal frequency division multiplexing (OFDM) based radio access is adopted in LTE TDD system downlink because of its high spectral efficiency, high user throughput and inherent robustness against multipath interference [2]. However, all of these advantages of the OFDM can be guaranteed when the orthogonality between sub-carrier is maintained. If the orthogonality is destroyed, serious deterioration of system performance may occur due to inter-carrier interference (ICI) and inter-symbol interference (ISI) caused by symbol timing and carrier frequency offset in the receiver.

Various existing synchronization algorithm to estimate symbol timing and carrier frequency offset can be classified into data-aided [3-5] and non-data-aided [6-9] estimation methods in LTE TDD system. The former methods mainly use reference signal, such as cell-specific reference signals, and PSS inserted in transmitted signal to estimate time and frequency synchronization, this kind of approach have high estimation accuracy, but low spectral efficiency. The latter methods utilize periodic nature of time-domain signal by using cyclic prefix (CP), this kind of estimation approach have high spectral efficiency, but low estimation accuracy. The algorithm of non-data-aided estimation works well in additive white Gaussian noise (AGWN) channel, but the performance of this algorithm is deteriorated in multipath fading

channel due to the correlation of CP and the last L(the length of CP) samples of the body of the OFDM symbol is destroyed.

In this paper, we present a method of symbol timing and fractional frequency offset (FFO) synchronization in LTE TDD system over multipath fading channel. The energy differential function of received signals and correlation of PSS are used to obtain accurate symbol timing synchronization and fractional carrier frequency offset in time domain, respectively. The process of time and frequency estimation is done in time domain, so the synchronization delay of receiver is reduced, because of eliminating the conversion to the frequency domain by DFT.

The remainder of this paper is organized as follows. Section 2 describes the system model. The algorithm of symbol timing detection and fractional frequency offset estimation are analyzed in Section 3 and 4, respectively. Simulation results are discussed in Section 5. Finally, conclusions are given in Section 6.

2   LTE TDD System Model

In LTE TDD system downlink, the OFDM symbol is generated by taking N-point inverse fast Fourier transform (IFFT), and represented as

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j2\pi kn/N}, \ n = 0, \cdots, N-1 \tag{1}$$

Where $X(k)$ is the modulated data on the kth sub-carrier, k refers to sub-carrier number. After multipath fading channels, the received signal under ideal synchronization case can be expressed as

$$y(n) = \sum_{m=0}^{J-1} a_m x(n - \tau_m) \tag{2}$$

Where J is the number of multipath. The channel fading coefficient and delay sample points of mrd multipath are denoted by a_m and τ_m, respectively.

During propagation, the transmitted signal is corrupted by multipath fading, AGWN, symbol timing and carrier frequency offset. The actual received time domain signal can be written as:

$$r(n) = y(n - \theta) e^{j2\pi \varepsilon n/N} + w(n) \tag{3}$$

Where θ is the integer-valued unknown arrival time of a symbol. ε denotes the difference in the transmitter and receiver oscillators as a normalized fractional frequency offset. $w(n)$ is white Gaussian noise with mean zero and variance $\delta^2 = E\{|w(n)|^2\}$.

3   Symbol Timing Detection

Since the appearance of a cyclic prefix yields a correlation between some pairs of samples that are spaced N samples apart, so we can get accurate symbol timing and carrier frequency synchronization, i.e. [6]. However, this correlation is destroyed under multipath fading channel. Assuming that the length of CP is larger than maximum

multipath delay in LTE TDD system, it still has good correlation in the tail of CP. So using energy differential function of received signal to estimate symbol timing synchronization location could eliminate the affects of multipath fading channel and carrier frequency offset.

Make the received signal $r(n)$ multiply its conjugate signal $r^*(n)$. The energy of received time domain signal can be represented as

$$r_{eng}(n) = r(n) \times r^*(n) = |r(n)|^2, \ 0 \le n \le N + L - 1 \tag{4}$$

Where L is the length of cyclic prefix. Make the received signal $r(n)$ with N sampling point delay $r(n+N)$ multiply its conjugate signal $r^*(n+N)$. The energy of received time domain signal with N sampling point delay can be derived as

$$r_{eng}(n+N) = r(n+N) \times r^*(n+N) = |r(n+N)|^2, 0 \le n \le N + L - 1 \tag{5}$$

The energy differential function of received time domain signal can be written as

$$r_{dif}(n) = r_{eng}(n) - r_{eng}(n+N) = |r(n)|^2 - |r(n+N)|^2, \ 0 \le n \le N + L - 1 \tag{6}$$

Obviously, equation (6) eliminate the influence of carrier frequency offset. Averaging G OFDMs symbols can be expressed as

$$r_{ave}(n) = \frac{1}{G} \sum_{g=0}^{G-1} || r[n + g(N+L)]|^2 - | r[n + g(N+L) + N]|^2 | \tag{7}$$

Where G is the number of OFDM symbol to average.

Construct the timing measurement function:

$$r_{div}(n) = r_{ave}(n) / r_{ave}(n-1) \tag{8}$$

Because the value of $r_{ave}(n)$, in the absence of ISI region transition to the data area, jumps from a small positive number (ideally equal to 0) to a very large number, so the value of $r_{div}(n)$ tends to infinity at the end of no ISI region in ideal case. By detecting the peak, we can get accurate symbol timing synchronization location.

4   Fractional Frequency Offset Estimation

Carrier frequency offset (CFO) is mainly caused by the Doppler shift, as well as frequency differences between the local oscillators in the transmitter and the receiver. Usually, it can be classified into fractional frequency offset (FFO) and integer frequency offset (IFO). In LTE TDD system, the FFO would destroy the orthogonality among subcarriers, resulting in ICI and causing system's error rate increase. The PSS uses sequences known as Zadoff-Chu. This category of sequences is widely used in LTE TDD system, including for random access preambles and the uplink reference symbols in addition to the PSS, due to its ideal cyclic autocorrelation in time domain. According to symbol timing synchronization location, the received PSS signal with FFO can be written as

$$r(n) = a_n \exp(j2\pi\Delta f n T) + w(n), \ 0 \le n \le N-1 \tag{9}$$

Where a_n and Δf are PSS complex time domain sequence of transmitting end and carrier frequency offset. In LTE TDD system, $T = T_s = 1/(15000*2048)s$ is the sampling period. The cross-correlation of the received PSS signal with FFO and the PSS complex time-domain sequence of transmitting end can be expressed as

$$y(n) = a_n^* a_n \exp(j2\pi\Delta f n T) + w'(n), \ 0 \le n \le N-1 \tag{10}$$

The maximum likelihood (ML) formula of Δf is given by

$$\Delta \widehat{f} = \arg\max_{\Delta f} \left\{ | \sum_{n=1}^{N} y(n) e^{-j2\pi\Delta f n T} |^2 \right\} \tag{11}$$

By detecting the peak of maximum likelihood formula (11) could obtain accurate FFO. Where $\Delta \widehat{f}$ is the ML estimation value of Δf. The normalized FFO is obtained through formula $\widehat{\varepsilon}_f = \Delta \widehat{f} / 15000$. According to maximum likelihood formula (11), it is hard to get simple closed-loop solution of formula (11), so it is necessary to use numerical methods to get the accurate ML estimation value $\Delta \widehat{f}$. Assumption that the range of frequency offset of $\Delta \widehat{f}$ is $\Delta \widehat{f} \in [-f_{max}, f_{max}]$, taking f_T as a length of stride to sample, i.e. $\Delta \widehat{f} = -f_{max}, -f_{max}+f_T, -f_{max}+2f_T, \cdots, f_{max}-f_T, f_{max}$. f_{max} is the maximum frequency offset. It is easy to get the corresponding likelihood function value by taking the sampling points value into the formula (11), the ML estimation value $\Delta \widehat{f}$ of frequency offset is sampling points corresponding to ML function value by comparing all the likelihood function values.

5   Simulation Results and Discussion

The major simulation parameters are based on 3GPP LTE standards as represented in Table 1 [10-11]. For the multipath fading channel environments, the extended typical urban (ETU) model with 70Hz and 300Hz Doppler frequency, extended pedestrian A (EPA) model with 5Hz Doppler frequency, extended vehicular A (EVA) model with 70Hz Doppler frequency which is recommended by 3GPP are considered [11].

Fig. 1 shows the detection performance comparison between proposed algorithm, also called energy differential algorithm, in this paper and classical ML algorithm [6] to get accurate symbol timing synchronization location under $G=70$, SNR=5dB. The symbol timing synchronization of proposed algorithm exactly locates in 142 sampling points which is completely consistent with the assumed value under ETU multipath fading channel with 70Hz Doppler frequency, while the classical ML algorithm locates in wrong place, approximately in 1000 sampling points, due to affected by multipath fading channel.

Table 1. Simulation parameters

Parameter	Value
Carrier Frequency(Hz)	2.6GHz
Bandwidth(Hz)	1.25MHz
CP Type	Normal CP
Number of OFDM Symbol	70
FFT Size	128
PSS Sequence	Root index u=25
Normalized FFO	0.1
Symbol Timing Offset	142 Sampling Points
Channel Model	EPA 5Hz/ EVA 70Hz/ ETU 70Hz/ ETU 300Hz
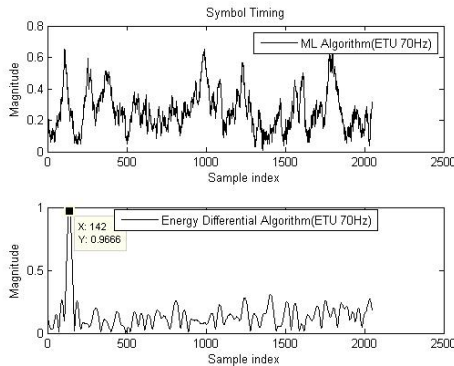
Fig. 1. Symbol timing synchronization of ML and proposed algorithm
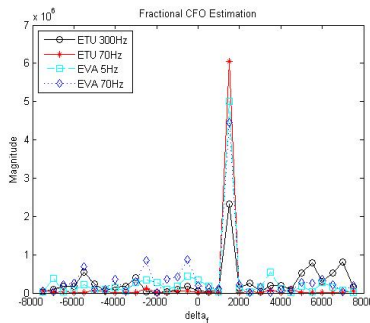
Fig. 2. FFO estimation performance of proposed under multipath fading channel

Fig. 2 depicts the proposed FFO algorithm to estimate FFO under four different multipath fading channels. Supposed that the normalized FFO, SNR, the range of frequency offset and the length of stride are 0.1, 5dB, $\Delta f \in [-7.5KHz, 7.5KHz]$ and $f_T = 500Hz$, respectively. Because the normalized FFO is 0.1, the actual frequency offset is 1500Hz. According to numerical methods, we can get frequency

offset ML estimation value $\Delta \hat{f} = 1500Hz$ by taking every possible separate value into formula (11). The FFO ML estimation value is completely consistent with the assumed value, so the proposed algorithm has good robustness against multipath fading channel.

Fig. 3 represents system bit error rate (BER) performance between the proposed algorithm in this paper and classical ML algorithm under ETU multipath fading channel with 70Hz Doppler frequency. No matter how SNR increased, the BER of classical ML algorithm still above 10^{-2}, due to the correlation between CP and the body of OFDM symbol is destroyed under multipath fading channel. But, as the SNR increasing, the BER of proposed algorithm decline quickly. It is obvious that the BER of proposed algorithm below 10^{-3} around SNR=14dB.
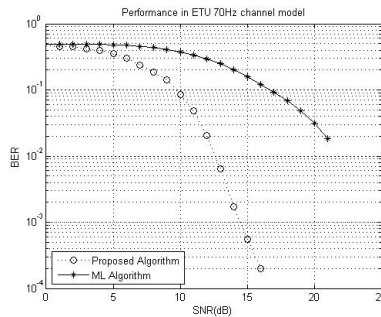
Fig. 3. BER between ML and proposed algorithm under multipath fading channel

6   Summaries

Based on literature [5] and [9], a new method about symbol timing and fractional frequency synchronization in time domain under multipath fading channel is presented in this paper. Simulation and theoretical analysis show that the proposed algorithm has good robustness against multipath fading channel interference, however, there exists IFO in actual LTE TDD system, so it need estimate IFO before using the proposed algorithm in this paper.

References

[1] Wang, Y.-M., et al.: TD-LTE Principles and System Design. Post Telecom Press, Beijing (2010)
[2] Stefania, S., Issam, T., Matthew, B.: LTE—The UMTS Long Term Evolution: from theory to practice. A John Wiley and Sons, Ltd. (2009)
[3] Coulson, A.J.: Maximum likelihood synchronization for OFDM using a pilot symbol: analysis. IEEE Journal on Selected Areas in Common 19(12), 2495–2503 (2001)
[4] Wang, Q., Mehlfuhrer, C.: Carrier Frequency Synchronization in the Downlink of 3GPP LTE. In: Proc. 2010 IEEE 21st International Symposium on Personal Indoor and Mobile Radio Communications (2010)

[5] Chen, F.-T., Zhen, J.-H.: A Method for Automatic Frequency Control, 200310110835.4 (May 4, 2005)

[6] van de Beek, J.-J., Sandell, M.: ML Estimation of Time and Frequency Offset in OFDM Systems 45(7), 1800–1805 (1997)

[7] Lee, J., Lou, H., Toumpakaris, D.: Maximum likelihood estimation of time and frequency offset for OFDM systems. Electronics Letters 40(22) (October 2004)

[8] Takahashi, K., Saba, T.: A Novel Symbol Synchronization Algorithm with Reduced Influence of ISI for OFDM System. In: IEEE GLOBECOM, San Antonio, Texas, USA, pp. 524–528 (2001)

[9] Hu, Y., Luo, Z.-N.: A Novel Timing Synchronization Algorithm Based on Energy Difference over Multi-path Fading Channel. Computer Simulation 26(12), 294–297 (2009)

[10] 3GPP. 3GPP TS 36.211 V9.0.0, Physical Channels and Modulation (Release 9) (May 2009)

[11] 3GPP. 3GPP TS 36.101 V9.0.0, User Equipment (UE) radio transmission and reception (Release 9) (June 2009)

High-Speed Optical Fiber Data Exchange Card Based on PCI-Express

Kun Liu[1], Wei-Chao Zhou[1], and Kun Liu[2]

[1] Institute of Optics and Electronics Chinese Academic of Sciences, Chengdu, Sichuan, China
[2] Graduate University of Chinese Academy of Sciences
Beijing, China
liukustc@gmail.com

Abstract. In this article, we introduce a high-speed optical fiber data transmission card based on single-chip FPGA, using a custom protocol to support optical fiber transmission, using the DDR2 for data cache, using program-controlled manner by the host through the PCI-Express interface to extract Data from memory, last sending out the data through the PCI-Express interface. Test results show that the design of high transmission card is high speed, high bandwidth, and scalability advantages.

Keywords: FPGA, high-speed, optical data communications custom protocol, PCI-Express.

1   Introduction

In meteorology, aerospace, data communications, the field signal has an important role; these signals have the following features: real-time, high data rate, and data volume, complex and so on. Therefore, high-speed data acquisition and transmission in the field of engineering practice is a very important branch in this paper, we implement a large-capacity and high-speed serial data transfer card for collecting real-time image signal. As digital optical fiber communications are more incomparable than traditional transmission media in immunity, transmission bandwidth, reliability and other indicators, so it has a significant and broad prospect. Compared with PCI bus, PCI-Express high-speed serial system becomes the preferred way because of its structure, channel link, scalability, up to 2.5Gbps (PCI-Express 1.0 specification) of the bus frequency. This card includes optical fiber transmission system design technology, PCI-Express and DDR2 cache transfer technology, fusing these several technologies in the FPGA, completing high-speed incoming data from optical fiber transmission to the board, then pushing the data into the memory cache, last the host reads and writes data from the memory through the PCI-Express bus.

2   Overall Design of the Data Exchange Card

The card is based on Xilinx's Virtex-5 family of FPGA-XC5VL110T. Firstly, fiber optic data is pushed into the fiber optic module, transformed the optical information

into electrical information, entered into the high-speed FPGA transceiver Rocket IO, and then extracted by a custom data transfer protocol, stored in the card memory. Then the computer sends read and write requests through PCI-Express, and accesses data from the memory card. To design the transfer card process has the following difficulties: storing the collected data in memory through custom protocol, sending commands to read the data from memory to FIFO through the PCI-Express IP, changing the FIFO data into the certain data packet, sending it to Xilinx's PCI-Express core, which transmits data to the host. The entire block diagram of data transmission card is as follows:
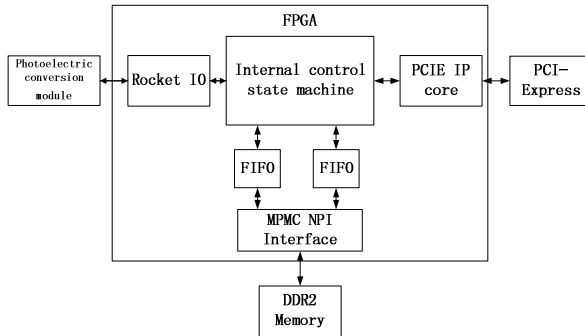
Fig. 1. Block diagram of system

3   The Implement of High-Speed Data Transfer Card

3.1   Design of Optical Fiber Data Transmission Protocol

Optical fiber module and the chip (FPGA) are used in high-speed serial interconnect technology to achieve data transmission. Based on the needs of high-speed transmission, the Virtex-5 Xilinx chip integrates the Rocket IO modules to receive and send high-speed data. The use of CML, CDR, line coding (8b/10b) and pre-emphasis of the Rocket IO hard core can greatly reduce clock distortion, signal attenuation and line noise on the receiver performance, and then the transmission rate is greatly improved.

The transmission rate of Rocket IO in this design is 2.5 Gb/s; the reference clock is 125MHz; the accuracy is 50ppm; the clock input is the differential input; BREFCLK is a reference clock input of Rocket IO, using static control mode. We choose 50 Ohm impedance at terminal; using 8B/10B encoding static control mechanisms; selecting 10% pre-emphasis; selecting 500mV as the output voltage swing. Pre-emphasis and the output voltage swing are not the higher the better, because the higher pre-emphasis affects output voltage swing on high overshoot, which leads to increase transmission link error rate. Therefore, we need to go through experimental tests to adjust the value of the output voltage swing, making the transmission link requirements to achieve stability.

The entire data packet transmission format is as follows:

Packet head (two bytes)	Packet length (two bytes)	valid data	Packet serial number (two bytes)	Packet end (two bytes)

Fig. 2. Block diagram of data packet format

By using the state machine, the first header in the initial state receives the start signal, and then jumps to read the data packet length, after that reads the data length, and stores the valid data in memory based on data length and packet number, last receives end packet signal to reset all signals, jumping back to the initial state, waiting for the arrival of the next packet.

3.2 PCI-Express Protocol Logic Design

PCI-Express protocol hierarchy from low to high has total of three layers: physical layer, data link layer, transaction layer. Transaction layer is the upper of PCI-Express architecture, and its main function is to accept, buffer, distribute TLP. TLP are using memory, IO, configured information and message services to transmit information. Transaction Layer Data Link layer is the intermediate level between physical layer and transaction layer to provide data exchange (TLP), error detection and recovery, initialization services and DLLP the generation and use services. Physical layer receives the TLP, DLLP types of data from data link layer, and transmit data through the link.

Xilinx's Virtex-5 family of FPGA-XC5VL110T integrate a hard core of Endpoint Block Plus for PCI-Express, which meets the PCI-Express Base Specification 1.1 standard and can support x1, x4, x8 link. Endpoint Block Plus for PCI-Express uses Rocket IO to implement the physical layer, and makes data link layer and part of transaction layer to be an efficient framework PCI-Express port. Its interface and the schematic diagram are as follows:
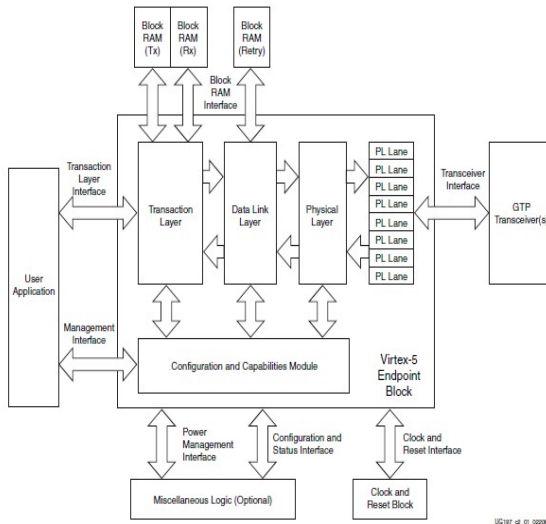
Fig. 3. Block diagram of PCIE IP CORE

We configure PCI-Express IP core and then follow the agreed timing to push data into the IP core; Endpoint Block Plus will automatically complete the other agreement, sending the data to the host. By the needs of the design, we configure the PCI-Express IP core for the x4 link, using frequency 125MHz, configuring Bar0 as 4M of 32-bit address space, selecting Max Payload Size as 512 bytes, and choose the rest as the default settings. We configure memory controller through the MPMC IP core of the EDK, selecting the working frequency of 100MHz. We introduce 125MHz clock from the outside, and use the DCM in the EDK to output 100MHz for the use of the internal logic. After completion of the configuration of EDK, We can call it directly through the ISE.

3.3   Design Internal Control State Machine for Data Exchange

3.3.1   The Overall Control Structure

We car already see the data transmission path in the photograph. First, it stores the data from the optical fiber to the memory. PCI-Express host sends a read instruction through PCI-Express, and then the receiving state machine receives corresponding read command from the IP core and parses out the address, read byte length, byte enable and other related information, and modifies the NPI control signal. At this point MPMC control state machine finds that control signal changes, so starts the state machine and reads the corresponding data from memory to store it in FIFO. FIFO state machine is responsible for FIFO read and write, and modifies the appropriate control word of PCIE. PCIE state machine detects the control signal changed and extracts data from the FIFO, then send is to the PCIE core by the appropriate timing. Last PCIE core send the date to the link. Similarly, the corresponding host writes data into memory and then sends through the fiber process, which is contrary to the above; we no longer go into detail.
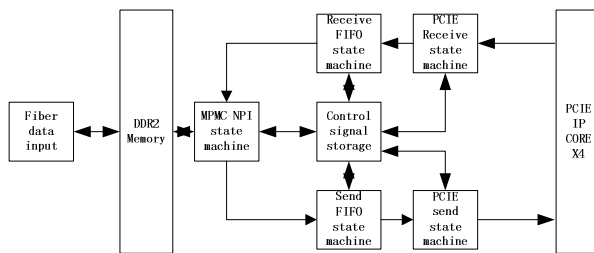
Fig. 4. Diagram of whole state machine

3.3.2   Receive State Machine Design Logic

RX state machine in accordance with the design requirements is as shown
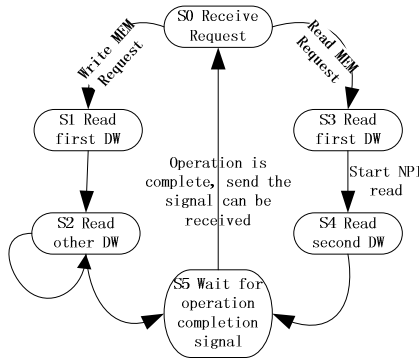
Fig. 5. Block diagram of Rx state machine

S0 state is the reset state, when receiving the data request of the PCIE core to jump to the next state. When reading the memory request S0 Jump to S1, and read the first double word to extract the write bytes length, write byte enables and other information. And then jump to the S2, reading write address and corresponding data until the end of the PCIE core instructions, and then jump to the S5, waiting for write operation completion signal valid, and finally transferred to S0 to reset. When the S0 accept read memory requests, Jump to S3, reading the first double word to extract the length of bytes read, read byte enables and other information. S3 Jump to S4 to read the memory starting address. Because of 32-bit addressing, so a total of read request signal is 3DW long, S4 Jump to S5, wait the effective time of the end of the signal, and then jump S0 reset.

3.3.3 MPMC NPI Control State Machine

S0 state is the reset signal, which starts to jump when they receive control signal effective. When it receives 32DW write request, then jump to S1 state and start the
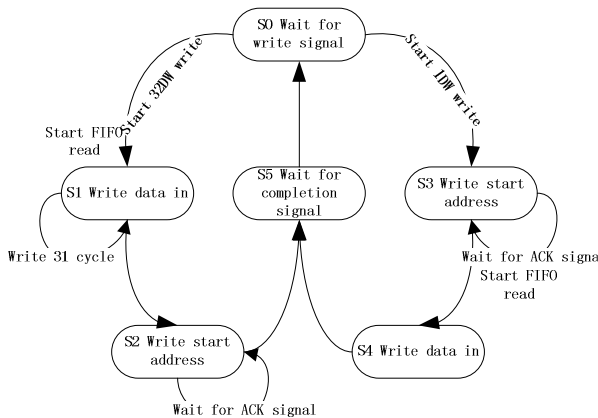
Fig. 6. Diagram of NPI write control state machine

FIFO read operation at the same time. In the S1 state, we push the data from the FIFO into the NIP interface, and jump straight to S2 state after write 31DW. In the S2 state, we write start address and push last a double word. At the same time, we monitor ACK signals, when ACK is high, indicating that the address is entered correctly, and then we reset all the signal and jump to state S5. In the S5 state, we wait for the operation of complete signal effectively, and finally jump to the reset state S0. When we receive 1DW write request, jump to state S3 and write the starting address, and then monitor the ACK signal, when ACK is high, jump to the S4, and at the same time start the FIFO read operation. In the S4 state, push the data from the FIFO into the NIP interface, jump to S5 state after write the 1DW, and then reset all signals. In the S5 state, we wait for the operation of complete signal effectively, and finally jump to the reset state S0.
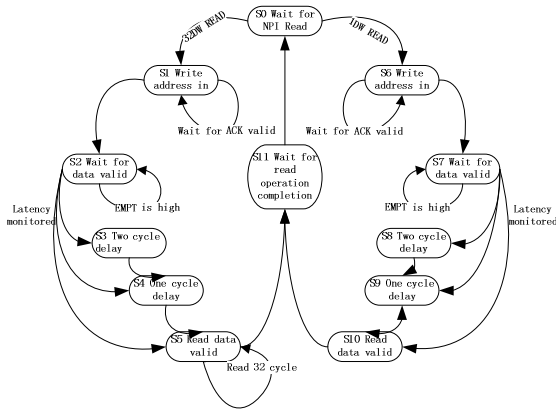
Fig. 7. Diagram of NPI read control state machine

S0 state is the reset signal, when we receive the request signal effectively to jump. When we receive 32DW read signal and jump to the S1 state. In the S1 state, write the starting address, and ACK signal is high to show that the address is written correctly. When the ACK is high, we jump to the S2 state. In the S2 state, we monitor the EMPTY signal, when the EMPTY signal is low, showing the data has been read out to the NPI data cache. Observe latency signal, when the latency = 2, jump to the S3, latency = 1 Jump to S4, when latency = 0, jump to the S5. The signals of S2, S4, S5 are used to represent the data valid delay. In the S5 state, pop the data of 32 consecutive cycles, and then jump to S11. In the S11 state, wait for the complete signal. When the signal is valid, and then jump to the reset state S0. Similarly, 1DW read and 32DW read are the nearly same, which are different in the number of data read out, and we do not describe it in detail.

3.3.4   Send Logic State Machine Design

The timing of sending state machine is simple. S0 state is reset, when receive the read start signal to jump to the S1 state, while send a sign of read FIFO signal at the same

time. In the S1 state, send the first double word of the data, including byte enables flags and other information, and then jump to the S2 state. In the S2 state, send the second DW first, which contains valid data and address. If valid data is only one DW, jump directly to the S3 waiting period; if the valid data is more than a DW, continue to send data until the effective data transmission complete, and then enter S3. In the S3 state, we reset all the signal and send read complete signal flag, last jump to the S0 state to wait for the next sending request.
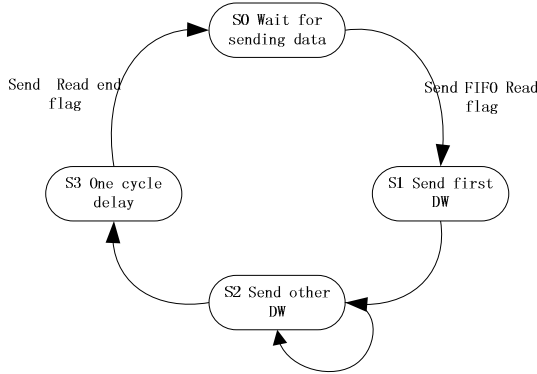
Fig. 8. Diagram of Send logic state machine

4   Experiment Test

Use ChipScope to capture the sending and receiving data.The diagrams of experimental results are shown in Fig.9 and Fig.10.
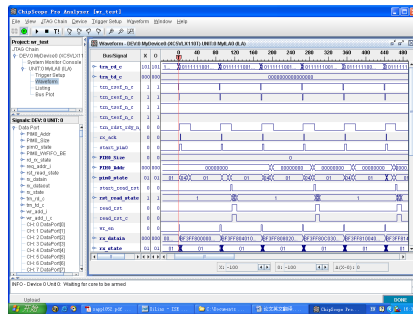
Fig. 9. Diagram of sending timing

In the picture, we can see that we can send one writing request in 50 clock cycles of 125MHZ, which writes 128Bytes, so the data sending and receiving rate is 125M/50 * 128Byte = 320MB.

Fig. 10. Diagram of receiving timing

In the picture, we can see that we can send one reading request in 100 clock cycles of 125MHZ, which reads 128Bytes, so the data sending and receiving rate is 125M/100 * 128Byte = 160MB. Because of the use of ping-pong cache, so we can transfer the rate of optical fiber is 320MB.

5   Summaries and Future Work

This card includes optical fiber transmission system design technology, PCI-Express and DDR2 cache transfer technology, fusing these several technologies in the FPGA to send and receive date. This card has greatly improved reliability and transmission bandwidth than the data cards based on PCI.

PCIE link width can continue to expand and improve the clock frequency to 250MHz, DDR2 200MHz clock can also be used, so that bandwidth will be doubled. Because the drive of this test is a simple code and hardware interactions are not very good. We can further modify the driver, so reading and writing rate can also be up to a lot.

References

[1] Liu, J.: Design and implement of High-speed Data Transfer System Based on PCI-Express and DDR2 SDRAM. Nanjing University of Information Science and Technology, Nanjing (2009) (in Chinese)
[2] LogiCORE^TM IP Endpoint Block Plus v1.14 for PCI Express (April 2010)
[3] Xilinx. Virtex-5 libraries Guide for HDL Designs (2008)
[4] Xilinx. Virtex-5 User Guide (September 2007)
[5] PCI-Express Base Specification rev1.1 (March 2005)

The Application of IPv6 Transition Technology in the Construction of Campus Network

Shi Kai

Ningbo University
Ningbo, Zhejiang, China
`shikai@nbu.edu.cn`

Abstract. At present, the rapid developments of IPv6's next-generation internet technology make many problems which basing on current IPv4 internet could be solved. How to make a perfect network transition from IPv6 and IPv4 becomes the focus of current researches. This paper first describes the mainstreams of IPv6 and IPv4 network transition technologies; it proposes campus's demands on IPv6 protocols basing on analyzing several issues exist in campus network. It also conducts researches and explorations on universities' construction of IPv6 networking model and upgrade mode of the old campus network.

Keywords: IPv4, IPv6, Campus Network, Transition Technology.

1   Introduction

Along with the rapid developments of internet, the network services are growing day by day and the number of IP network users is increasing dramatically. as early as 1992, IETF (Internet Engineering Task Force) bases on IPv4 and makes definition on the internet protocol of next generation, which is known as "IPv6". Through 10 years of developments, IPv6 has gradually become a new internet protocol which is an alternative for current IPv4 protocol. In 2003, China launched Next Generation Internet (CNGI) and planed to completely replace IPv4 with IPv6 in 2015.

2   Network Transition Technology from IPv4 to IPv6

Because IPv6 is not downward compatible with IPv4, it shall be totally established on the network before IPv6 protocol. First we must deal with the relationship between existing IPv4 network and IPv6 network in the future. We shall make IPv6 protocol stack network can conduct normal communications with the one which is supported by current IPv4 and ultimately achieve smooth transition from IPv4 to IPv6. Therefore, if we can make successful transition from IPv4 to IPv6 becomes the key for whether IPv6 network can achieve success in the future. At present, many people have studied various transition technology and interoperability solutions to provide a series of tools in addressing the transition issues [1].

The transition process from IPv4 to IPv6 is gradual and controllable. The transition period will be quite long.

IPv6 provides many transition technologies to achieve this gradual process. These transition techniques mainly solve two major categories of problems: IPv6 isolated island working face to achieve the interoperability between IPv6 networks and IPv6 networks; IPv6 and IPv4 inter-working technologies can achieve access of mutual visits from two different networks. At present there are 3 basic technologies to solve the transition problems: Dual-stack technology, tunneling and translation technologies [5].

Fig. 1. Network transition technologies from IPv4 to IPv6

2.1   Dual-Stack Technology

Dual-stack technology is the most direct way to make node compatibility between IPv6 nodes and IPv4 nodes. It refers to there are two protocol stacks of IPv6 and IPv4 in network elements and terminals. When IPv6 node which supports dual-stack communicates with IPv6 nodes, it makes use of IPv6 protocol stack; when it communicates with IPv4 nodes it makes use of IPv4 protocol stack. When the equipment is upgraded to IPv6 it also retains the supports on IPv4, it can also simultaneously access to IPv6 and IPv4 device, in which the application program relies on DNS address to identify the types of returned address to decide what kind of protocol stacks shall be adopted.

However, this approach provides fully compatibilities for IPv4 and IPv6. Since it needs dual-route infrastructure, increase the complexity of network, so it still can not resolve the depletion problem of IP address.
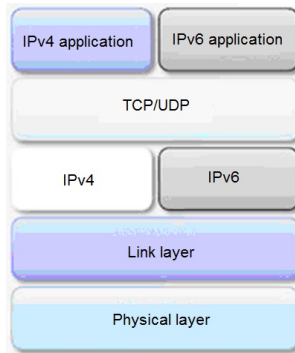
Fig. 2. Dual-stack technology

2.2   Tunnel Technology

Tunnel technology is a data technology which makes use of one protocol to transmit another protocol; it is a quite important transition method. The requirements of tunnel technology need all nodes on package and package-releasing shall have the function of IPv4/IPv6 dual-stacks, it provides an existing IPv4network to transfer data of IPv6. It takes the groups of IPv6 as non-significant data and encapsulates them in IPv4 packets. It is transmitted by IPv4 network and then re-opened at the destination to get IPv6 packet [2].

    Tunnel technology is divided into manual configuration and automatic configuration these two kinds. Tunnel technology which is manual-configured refers to configure manually of tunneling end to a specific IPv4 address. While the establishment and demolishment of automatic-configured tunnel is a dynamic status, IPv4 address on the tunnel endpoint is included in the data packets which have destination address of IPv6, such as IPv6 to IPv4 tunneling.

    It indicates that the tunnel technology makes clever application of the existing IPv4 network. Its significance is that it provides a way for making nodes of IPv6 can communicate with each other during transition period, but it can not solve the problem of fluent communication between IPv6 nodes and IPv4 nodes.
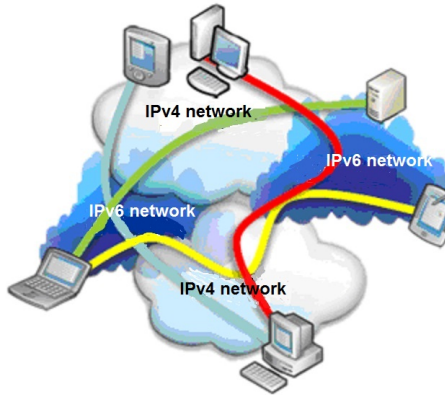
Fig. 3. Tunnel technology

2.3   Translation Technology

Protocol translation can be divided into state protocol translation and stateless protocol translation. SIIT (Stateless IP / ICMP Translation) (Figure 4 (a)) means stateless translation, that is, equipment simply conducts one to one address translation and message format translation. The device does not record the corresponding relationships and there are no resource occupancy situation, but these also restraint the application of address space. Stateless protocol address translation can not deal with sensitive protocols; state protocol translation can solve this problem, namely going through the application protocol gateway.
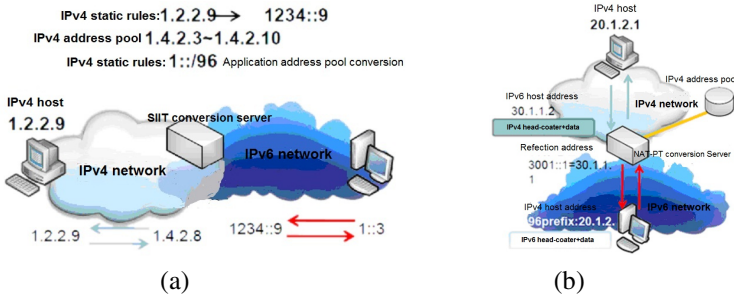
Fig. 4. Translation technology

The above-mentioned three kinds are the main technologies which are recommended by IETF. In addition, there are many transition mechanisms; they have their own advantages, disadvantages and different application scope. Only we choose the suitable ones according to local conditions, conduct scientific analysis and identify different types, scope and systematic categories of transition phase, make choice of a reasonable conversion mechanism can we realize IPv4 network's transition to IPv6 at small costs.

3   The Application of IPV6 in the Construction of Campus Network

The campus network which is based on IPv6 technology mainly considers campus network upgrade IPv6 supporting business and also adopts IPv6/IPv4 network devices to conduct the construction of new campus construction these two cases.

3.1   Upgrading the Existing Campus Network

There are a lot IPv4 equipment in original campus network, they do not have IPv6 functions or they can not be upgraded to IPv6. The transformation of network quickly into IPv6 will take a long time and huge costs. In order to protect the original IPv4 investment, while allowing new users to use IPv6, you can adopt tunnel mode. The original IPv4 need not be modified; you only need to increase one dual-stack and three layers of switches in the core region and establish ISTAP tunnel to provide services. This will achieve interoperability with IPv6 network and make data going through the existing IPv4 network. At the same time, through the exchanges of technology it can realize interoperability with original IPv4 and get shared access to the existing IPv4 network resources. However, this upgrade still does not solve the poor performance of IPv6 forwarding in dual-stack network, since the forwarding of IPv6 forwarding can be achieved by router software, so the poor performance can not be avoided.

Base on the original campus network to construct a new generation of campus network, you need to re-build core layer and aggregation layer which support IPv6 business. You shall reach dual-core aggregation network. IPv4 business can be forwarded via original network; IPv6 can be forwarded by the new core. If you need to

implement the interoperability between IPv6 new campus network and the original IPv4 campus network, you need to start NET-PT function on the dual stack core switch. After the transformation, the core of whole network convergence layers is forwarded by high-performance hardware so that each information point on whole network has the ability to be accessed by dual-stack on client end [3].

3.2   Construct New IPv6 Campus Network

Construct new campus network bases on IPv6, it is recommended to construct by supporting IPv4/IPv6 network equipment and then achieve communication and exchange of two kinds of business supporting by next generation campus network platform. The campus network solution program which is supported by IPv6 makes use of tri stacks switch of network processor as the core equipment. The convergence layer adopts IPv6 switch or dual-stack router to achieve the function and export adopts high-efficiency dual-stack routers to achieve interoperability between IPv4 network and IPv6 network. This program supports both IPv4 and IPv6 traffic these two business currents. And it will not affect main IPv4 services in all major universities. The wire-speed forwarding capability, multiple tunnels and interoperability technology of IPv4/IPv6 which have hardware can meet the IPv6 applications business service of various universities under different network environments.

4   Summaries

The transition from IPv4 to IPv6 is a process which can not be overstepped to achieve global Internet process. The transformation from IPv4 to IPv6 requires a fairly long transition period, during this transition period it requires coexistence of IPv4 and IPv6, and both sides make perfect compatibility adaptations with each other. They shall gradually move smoothly and evolve, eventually all the network nodes will run IPv6 and play IPv6's full advantages in address space, performance and security. We are looking forward to qualitative changes will be brought to our life, study and working manners.

References

[1] Lu, J., Zhou, J., Ren, Q., Fan, J.: IPv4 to IPv6 transition technologies and network transition program. Modern Computer, 9–11 (2005)
[2] Cover Lingyun: Tao fly. based on the IPv4 to IPv6 tunneling transition. China Education Information, 3–4 (2007)
[3] An account Yong, Du, Z.: Ipv6 in the campus network application. Computer and Information Technology, 6–8 (2007)
[4] Leaves New. IPv6 protocol application analysis and research. Computer Knowledge and Technology, 21–25 (2009)
[5] Jin, X., Chen, X.: IPv4 to IPv6 transition technology. Computer Knowledge and Technology, 30–32 (2008)
[6] Chu, Y.: He Xin. from IPv4 to IPv6 transition mechanisms of the tunnel. Fujian Computer, 9–12 (2006)

A MAC Protocol for Underwater Acoustic Network

Liu Xun, Li Yu, Fang Dong, Zhang Chun-Hua, and Huang Hai-Ning

Institute of Acoustics, Chinese Academy of Sciences, Beijing, China

Abstract. MAC protocol is an important problem in Underwater acoustic network. Many terrestrial wireless MAC protocol can't work well in underwater because of long latency and limited bandwidth. In this paper, we modify the quiet duration in MACA-U using node position. The proposed MAC protocol is known as Geographical MACA(G-MACA). Simulation results in NS2 show that the throughput performance of G-MACA is much higher than that of MACA-U.

Keywords: underwater acoustic network, MAC protocol, throughput, MACA-U.

1 Introduction

The character of underwater channel pose great challenge in the design of MAC protocol for underwater acoustic network[1][2]. Traditional wireless MAC protocol such as TDMA and FDMA can't work well in underwater because of high latency and narrow bandwith.

Recent researches of MAC protocol for underwater focus on contention based MAC protocols which include Aloha based protocols and handshaking based protocols [3][4][5][6]. In [4], a MAC protocol named MACA-U has been proposed. MACA-U improves the state transition of MACA and gets better throughput performance than MACA. However, MACA-U hasn't consider over the real latency in quiet rules.

While many Geographical routing protocols have been proposed for underwater[7], there are no MAC protocols using location of nodes.

In this paper, we proposed a geographical MAC protocol for underwater which is named G-MACA. G-MACA uses location of sensor nodes to optimize the quiet duration. The throughput performance of G-MACA is much higher than that of MACA-U.

2 Description of G-MACA

2.1 MACA-U Overview

In MACA-U[4], node has five distinct states, namely, IDLE, CONTEND, WFCTS, WFDATA and QUIET. A source node goes to CONTEND state from IDLE state when it has packet to send. When CONTEND state ends, the source node transmits a RTS and goes to WFCTS state. Meanwhile, the source node sets its timer to $2\tau_{max} + T_{cts}$, where τ_{max} is the maximum propagation delay and T_{cts} is the CTS duration. Once the receiver node receives the RTS, it transmits a CTS and goes to WFDATA state. Similarly, the

receiver node sets its timer to $2\tau_{\max}+T_{data}$,where T_{data} is the data packet duration. Other nodes which receive a RTS or CTS will go to QUIET set their timers to $QUIET\_RTS$ or $QUIET\_CTS$.

In order to improve the throughput, MACA-U proposes following transition rules:

In WFCTS state, a source node disregards any RTS or xRTS(RTS intended for other nodes). If it overhears a xCTS(CTS intended for other nodes), it will goes to QUIET state.

In WFDATA state, a receiver node disregards any RTS, CTS, xRTS and xCTS.

In QUIET state, a node remains QUIET state when it overhears xRTS or xCTS. The node computes $\max\{Q_{lo},Q_{ov}\}$,where Q_{lo} is the local quiet duration, and Q_{ov} is the overheard control packet's quiet duration.

2.2   G-MACA

In our analysis，D is the maximum propagation distance and c is the sound speed, so $\tau_{\max}=D/c$.

Propagation delay isn't important in wireless network because it is much smaller than packet duration. However, in underwater network, propagation delay is so high that it can't be neglected. When distance is very long or the packet duration is very short such as RTS, propagation delay will be larger than packet duration. So propagation must be considered in design of MAC protocol.

As in Geographical routing protocols, we premise that underwater node can know its own location and its neighbor's location by exchanging beacon periodically in G-MACA. We propose following modifications to improve G-MACA's throughput efficiency.

2.2.1   Quiet Rule

In MACA-U, if a node receives xRTS when its state is IDLE, CONTEND or QUITE, it will set $QUITE\_RTS=2\tau_{\max}+T_{cts}$ so that its next transmitting packet can't arrive at source node while source node receive CTS. If a node receives xCTS when its state isn't WFDATA, it will set $QUITE\_CTS=2\tau_{\max}+T_{data}$ so that its next transmitting packet can't arrive at receiver node while receiver node receive DATA packet. When $\tau_{\max}$ is high, quiet time is so large that much time is wasted in some scenes. For example, in the scene of fig.1, node A, S and D are in a straight line. The distance of A and S is D while the distance of S and D is $D/2$. If node S sends RTS to node D at t_0 , node S will receive CTS at $t_0+D/c+T_{cts}+T_{rts}$. Meanwhile, node A will receive xRTS at $t_0+D/c+T_{rts}$ and set $QUITE\_RTS=2D/c+T_{cts}$. So node A can't send a packet before $3D/c+T_{cts}+T_{rts}$. But in fact, if $D/c>>T_{rts},T_{cts}$, the packet will arrive at node S after node S receives CTS even if node A sends a packet at $t_0+D/c+T_{cts}+T_{rts}$. So node A doesn't need to quiet.
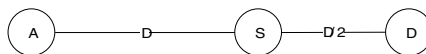
Fig. 1. A scene in which overhear node don't need to quiet

The purpose of quiet is that a node which overhears a packet can't send a packet to disturb current communicate process. We propose quiet time modifications in following three scenes to make quiet duration more appropriate.

1) Overhearing node is source node's neighbor but not receiver node's neighbor
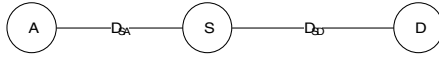
Fig. 2. Node A is only the neighbor of node S

In fig.2, node S is source node and node D is receiver node. Node A is only the neighbor of node S. We should only compute G_QUITE_RTS of node A because it can't overhear any CTS from node D.
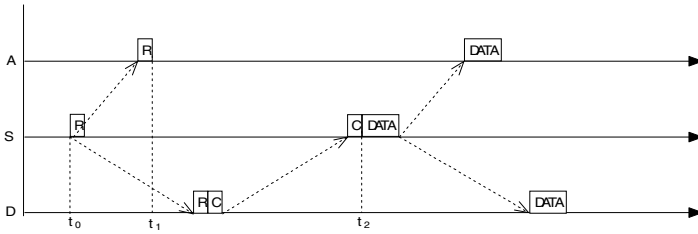
Fig. 3. Time diagram in scene of fig.2

In fig.3, node A receives xRTS at t_1. Node A should set a quiet duration so that the next packet which node A may send can't arrive at node S before t_2. So the quiet duration should satisfy following inequality:

$$G\_QUITE\_RTS \geq t_2 - t_1 - \tau_{SA} \tag{1}$$

where $\tau_{SA} = D_{SA} / c$. Then,

$$\begin{cases} t_1 = t_0 + \tau_{SA} + T_{rts} \\ t_2 = t_0 + 2\tau_{SD} + T_{rts} + T_{cts} \end{cases} \tag{2}$$

where $\tau_{SD} = D_{SD} / c$. Then,

$$G\_QUITE\_RTS \geq 2(D_{SD} - D_{SA}) / c + T_{cts} \tag{3}$$

2) Overhearing node is source node's neighbor but not receiver node's neighbor.

In fig.4, node S is source node and node D is receiver node. Node B is only the neighbor of node D. We should only compute G_QUITE_CTS of node B because it can't overhear any RTS from node S.

Fig. 4. Node B is only neighbor of node D
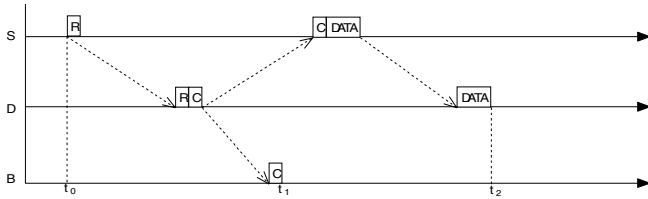
Fig. 5. Time diagram in scene of fig.4

In fig.5, node B receive xCTS at t_1. Node B should set a quiet duration so that the next packet which node B may send can't arrive at node D before t_2. So the quiet duration should satisfy following inequality:

$$G\_QUITE\_CTS \geq t_2 - t_1 - \tau_{DB} \tag{4}$$

where $\tau_{DB} = D_{DB} / c$. We can get following equations in (5) from fig.6:

$$\begin{cases} t_1 = t_0 + \tau_{SD} + \tau_{DB} + T_{rts} + T_{cts} \\ t_2 = t_0 + 3\tau_{SD} + T_{rts} + T_{cts} + T_{data} \end{cases} \tag{5}$$

where $\tau_{SD} = D_{SD} / c$. Then,

$$G\_QUITE\_CTS \geq 2(D_{SD} - D_{DB}) / c + T_{data} \tag{6}$$

3) Overhearing node is not only source node's neighbor but also receiver node's neighbor.
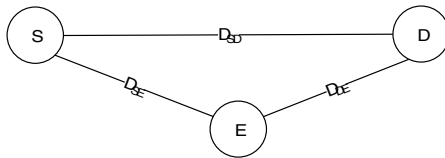
Fig. 6. Node E is neighbor of both node S and node D

In fig.6, node S is source node and node D is receiver node. Node B is not only neighbor of node S but also neighbor of node D. We should compute not only G_QUITE_RTS but also G_QUITE_CTS of node E because it can overhear both RTS from node S and CTS from D.

In fig.7, as in scene of fig.2, the G_QUITE_RTS should satisfy:

$$G\_QUITE\_RTS \geq 2(D_{SD} - D_{SE}) / c + T_{cts} \tag{7}$$

However, node E is also the neighbor of node D, so node E can receive CTS from D. If node E can't receive CTS from D, it is possible that node D can't communicate with node S. So if node E can't receive CTS from node D at t_2, node E could end its quiet timer. So the G_QUITE_RTS should satisfy:

$$G\_QUITE\_RTS \geq t_2 - t_1 \tag{8}$$

We can get following equations in (10) from fig.8:

$$\begin{cases} t_1 = t_0 + \tau_{SE} + T_{rts} \\ t_2 = t_0 + \tau_{SD} + \tau_{DE} + T_{rts} + T_{cts} \\ t_3 = t_0 + 2\tau_{SD} + T_{rts} + T_{cts} \\ t_4 = t_0 + 3\tau_{SD} + T_{rts} + T_{cts} + T_{data} \end{cases} \tag{9}$$

where $\tau_{SD} = D_{SD}/c, \tau_{DE} = D_{DE}/c, \tau_{SE} = D_{SE}/c$. Then,

$$G\_QUITE\_RTS \geq (D_{SD} + D_{DE} - D_{SE})/c + T_{cts} \tag{10}$$

Because $D_{DE} + D_{SE} \geq D_{SD}$, G_QUITE_RTS should only satisfy inequality (10).
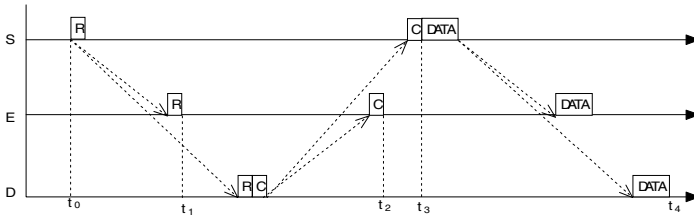
Fig. 7. Time diagram in scene of fig.6

If node E receives CTS from D, it should reset quiet timer so that the next packet which it may send can't disturb the DATA receiving of node D. As in scene of fig.4, G_QUITE_RTS should satisfy following inequality:

$$G\_QUITE\_CTS \geq 2(D_{SD} - D_{DE})/c + T_{data} \tag{11}$$

2.2.2   Packet Modify

Because overhearing node can only know the location of its neighbor, it can't compute the quiet duration when it is only neighbor of source node or neighbor of receiver node. From expression(3) and (6), we can find that only D_{SD} is unknown. So we add the distance of source node and receiver node in RTS and CTS as in table 1. The Distance of source and destination can be stored in four bytes.

Table 1. Packet format of RTS and CTS in G-MACA

Frame type	Source ID	Destination ID	Distance of source and destination	......

3   Simulation and Analysis

G-MACA and MACA-U are simulated in ns2 and compared. The acoustic propagation speed is 1500m/s. A multi-hop network topology comprised of 25 static nodes is set in our simulation as shown in Fig.9. Each node assumes static routing with the route denoted by the arrows in Fig. 9. The 16 nodes residing on the outer circle generate random packets with constant length and Poisson distribution independently, which are

forwarded by the 8 nodes on inner circle and are received by the center node. Throughout vs offered load is simulated for various data packet's lengths, bit rates and grid sizes. Offered load and throughput are computed as following equations:

$$\text{Offered Load} = \frac{\text{Packets Sent}/\text{Simulation Time}}{\text{Bit Rate}/\text{Packet Length}} \qquad \text{Throughput} = \frac{\text{Packets Received}/\text{Simulation Time}}{\text{Bit Rate}/\text{Packet Length}}$$

Throughput for packet lengths of 128, 256 and 512 bytes is shown in Fig.10 when the bit rate is 1000bps and grid size is 800m. The maximum transmission range is 1500m. The solid lines refer to throughput of G-MACA while the dashed lines refer to that of MACA-U. Both protocols exhibit stable throughput for all packet lengths at last. But the throughput of G-MACA is clearly higher than that of MACA-U for same packet length. Throughput increases as packet length increases because larger packet duration results in less influence of propagation delay.

Throughput for various bit rates of 1000bps, 2000bps and 4000bps is shown in Fig.11 when data packet length is 128 bytes and grid size is 800m. The maximum transmission range is 1500m. Similarly, the throughput of G-MACA is higher than that
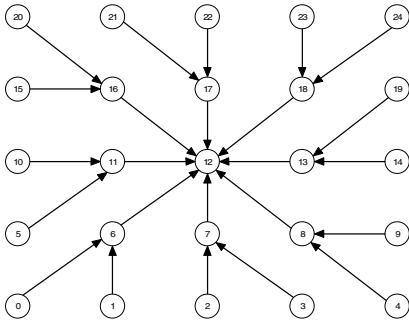
Fig. 8. Network topology
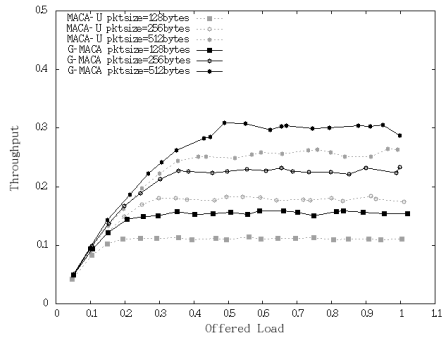
Fig. 9. Throughput vs offered load for various packet lengths
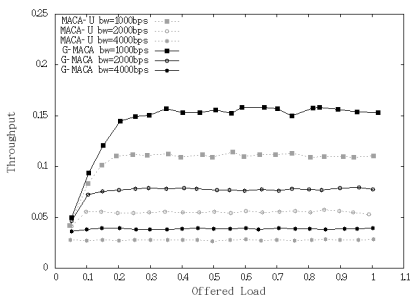
Fig. 10. Throughput vs offered load for various bit rates
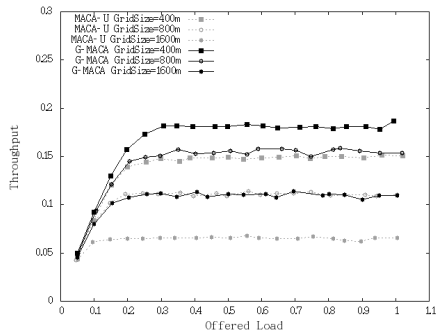
Fig. 11. Throughput vs offered load for various grid sizes

of MACA-U for same bit rate. Packet duration becomes smaller when bit rates increases, so the throughput decreases.

Throughput for various grid sizes of 400m, 800m and 1600m is shown in Fig.12 when data packet length is 128 bytes and bit rate is 1000bps. Meanwhile, the maximum transmission range is 1.875 times of grid size. Similarly, the throughput of G-MACA is higher than that of MACA-U for same grid size. Throughput decreases as grid size increases because propagation delay increases.

4   Summaries

In this paper, we design a MAC protocol named G-MACA for underwater network. In G-MACA, we compute the quiet duration by using node's real propagation delay instead of maximum propagation delay which is used in MACA-U. So node can save much more time to send or receive packet than MACA-U. Simulation results show that the throughput of G-MACA is much higher than that of MACA-U in various packet lengths, bit rates and grid sizes. As distance of nodes increases, G-MACA is more and more suited for underwater network than MACA-U. In G-MACA, we should know the location of node. In next work, we will research for using G-MACA without knowing location of node.

References

[1] Akyildiz, F., Pompili, D., Melodia, T.: Underwater acoustic sensor networks: research challenges. Elsevier's Journal of Ad Hoc Networks 3(3), 257–279 (2005)
[2] Chitre, M., Shababuddeen, S., Stojanovic, M.: Underwater acoustic communications and networking: recent advances and future challenges. Marine Technology Society Journal 42(1), 103–116 (2008)
[3] Karn, P.: MACA-A new channel access method for packet radio. In: Proc. ARRL/CRRL Amateur Radio Ninth Computer Networking Conf. (September 1990)
[4] Ng, H.-H., Soh, W.-S., Motani, M.: MACA-U: A media access protocol for underwater acoustic networks. In: Proc. IEEE Globecom, New Orleans, Louisiana, USA (December 2008)
[5] Guo, X., Frater, M.R., Ryan, M.J.: Design of a Propagation-Delay- Tolerant MAC Protocol for Underwater Acoustic Sensor Networks. IEEE Journal of Oceanic Engineering 34(2), 170–180 (2009)
[6] Fang, D., Li, Y., Huang, H., Yin, L.: A CSMA/CA- Based MAC Protocol for Underwater Acoustic Networks. In: 2010 6th International Conference on Wireless Communications Networking and Mobile Computing (WiCOM), Chengdu, September 23-25 (2010)
[7] Chirdchoo, N., Soh, W.-S., Chua, K.C.: Sector-based Routing with Destination Location Prediction for Underwater Mobile Networks. In: International Conference on Advanced Information Networking and Applications Workshops, WAINA 2009, Bradford, May 26-29 (2009)

Computer Network Environment of Network Teaching

Yu Du

School of Information
Linyi University
Linyi, Shandong, China, 276005
`duyu@lyu.edu.cn`

Abstract. Network development creates a new network culture. Many schools need to build a campus network connected to the Internet; the network began to try teaching. Teachers in the teaching process, teaching how to network the different mode of teaching, play the leading role of teachers and inspire students to become self-learning so that students understand the subject. It is necessary to understand the online teaching mode, characteristics, issues and responses, and in teaching Work, continuous practice and innovation.

Keywords: computer network, online teaching, Problems and Solutions.

1 Introduction

Network development, especially in the development of the Internet information age, social cell (multi-media computers and computer   technology to master the people) even become one, realized the networking of human wisdom, and thus creates a new network culture. The times, many schools need to build a campus network, and connected to the Internet, the network began to try teaching.

The so-called use of online teaching is a teacher teaching computer networks to pass information to the students, or students access to knowledge through computer networks. As the network has a time of teaching, sharing, interaction and individual characteristics, teachers in the teaching process, teaching how to network the different mode of teaching, to play the leading role of teachers and inspire students to become self-learning so that students understand the subject, which Necessary to understand the online teaching mode, characteristics, issues and responses and to constantly practice in teaching and innovation.

2 The Network Teaching Characteristics

Network under the guidance of teachers teaching students to construct learning, is based on network technology, multimedia technology under the new learning theory. This theory is the learning of students as active constructors of knowledge, external information environment through the organization of teacher, guide, students take the initiative to construct, transformed into their own knowledge. This openness,

innovative ideas and models of education, help to overcome the drawbacks of traditional teaching model and solid diseases.

Network environment based on computer network teaching, the main features of the following areas:

(1) Network interactive teaching process. Network is its most important feature of teaching real-time interactivity. Real-time interaction refers to the various terminals on the network can be implemented immediately answered. The way interacts between students and teachers, among students. Through the courseware can be effectively obtain maps, text, sound, and Mao as educational information, the interaction between teachers and students can get the guidance of teachers and students can interact between the collaborative learning. This two-way interaction so that students not only through sight and hearing means of access to teaching information, and it represents a student can receive, and the unprecedented interest. The process of learning in the classroom, students can interactively timely information on their progress and shortcomings, learning to adjust as required, thereby greatly enhancing the quality and efficiency of learning.

(2) The sharing of online teaching resources. Students through the network can share resources, including the sharing of hardware and software sharing; the core is the sharing of software. Online resource rich and colorful, illustrated, phonetic skills, learners or teachers under the guidance of the link, you can ease in the surf in the ocean of knowledge. Inexhaustible information resources, the magic of network environment, the formation of individualized education, the cultivation of creative thinking of students, to achieve change in the relationship between elements of the teaching process, from the examination-oriented education to promote the transition to quality education will have significant and far-reaching impact.

(3) A comprehensive multimedia information. Multimedia Hypermedia and Hypertext financial integration, set graphics, images, graphics, sound, and text in one will help students participate in a variety of sensory awareness promote student construction of knowledge; help to stimulate student interest, to develop students Emotion; is conducive to implementation of quality education.

(4) Advanced teaching methods. The center of the traditional teaching of teachers in teaching, online teaching is the student learning center. In online teaching, teaching is no longer the Man Tang guan; students should be under the guidance of teachers to learn according to their needs. The main function of teachers is to organize, control, and job evaluation methods to help students.

(5) The diversity of teaching goals. Individual differences of students, such as student learning, learning styles, and so determines the starting point of the diversity of teaching objectives. In the traditional "teaching lines" it is difficult to achieve individualized, online teaching to overcome these drawbacks. Multimedia network teaching, brought many real situations, will be living in the bright lights of the phenomenon into the classroom, students will be increased efficiency, is conducive to physical and mental health of students.

(6) Teaching the richness of content. Teaching content online is not only our subject teachers prepared CA I courseware, but also from different types of schools in different parts of the courseware content. Students learn by comparison, select those that best organized content structure of Web sites or courseware to learn, to improve the efficiency of learning per unit time.

3   The Network's Basic Teaching Mode and Teaching Methods

Central task of the school classroom teaching, classroom teaching is the main front of quality education, schools should play a real performance for the training of a new generation to be started from the main front of the classroom. Modern constructivist theory that "scenario ", "cooperation", "session "and "construction of meaning" are the four elements of learning environment. In the constructivist learning environment, classroom teaching center shifted from the teacher learner, the core of the learner take the initiative to new information and experiences with their existing knowledge base, the process of integration, so the network of teaching is adapted to the needs of the Theory of Constructivism.

Online teaching mode can be roughly divided into:

(1) Collective Teaching. Teachers first taught using conventional teaching methods in the creation of scenarios to stimulate student interest while the concepts, principles, etc. about the clear and provide an avenue for students to gather information on the computer or group discussions, collaborative exchanges, while briefing given scenarios. Rapid and timely feedback is the use of computer characteristics, providing of hard to clear up the doubts. Let go and allow students to self-practice, students difficult problems arise, can be e-hands. Teachers by monitoring the management functions, some of the selected students have problems, the formation of a specific study group. Students can use voice chat, BBS, chat room, bulletin boards and other discussion, exchange of experience. Teachers also accepted the advice of students, guide students to eliminating, to help. The use of radio functions, the typical error can be broadcast in the group or class, quickly corrected to attract student's attention. This teaching model for students are to learn in a relaxed environment, new knowledge, to work together to promote the awareness of students. For a teacher, quickly and effectively collect feedback from students and make teaching more in line with the actual content of the regulation changes and teaching programs.

(2) Self-learning mode. The model is the use of server functions, full use of hardware resources, so that the limited equipment to maximize efficiency. Students to use the server site, choose the difficulty of learning content, and keep pace with the teachers and students interact. A server site, the equivalent of a small library, students can share a wealth of information on the site server, in a network environment for individual learning, so that gradually away from the traditional classroom teacher center model, students are forced from the traditional learning into active learning.

Eight ways of teaching the current network:

(1) Video Broadcasting: the network management center by playing the video (real-time video or video). This form of the TV or playing video is not much difference, the students can not control, is actually a live classroom teaching. Its advantage is that the number of learners from the constraints of geography, and the bandwidth is small (only 1. 37M). Used for teacher instruction, academic reports, broadcast and other important meetings.

(2) Video on demand: the learner as needed on-demand video server. Teaching film, television content can also be a live video classroom. As a non-real-time on-demand, so we can carefully design their (insert graphics, video, animation, etc.).

(3) Video conferencing: real-time transmission of video and audio in the form of multi-directional.

The equipment is expensive, the application is not very common, and more for teachers answering.

(4) WEB Book: the teaching content made in the form of web pages.

The advantage is that preparation is not difficult, easy operation, so use common.

(5) Multimedia courseware: the use of multimedia courseware development tools to develop language or teaching materials, usually by downloading to the local operation.

(6) BBS Forum: between teachers, between learners in the form of an electronic bulletin board exchange and collaboration.

(7) Chat rooms (Teaching Forum): between teachers, between learners through text, voice and other forms of remote real time.

(8) E - mail: between teachers, between learners in the form of e-mail exchange.

4 The Main Problems and Solutions

(1) Class size problem --- the rational utilization of network resources.

To increase teaching capacity, the officer placed a great deal on the network of information resources, but it has had a serious consequence, that is: a wealth of information for students to get confused, but also the background information, and an electronic blackboard, but also a picture, but also the Related Links information.

Students even have time to visit it again, let alone take notes, analysis, reflection, opinion, absorption, synthesis, digestion, and resulting in a student can not start, aimlessly, reading random phenomenon. This is actually the teaching of the network environment is a serious out of control. In the process of solving this problem is found: the capacity of our understanding of classroom teaching into the errors. We know that the classroom teaching capacity per unit time is not unlimited expansion. Students in the unit time to accept, to digest the information capacity is limited. Therefore, increasing the information capacity of the network should not be used as teaching goals, improve teaching efficiency is the goal. As a teacher, properly handle the relationship between networks and teaching materials.

If confined to the existing network knowledge materials, it becomes an electronic textbook; network knowledge if completely out of the materials, it becomes a source of water. Network is just teaching the tools and instruments, in the use of their teaching, we must carefully analyze, in-depth study materials, clear teaching objectives, focus on the theme, focus, and difficult, not to the interest of students and arbitrary. Therefore, the only reasonable use of network resources, classroom capacity is to master in order to improve teaching efficiency.

(2) Independent learning issues --- the role of teachers re-positioning.

In order to facilitate independent learning, to reduce the teaching time of teachers, increase student independent reading time, in the reading process without interference. Students to enjoy reading this should not interfere with it? This will affect the student self-learning? This is confusing the practice of online teaching another problem. Through theoretical study and practical demonstration, online teaching, students are learning the subject, the learning process is the learner to

complete the minds of their own process of meaning construction. Teachers in the teaching process in the network from the traditional role of imparting knowledge, transformed into a facilitator of student construction of meaning, and the resources to help the organizers of the role. Network teaching interaction if it is benign, not only will not weaken the role of teachers, but will be enhanced.

For example, in reading, discussion can help teachers adjust the reading progress of individual students, the course of the discussion the concept of time to correct the error and found that high school students to discuss the arguments of the process to remind everyone's attention, online guide to solve the students questions and more.

(3) Emotional interaction problem --- a necessary complement to traditional means.

Encountered in practice but also a more difficult issue is how to solve the emotional interaction in online teaching issues. Ideological education feature in the network environment, how to embody it? By exploring, I believe that the advanced network technology and traditional teaching methods are not mutually exclusive, have you not my relationship. Organic combination of the two to maximize the mobilization of the student's thoughts and feelings, emotional sublimation will sound and the silence in between the dark potential of AIDS rose. Such as audio and video data playback of a class before, interspersed with a vivid lesson about, among students through language classes start the end of the fierce confrontation of ideas, etc., can play the role of emotional interaction.

(4) Courseware general problem of optimal design template.

Basically, under normal circumstances courseware is to a section designed for specific class examples, so versatility is poor. Courseware is to investigate the versatility of the efficiency and effectiveness of the most important criteria. If a courseware highly versatile, making it not only the teacher, the school can be used, the country could use them.

The best way to improve the versatility of some of the content is to design simple and focused network courseware templates. Teachers, according to the different needs of class-based or embedded systems or embedded messaging system are to discuss and so on, to play the computer's interactive features. In addition, the production of courseware can be launched by the students, Internet search for information, etc. to teach students to do things. Students in the production of courseware in the learning process have been, and are a more demanding learning initiative.

(5) Network teaching equipment, technology, demanding, difficult to promote the problem --- step by step, a clear division of labor. Online teaching has certain requirements for the equipment, the school does not a certain number of computers, no Internet, and network access will not be able to carry out teaching activities. However, our understanding of the development of network teaching must go through a process of new learning methods can not change overnight, this change is a process of quantitative change to qualitative change. We can use a very low cost to build a small network environment began the practice of Teaching. Of course, as we practice deep, with the further development of social economy, the network will be teaching physical conditions improved. Even if limited by economic factors, can not carry out large-scale network of teaching, small-scale networks as long as we explore the process of teaching the practice of changing the concept of education to improve the teaching level, this network can also be regarded as the beneficial effects of practice

of the course. Multimedia Technology in Teaching and compared to the school level, the network teaching reliance on technology is not high.

True multimedia production, teaching programs involving the design, text, script writing, graphic design (art) creative, 2D3D animation, capturing and editing video and audio material, programming, etc., on human resources requirements of a high technical level of production platforms the high hardware requirements, the client device has certain requirements. If we make full use of existing public network resources, make full use of existing authority professional website, personal characteristics of sites, full use of common software, technical complexity of individual professionals to do things, just use Word to write a teacher program of learning activities. Teachers focus more on the program of activities in the design of inquiry learning, and more on learning activities, guidance on the learner.

5 Summaries

Vast network of information resources, excellent multimedia features and more functions to interact with the improvement of teaching quality and efficiency possible. We face is a web service is improving times, faced with a highly information-oriented education in the era of era of online teaching is the need for a school teaching the inevitable choice. We must rely on the campus network, to create a digital campus environment, the comprehensive utilization of various teaching methods, give full play to the advantages of online teaching to get the best teaching results.

References

[1] Ji, P.: Education and information technology into a new stage of development. China Educational (March 2000)
[2] Xiuying: The Century in Retrospect – from the media to see the evolution of the development of audio-visual. China Educational (October 1999)
[3] Tian, Y.: On the network of schools teaching the Information Age. Educational Technology Center (October 24, 2004)
[4] Ren, C.: Inquiry Learning: 18 Principles. Educational Theory and Practice 1, 2 (2002)
[5] Zhu, Z.: Modern Education Technology - About Information Education. Higher Education Press (2001)
[6] Liu, D.: The definition of online teaching. But Keep Education Laboratory (April 2002) (on - text - Net)

Design of a Wireless Sensor Network Platform for Real-Time Multimedia Communication

Yong Fu, Qiang Guo, Changying Chen, Jialiang Lv, and Xiaoshi Zheng

Shandong Provincial Key Laboratory of Computer Network
Shandong Computer Science Center
Jinan, China
fuy@keylab.net

Abstract. The frame formats of wireless packets for real-time burst multimedia communications are analyzed and designed at first. Then we theoretically study the network management and transmission protocol in detail. We also design and analyze a clock synchronization method for multimedia communications. At last we set up a wireless sensor network platform for real-time audio communications. This platform is a low power, high performance and low cost system. It suggests possible applications to Internet of Things, smart home, virtual environment, etc.

Keywords: Wireless sensor network, Internet of things, Time slot, Multimedia.

1 Introduction

A Wireless Sensor Network (WSN) [1, 2] consists of spatially distributed autonomous sensor nodes to monitor physical or environmental conditions, and to automatic networking each other by wireless communications. In the last ten years, WSN technologies have developed from environment parameter collection and monitoring to automatic control and data communications. These new type WSNs have potential applications on Internet of Things (IOT), smart home, virtual environment, short range data communication system, etc. Many researches have been carried out on WSN protocol, algorithms and hardware platform designs. Recently multimedia data transmission in the WSN has received considerable attention from scientist and engineers[3-6]. Multimedia data, which are audios, videos and images, are important medium, they contain rich and intuitive information[7-9]. However, little multimedia communication is applied in WSNs because of the limit of bandwidth and transmission mechanism. Multimedia data have these characters: larege volumn of data, long duration and point-to-point. highly real-time. For videos and images, we request the transmissions are real-time, and some packet loss and data error are acceptable. In most cases multimedia communication are burst, we can't forecast when it begin and when it finish. It can be seen the multimedia transmission has many differences with sensor data transmission.

In this paper, we first design packet format for real-time burst data communica-tions. Then we theoretically study the network management and transmission protocol in

detail. We also design and analyze a clock synchronization method for our application. Next we design hardware circuits for multimedia nodes and sensor nodes. At last we set up a wireless sensor network platform for real-time audio communications. The currents are measured and analyzed too.

2   Protocol Design

Packets are the base of a WSN. This paper designs packet formats for real-time burst data communications at first. Multimedia data is treated as a new data type in our design, and the WSN identify different data type and handle them accordingly. We divide WSN packets to 3 packet types: network management packet (NMP), sensor data packet (SDP) and burst data packet (BDP). NMP and SDP are traditional WSN packet types. NMP is used to establish, manage and maintain WSN, it's sent by center nodes usually. SDPs are main packets in traditional WSNs, they are uploaded to center nodes periodically. BDPs are used to transmit multimedia data, these new type of packets occupation much network resources since multimedia communications begin. The frequent data sending and receiving operations request more precise clock and more effective management than traditional WSNs.[10]

In our designed WSN, packets are packaged with address information for effective management of wireless data. We will discuss it in details in next section. We also design different packet formats and priorities for different packets. NMPs have highest priority; BDPs have lowest priority. NMPs and SDPs have check bytes and using ACK to ensure reliable communication. BDP's format is simplest to ensure real-time communication.

Time slot design is the base of the protocol. All operations should be done in specified slots, and only one node sends RF packet in a time slot. And the most important thing is all nodes in the network must be clock synchronization. We designed a simple but effective strategy to realize precise clock synchronization:

First, when a client node joins the network, it must adjust clock to be synchronized with the center node's clock.

Second, the center node send a polling command every ten cycles, this operation has two effects: checking client node's status and synchronizing clock.

At last, the clock accuracy is selected as 32us, which is good enough for our application, it can be generated by 16MHz crystal and can be used for multimedia sampling period. Also a time gap about 160us is introduced, when nodes wake up and enter slots, they have to wait 160us before they send RF packets. If time error of a received packet is less than 400us, it can be treated as clock synchronized.

It's important to get precise packet processing time for a time-slot based system. We introduce some parameters to fix ideal packet processing time model: processing time of MCU T_m, frame head and frame end length of a packet N_0, sending processing time T_s and receiving time T_r of RF chip, assuming a wireless packet length is N, total processing times are:

$$\begin{cases} T_{send} = T_s + T_m + \dfrac{N + N_0}{p} \\[2ex] T_{recv} = T_s + T_r + \dfrac{N + N_0}{p} \end{cases} \tag{1}$$

Where p is wireless transmission speed, the unit is Bps, T_{send} is packet sending time, T_{recv} is the packet receiving time. In our platform, the network is designed for both 250kbps and 500kbps RF speed applications, the RF packet sending time for different MCU settings are measured:

Table 1. RF packet sending time

MCLK	SPI speed	RF speed	Packet length	Time Spent*	Equ. speed
16M	4Mbps	500Kbps	5 bytes	~1.08ms	4.6 Kbps
16M	4Mbps	500Kbps	13 bytes	~1.23ms	10.6 Kbps
16M	1Mbps	500Kbps	13 bytes	~1.62ms	8 Kbps
16M	4Mbps	250Kbps	46 bytes	~2.8ms	16.4 Kbps
16M	4Mbps	500Kbps	46 bytes	~1.9ms	24.2 Kbps

*Time spent includes synthesizer calibration time 721us.

It can be seen that longer packet length and higher RF speed bring high equivalent speed. In our system, equivalent speed of 24.2kbps meets most requirements for audio and image transmission. Video transmission request higher equivalent speed, so packet length should be longer, RF speed and MCLK should be higher. The time slot is set to 2.4ms. The time slot for command and data is 3ms, which is longer than the multimedia slot, it's because data nodes don't have precise high speed crystal, and their MCLKs are much lower than multimedia nodes or center node, so we must left enough time to cancel out the time error. At last, the length of RF packets should be less than 32 bytes in order to make sure the packet can be transmitted in given time slot, if there are more data to be transmitted, the nodes must wait for the next time slot. A processing time period T consists several time slots depending on application. The period is usually less than 100ms. A cycle consists several T, which is the basic unit for DNODEs, the time is from several seconds to hours.

In our network, client node which transmits data only is called DNODE. No high speed crystal or display component is used for DNODE in order to reduce power consumption. DNODEs wakeup one time (about 3ms) every cycle, other times are in sleep.For multimedia nodes (MNODE), the multimedia codec and related modules are closed down when MNODEs are not in multimedia communication modes. MNODEs wake up every processing time period to send audio packet, other times are in sleep.

Address check in used to simplify data processing. When packet address is match, the packet is unpacked and handled; otherwise the packet will be abandoned. The processing time for receiving a packet of 32bytes is reduced to 7%of that complete reception. One or two bytes are added to packets. Though power consumption of
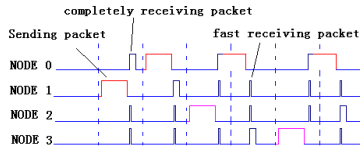
Fig. 1. Wireless packets processing diagram

single node increases slightly, power consumption of whole WSN reduces greatly. It is seen in Fig. 1.

3   Network Management

The network is startup by the center node, and the startup process is as following:

(1) CNODE sets frequency.
(2) Check the frequency.
(3) If the frequency is available, Startup the Network, go to (4).
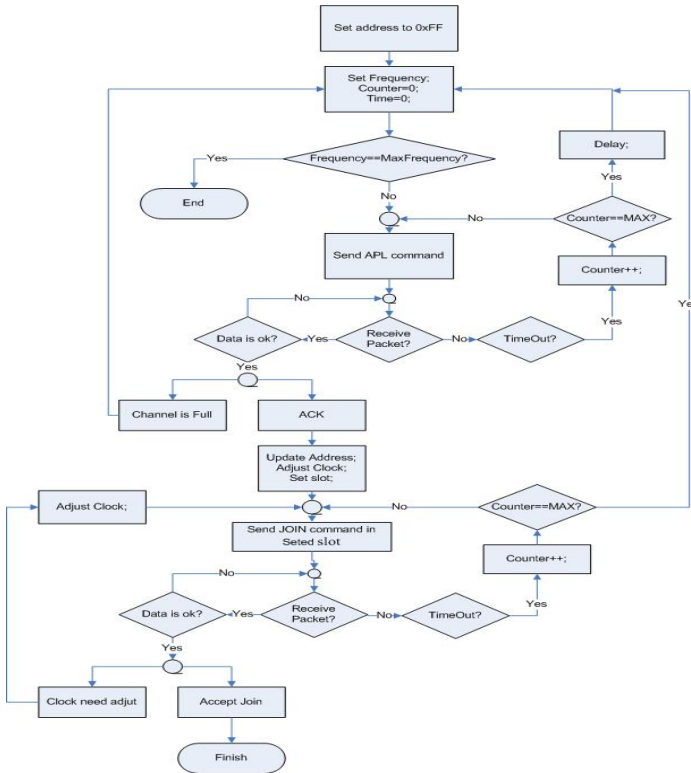
Fig. 2. Flow chart of client nodes join –client node part
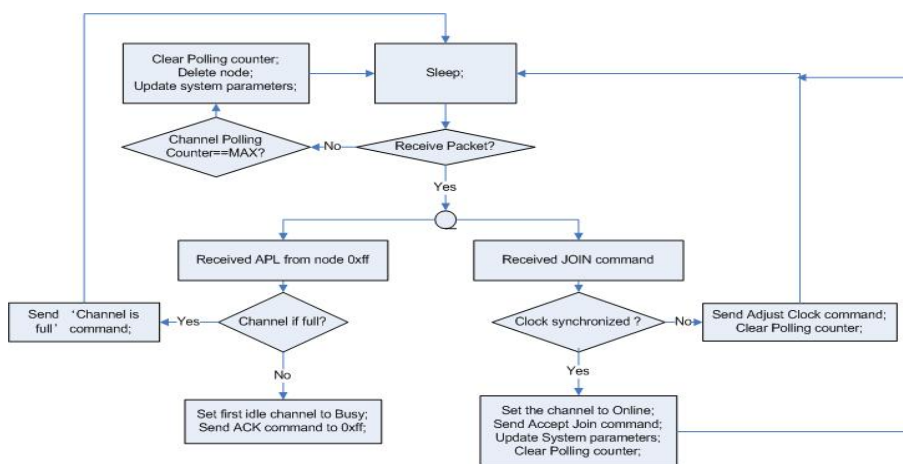
Fig. 3. Flow chart of client nodes join –center node part (simplified)

Else if all frequencies are used, go to (5), else update frequency, go to (1)

 (4) Main loop, waiting for RF packet and system event.
 (5) Halt application.

After the network started, client nodes can join the network, the join flow as shown in Fig. 2 and Fig. 3.

 In the slot-based system, the slot resource is limited, so it's important to withdraw the unused slots as soon as possible. The CNODE sending a polling command every 10 cycles (about 30 second), all client nodes should ACK the command, if Continuous three times no ACK is received, the node will be treated as offline and be deleted, the slot will be withdrew.

4 Hardware Platform

We have established an experimental platform(Fig. 4). This platform consists 6 audio nodes, 16 data nodes and 1 center nodes. MCU is MSP430f2618 for audio nodes and center node, MSP430f149 for data nodes. RF transceiver is CC2500. Audio codec is ADPCM 16kbps. The RF frequency is 2.4Ghz. RF speed is 500 kbps. Audio sampling period is 20ms, it consists of 8 time slots. The capacity is 3 pairs of real-time 2-way audio communications at the same time. Platform operation currents are shown in table 2. Average current 1 are the currents for audio communications and average current 2 are the currents when nodes are in standby or sleep mode. The platform working time is designed for more than half a year by using 3 AAA batteries for multimedia and center nodes and 2 AAA batteries for sensor nodes.
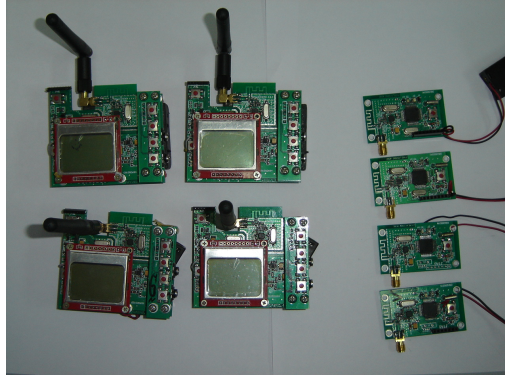
Fig. 4. Experimental WSN platform for real-time audio communication

Table 2. Current summary

nodes	MCLK	TX	RX	audio	average current 1	average current 2
Audio node	16Mhz	21.4mA	19.6mA	27mA	52mA	500uA
Data node	~1Mhz	21.4mA	19.6mA	-	22mA	300uA
Center node	16Mhz	21.4mA	19.6mA	27mA	53mA	600uA

5   Summaries

In conclusion, we have studied the network management and transmission protocol in detail. We also design and analyze a clock synchronization method for our application. At last we set up a wireless sensor network platform for audio communications. This platform is a low power, high performance and low cost system. It suggests many possible applications: such as emergency communication, short-range audio network and IOT.

References

[1] Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks 38, 393–422 (2002)
[2] Yick, J., Mukherjee, B., Ghosal, D.: Wireless sensor network survey. Computer Networks 52, 2292–2330 (2008)

[3] Akyildiz, I.F., Melodia, T., Chowdhury, K.R.: A survey on wireless multimedia sensor networks. Computer Networks 51, 921–960 (2007)

[4] Ma, H.-D., Tao, D.: Multimedia Sensor Network and Its Research Progresses. Journal of Software 9 (2006)

[5] Misra, S., Reisslein, M., Xue, G.: A survey of multimedia streaming in wireless sensor networks. IEEE Communications Surveys & Tutorials 10, 18–39 (2008)

[6] Akyildiz, I.F., Melodia, T., Chowdury, K.R.: Wireless multimedia sensor networks: A survey. IEEE Wireless Communications 14, 32–39 (2007)

[7] Zhu, X., Ying, S., Ling, L.: Multimedia sensor networks design for smart home surveillance. In: Control and Decision Conference, pp. 431–435 (2008)

[8] Kopetz, H.: Internet of Things. Real-Time Systems, 307–323 (2011)

[9] Gershenfeld, N., Krikorian, R., Cohen, D.: The Internet of Things. Scientific American 291, 76–81 (2004)

[10] Mock, M., Frings, R., Trikaliotis, S., Nett, E.: Clock synchronization for wireless local area networks. Euromicrorts, 183 (2000)

Application of the ZigBee Wireless Communication Technology on the Endless Rope Continuous Tractor Derailment Monitoring System

Li-Rong Wan, Yi Liu, Liang Wang, and Zhi-Hai Liu

College of Mechanical and Electrical Engineering, Shandong University of Science and Technology, Qingdao 266510, China
{Wanlr666,skdliuyi,6057079,zhihliu}@163.com

Abstract. During operation, the endless rope continuous tractor system is affected by the ups, downs and turnings of coal mine tunnel and other transportation conditions, shuttle car derailment accidents often occur. To solve significant personal injury and death issues caused by that the discovery of derailment is not timely, there is a great need for the development of derailment monitoring system in the endless rope continuous tractor system. This paper presents a implementation of derailment monitoring in endless rope continuous tractor system based on ZigBee wireless communication technology, it can achieve real-time monitoring of shuttle car running, detect shuttle car derailment timely, alarm and park, which is beneficial to ensure the safe running of the endless rope tractor.

Keywords: ZigBee, Endless rope continuous tractor, Derailment, Monitoring, Wirless communication.

1   Introduction

In coal mine auxiliary transportation system, the endless rope continuous tractor system is one of the main rail transportation, which has the advantages of fast transportation, long distance transportation and adaptation in to changes roadway. The endless rope continuous tractor system consists of winch, tensioning device, shuttle car, rear wheel, pressure wire wheels, tow rope wheels and steering (Fig.1). Winch is the power source of the endless rope continuous tractor, which is powered by a three-phase asynchronous motor, it transmits power to the winch drum by reducer or gearbox. The friction between the rope and the rope drum makes the rope moving, shuttle car and trucks are fixed on the rope and move with the rope to complete auxiliary transportation tasks.

During running of the endless rope continuous tractor system, the wire rope tension can reach dozens kN, and shuttle car derailment accidents often occur which is affected by the ups, downs and turnings of coal mine tunnel and other transportation conditions. If the discovery of derailment is not timely, it will cause the wire rope and tractor equipment damaged, even broken rope accident occurred, which will cause casualties. To solve significant personal injury and death issues caused by that the

discovery of derailment is not timely, there is a great need for the development of derailment monitoring to detect shuttle car derailment timely, alarm and park to ensure the safe running of coal mine auxiliary transportation system.
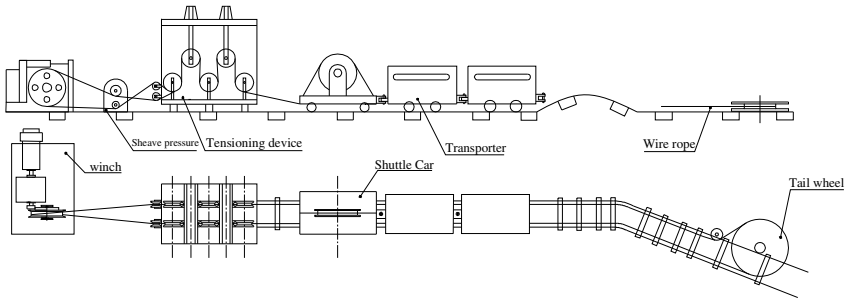
Fig. 1. Composation of Endless Rope Series Tractor system

2   Program Proposation of Derailment Monitoring in Endless Rope Continuous Tractor System

During running of the endless rope continuous tractor system, shuttle car moves along the tracks, and changes positions at all times. Derailment case will occur on any position in the orbit. So the derailment detection sensors must be installed in the shuttle car, and detecting signal can not be transmitted to signal processing module by fixed cable, which bring difficulties to the development of derailment monitoring system. In addition, the detecting sensors which are installed in the shuttle car can not be powered directly through the fixed cable**.** Therefore, we need a low-power, low cost solution. Based on the above two reasons, at present, there is not a complete solution in development of derailment monitoring of the endless rope continuous tractor system at home and abroad. Through comprehensive analysis, combined with site-specific conditions, we use wireless signal transmission technology to solve transmission problems of derailment detection sensor signal.

3   Analysis of Wireless Signal Transmission Technology

The current industry standards for wireless networks are the following:

(1) Bluetooth: Bluetooth is a short-range wireless communications technology standard. Bluetooth uses distributed network structure, and fast frequency hopping and short packet technology, supports point to point and multipoint communications. It works on global 2.4GHz, and the data rate is 1Mbps. Bluetooth technology is mainly applied for three categories: voice / data access, Internet of peripherals, and personal area network.

(2) Wi-Fi: The so-called Wi-Fi, in fact, is another name of IEEE802.11b, it is a short-range wireless transmission technology, supporting the radio signals access within a few hundred feet. It works on global 2.4GHz, the biggest data rate is 11Mbps. The advantages of Wi-Fi technology are mature, high transmission

speed, network capability, wide coverage, but high power consumption, technical complexity, high cost of implementation are its' disadvantages.

(3) ZigBee: ZigBee is an emerging short-range wireless network technology which is low power, low data rate, low cost, low complexity. It is a wireless data-transmission network platform which is available to contain 65,000 more wireless modules and is very similar to the mobile communication CDMA / GSM network. Each ZigBee network data transmission module is similar to a mobile base station. In the context of the entire network, they can communicate with each other, and the distance between each network node can be from the standard 75 meters, to the extended several hundred meters, even a few kilometers. The key application areas include wireless data collection, wireless industrial control, consumer electronic, automotive automation, home and building automation, medical device control, remote network control and so on.

Through comprehensive analysis of ZigBee and other industries wireless communication technology, in low SNR environment, ZigBee has a strong anti-interference ability. In the same environment, its' anti-jamming performance is far better than the Bluetooth and WLAN. Comparison of several wireless communication technologies is shown in Table 1.

Table 1. Comparison of the performance of some wireless network standards

Market/ Name Standard	Wi-Fi^{TM} 802.11b	Bluetooth^{TM} 802.15.1	ZigBee^{TM} 802.15.4
Applications	Web,Email, Images	Cable replacement	Monitoring and Control
System resources	1MB+	250KB+	4-32KB
Battery life (/days)	0.5~5	1~7	100~1000+
Network capacity	32	7	255/65,500
Bandwidth (KB/s)	11,000+	720	20~250
Transmission distance (/m)	1~100	1~10+	1~100+
Advantages	Speed, Flexibility	Cheap, Convenient	Reliable, Low-power, Cheap

By comparing, we can see that the ZigBee wireless communication technology is particularly suitable for the coal mine endless rope continuous tractor derailment monitoring system with low power consumption, low cost, taking up less system resources, convenient networking, transmission distance and other advantages.

4   Design of Monitoring System Based on ZigBee

The coal mine endless rope continuous tractor derailment monitoring system is shown in Figure 2.

It is mainly composed by detection and transmission module, receiving and transmission module, receiving and alarm displaying module. The derailment detection module which is installed on the endless rope continuous tractor shuttle car detects the signal, then sends the signal to receiving and transmission module placed

on the wall where the tractor system on by ZigBee wireless transmission chip. Then, receiving and transmission module sends signal to display and alarm displaying module by wireless or wired communication, or sends warning signals by wiring. Receiving and alarm displaying module do module do the alarm display to prompt the operator to stop the running tractor, and automatically cuts off power to stop tractor's run and blocks, to prevent damages and impact derailed because of the vehicle derailment.
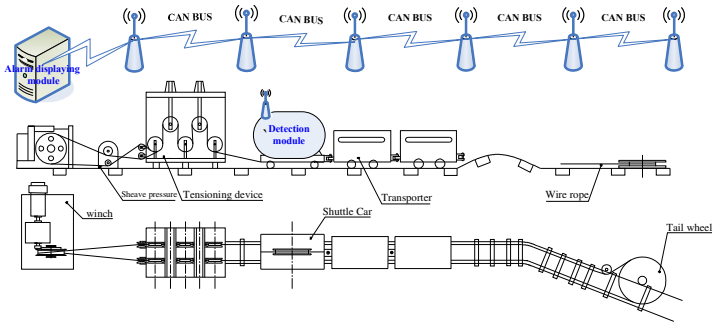
Fig. 2. Coal mine endless rope continuous tractor derailment monitoring system based on ZigBee technology

(1) Detecting and transmitting module: this module consists of proximity switches, wireless transmission devices, battery components. Two proximity switches are separately fixed on the corners near the two relative wheels of the shuttle car, and the sensor surface closes to the rail to induce whether wheels have derailed. So that regardless of what kind of derailment situation, there is always a proximity switch which can sense and output signals at least. Wireless sending unit is fixed on the shuttle car. Delaying a short time after it receives and pretreats the derailment signal sent by the proximity switch, it drivers ZigBee module to send alarm signals to the receiving and transmission module. Battery pack is installed on the shuttle car, providing power to the proximity switches and wireless transmission devices.(Fig.3)
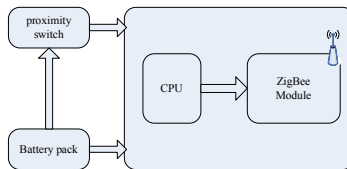
Fig. 3. Composition of detecting and transmitting module

(2) Receiving and transmission module: This module is composed by the wireless receiver, power supply module. Radio receivers which are installed on the wall of the roadway by regular intervals receive derailment alarm signal and battery power which are sent by the detecting and transmitting module. Then, the module sends the signal

to receiving and alarm displaying module via CAN bus or by wireless transmission. Power supply module is installed next to the wireless receiver to convert the AC power to DC power for the radio receivers device.(Fig.4)
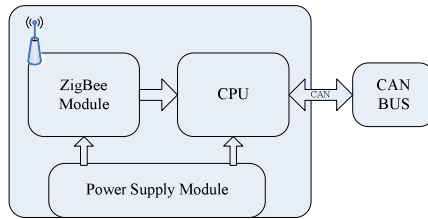
Fig. 4. Composition of receiving and transmission module

(3) Receiving and alarm displaying module: This module receive data signal sent from the receiving and transmission module by CAN-Bus. That analyzing the derailment situation determine the battery power after analysis of the data, then do corresponding alarm display processing.(Fig.5)
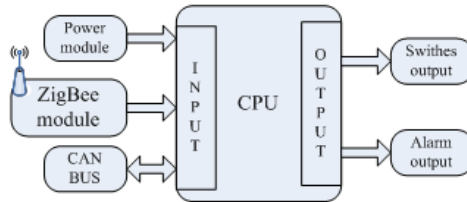
Fig. 5. Composition of receiving and alarm displaying module

5   Conclusions

That using ZigBee wireless communication technology on the coal mine endless rope continuous tractor derailment monitoring system can meet system for low power, low-cost wireless communications need. this system has the advantages of low-power, low cost, networking implicit, and anti-interference ability, Which is particularly suitable for rail transport system in the coal mine monitoring test. After application in a endless rope continuous tractor system on a mine working face, we verified the reliability and stability of the system. Such monitoring systems can also be used on other occasions of inspection and monitoring rail transport.

References

[1] Pang, H.: ZigBee Technology in Mine personnel management system (2010)
[2] Tang, Y.: ZigBee-based wireless data acquisition system design and implementation (2009)
[3] Yan, M.: ZigBee protocol stack analysis and design (2007)
[4] Wang, X.: Research and application of ZigBee wireless sensor network (2007)
[5] Zeng, Z.: Design of flow transmitter based on WirelessHART protocol (2010)
[6] Li, B.: Mine personnel Positioning System Status and Development (2008)

Author Index