Mingmin Zhang
Guest Editor

Transactions on
Edutainment XII

Zhigeng Pan · Adrian David Cheok · Wolfgang Müller
Editors-in-Chief

Springer

Lecture Notes in Computer Science 9292

Zhigeng Pan · Adrian David Cheok
Wolfgang Müller · Mingmin Zhang (Eds.)

Transactions on Edutainment XII

Springer

Editors-in-Chief
Zhigeng Pan
Hangzhou Normal University
Hangzhou
China

Wolfgang Müller
University of Education
Weingarten
Germany

Adrian David Cheok
City University London
London
UK

Guest Editor
Mingmin Zhang
Zhejiang University
Hangzhou
China

Editorial

This issue comprises four sections on different topics. In the first section on games, there are four papers. In the first paper, by Andreea Molnar et al., the authors present an interactive digital storytelling (IDS) formalization that captures the IDS logic constraints. This was implemented in a game engine and an authoring tool was created to facilitate the development of IDS-based games. In the second paper, by Qing Wang et al., a methodology is proposed for integrating knowledge with applied games combining recreational functions and pedagogic effects. For this purpose, a unified representation based on ontology is employed for game elements, game challenge models, and knowledge expression with which the relationship-mapping mechanism from diverse categories of knowledge of game atomic elements, especially primary rule, is constructed. In the third paper, by Haiying Zhao et al., the authors apply AR technology to develop an AR basketball smart phone game. Users can control the interaction of their counterparts in the virtual world through an image characteristics identification card, which is interactive and playable. The fourth paper, by Chenyang Cui, aims to articulate the cultural characteristics of digital games from three aspects: mass media, mass culture, international communication.

In the second section on human–computer interaction there are four papers. The first paper, by Shichang Feng et al., deals with the problem of finding 3D hand models corresponding to the user's 3D hand pose in the initial frames. Blending computer interaction techniques and cognition theories, a novel initialization approach for 3D hand models is put forward to make the initialization process more human-oriented. In the second paper, by Benyue Su et al., the authors develop a method for recognition of human daily actions by using a wearable motion sensor system. It can recognize 13 daily actions with the data in WARD1.0 efficiently. In the third paper, Hongpu Liu et al. propose an intelligent rehabilitation supporting system based on the Kinect device for the physical rehabilitation of stroke patients with apoplectic hemiplegia. In the fourth paper, by Hui Liu et al., the authors present the concept of a virtual interface, which narrows the effective area of hand activity to a specific region.

In the third section on image and graphics, there are five papers. In the first paper, Xiaoyan Li et al. propose a novel pixel-level change detection method based on the fuzzy distance for the interest building areas. Based on the imaging characteristics of synthetic aperture radar images in building areas, the variogram texture and gray feature are combined to form the feature vector in order to describe the image pixel feature clearly. In the second paper, by Jia Liu et al., a fast plant modeling method based on BP neural networks is proposed. The input is a plant image. Users can sketch the main branches and crown silhouettes on the image. Through branch copying, rotation, and adjustment, the 3D main branches are obtained. In the third paper, Jinhui Huang et al. implement the depth of the camera based on the actual situation of augmented reality systems with seamless integration of interactive design. In the fourth paper, by Xiangjiu Che et al., the authors present an image pre-processing tool for ant

colony optimization (ACO) with a distributed adaptive threshold strategy (DATS). They propose an interactive construction method of 3D objects from Chinese ink paintings for the challenging problem of generating the Chinese ink animation. The fifth paper, by Yunliang Jiang et al., presents a novel vertex mesh simplification algorithm based on curvature and distance metrics, using vertex curvature on the boundary type classification to get the alternative vertex of the folded edge through the LOOP subsidiary patterns.

In the fourth sections on applications, there are four papers. In the first paper, Jianwen Song et al. present the color analyzing and designing toolkits of the project product, which not only accords with the designer's intelligence and experience, but also proposes a new method for color analyzing and teaching. In the second paper, by Ruisong Ye, the chaotic nature of affine transformation is studied. The affine transformation is then utilized to generate pseudo-random number sequences and is applied to shuffle the host image during the preprocessing of watermarking. The watermarking is performed in the spatial domain, where the watermark bits are also encrypted by the skew tent map and then embedded in the shuffled host image. In the third paper, Wanxing Sheng et al. introduce a novel method – three-dimensional likelihood ratio test (3D-LRT) – to detect emergencies in large-scale power systems, which is a typical work that leverages spatiotemporal behavior characteristics to identify anomalous patterns in power systems. In the fourth paper, Shunting Wang et al. propose a method of generating birthday couplets fitting to gender, age, and birthday time. The birthday couplets generated by this system not only can fit the main features of birthday couplets, but can also embed the user's names.

<div align="right">Mingmin Zhang</div>

Transactions on Edutainment

This journal subline serves as a forum for stimulating and disseminating innovative research ideas, theories, emerging technologies, empirical investigations, state-of-the-art methods, and tools in all the different genres of edutainment, such as game-based learning and serious games, interactive storytelling, virtual learning environments, VR-based education, and related fields. It covers aspects from educational and game theories, human–computer interaction, computer graphics, artificial intelligence, and systems design.

Editorial Board

Daniel Thalmann   EPFL, Switzerland
Kok-Wai Wong   Murdoch University, Australia
Gangshan Wu   Nanjing University, China
Hyun Seung Yang   KAIST, Korea
Xiaopeng Zhang   IA-CAS, China

Contents

Interactive Digital Storytelling Based Educational Games: Formalise, Author, Play, Educate and Enjoy! - The Edugames4all Project Framework

Andreea Molnar[1(✉)] and Patty Kostkova[2]

[1] University of Portsmouth, Portsmouth, UK
andreea.molnar@port.ac.uk
[2] University College London, London, UK
p.kostkova@ucl.ac.uk

Abstract. In this paper we present an Interactive Digital Storytelling (IDS) formalisation that captures the IDS logic constraints. This was implemented in a game engine and an authoring tool was created to facilitate the development of IDS based games. We created several IDS based games using the proposed engine and authoring tool. They were used as case studies in this research. We also present some of the challenges we encountered once the games were implemented and how we addressed them.

1 Introduction

In recent years, Interactive Digital Storytelling (IDS) based games have been growing in popularity [1] and recognized as providing a motivating and engaging experience for players [2, 3]. Moreover, when storytelling is integrated in the game, it has been shown to help players make sense of their experience, improve their problem solving abilities, and ability to organise knowledge [2]. Due to their dynamical nature, they also provide the player with a better feeling of being able to control the game and how the story unfolds.

However, the design of interactive digital storytelling (IDS) [4] in particular, and of games in general [5] is increasingly complex. One of the challenges faced consists of a lack of "development of formalism supporting narrative description" [6]. As game story writers most often do not come from a computing background there is a need for both a logic design formalism for IDS based games and authoring tools to facilitate IDS creation. The first part of this research provides a solution to this problem by providing a platform that includes a Finite State Machine based formalisation of the IDS. The formalism is an important step in achieving a shared understanding of the field as it establishes shared requirements that can be used by future researchers. It can also be a step forward towards a fully automatic approach. The proposed formalism was used to develop a game engine.

This paper also presents the authoring tool that enables easier development of IDS based games. Furthermore, in this research we also investigated three challenges we

encountered once the games were implemented: high drop-out rate (despite a user centric design being followed through the implementation), usability and educational content assessment issues. We present our proposed solutions to this problems.

The rest of the paper is organised as follows. Section 2 presents the state of the art in this area and pins down where this research stands. The IDS FSM formalisation is defined in Sect. 3. Section 4 introduces the game engine implementation and how it connects with the authoring tool. It also presents examples of implementations using the authoring tool and a brief introduction into the edugames4all project. The following section presents the challenges encountered when evaluating the games. The next section discusses the results of this research and presents the work in progress. The paper ends with the conclusions presented in Sect. 7.

2   Related Work

The work presented in this paper spans across the following research domains: usage of artificial intelligence (AI) techniques in interactive digital storytelling, usage of interactive storytelling in educational games and educational content evaluation in educational games.

2.1   Figures Artificial Intelligence Techniques in Interactive Digital Storytelling

IDS is dynamic and allows players to act as they are the main characters in the story, to select among multiple paths, and basically to modify the story based on their actions. Despite the potential of a formalisation to "allow the deduction of results which throw a specific light at research applications" [7], there is little research available in this domain on the development of a formalization for the narrative description [6].

AI techniques have been used in IDS mostly to adapt the narratives to user's actions [8] or in the implementation of different aspects of storytelling generation. Among these techniques one can count case based reasoning [9], genetic algorithms [10], goal net and fuzzy cognitive maps [11], natural language processing [12], search algorithms [13], and FSM [14]. FSMs have been increasingly used for modeling dynamic systems [14] and in this research we use them for the formalisation of the IDS.

The research done so far on formalising IDS was based on logical reasoning [6, 15] and constraints [16]. Constraint planning aims to generate a good or optimal solution given a set of possible goals [17]. Our aim is not to determine the optimal but to allow users the possibility of exploring as many solutions as they want as long as given the path they are taking there is a way to reach a resolution. Although many aspects of the IDS can be formalised through logical reasoning and constrains these are not necessarily suitable for all aspects of IDS as shown by [7]. We propose to add to the state of the art in the field and to use an AI based approach formalisation based on Finite State Machines (FSM).

FSM allows an expressive representation and it allows modelling the IDS states (FSM states) and user actions (the transition function). It allows separation between the

states and the transition function that could lead to the implementation of an authoring tool where the game designers could define the game states and the transition function could be implemented separately from the game engine. This allows the generation of games that make use of interactive storytelling without the game designers needing to get into the game engine code. The formalism allows to create a sequence of causal events, recombine events in the story and determine whether a sequence of events leads towards a resolution of the story.

2.2 Usage of Storytelling in Educational Games

Digital storytelling is said to help in acquiring knowledge, improving media literacy and user attitude and behaviour [18]. When used in educational games it provides a "gateway to immersive experience" [19]. It has been shown to engage players [2], improve the learning experience by allowing the players "to make sense of experience, organise knowledge, sparking problem-solving skills and increase motivation" [2]. The story provides a context to learning [20], while the interaction actively involves the players in the building of the story [2]. This stimulates imagination and curiosity and leads to unintentional learning [2]. Despite their educational potential [21, 22] most studies have been focused on the simulation aspect neglecting the storytelling [23]. This study addresses this gap by focusing on the storytelling aspect of educational games.

2.3 Evaluation of the Educational Content Integrated in Games

Assessment is an essential part of any educational experience. Traditionally, assessment of learning has taken place outside of the actual game. Pre and post-play assessments of knowledge remove the task of assessment from the educational game itself and allows for a summating assessment of player knowledge and/or attitude change. Although the evaluation is considered to be one of the most important parts in educational games [24], little research has been dedicated to seamlessly integrating the assessment in the game ow. Most of the assessment is done as a separate part of the educational game, or when integrated it is often done in an intrusive manner affecting the game ow and the player experience [25]. Among the studies that proposed an assessment integrated as part of the game play one can count: [26–30].

[26] integrate multiple choice questions as a part of a point-and-click game-like simulation. The player is provided feedback if the answer provided is wrong. The game was assessed in terms of the educational potential, but the integrated assessment and the way it was perceived by students was not evaluated.

[27] propose a similar approach to [26] in which questions are being asked in a simulation like game. The assessment is designed to be as much as possible similar to the traditional pen and paper alternative, and the students get feedback only at the end of the evaluation. In our research, we want to have the assessment integrated as seamlessly as possible in the game so that the player immersion is not affected. Moreover, our proposed approach allows for feedback to be provided immediately after the assessment, if desired, and this has the potential to improve learning achievements [31, 32].

EducaMovil [28] is a suite of open source mobile learning games, through which snippets of educational content and quizzes are integrated. A certain action of the game triggers a certain lesson to open, and a question is given to the learner after the educational content snippet is shown. Feedback, either positive or negative is offered immediately after the lesson is finished. The authors showed that game performance could be used as a proxy for measuring student classroom performance, on a study performed with second level students learning mathematics.

[30] propose the assessment to be integrated through a quiz, similar in style to "How to be Millionaire". They used as a case study a platform game, where the quiz is inserted between the levels of the game. In this game, a virtual character asks questions and the player has to choose among three options as answers [33]. Feedback is given after the results are submitted. A total of 63 % of the participants involved in the study reported preferring this type of assessment as opposed to taking a "traditional quiz" [30]. Moreover, the integrated assessment has not shown to affect the participants enjoyment while playing the game [30].

In Crystal Island [29] a narrative game similar to our game, players are asked to fill quizzes and worksheets through the game. This study does not consider the students perspective on the integrated assessment and whether this is more effective than when the assessment is not integrated.

There are several aspects in which our research is novel, compared to the previously mentioned ones. As oppose to previous research we are focusing on formalisation and we show how an integrated assessment can also be included as part of the formalisation.

3   IDS Formal Notion

This section will first introduce the five usual layers of interactive storytelling Presentation Layer, Conversation layer, Quest Layer, Mission Layer, and Education Layer and afterwards introduces the formalisation of an IDS as FSM. The layered division dawns from the work done by [34] but expands the notion to cover educational remit and designs the actual FSM formalism. A detailed definition of each IDS layer and how they map into the FSM is described below. In doing so, we follow a bottom up approach in which the lower layers are described first (Fig. 1).

3.1   Presentation Layer

The Presentation Layer contains the assets/animations needed to deliver the interactive story. It consists of images and animations for characters, rooms, items, in the game, and the motion models that are used to describe how the virtual characters move or behave. They form the interactive story animation function.

Hotspots. Hotspots are places where the player can click and invoke an action.

Definition 1 (Hotspot). *Let H be a finite non empty set of hotspots, a hotspot, $h \in H$ is defined as a tuple:*

   $h = (hotspotImage, action),$
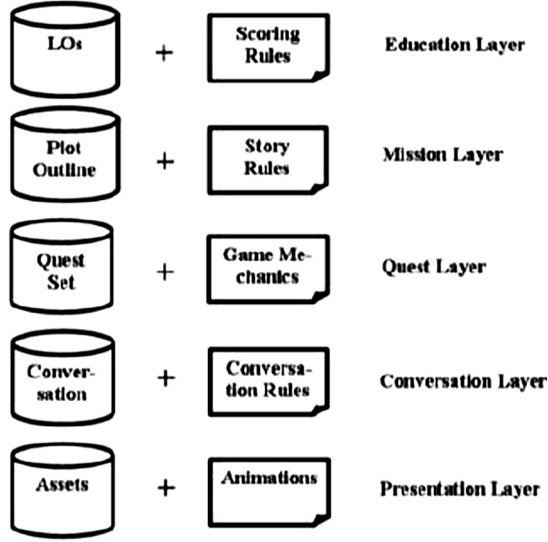
Fig. 1. Interactive storytelling layers

where the hotspotImage is a clickable area in an image, and action is the action to be performed as a result (see Definition 7 for more details about game actions).

Item. An item is a virtual object that can be collected by the player. The item collection is triggered in response to certain actions performed by the player.

Definition 2 (Item). *Let I be a finite non empty set of items, an item, i ∈ I, and is defined as a 3-tuple:*

$$i = (itemName,\ itemImage,\ hint),$$

where itemName is a text representing the name of the item, itemImage is a thumbnail image of the item representation, and hint is text to give the player clues about the item (it could be nil).

Rooms. Rooms are specific locations in the story. They could be interior rooms, outdoor areas, or sections of a room/outdoor area.

Definition 3 (Room). *Let R be a finite non empty set of rooms, a room, r ∈ R is defined as a 4-tuple:*

$$r = (roomImage,\ thumbnailImage,\ roomName,\ \{hotspot1,\ ...,\ hotspothr\}),$$

where {hotspot1,..., hotspothr} ∈ H, hr ∈ N, hr ≥ 0 defining the number of hotspots in room r, roomImage is the location image and thumbanilImage is roomImage in thumbnail. The thumbanilImage can be nil.

Characters. Characters are the various story personalities and the player's character. The virtual characters, agents, help the users to accomplish a task [35]. In an educational IDS the virtual characters can act as pedagogical agents by helping the player increase his effectiveness by delivering customised advice [36]. This definition is not targeted towards antagonistic characters.

Definition 4 (Character). *Let C be a finite non empty set of characters, CN a finite set of conversation nodes (see Definition 5), a character, $c \in C$, is defined as a 5-tuple:*

$$c = (name, \ characterImage, \ thumbnailImage, \ cn_c, \ presentCN_c),$$

where $\{cn_c, presentCN_c\} \in CN$, characterImage is the image representing the character, thumbnailImage is characterImage just thumbnail format.
    The player is a special 'player character' $pc \in C$,

$$pc = (You, \ nil, \ nil, \ cnYou, \ nil).$$

3.2   Conversation Layer

The layer on top of the Presentation Layer is the Conversation Layer. In IDS, conversation provides the means of interaction and content presentation [37]. This layer consists of Conversation Nodes and Conversation Rules. A Conversation Node (CN) is a line of text or a sentence recited by a character.

    The Conversation Rules define which character is saying what, what should happen after the player has read this line of dialogue, and the context in which this happens. For example, a rule could be that a virtual character greets the player at the beginning of the story. Another rule dictates whether the virtual character should greet the player as soon as they meet or wait for the player to initiate conversation. This part has the power to generate adaptive conversations based on the player's previous actions.

Definition 5 (Conversation Node). *Let CN be a non empty set of conversation nodes, C a set of characters, then a conversation node, $cn \in CN$, is defined as a tuple:*

$$cn = (c, \ Text), \ where \ c \in C.$$

    The evaluation of the game LOs could be integrated in IDS as a part of the conversation [38]. In order to do so CNs through which the evaluation questions and feedback were delivered and three conversation rules need to be added as explained in [38]. Basically a question will consist of several CNs: the question (each question corresponds to a certain LO), the options the player has to select among and the feedback given (each selected option will trigger a different feedback). The rules are meant to ensure that if the player's prior knowledge is evaluated then the question has to be asked prior to the player being exposed to the LO while if the player is exposed to the game mechanics the post knowledge is evaluated. The rules also ensure that depending on the selected option the player will get the appropriate feedback. As scoring is also supported

by the engine the rules are also meant to ensure that depending on the option selected, the user score will increase/decrease with the appropriate amount. This enables real time feedback for the player.

3.3   Quest Layer

The *Quest Layer* contains the *Quest Set* and the *IDS Mechanics*. Quest in the context of this research refers to any story element of the game that requires activation (and later, deactivation) when certain conditions are met. In the context of the IDS FSM it represents a series of events (which could be conversation nodes or actions the player performs) visited according to the transition function and finished when certain conditions are met (Fig. 2). The *IDS Mechanics* component determines the operations of the game world and deals with the player interactions with the game.

Quest. A quest implies achieving a task, exploring a room, collecting evidence, etc. Quests can include other quests, and multiple quests could be opened at the same time.

Definition 6 (Quest). *Let T be a set of states, Q a set of states, t $\in$ T, is defined as a sequence, Q, of states of the IDS FSM starting at q_1 and finished by reaching state q_n following the transition function δ (see Definition 7),*

$$t = \{q_1, \ldots, q_n\}, \; 8 \; i \in N, \forall : q_i \to q_{i+1}$$

where, Q is a set of states, P be a set of game specific conditions, C a set of characters, R a set of rooms, {q_1, q_n} $\in$ Q are defined as the 4-tuples:
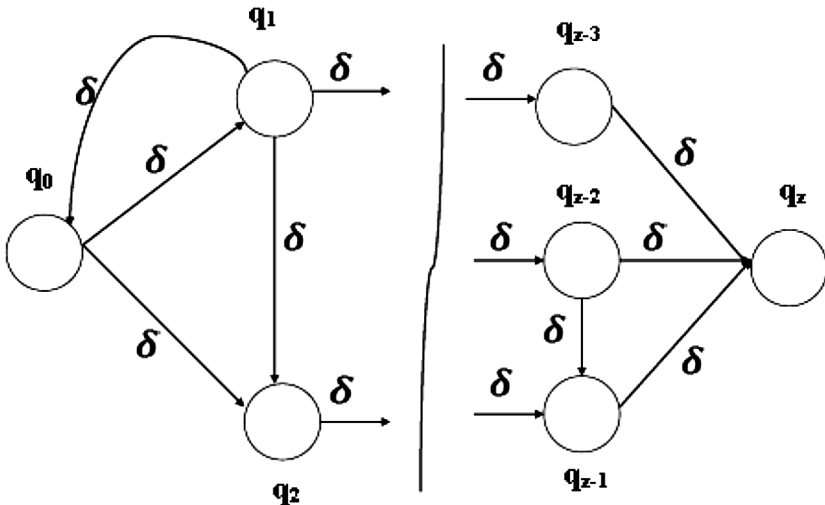
Fig. 2. Finite state-machine representation

$$q_1 = (p_1, r_1, c_1, cn_1),$$
$$q_n = (p_n, r_n, c_n, cn_n),$$

where $\{p_1, p_n\} \in P$, p_1 includes t is open, p_n includes t is closed, $r \in R$, $\{c_1, c_n\} \in C$, and $\{cn_1, cn_n\} \in CN$, and q_1 is an opening quest state and q_n is a closing quest state.

Action. An action is the transition function δ defined as below:

Definition 7 (Transition Function δ (Action)). *Let P be a set of game specific conditions, R a set of rooms, C a set of characters, and CN a set of conversation nodes. The transition function (game action), δ, is defined as:*

$$\delta : (p, r, c, cn0) \rightarrow (p', r', c', \{cn_1, \ldots, cn_i\}),$$

where $\{p, p\} \in P$, p representing a set of pre-conditions for an event to occur, while p' a set of post-conditions for the event to finish, $\{r, r'\} \in R$, where r can be the same as r', $\{c, c'\} \in C$, where c can be the same as c', and $\{cn_0, \ldots, cn_i\} \in CN$ and $\{cn_0, \ldots, cn_i\}$ is a set of conversation nodes that follow after cn_0, typically i = 1 but it can be empty.

An action could lead to a change in the room the player is in, displaying multiple conversation nodes, etc. A pre-condition to a game action could be that the player clicks on a hotspot. This leads to a change in the FSM states (Fig. 2). Unlike in generic FSM in IDS, the δ is defined by several predefined actions implementing the IDS mechanics. Several predefined actions were implemented by the IDS engine:

Let P be a set of IDS specific conditions, $\{p, p'\} \in P$, p representing a set of pre-conditions for an event to occur, and p' a set of post-conditions, R a set of rooms, $\{r, r'\} \in R$, C a set of characters, $\{c, c'\} \in C$, and CN a set of nodes, $\{cn_0, \ldots, cn_i\} \in CN$.

– enter room r:

$$\delta : (p, r, c, cn_0) \rightarrow (p', r', c', \{cn_1, \ldots, cn_i\}),$$

where $r \neq r'$, p and p' can be the same state of conditions, c and c' can be the same character, $cn_0, \ldots, cn_i$ can be equal.
– put character c in room r':

$$\delta : (p, r, c, nil) \rightarrow (p', r', c, nil),$$

– start conversation cn_c

$$\delta : (p, r, c, nil) \rightarrow (p', r', c, cn_c),$$

– continue conversation cn_{i-1}, where the transition function is:

$$\delta : (p, r, c, cn_{i-1}) \rightarrow (p', r, c', cn_i)$$

– end conversation cn_c:

$$\delta : (p, r, c, cn_c) \rightarrow (p', r', c, nil),$$

– open quest t:

$$\delta : (p, r, c, cn_0) \rightarrow (p', r', c', \{cn_1, \ldots, cn_i\}),$$

where p' is a set of conditions including t open, r, r' can be the same room, c and c' can be the same character, $cn_0, \ldots, cn_i$ can be equal.
– close quest:

$$\delta : (p, r, c, cn_0) \rightarrow (p', r', c', \{cn_1, \ldots, cn_i\}),$$

where p' is a set of conditions including t = close, r, r' can be the same rooms, c and c' can be the same character, $cn_0, \ldots, cn_i$ can be equal.

3.4   Mission Layer

The top level layer contains the overall dramatic outline [34]. The story mission is an ultimate quest starting with the game initiation state and finishing when the story is finished by reaching the finish state. It has the highest level of abstraction of the interactive storytelling.

Definition 8 (Game Mission - Finite State Engine). *Let Q be a set of states, P be a set of game specific conditions, R a set of rooms, C a set of characters, and CN a set of nodes. An IDS FSM Game Mission is defined as a 4-tuple:*

$$(Q, \delta, q_0, q_z),$$

where $\{q_0, q_z\} \in Q$, $q_0 = (p_0, r_0, c_0, cn_0)$ is the initial state - the start of the mission quest, and $q_z = (p_z, r_z, c_z, cn_z)$ is the final state - the end of the mission quest; $\{p_0, p_z\} \in P$, $\{r_0, r_z\} \in R$, $\{c_0, c_z\} \in C$, $\{cn_0, cn_z\} \in CN$.

A FSM also defines an alphabet, Σ. In our case the alphabet is included in the states, rooms, and characters.

3.5   Educational Layer

The Education Layer consists of educational content (LOs that has to be integrated in the game) and *Scoring Rules* (how much credit would the player get for answering a certain question or performing a certain action in the game).

The LOs contain a high level description of the LO delivered through the game. For example a LO could be: One should only use antibiotics with a doctor's permission, whereas the CNs contextualize the LOs for the given story. The Scoring Rules consist of rules describing how the LO evaluation, contributes to the player's score.

Fig. 3. The connection between the authoring environment, engine and IDS generation

4   Game Engine and Authoring Environment

Based on the proposed formalisation we design and implement a game engine. Authoring the IDS game in the FSM formal notation can be difficult. For example, a definition of Alyx, a *character* who says *Hello* and continues saying *Welcome to edugames4all* is noted according to Definition (4) A = *(Alyx, image of Alyx, thumbnail image of Alyx, (Alyx, Welcome to edugames), (Alex, Hello))* and requires to express the logical constrains manually. To address this we developed the authoring environment to ease the burden.

With this in mind, the IDS FSM δ action is implemented in the game engine while the authoring tool lets one define the states of the IDS (Fig. 5). The editor generates files used by the engine to create the IDS. Each set from the IDS FSM is represented by one or more worksheets in the authoring user interface. This allows a more modularised design, the defined semantics becoming more clear and simplifying the game designer's job. The worksheets for each of the layers are:

– Presentation Layer: Rooms, Hotspots, Items, Characters
– Conversation: Conversation Nodes
– Quest: Game Actions, Quests
– Mission: Initial State.

Room Id	Room Image	Thumbnail Image	Room Name	Hotspots
1	ebug_hq.swf	ebug_hq_thumb.swf	e-Bug HQ	
2	helllab.swf	helllab_thumb.swf	e-Copter	
3	barbeque.swf	barbeque_thumb.swf	Barbeque	Punch Table - Dining Table - Activate Microbe View
4	punch_table.swf		Punch Table	Back to Barbeque - Salad Bowl - Punch Bowl - Activate Microbe View
5	dining_table.swf		Dining Table	Back to Barbeque - Hugh's Plate - Activate Microbe View
6	kitchen.swf	kitchen_thumb.swf	Kitchen	Sink Area - Chopping Board - Activate Microbe View
7	kitchen_sink.swf		Kitchen Sink	Back to Kitchen - Temperature Guage - Soap above Sink - Activate Microbe View
8	bathroom.swf	bathroom_thumb.swf	Bathroom	Portrait - Pills - Toilet - Toilet Area - Door Area - Sink - Mirror - Activate Microbe View
9	bathroom_door.swf		Bathroom Door	Back to Bathroom - Door Handle - Activate Microbe View
10			MV e-Copter	
11	mv_barbeque.swf		MV Barbeque	Punch Table in MV Mode - Dining Table in MV Mode - De-activate MV Mode
12	mv_punch_table.swf		MV Punch Table	Back to Barbeque in MV Mode - Salad Bowl in MV Mode - Punch Bowl in MV Mode - De-activate MV Mode
13	mv_dining_table.swf		MV Dining Table	Back to Barbeque in MV Mode - Hugh's Plate in MV Mode - De-activate MV Mode
14	mv_kitchen.swf		MV Kitchen	Sink Area in MV Mode - Chopping Board in MV Mode - De-activate MV Mode
15	mv_kitchen_sink.swf		MV Kitchen Sink	Back to Kitchen in MV Mode - Temperature Guage in MV Mode - Soap above Sink in MV Mode - De-activate MV Mode
16	mv_bathroom.swf		MV Bathroom	Painting in MV Mode - Pills in MV Mode - Toilet in MV Mode - Toilet Area in MV Mode - Door Area in MV Mode - Sink in MV M
17	mv_bathroom_door.swf		MV Bathroom Door	Back to Bathroom in MV Mode - Door Handle in MV Mode - De-activate MV Mode

Fig. 4. Rooms worksheet

Each set is represented in the worksheet with all elements defined as in the IDS FSM. For example, the room has a room image (*roomImage*), a thumbnail version of the room image (*thumbnailRoomImage*), the name of the room (*roomName*) and a set of hotspots (*{hotspot1,…, hotspothr}*). Figure 4 shows the Rooms worksheet. Each room has its own id, a .swf file containing the room image (the generated IDSes are Flash based), a thumbnail image (the thumbnail image is not mandatory and can be ignored if there is no use for it during game play), the room name, and a set of hotspots that link to the *Hotspots* worksheet. All the rooms used in the game are defined in this sheet, and when a room is cross-linked from another worksheet it is done by using the id column.

To give some examples on how these are cross-linked, the Hotspots worksheet has references to the actions triggered when a player will click on the hotspot, therefore the action is cross-linked in the Hotspots; the Actions worksheet cross links *Rooms*, *Items* and the *Conversation* Nodes worksheet, in order to specify different actions (e.g. travel between rooms, collecting evidence, starting a new conversation). The main file of the authoring process is the *Quest* worksheet, which is entirely dependent on the others.

Figure 5 describes the connection between the engine and the authoring environment. Practically every IDS has different assets, either the images, animations, sound files, etc that are being used to create the IDS. These, together with the IDS ow are described in the authoring editor.

There are several advantages of separating the IDS FSM states (defined in the authoring tool) from the transition function (implemented by the engine). First, designers do not need to modify the engine but can instead focus on defining the user experience in the authoring environment. Second, after the initial investment of building the data-driven engine, any new IDS developed this way requires less effort and investment of time. It is also easier to modify since just the files created with the authoring tool need to be changed. Moreover, from an internationalisation and localisation perspective it is more exible. For example, when translating the dialogue, the translator needs to be given only the file containing the conversation nodes. Third the FSM focuses on the progression rather than on levels, leading to less inconsistency than when the focus is solely on the levels.

The engine and authoring environment were used to create five different IDS based games that can be found at edugames4all. The games we created were constructed with the aim of creating awareness about issues related to health in an enjoyable manner.

The translation and localisation were done using the authoring tool by public health specialists and a microbiologist from the country of the language in which the story was translated. None of the public health specialists or microbiologists had any computing background. The games and the translations were created entirely using the authoring tool and the generated files were given to the game engine at run-time to generate the games.

The games are part of the edugames4all project which has followed e-Bug, the project under which some of the games were originally created. e-Bug was an European project in which researchers and practitioners from 18 European countries were partially funded under The Directorate-General for Health and Consumers (DG SANCO) of the European Commission [39]. The project aim was to develop and disseminate educational resources that teachers would use to teach junior and senior school children [40]. With this aim an EU wide survey was distributed to assess children knowledge of hygiene to set the LOs at a common denominator [41]. Two types of games were created from LOs found as a result of the survey: a platform game for junior kids [42] and an IDS based game for 13–15 years old [21].

5   Challenges and Proposed Solutions

As the previous sections addressed the formalisation and authoring of IDS based games, this section presents some of the challenges encountered when the games were already implemented and how we addressed them. These are grouped in three sections addressing the drop-out rate, game usability assessment and educational content evaluation.

5.1   Drop-Out Rate

Drop out rates are a known issue for computer games as reported by academic studies of educational games [43] as well as in industry settings [44]. In our case, IDS based games were played by over 60.000 users in 2010 [45]. However in all cases when the games were introduced in the classroom it was noticed that there was a dropout in the number of students that finish playing relatively to those that started [46]. One of the problems was that not all the students played a similar game before. A text tutorial was integrated in the game but was rarely consulted by the children. To help alleviate this problem, another mission called Training Mission [46], was implemented to teach the students the mechanics of the game. The mission is shorter than the other missions and a lot of guidance is provided by the non-player character. The player is shown how to use travel between different locations, how to collect and analyse evidence, and the usage of microvision, elements pertinent to solving the mystery in the rest of the missions (Fig. 3).

The usability of this mission was initially tested with 49 students at a summer school and assessed through five open questions targeting the mission graphics, easiness to follow directions through the game, willingness to continue playing and improvements to be made [46]. Most of the participants liked the graphics, found the
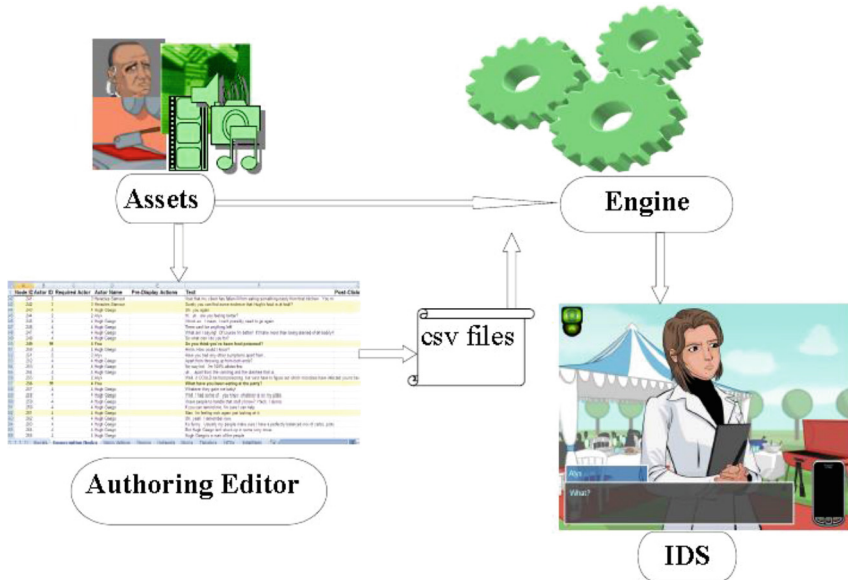
Fig. 5. The virtual character providing guidance

Training Mission easy to play and they wanted to continue playing. A second study of the Training Mission took place with 15 students divided into two groups. One group played one of the missions without the Training Mission beforehand and the other one played the Training Mission and afterward a mission of the game. The researcher taking part in the study observed the children and reported that the children who played the Training Mission beforehand asked less for help and more students completed the game in this scenario as opposed to the previous one [46]. A second evaluation was performed to determine whether playing the training mission before playing the other missions affects the usability of the other missions [47]. Although the results showed that when playing the training mission the perceived usability of the other missions is increased, no statistically significant difference was obtained except for the fact that the children perceived the missions as less awkward when they play the training mission beforehand.

5.2 Game Usability Assessment

System Usability Scale (SUS) [48] is a questionnaire that has been extensively used in testing usability. Because the SUS questionnaire has been designed for analysing systems in general, the questionnaire was adapted to our situation by changing the word "system" to "game" and "use" with "play" (see Table 1 for the full modified questionnaire). Replacing similar terms in a SUS questionnaire has been found not to affect its reliability and validity [49, 50].

Table 1. SUS questionnaire adapted for the game

No.	Question
1	I think I would like to play this game frequently
2	I found the game unnecessarily complex
3	I thought the game was easy to play
4	I think I would need the support of a teacher or other expert to be able to play this game
5	I found various functions in this game were well integrated
6	I thought there was too much inconsistency in this game
7	I would imagine that most people would learn to play this game very quickly
8	I found the game very awkward to use
9	I felt very confident playing the game
10	I needed a lot of help before I could get to play this game

SUS is a short survey, consisting of 10 questions. It indicates the usability of a system in general, and a game in this particular case, on a scale from 0 to 100, where 100 is the best score that can be obtained. For each question the subject has to choose an option on a 5 point Likert scale to indicate the degree of agreement or disagreement with the statement in the question. In this case 1 stands for Strongly Disagree and 5 stands to Strongly Agree. To compute the SUS score, the following steps are required. For the odd questions one has to subtract 1 from the option the subject chooses on the Likert scale, while for the even questions the chosen scale position is subtracted from 5. In this way each of the questions contributes with a value from 0 to 4 to the final score. The sum of the results computed in this way is multiplied by 2.5 and divided by the number of subjects. The number obtained in this way is the SUS score. The SUS score for our game is 66.78 out of 100 which is considered to be a Good usability score [49].

As SUS is just a measure of quantifying the usability and not a diagnostics tool, several other questions were added in the survey to determine areas for improving the game and when the game was played in schools, focus groups with the children and the teachers took place to further drill down on the issues that need improvement (some of the teachers were eager to play the games themselves and they also provided feedback).

Fig. 6. Pre-knowledge evaluation

In the survey the players were asked about having difficulties on specific parts of the game and most of them reported on not having any. When asked what they liked most about this game, the players said: the educational nature of the game, characters, that it was funny, intuitive etc. Among the parts disliked was the fact that there was too much text to read, the slow curve to learn the game, the story etc. There were some comments to which the players answers contradict, some subjects disliking that the game is too short while others consider it is too long. When asked in the survey, as a separate question if they consider the length of the game appropriate, 85% considered it was the appropriate length.

5.3   Assessing Educational Potential

Assessment of educational content integrated in the game is unpopular with game users and especially children. Seamless evaluation was implemented to overcome this issue. The questionnaire assessing the knowledge assessment was basically integrated in the IDS story in such a way that most of the players did not realise they were being evaluated [38]. The questions are asked at different stages of the game attention being paid that a first round of questions is asked before the player is exposed to the learning objectives and another one after the player is exposed to the game mechanics. For example, we integrated this questionnaire in one of the missions, having one of the virtual characters ask the player questions during the investigation - questions that will be part of the pre-test of the players knowledge, (Fig. 6) and the post-test questions were asked when the player returns to the agency during the debriefing (Fig. 7). This allowed

Fig. 7. Post-knowledge evaluation

us to assess the players knowledge without affecting the game experience and also without asking the children to complete a questionnaire when playing the game.

6 Discussions and Work in Progress

Although a user centred design was followed through the game's design and implementation, where teachers and students were consulted from the initial stages of the project and during the implementation to iteratively improve the games [51] a high drop out rate was noticed. One reason for the drop out rate were the students low reading skills. Due to the high textual content of the game, sometimes the children found it tiring reading through the game. We are currently working on also having a version of the game that will contain more audio, the characters being able to speak to the user through voice rather than through text as it is currently implemented.

We are planning on adding gamification elements to the website hosting the games, such as league tables with the players that get bigger scores in the game. We want the children to be able to easily share their score in the game with others. We hope that this will also help with the retention in playing the games and that the children will attempt to improve their scores, scores that are computed by taking into account the LOs evaluation. Moreover we are currently working on having the games available on mobile phones as mobile apps.

Another challenge that is undergoing is setting the right level of complexity in the game not only by adjusting the player knowledge but also the game mechanics. We want to determine during the game play if the player needs help without him requesting it. Player knowledge is difficult to assess in an online game, played on-line by children from different countries. We addressed this by assessing the LOs across European countries to be able to cover the core LOs. Moreover, the game has some level of adaptation based on the previous user knowledge (e.g. depending on the LOs displayed in a certain part of the game, the LOs in some other part of the game may or may not be displayed). Currently we are working on improving the knowledge adaptation.

7 Conclusions

IDS based games have been growing in popularity both in entertainment but also as a means of delivering educational content in an enjoyable manner to children. As game story writers do not always come from computing backgrounds, there is a need for a user-friendly authoring tool to help in alleviating the authoring process. In this paper we presented how IDS can be formalised using FSM and a game engine was implemented based on this formalisation. The game engine takes input files generated by an authoring tool. This engine and authoring tool were used in implementing several IDS based games as part of the e-Bug/edugames4all project with a number of authors in different countries who used the tool for localisation and translation. We also presented a standard solution to three common game play challenges encountered when evaluating the games.

References

1. Park, N., Min Lee, K., Annie Jin, S.A., Kang, S.: Effects of pre-game stories on feelings of presence and evaluation of computer games. Int. J. Hum. Comput. Stud. **68**(11), 822–833 (2010)

2. Hodhod, R.A., Cairns, P.A., Kudenko, D.: Innovative integrated architecture for educational games: challenges and merits. In: Pan, Z., Cheok, A.D., Müller, W., Yang, X. (eds.) Transactions on Edutainment V. LNCS, vol. 6530, pp. 1–34. Springer, Heidelberg (2011)

3. Lee, S.Y., Mott, B.W., Lester, J.C.: Director agent intervention strategies for interactive narrative environments. In: Si, M., Thue, D., André, E., Lester, J., Tanenbaum, J., Zammitto, V. (eds.) ICIDS 2011. LNCS, vol. 7069, pp. 140–151. Springer, Heidelberg (2011)

4. Smith, G., Anderson, R., Kopleck, B., Lindblad, Z., Scott, L., Wardell, A., Whitehead, J., Mateas, M.: Situating quests: design patterns for quest and level design in role-playing games. In: Si, M., Thue, D., André, E., Lester, J., Tanenbaum, J., Zammitto, V. (eds.) ICIDS 2011. LNCS, vol. 7069, pp. 326–329. Springer, Heidelberg (2011)

5. Lewis, C., Whitehead, J.: The whats and the whys of games and software engineering. In: Proceedings of the 1st International Workshop on Games and Software Engineering, New York, NY, USA, pp. 1–4. ACM (2011)

6. Cavazza, M., Champagnat, R., Leonardi, R.: The IRIS network of excellence: future directions in interactive storytelling. In: Iurgel, I.A., Zagalo, N., Petta, P. (eds.) ICIDS 2009. LNCS, vol. 5915, pp. 8–13. Springer, Heidelberg (2009)

7. Jantke, K.P.: Logical formalization and reasoning for computerized interactive storytelling. In: 2010 IEEE International Conference on Progress in Informatics and Computing (PIC), vol. 2, pp. 851–857. IEEE (2010)

8. Magy, S.E.N., David, M., Tony, M.: Experiencing interactive narrative: a qualitative analysis of façade. Entertainment Comput. **4**(1), 39–52 (2013)

9. Peinado, F., Gervás, P.: Transferring game mastering laws to interactive digital storytelling. In: Göbel, S., Spierling, U., Hoffmann, A., Iurgel, I., Schneider, O., Dechau, J., Feix, A. (eds.) TIDSE 2004. LNCS, vol. 3105, pp. 48–54. Springer, Heidelberg (2004)

10. Ong, T., Leggett, J.J.: A genetic algorithm approach to interactive narrative generation. In: Proceedings of the 15th ACM Conference on Hypertext and Hypermedia, HYPERTEXT 2004, New York, NY, USA, pp. 181–182. ACM (2004)

11. Cai, Y., Miao, C., Tan, A.H., Shen, Z.: Fuzzy cognitive goal net for interactive storytelling plot design. In: Proceedings of the 2006 ACM SIGCHI International Conference on Advances in Computer Entertainment Technology. ACM, New York, NY, USA (2006)

12. Mateas, M., Stern, A.: Natural language understanding in façade: surface-text processing. In: Göbel, S., Spierling, U., Hoffmann, A., Iurgel, I., Schneider, O., Dechau, J., Feix, A. (eds.) TIDSE 2004. LNCS, vol. 3105, pp. 3–13. Springer, Heidelberg (2004)

13. Cavazza, M., Lugrin, J.L., Pizzi, D., Charles, F.: Madame bovary on the holodeck: immersive interactive storytelling. In: Proceedings of the 15th International Conference on Multimedia, MULTIMEDIA 2007, New York, NY, USA, pp. 651–660. ACM (2007)

14. Mahmud, N., Papadopoulos, Y., Walker, M.: A translation of state machines to temporal fault trees. In: Proceedings of the 2010 International Conference on Dependable Systems and Networks Workshops (DSN-W), DSNW 2010, Washington, DC, USA, pp. 45–51. IEEE Computer Society (2010)

15. Dang, K.D., Champagnat, R., Augeraud, M.: Modeling of interactive storytelling and validation of scenario by means of linear logic. In: Aylett, R., Lim, M.Y., Louchart, S., Petta, P., Riedl, M. (eds.) ICIDS 2010. LNCS, vol. 6432, pp. 153–164. Springer, Heidelberg (2010)

16. Porteous, J., Cavazza, M.: Controlling narrative generation with planning trajectories: the role of constraints. In: Iurgel, I.A., Zagalo, N., Petta, P. (eds.) ICIDS 2009. LNCS, vol. 5915, pp. 234–245. Springer, Heidelberg (2009)

17. Gerevini, A., Long, D.: Plan constraints and preferences in PDDL3. The Language of the Fifth International Planning Competition. Technical report, Department of Electronics for Automation, University of Brescia, Italy, vol. 75 (2005)

18. Di Blas, N., Paolini, P.: Digital storytelling and educational benefits: evidences from a large-scale project. In: Pan, Z., Cheok, A.D., Müller, W., Iurgel, I., Petta, P., Urban, B. (eds.) Transactions on Edutainment X. LNCS, vol. 7775, pp. 83–101. Springer, Heidelberg (2013)

19. Lim, T., et al.: Narrative serious game mechanics (NSGM) – insights into the narrative-pedagogical mechanism. In: Göbel, S., Wiemeyer, J. (eds.) GameDays 2014. LNCS, vol. 8395, pp. 23–34. Springer, Heidelberg (2014)

20. Kapp, K.M.: The Gamification of Learning and Instruction: Game-Based Methods and Strategies for Training and Education. Wiley, Hoboken (2012)

21. Molnar, A., Kostkova, P.: 13 learning through interactive digital narratives. In: Interactive Digital Narrative: History, Theory and Practice, pp. 200–210 (2015)

22. Paliokas, I., Arapidis, C., Mpimpitsos, M.: Game based early programming education: the more you play, the more you learn. In: Pan, Z., Cheok, A.D., Müller, W., Liarokapis, F. (eds.) Transactions on Edutainment IX. LNCS, vol. 7544, pp. 115–131. Springer, Heidelberg (2013)

23. Padilla-Zea, N., Gutiérrez, F.L., López-Arcos, J.R., Abad-Arranz, A., Paderewski, P.: Modeling storytelling to be used in educational video games. Comput. Hum. Behav. **31**, 461–474 (2014)

24. Liu, S., Ding, W.: An approach to evaluation component design in building serious game. In: Chang, M., Kuo, R., Kinshuk, Chen, G.-D., Hirose, M. (eds.) Learning by Playing. LNCS, vol. 5670, pp. 141–148. Springer, Heidelberg (2009)

25. Wendel, V., Göbel, S., Steinmetz, R.: Seamless learning in serious games-how to improve seamless learning-content integration in serious games. In: CSEDU, pp. 219–224 (2011)

26. Torrente, J., Moreno-Ger, P., Fernández-Manjón, B., del Blanco, Á.: Game-like simulations for online adaptive learning: a case study. In: Chang, M., Kuo, R., Kinshuk, Chen, G.-D., Hirose, M. (eds.) Learning by Playing. LNCS, vol. 5670, pp. 162–173. Springer, Heidelberg (2009)

27. McAlpine, M., van der Zanden, L., Harris, V., Authority, S.Q.: Using games based technology in formal assessment of learning. In: Proceedings of the 4th European Conference on Games-Based Learning, The Danish School of Education, Aarhus University, Copenhagen, Denmark, p. 242. Academic Conferences Limited, 21–22 October 2010

28. Molnar, A., Virseda, J., Frias-Martinez, V.: Insights from EducaMovil: involving teachers in creating educational content for mobile learning games. J. Interact. Learn. Res. **26**(2), 209–221 (2015)

29. Spires, H.A., Rowe, J.P., Mott, B.W., Lester, J.C.: Problem solving and gamebased learning: effects of middle grade students' hypothesis testing strategies on learning outcomes. J. Educ. Comput. Res. **44**(4), 453–472 (2011)

30. Kostkova, P., Molnar, A.: Educational games for creating awareness about health issues: the case of educational content evaluation integrated in the game. In: Medicine 2.0 Conference. JMIR Publications Inc., Toronto

31. Higgins, R., Hartley, P., Skelton, A.: The conscientious consumer: reconsidering the role of assessment feedback in student learning. Stud. High. Educ. **27**(1), 53–64 (2002)

32. Persky, A.M., Pollack, G.M.: Using answer-until-correct examinations to provide immediate feedback to students in a pharmacokinetics course. Am. J. Pharm. Educ. **72**(4), 83 (2008)

33. Molnar, A., Kostkova, P.: On effective integration of educational content in serious games: text vs. game mechanics. In: 2013 IEEE 13th International Conference on Advanced Learning Technologies (ICALT), pp. 299–303 (2013)

34. Spierling, U.: Interactive digital storytelling: towards a hybrid conceptual approach. In: DIGRA Conference (2005)

35. Sklar, E.: Agents for education: when too much intelligence is a bad thing. In: Proceedings of the 2nd International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 1118–1119. ACM (2003)

36. Conati, C., Zhao, X.: Building and evaluating an intelligent pedagogical agent to improve the effectiveness of an educational game. In: Proceedings of the 9th International Conference on Intelligent User Interfaces, pp. 6–13. ACM (2004)

37. Spierling, U.: Introducing interactive story creators to conversation modelling. In: Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology. ACE 2011, New York, NY, USA, pp. 38:1–38:8. ACM (2011)

38. Molnar, A., Kostkova, P.: Seamless evaluation integration into IDS educational games. In: FDG, pp. 322–329 (2013)

39. McNulty, C.A., Lecky, D.M., Farrell, D., Kostkova, P., Adriaenssens, N., Herotová, T.K., Holt, J., Touboul, P., Merakou, K., Koncan, R., et al.: Overview of e-bug: an antibiotic and hygiene educational resource for schools. J. Antimicrob. Chemother. **66**(suppl. 5), v3–v12 (2011)

40. Lecky, D.M., McNulty, C.A., Adriaenssens, N., Herotová, T.K., Holt, J., Kostkova, P., Merakou, K., Koncan, R., Olczak-Pienkowska, A., Avô, A.B., et al.: Development of an educational resource on microbes, hygiene and prudent antibiotic use for junior and senior school children. J. Antimicrob. Chemother. **66**(suppl 5), v23–v31 (2011)

41. Lecky, D.M., McNulty, C.A., Adriaenssens, N., Herotová, T.K., Holt, J., Touboul, P., Merakou, K., Koncan, R., Olczak-Pienkowska, A., Avô, A.B., et al.: What are school children in Europe being taught about hygiene and antibiotic use? J. Antimicrob. Chemother. **66**(suppl. 5), v13–v21 (2011)

42. Farrell, D., Kostkova, P., Weinberg, J., Lazareck, L., Weerasinghe, D., Lecky, D.M., McNulty, C.A.: Computer games to teach hygiene: an evaluation of the e-Bug junior game. J. Antimicrob. Chemother. **66**(suppl. 5), v39–v44 (2011)

43. Frias-Martinez, V., Virseda, J., Gomero, A.: Mobilizing education: evaluation of a mobile learning tool in a low-income school. In: Proceedings of the 14th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp. 441–450. ACM (2012)

44. Snow, B.: Why most people don't finish video games. Online Publ. CNN **17**(August), 2011 (2011)

45. Weerasinghe, D., Lazareck, L., Kostkova, P., Farrell, D.: Evaluation of popularity of multi-lingual educational web games: do all children speak English? In: 3rd International ICST Conference on Electronic Healthcare for the 21st Century (2010)

46. Molnar, A., Kostkova, P.: If you build it would they play? Challenges and solutions in adopting health games for children. In: Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems, Let's Talk About Failures: Why was the Game for Children Not a Success (2013)
47. Molnar, A., Kostkova, P.: Gaming to master the game-game usability and game mechanics. In: 2014 IEEE 3rd International Conference on Serious Games and Applications for Health (SeGAH), pp. 1–7. IEEE (2014)
48. Jordan, P.W., Thomas, B., McClelland, I.L., Weerdmeester, B.A. (eds.): SUS: A Quick and Dirty Usability Scale, pp. 189–194. Taylor and Francis, Abingdon (1996)
49. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Int. J. Hum.-Comput. Interact. **24**(6), 574–594 (2008)
50. Lewis, J.R., Sauro, J.: The factor structure of the system usability scale. In: Kurosu, M. (ed.) HCD 2009. LNCS, vol. 5619, pp. 94–103. Springer, Heidelberg (2009)
51. Lazareck, L., Farrell, D., Kostkova, P., Lecky, D., McNulty, C., Weerasinghe, D.: Learning by gaming-evaluation of an online game for children. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 2951–2954. IEEE (2010)

A Methodology of Integrating Knowledge into Applied Game and Validation of Education and Entertainment Effects

Qing Wang[1], Hong Chen[1(✉)], Jinfeng Li[2], and Dehai Zhu[1]

[1] College of Information and Electrical Engineering,
China Agricultural University, 17 Tsinghua East Road, Haidian District,
Beijing 100083, People's Republic of China
wangqingait@sina.com, norman_chen@263.net
[2] China Agricultural University East Campus Kindergarten,
17 Tsinghua East Road, Haidian District,
Beijing 100083, People's Republic of China

Abstract. Applied game aims to help player accumulate knowledge, improve skills, or enhance emotion quotient through game playing process. Although many applied games have been developed in education and training fields, the solution on issues of balance between education and entertainment is still open. This study proposes a methodology for integrating knowledge with applied game combining recreational functions and pedagogic effect. For the purpose, the unified representation based on ontology is employed for game elements, game challenge model and knowledge expression with which the relationship mapping mechanism from diverse categories of knowledge onto game atomic element, especially primary rule, is constructed. As a result, a practical design technique that facilitates to import knowledge into applied game is achieved. In order to verify the practicability, a case of applied game, Qbaby Picture Matching game is designed and developed based on the proposed method. It has been running online and applied in a kindergarten. Moreover, an experiment is performed to identify the effects of education and entertainment. The results imply that the game can definitely produce recreational functions and pedagogic effect. The fact indirectly verifies that the methodology may effectively integrate knowledge with game thereby providing a referable solution for issues of education and entertainment balance.

Keywords: Applied game · Knowledge · Event ontology · Game challenge

1 Introduction

Applied game, might be called educational game or serious game, is defined as a digital game integrating cognitive process with game challenge which helps learners accumulate knowledge, improve skills, or enhance emotion quotient. For it educative objective, one of the key points of designing applied game is investigating balance mechanism between entertainment and education aspects which gives instructions for designing a game with pedagogic function and recreational feature. According to the

characteristics of learning software and game [1], it is essential to analysis the relationship between cognitive procedures and game playing sessions and to give a methodology for integrating two activities. As a new element to be imported, knowledge plays a very important role in applied game. Study on a model of unified representation shared by knowledge and other game elements game may provide a basis to analyze their relationships and thus propose an alternative reference to solve the balance issues mentioned above.

2 Related Work

For the purpose er model and learner model identifying his or her behaviors and knowledge states respectively, is applied in an educational game to implement personalization and adaption [3]. It facilitates to track the player's knowledge state according to which apof education and entertainment in designing applied game, many researchers propose several models or methodology. EFM is a model which connects motivation of learning, flow experience, effective learning environment with educational game [2]. It gives guideline and ideas for designing educational game such as to set up goals according knowledge and skill, process and method, attitude and values objective, etc. NGLOB (Narrative Game-based Learning Objects) method, which combines play propriate game structure is adapted in personalized and adaptive manner. But, as it states, NGLOB-based method is mainly aimed on personalization and adaption. The performance of combining knowledge with game is not addressed. Besides these studies, some researchers propose idea of knowledge-centered and try to make knowledge structurized in games. Bruno et al. [4] created a cognitive model based on analysis of knowledge natures, knowledge classification and pedagogical objectives specification, implementing the connection between knowledge and game. The model applies "object-oriented modeling" technique and refers to meta-knowledge model which describes learning activities and leaner's mental processes. The approach helps to give a reference when design a game with the objective of combining cognitive processes and game playing sessions. Another method using knowledge modeling is given by Minović et al. [5]. The model constructs pieces of knowledge resources based on learning objects to establish the balance between knowledge and its reusability. In his presented methodology an educational game metamodel and knowledge metamodel are designed using UML and XML. The approach is aimed to step towards a unified framework for development of educational games.

The methodologies and models above contribute to design and develop applied games with functions of entertainment and education. However, works on relationship and mapping mechanism between game elements and knowledge are not enough to support a framework which specifies the details of knowledge importing. Therefore an unified representation of elements in game including knowledge maybe a feasible approach for analyzing the relationships between game and knowledge and contributing to construct the specification of knowledge-game integration.

3   The Methodology for Importing Knowledge into Game

The paper proposes a Game General Elements Model (GGEM) specifying a unified representation for all game elements, and Game Challenge Model (GCM) which is based on the former with the purpose of identifying the essential factors that make game attractive. The two models contribute to construct a mapping mechanism between knowledge and game elements accordingly implement fully integration at particle level.

3.1   Game General Elements Model (GGEM)

No matter what category a game belongs to, it consists of some general elements such as mission, character, tool, scene, etc. In this section, a model is constructed to describe all general elements according to practical experience.

Game elements can be classified as two categories: explicit elements and implicit elements. The former can be viewed as entities in game that corresponding to objects or facts in real word and is normally visible in interface. The latter refers to invisible factors such as rules, parameters, or event that only be perceived through hint or feedback when player interacts with game environment. The model is based on concept of ontology and implemented by protégé. Figure 1 shows the framework of the model. As it implies, GameExplicitElem and GameImplicitElem stand for the explicit elements and implicit elements respectively. GameSubject and GameObject are two main typical explicit elements. GameSubject refers to the entities in game that can be controlled directly by player, while GameObject is the one that cannot be controlled directly by player and only impacted through interaction with subjects. Among implicit elements, GameRule is the key element of a game which regulates all other elements' behaviors and interaction mode between player and game environment. Therefore it is also regarded as core of a game which specifies the game nature. GameKnowledge is another implicit element which is one of the main factors to distinguish applied game with other entertainment game. Knowledge cell in game may be a concept, fact, procedure, technique, formula, or method. They may be diverse patterns according to the knowledge types and learning objective.

3.2   Game Challenge Model (GCM)

Game should be fun and attractive. But what is the essential mechanism to make played attracted by a game? The question can be converted to another issue: what is the origin of gameplay [6]. Some researchers propose ideas like dividing game into elements or units which act as block to construct a game for analyze game natures [7]. Huang states that playing experience is the core of a game [8]. It refers to synthetical perception including art enjoy, thinking process and mental feeling through playing a game. From the aspect of interaction, playing experience is affected by feeling of sense organs and body, such as immersion of 3D environment, easy controlling method, or coordination of eyes hands. From the psychology and emotion of player, experience may come from
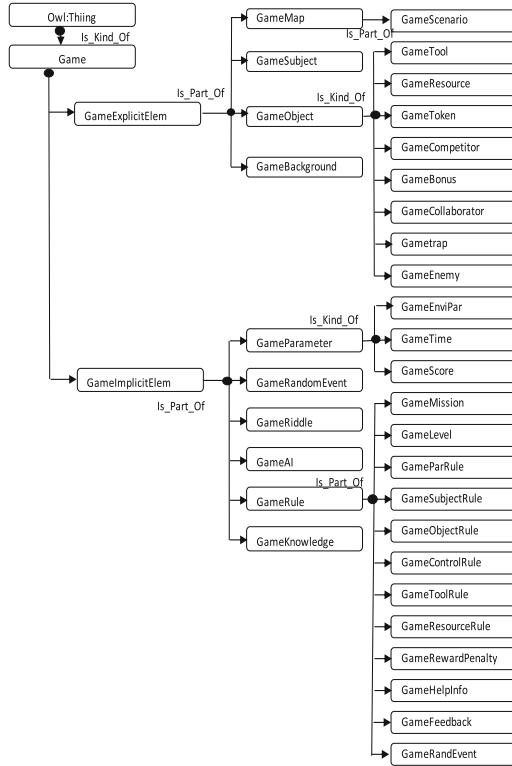
Fig. 1. Framework of game general elements model

delighted mental feeling, such as sense of achievement after a large amount of effort, or enjoyment from a splendid image and melodious music, or gratification from an exciting and tense atmosphere [9]. Therefore, game playing experiences mainly relate to two contents: one is interacting process based on rules; the other is mental feeling or body sense. Interaction process is the procedure of game challenging while mental feeling or sense may partially generated from challenging. Consequently, study on game challenge can help to identify the origin of gameplay.

In this paper, Game Challenge Model (GCM) based on GGEM is proposed to gives configuration for game challenge, aiming to address the essential mechanism of a game which makes player immersed in virtual environment. As discussed above, game challenge involves interaction process and mental feeling or sense caused by challenge, the model is focused on the interaction specifics through description of game elements behaviors and the key factors which motivate positive motions or facilitate mentality training.

Challenge comes from difference or distance from player initial state to the objective capability which is necessary for achieving game mission. That means player should spend a moderate amount efforts to reach the goal. Due to diversity of challenges of game, a hard action, tough puzzle, complicated tactics or strategy, even a

simulating activity can form challenge content. Thereby the model can be classified as ActChallModel, PuzChallModel, StratChallModel, and SimlChallModel which specify tasks as performing actions, solving puzzles, making strategy, and experiencing simulation activities respectively.

A game challenge model may not compulsorily be corresponding to one specific game. Conversely, a game is not mandatorily limited to contain one challenge model. Many games are design to be synthesis of several challenges of actions, thinking, creation or other aspects. The model merely gives fundamental blocks to constitute a complex game based on the GGEM.

For any category of challenge model, it consists of main elements as illustrated in Fig. 2: PrimRule describing the key rule which regulates main elements behaviors; Mission interpreting what a player pursues; InterMode indicating how a player controls game; ChallFactor elaborating the essence that makes player attracted and kept trying to approach the goal.
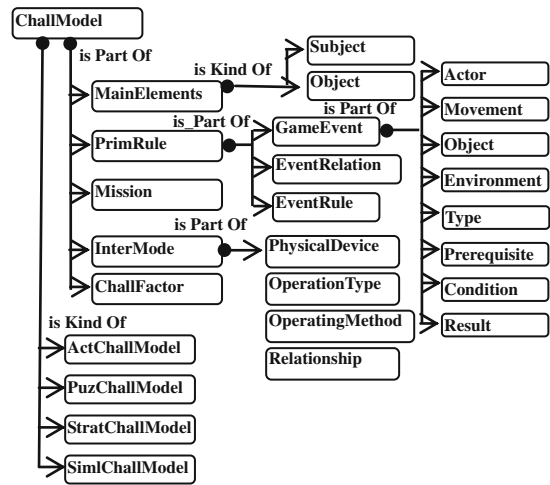
Fig. 2. Composition of game challenge model

To be in accord with GGEM, ontology is employed to construct the game challenge model. In the model, MianElements lists subject and main objects with their primary properties which can be obtained directly from game elements model. PrimRule is generalized regulations and can be viewed as valid actions set or possible events sequence. In the paper, event ontology is introduced to represent events. It normally involves dimensions of participant, action, type, occurrence, spatial or temporal information, etc. According to the definition of event ontology [10], triple-tuple expression is used for complete and efficient representation of PrimRules: {GameEvent, EventRelation, EventRule}. GameEvent is a class consisting of eight tuples referring to the six-tuple definition of event given by Liu et al. [10], it is modified and added supplementary factors for describing game event in virtual environment as {Actor, Movement,

Object, Environment, Type, Prerequisite, Condition, Result}. Actor usually refers to subject of game; Movement stands for subject's movements or actions set, including description of mode, pattern, duration, intensity; Object are the one who accepts the action; Environment identifies location, time, or parameters of virtual game environment; Type indicates what category an event belongs to, such as continuous process with prosperity of duration or discrete movement occurring instantaneously; Prerequisite is trigger that initials an event; Condition means the requirements that maintain event; Result describes consequence an event produces. EventRelation refers to relationship among all possible events in the model, such as Causal or Sequential, and is used to imply possible event sequences generated by the model.

Mission specifies the objective state of game elements which indicates direction for player action. It also describes the requirements for game continuously running or thresholds for entering next level. Mission is represented as key concerned game element with definitive relationship and rules.

InterMode determines how player controls game elements. With the advancement of technologies, human-computer interaction is getting more natural and approaching to manipulation in real world. Generally, interacting operation can be classified as: (a) acquiring information from game environment including referring to help, viewing map, or reading values of parameters; (b) manipulating individual element, such as moving, rotating, creating an object, or controlling an avatar to jump and fight; (c) triggering or terminating events, for instance, creating a building, buying, or hiring employee. These operations are implemented through pressing buttons on keyboard, typing commands, using menu, icon, or tool bar, dragging or clicking entities in virtual game environment, writing or drawing objects, touching screen, converting voice into instruction or using somatosensory devices. These diverse methods give player entirely different interacting experience which greatly impacts on playability of game. For involvement of forementioned contents, InterMode is composed of three subclasses: PhysicalDevice which may be assigned as keyboard, mouse, touchscreen, force_feedback, microphone, somatosensorydevice, etc., OperationType can be InforAcq, IndivManipl, EventTrig corresponding to the three operations discussed above, OperatingMethod is specific means embodied by key, command, menu, icon, pointer, touch, voice, movement. Besides, relationships among these subclasses are needed for detailed description of how every single interacting operation is performed. Moreover, when player uses key or command for OperatingMethod, there is always no visual indicator (cursor) or physical touching directly acting on game elements. That means player need create mapping from controlling actions onto game elements behaviors through many practices to master interacting skills. Menu and icon are normally used for information acquiring and event initiation. Pointer may give a visual instruction indicating what objects to be controlled and make interaction more convenient. Touch method means physical touching to identify target. It approaches natural interactions in real world and provides more actual feeling. Voice and movement are another two untouched interaction methods and make player completely participated with his whole body.

In most cases, motivations of playing game come from curiosity for unknown world, desire for recreation, satisfaction or fulfillment from a hard work, etc. Therefore game challenge is concerning with player's psychological demands. In the model,

ChallFactors summarized difficulties or novelty of game mission and psychological enjoyment achieved from the challenge. Like other components of the model, it uses event ontology with description of psychological situation during playing process.

For ActChallModel, PuzChallModel, StratChallModel, and SimlChallModel, sub classification is necessary for more detailed expression of each component. Section 4 will give a detailed instance for subclass of ActChallModel using GCM representation.

3.3   The Methodology of Mapping Knowledge onto Game Rule

Because of the educative purpose, applied game should involve knowledge leaning, skill training, or emotion cultivating. Based on the analysis of cognition progress [11–14] and game procedure [15], these two activities can be combined and therefore make a player complete a cognition task through a uninterrupted game experience rather than perform a learning activity such as reading, writing, or answer questions. For jointless integration and effective pedagogic results, imported knowledge is expected to enter game rules and to be essential parts for implementation of game mission. Generally speaking, knowledge representation [16] can be fallen into three types: Declarative knowledge which can be used to give definitions, attributes, states description of an object or a fact; Procedural knowledge mainly telling how it is applied with a manner of procedure or progress; Meta knowledge is knowledge of knowledge. In GCM, knowledge shares the unified representation to make it possible to map knowledge onto game rules. Due to the diversity of knowledge, this section proposes ontology GameKnowledge to denote knowledge in game (Fig. 3).
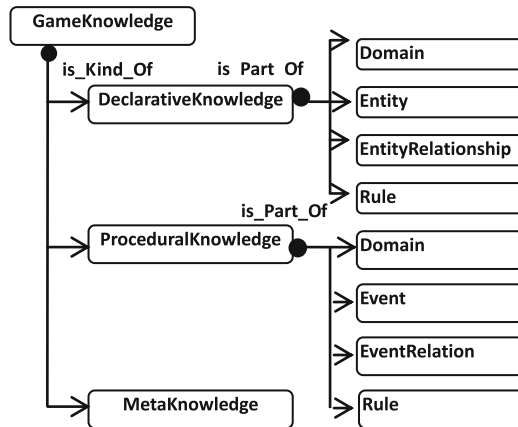
Fig. 3. Game knowledge ontology

For DeclarativeKnowledge, Domain indicates which field knowledge belongs to based on the standard classification of discipline. Entity is defined as class with attributes, used for definition, conception, or description of an object or a fact. EntityRelationship

specifies spatial, temporal, topological, or taxonomic relationship among entities. Rule describes inference on entities. ProceduralKnowledge deals with knowledge of procedure, function, method, technique, etc. which generally relates to dynamic events. Event ontology is employed here with the same definition and structure of GameEvent in challenge model.

On the basis of the representations of game elements, mapping knowledge onto game rules can be depicted as follows:

(1) Declarative knowledge:
    As given in Fig. 4 Entity describes object or fact which can be simulated as actor or object in GameEvent. Therefore their attributes can be presented through actor or objects behaviors. Moreover, EntityRelationship and rule explicate relationships of entities and events which possibly occur when an entity interacts with others. This part of knowledge consequently can be imported to GameEvent to facilitate describing prerequisite, condition or result of an event. Once they come to be portion of PrimRule, the knowledge is viewed as an essential mechanism to support mission implementation.
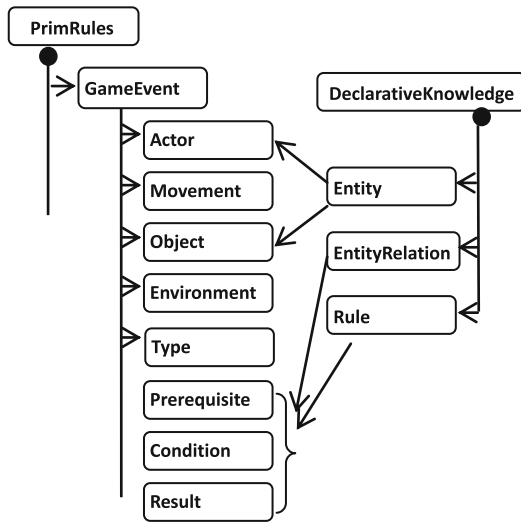
Fig. 4. Mapping declarative knowledge onto rule

(2) Procedural knowledge:
    Procedural knowledge uses event ontology to explain program, method, procedure or skill which generally involves one or more dynamic events. Like GameEvent in GCM, the event in ProceduralKnowledge shares the same structure as shown in Fig. 5. Their difference may be determined by proximity from virtual game to the real world. For actually and correctly delivering knowledge, the corresponding classes in Fig. 5 should have high similarity. Consequently mapping knowledge onto game rule is transformed into constructing a similar virtual world composed of analogue contents of knowledge based on the structure of
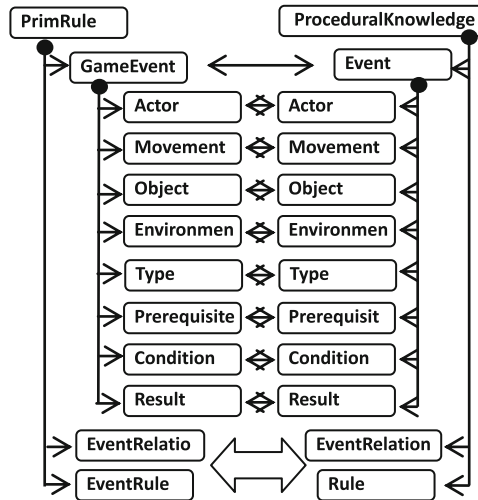
Fig. 5. Mapping procedural knowledge onto rule

appropriate GCM. Obviously, different types of challenge model have distinct detailed architecture and provide more valuable instructions for construction of virtual events and environment.

To sum up, designing applied game consists of following steps: prescribing applied scope and audiences; determining knowledge contents; arranging knowledge organization or creating knowledge ontology; selecting an appropriate GCM according to knowledge category and characteristic; mapping knowledge ontology onto game elements; improving game design.

4   A Case of Applied Game: Qbaby Picture Matching Game

To demonstrate practicability and effectiveness, an applied game, Qbaby Picture Matching game is designed according to the proposed methodology and developed using unity 3D.

The game aimed to help preschool-age children to complete cognition of color, shape, symbol, and other scientific or social knowledge which mainly are declarative knowledge contents and have nothing to do with dynamic events or procedures.

4.1   Creating Knowledge Ontology for the Game

The game is expected to contain knowledge listed as follows:

(1) Concepts related to color, shape, position, geometry; skills for identifying same colors, similar shapes, symmetric figures through comparison; skills for indicating relationship of whole and parts.

(2) Symbols of English alphabets; skills for indicating correlative symbols.
(3) Animals and corresponding food, living environments.
(4) Characterized elements of nations, countries such as building, costume and national flag.

Knowledge for children is defined as: ChlKnowledge = {Class, Relation} in which Class stands for entity or concept with attributes and Relation defines relationships between classes. Therefore knowledge framework can be simplified as following ontology in Fig. 6.
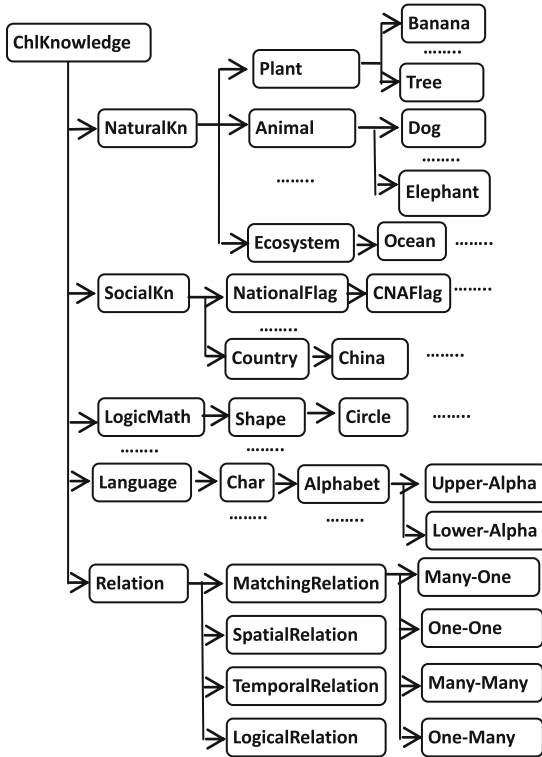
Fig. 6. Knowledge framework for children cognition

4.2   Selecting Challenge Model

For easy interaction, a subclass of PuzChallModel, MatchingChallModel is selected for simulating the learning environment and will be rewritten with more special contents based on the structure proposed in Sect. 3.2 as shown in Fig. 7. The game challenge can be described as: selecting identical or correlated icons if they can be connected using less than three straight lines which results in that the two selected icons disappear. The MainElements is Icon in game. There are generally multiple icons in one game level. So index is needed for identifying every single icon. GameEvent can be

defined as Remove. Prerequisite can be stated as: two corresponding icons are selected and their positions make it possible that less than four straight lines form path from one icon to the other. A formal representation of Prerequisite is written as:

$$IsActived(IconN1) \wedge IsActived(IconN2)$$
$$\wedge \, Relation(IconN1, IconN2) \wedge Connected - 3(IconN1, IconN2)$$

IsActive(Icon) is function to indicate if the icon is active. Relation(IconN1,IconN2) is function to tell if IconN1 and IconN2 have predefined relationship. Connected-3 is function to calculate possible paths consisting of less than four lines. It returns True if there exits such path(s) and False otherwise. Result of the event is altering icon state into removed and be written as AlterSta(IconN1, IconN2, removed).

Mission is the objective state of game. Suppose that there are m icons, it can be written as:
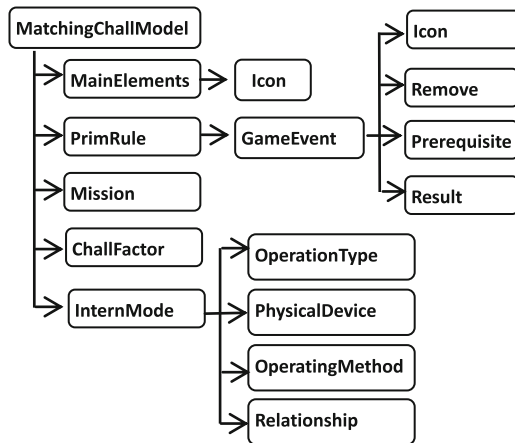
Fig. 7. Structure of MatchingChallModel

$$IsRemoved(IconN1) \wedge IsRemoved(IconN2) \wedge IsRemoved(IconN3)...$$
$$IsRemoved(IconNm) \wedge Terminate().$$

IsRemoved(IconN) is function to tell if the icon's state is removed. Terminate() function will returns True if the game terminates. The mission motivates player in constantly trying to make all icons removed within game limited time. Therefore, ChallFactor can be described as: find the correlated icons, analyze their positions, and click them as soon as possible.

4.3   Mapping Knowledge onto the Game Rule

According to the methodology proposed in Sect. 3.3, the declarative knowledge can be mapped onto MatchingChallModel. Mapping from Entity of knowledge onto Subject or

Object is converted into mapping class onto icon in MatchingChallModel. It means that using visual or graphic icon to standing for a class or concept. Moreover, mapping EntityRelation and Rules onto event's prerequisite, condition and result can be specialized as mapping Relation of class in knowledge onto Prerequisite and result in MatchingChallModel. As given in Sect. 4.2, Prerequisite involve relationship description of icon, therefore Relation of class in knowledge can be integrated with function Relation(IconN1, IconN2) in Prerequisite expression. In brief, for mapping knowledge onto game rule means two tasks: employing graphic icon to illustrate a concept or class in knowledge; defining relationship of icon for Relation function according to the Relation of class in knowledge.

4.4   Game Implementation

The contents to be involved in the game are listed as follows: (1) Matching items of same color; (2) Matching objects of similar shape; (3) Matching animals and their food; (4) Matching whole and part; (5) Matching upper-case and lower-case letters; (6) Matching correlative items; (7) Matching notional flags and country. Each content contains twenty pairs of icons with one-to-one corresponding relationship and is designed with ten levels. The game runs on iPad using touching screen as interaction device. It has been developed by unity 3D game engine and been running online. Figure 8 gives interface screenshots of the game.

Fig. 8.   Interface of Qbaby game

5   Verification of Education and Entertainment Effects of Qbaby

For verifying educative and recreational effect of the game, it is applied in a kindergarten and forty-four children are selected whose age ranging from 5 to 6. They are divided into two parts, one part with 20 children for entertainment effect verification and the other part composed of 24 children for education effect assessment. Furthermore, the second part is further randomly divided into two groups with ten children for each. To simplify the experiment, only three contents of the game are tested by two

groups. They are Upper-case and Lower-case letters Matching, Correlative items Matching and Notional flags and Country Matching.

5.1 Verification of Entertainment Effect

For assessing playability and recreational effect of the game, another picture matching game, Fruit Matching, is selected for comparison. Control group is arranged to play Fruit Matching, while experimental group play Qbaby. Concerned data to be recorded contains duration of continuous playing of player's own accord, times of retrying after player fails, and final level player reaches. The experiment is performed without any limitation on time and times. Therefore, the combination of the data can imply if a player has continuous motivation on the game and how deeply he is attracted. This fact can give basis to analysis the game's playability and entertainment effect.

5.2 Assessment of Education Effect

On the other hand, for comparison educative effect of Qbaby with traditional learning materials, the two groups are arranged to adopt different learning approaches: the experimental group learn by Qbaby while the control group learn by learning materials using colorful pictures printed on paper (as shown in Fig. 9) for obtaining comparability with game interface.

Fig. 9. Learning materials using colorful pictures

(1) Pre-test for obtaining knowledge background

Before experiment, the children are arranged to have a questionnaire to test their knowledge background about the contents that to be tested. Questionnaire consists of ten pairs of items. They are separated and printed on two columns disorderly as shown in Fig. 10. The children should find the matching pairs and link them by drawing lines. The numbers of correctly linked pairs are recorded to identify their initial knowledge levels for next comparison with post-test results.

Fig. 10. Pre-test paper

(2) Post-test for verification of didactical effect

After two groups learning and playing, they are asked to have another test using the same questionnaire in pre-test. Similarly, the numbers of correctly linked pairs are recorded for analysis.

5.3   Data Results and Analysis

The recorded data includes Table 1 for entertainment effects verification and Table 2 for pedagogic effects assessment which contains pre-test and post-test results.

Table 1. Records for entertainment effect verification

	ID	Duration (min)	Retrying times	Level reached		ID	Duration (min)	Retrying times	Level reached
Control group	1	10	6	3	Experimental group	11	8	5	4
	2	9	5	3		12	9	5	4
	3	8	4	4		13	8	6	3
	4	8	5	3		14	9	5	3
	5	9	4	5		15	9	5	4
	6	10	5	3		16	10	5	3
	7	8	5	4		17	9	5	4
	8	9	6	3		18	8	6	3
	9	7	4	4		19	9	4	4
	10	10	5	3		20	8	5	4

To examine the deference between two groups, T-test is applied to identify if two samples have significantly different means using 0.05 as significance level which is generally accepted. For the data of duration (min), $P = 0.80062 (>0.05)$, which implies

that there is no significant difference on duration of playing between two groups. It shows that the two groups are basically equally interested in entertainment game (Fruit Matching) and the applied game (Qbaby Picture Matching). Similarly, for times of retrying, $p = 0.50555(>0.05)$, indicating that the individuals of two groups have basically same degree of motivation on the two games. Moreover, data of levels that the players reached produced $P = 0.72219(>0.05)$ using T-test shows the two games are more or less equally difficult for them. The above results jointly demonstrate Qbaby is definitely playable and fun enough to make player immersed for a while. Hence, the performance of entertainment is verified.

Table 2. Records of correctly linked pairs in the three contents

Group	ID	Content 1: letter matching		Content 2: correlated items		Content 3: national flag	
		Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
Control group: learning by materials	1	1	2	1	4	1	2
	2	0	4	0	4	0	4
	3	3	6	2	4	0	3
	4	2	4	1	5	2	4
	5	0	2	0	5	1	3
	6	1	5	0	4	1	3
	7	1	5	1	3	1	3
	8	0	4	2	5	3	6
	9	0	5	1	5	0	3
	10	1	4	2	6	2	4
	11	0	5	1	4	0	3
	12	2	5	0	2	0	2
Experimental group: learning by Qbaby matching game	13	0	8	1	5	1	5
	14	1	6	0	5	0	3
	15	3	8	1	5	2	4
	16	2	6	1	4	2	4
	17	1	4	0	4	0	5
	18	0	5	1	6	0	4
	19	1	6	1	5	3	5
	20	3	6	0	6	2	6
	21	0	4	2	5	0	3
	22	0	5	0	6	1	4
	23	1	6	1	6	1	4
	24	0	5	2	7	0	5

Like the entertainment verification, T-test is used for educative effect assessment. For clearly comparisons, the values of p for data in Table 2 are computed and listed in Table 3.

Table 3. Values of P for two groups using T-test

	Content 1: letter matching		Content 2: correlated items		Content 3: national flag	
	Pre-test	Post-test	Pre-test	Post-test	Pre-test	Post-test
P	0.85	0.004	0.79	0.006	0.843	0.01

In Table 3, P values computed with records of pre-test for three contents are 0.850, 0.790, and 0.843 (all >0.05). It shows that the two samples have same level background on these contents knowledge. After the experiment, two groups all gain enhancement on the number of correctly linked pairs according to comparison between pre-test and post-test for same content, as indicated in Table 2. It preliminary reveals Qbaby results in improvement on the knowledge and has more or less educative effects. But to identify its advantages, one-side T-test is performed to determine if the control group mean is larger than the experimental group mean. Table 3 gives the post-test P values are 0.004, 0.006 and 0.010 corresponding to the three contents respectively. The values are all less than 0.05. It indicates that the experimental group mean is larger than the control group and consequently verifies that learning by Qbaby can produce a better educative performance compared with learning materials.

6   Conclusions

This research is focused on the methodology of integrating knowledge into game to achieve an applied game combining effects of entertainment and education. With the purpose, knowledge is expected to enter primary rules so that player will gradually increase knowledge or enhance skills during repeatedly practicing game rules thereby makes cognition and game process merged. For construction mechanism of importing knowledge into game rules, a unified representation of game elements and knowledge based on ontology is proposed. As result, the relationships and mapping mechanism between various knowledge categories and game elements, especially game challenge factors, are analyzed and constructed to provide an approach for integration of knowledge with game rule. In order to verify practicability of the methodology, an applied game, Qbaby Picture Matching game which concerns with children cognitive contents is design based on the proposed method and developed using Unity 3D. The game has been running online. For further demonstrating that the methodology can help to design a game with educative effects but not lacking recreational function, Qbaby is applied by dozens of children in a kindergarten. Meanwhile, an experiment is performed to compare knowledge levels before and after the game and identify differences between Qbaby performance with other entertainment game and traditional learning mode. With the analysis of the recorded data, a conclusion is obtained that the applied game (Qbaby) can definitely provide entrainment function and education effect. The fact can be used as evidence to indirectly verify that the proposed methodology can provide a practical reference specification for integrating knowledge with applied game.

References

1. Connolly, T.M., Stansfield, M., Hainey, T.: An application of games-based learning within software engineering. Br. J. Educ. Technol. **38**(3), 416–428 (2007)
2. Song, M., Zhang, S.: EFM: a model for educational game design. In: Pan, Z., Zhang, X., El Rhalibi, A., Woo, W., Li, Y. (eds.) Edutainment 2008. LNCS, vol. 5093, pp. 509–517. Springer, Heidelberg (2008)
3. Göbel, S., Wendel, V., Ritter, C., Steinmetz, R.: Personalized, adaptive digital educational games using narrative game-based learning objects. In: Zhang, X., Zhong, S., Pan, Z., Wong, K., Yun, R. (eds.) Edutainment 2010. LNCS, vol. 6249, pp. 438–445. Springer, Heidelberg (2010)
4. Bruno, C., Boudier, V., Labat, J.: Knowledge management approach to support a serious game development. In: Proceedings of 2009 9th IEEE International Conference on Advanced Learning Technologies, ICALT 2009, pp. 420–422 (2009)
5. Minović, M., Milovanović, M., Starcevic, D., Jovanović, M.: Knowledge modeling for educational games. In: Lytras, M.D., et al. (eds.) WSKS 2009. LNCS, vol. 5736, pp. 156–165. Springer, Heidelberg (2009)
6. Gameplay definition. http://www.webopedia.com/TERM/G/gameplay.html. Accessed 14 Sept 2006
7. Joris, D.: A theory of fun for game design, a review. http://www.jorisdormans.nl/article.php?ref=theoryoffun. Accessed 24 Nov 2006
8. Huang, S.: Fundamentals of Digital Game Design, pp. 33–35. Tsinghua University Press, Beijing (2008)
9. Nicole, L.: Why we play games: 4 keys to more emotion. In: Game Developers' Conference. http://www.xeodesign.com/xeodesign_whyweplaygames.pdf. Accessed 19 Sept 2006
10. Li, Z.-T., Huang, M.-L., Zhou, W., et al.: Research on event-oriented ontology model. Comput. Sci. **36**(11), 189–192 (2009). Simon, H.A.: Comparison of game-theory and learning-theory. Psychometrika **21**(3), 267–272 (1956)
11. Freitas, S., Oliver, M.: How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? Comput. Educ. **46**(3), 249–264 (2006)
12. Gagne, R.M., Driscoll, M.P.: Essentials of Learning for Instruction, 2nd edn, pp. 65–80. Prentice Hall, London (1988)
13. Gagne, R.M.: The Conditions of Learning, 4th edn. New York, Holt, Rinehart and Winston (1985)
14. Kristian, K.: Digital game-based learning: towards an experiential gaming model. Internet High. Educ. **8**, 13–24 (2005)
15. Chen, W.-W., Chen, S.: Knowledge Engineering and Knowledge Management, pp. 9–15. Tsinghua University Press, Beijing (2010)

Augmented Reality Game Development and Experience Based on Intelligent Mobile Phone

Haiying Zhao[1]([✉]), Hong Chen[2], Jiongzhi Wang[2],
and Ruixuan Zhang[3]

[1] Mobile Media and Cultural Calculation Key Laboratory of Beijing
Century College, BUPT, Beijing 102613, China
zhy.yn@163.com
[2] College of Information and Electrical Engineering,
China Agricultural University, Beijing 100083, China
norman_chen@263.net
[3] Beijing JiuYiTongXing Technological Co. Ltd, Beijing, China

Abstract. The interactive mode and scene effect of games are facing more and more challenges with increasing attention given by game players. Augmented Reality produces a real-time 3D display effect by overlaying virtual information with the real world. Augmented reality technology is applied in this article to develop an AR basketball smart phone game. Users can control the interaction of their counterparts in the virtual world through image characteristics identification card, which is interactive and playable. The game not only combines entertainment with education, but also offers an immersive experience. More importantly, it is designed to promote the concept that teenagers are expected to change the world through their intelligence and wisdom as well as to have sports and exercise.

Keywords: Augmented reality · Gaming experience · Image feature identification card · Smart phone

1 Introduction

With apprehensive application of computer graphics technology, artificial intelligence technology, multimedia technology, network and communication technology, games are an important part of modern forms of entertainment. Game development involves traditional games, online games and virtual reality games in terms of design philosophy, and it also goes through stages from 2D to 3D technically. The development and attention given to the new technology in the field of online games, especially to 3D display technology, have promoted the rapid development of game industry. Facing

prominent contradictions between the focus on interaction and the lack of recreation, virtual reality games appeared. Virtual reality sublimates audio-visual experience to body experience, which enables users to be involved physically and to fully experience the joy brought by games. Responding to the focus on the user's experience in virtual reality games and based on the virtual reality technology, the article develops a basketball game with the rules that users can gain points by passing levels. The game not only combines entertainment with education, but also increases the fun of the game through the augmented reality technology.

2 Related Works

Augmented reality (AR) constructs a virtual space with a virtual-real combination by overlaying computer-generated virtual objects [1], scene or information suggested by system on the real scene, which realizes the augmentation of the real world [2–4]. The rise of small hand-held mobile devices (such as palm computer PDA and smart phones) provides a new technical method for the development of mobile augmented reality system [5, 6]. Most of the current mainstream mobile handheld devices have built-in megapixel camera and integrate high-speed wireless communication network equipment such as infrared communication interface, Bluetooth wireless interface and wireless LAN card [7, 8]. Choosing small handheld mobile devices as a new carrier of augmented reality technology has a strong advantage in aspects of mobility, portability and human-computer interaction. It has become an inevitable trend of the new generation AR technology to take smart phones as AR technology development and application platform [9–12]. Game development in AR system mainly deals with mobile video detection technology and virtual-real registration technology. The mainstream motion detection methods are optical flow method [13–15], background difference method [16] and interframe difference method [17]. AR virtual-real registration technology [18, 19] mainly includes: reference point method, template matching method, affine transformation method and the method based on image sequence. In template matching registration technology, augmented reality system is based on feature points whose accurate identification is the key to the virtuality and reality combination.

3 Game Development of Augment Reality Based on Smart Phone

Augmented reality game system is composed of a number of hardware and software. In order to simplify the design process, the entire augmented reality system is divided into four subsystems: image acquisition, feature recognition, scene generation and display enhancements. As is shown in the figure below, the mobile phone camera belongs to the image collection subsystem, feature recognition and feature tracking units belong to the feature identification subsystem, scenario generation computing cell belongs to the scenario generation subsystem, image synthesis computing cell and mobile terminal belong to display enhancement subsystem.
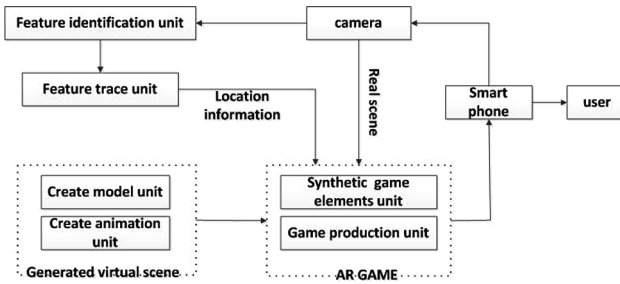
Fig. 1. System structure design drawing

Users turn on cell phone cameras to obtain video frame. The system will identify video frame data firstly. Then it will match video frame data, and track the video. At last, the video is overlaid with corresponding registered virtual scenes and then send it to the user terminal to display. In the whole design phase, virtual scene is generated by modeling and animation units, and AR games are implemented by game element synthesizing and game production units, as is shown in Fig. 1.

3.1 Construct Model

The steps of creating model based on 3Dmax

   Step1: Design the original painting
   Step2: Tpose model building
   Step3: Break up the UV map
   Step4: Draw the UV map
   Step5: Adjust the material
   Step6: attach texture onto material

According to the modeling method mentioned above, three models are built in the virtual reality game system: half a basketball court, a basketball board, a bind bones basketball figure, as is shown in the figure below (see Fig. 2).

3.2 Animation Production

In the development of augmented reality game, animation includes waiting animation, dribble animation, single finger spinning ball, crossover (two people), shooting animation, and turnaround jump shot. Technologies involved mainly include skeleton building and skin binding. Animation production process is as follows:

   Step 1: Earlier conceiving
   (1) Conceiving levels.
   (2) Conceiving levels and actions (before triggered, when triggered, after triggered).
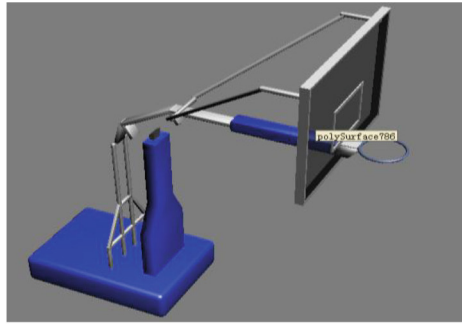
(a) Basketball Shelf

(b) Basketball court

(c) Characters of basketball figure

Fig. 2. Set up the virtual model

A relatively new operation method is designed that basketball players acting as main characters and triggers as well as scenes added to complete the level, users can play game by cards.

Step 2: Designing actions

After we have watched 20 matches and 3 2K12 matches (simulating PC basketball game), we select 6 representative, easy-to-do, and effective technical moves and standardize them in order to be more adaptable to the requirements of game material.

Step 3: The earlier preparation of animated characters

(1) Binding animation figures' skeletons

First load the player models within the view, using the biped tool to create a basic skeleton. Use figure mode in sports panel. Then adjust the position of bones and models for alignment by means of movement, rotation, and scaling. Next enter the modify command panel, select the physique modify command, and initialize the command. Lastly adjust the detail according to the parameters, making the whole skin binding results more precise.

(2) Skeleton binding test

In sports panel, import bip generated after the skeleton binding, and make fine adjustment to bones.

Step 4: Animation

Use premiere to clip the filtered action samples and export and save in the form of sequence of frames. Then according to each frame image, establish key frames in the 3D max, and adjust the character's joint displacement and rotation of each frame to match the action in the picture and complete the whole set of action.

3.3   Synthesis of Game Elements

In the Unity3D development environment, synthesize game elements in the 3D model, image feature recognition card and Qualcomm AR toolbox. And Qualcomm AR toolkit is mainly composed of several function modules: (1) the camera calibration and parameter collection, target labeling and tracking module. Calculate the distance and the location of the camera relative to the marking. First of all, according to the threshold set by the user, convert a collected single frame color image into a binary image (black and white); then analyze the connected domain of the binary image to find out all of the quadrilateral area as the candidate matching area. Matching each candidate area with templates in the template base, if match is produced, then the system will recognize a marker; calculate the camera position and posture relative to the known marker by the deformation of that labeled area; and finally according to the transformation matrix to realize the virtual- real registration. Based on this measure, complete the camera calibration, tag identification, 3D registration and other functions. (2) Video processing library will complete the real-time image acquisition function.

3.4 Game Production

Game production includes level, trigger action invocation and interception identification, among which interception identification determines card's identification state according to an enumeration variable provided by the development kit, 3D coordinate relative to the game scene, and the card's ID. In order to improve the interactivity of the game, the system designs 2 model collections: one is the model of the basketball player and the other is of basketball court. Each collection corresponds with an ID. The function of trigger action invocation is to triggers events when characters enter the special area, and the design flow is as follows (see Fig. 3):
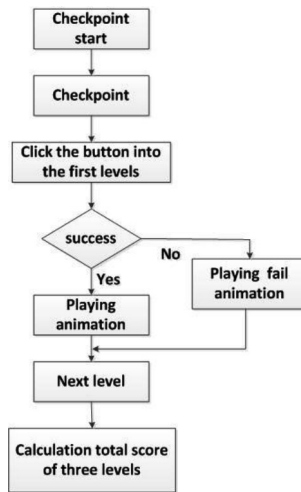
Fig. 3. Game production flow

4 Prototype System Implementation

In the applications of AR basketball, AR technology is embodied in two parts: (1) the selection process of basketball court; (2) the playing process of basketball players.

In the selection process of basketball court, if the image characteristics identification card is recognized when the development kit reads the video frames through cameras, calculate the position of identification card in the camera coordinate system and output the corresponding model of the basketball court on the location (as shown in Fig. 6). In the playing process of basketball players (as shown in Fig. 7), when the program recognizes identification card for basketball players, basketball game is completed by controlling the game (see Figs. 4 and 5).

Controlling trigger game, the basketball game based on AR technology is shown in Fig. 8.
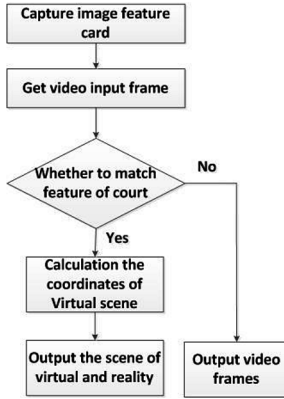
Fig. 4. Flow of selecting the basketball court
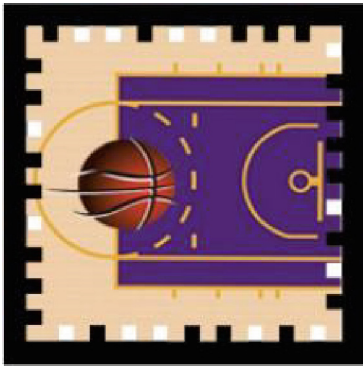
(a)The logo card with basketball   (b) The card represents basketball court
represents basketball player.

Fig. 5. Image feature identification card
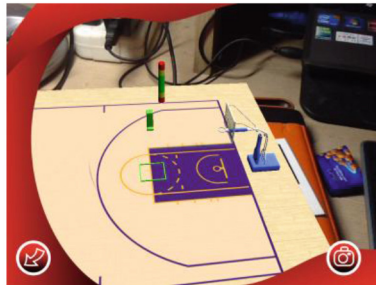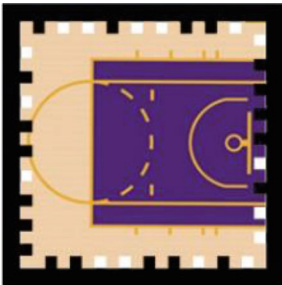
Fig. 6. The selected basketball model
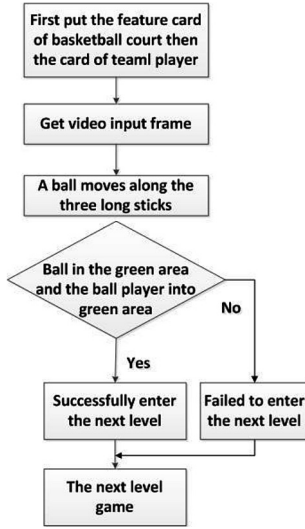
Fig. 7. Control flow of basketball game
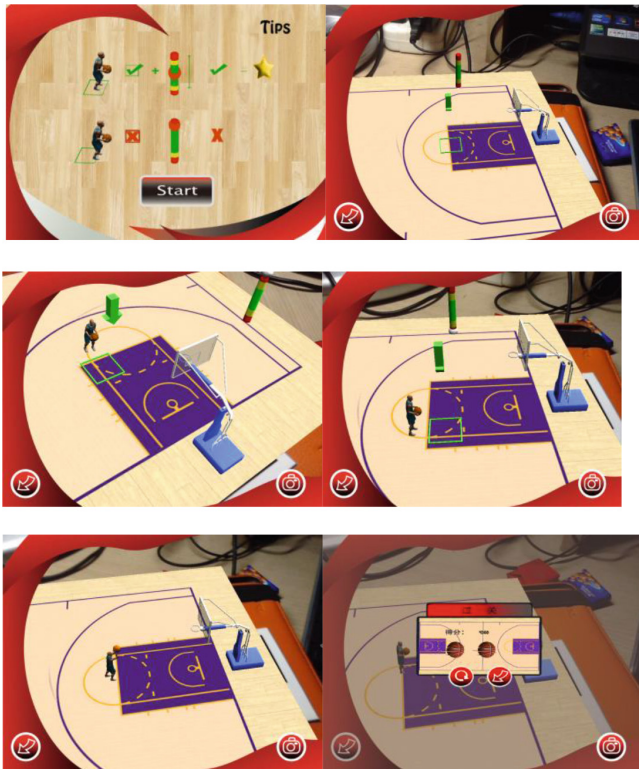
Fig. 8. Basketball game based on AR

5   Conclusions

The game scene of virtual-real combination provides immersive feeling to users, which is the advantage of using augmented reality technology in the game. With the ever-accelerated development and improvement of mobile terminal technology, mobile phones, as an indispensable part of daily life, will show greater superiority and function. The combination of AR technology and mobile terminals will certainly bring a new research direction in the field of AR. On the basis of AR technology, the article has designed and developed a set of basketball games which adds more fun to games.
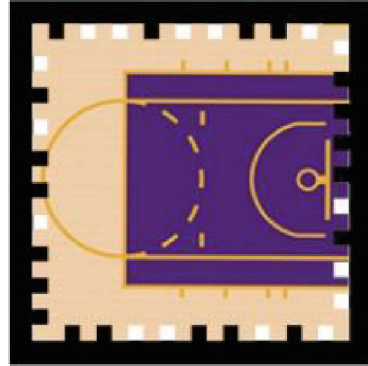
References

1. Qi, Y., Ma, H.: Augmented reality: characteristics, key technology and applications. J. Chin. Comput. Syst. **25**(5), 900–903 (2004)
2. Azuma, R.T.: A survey of augmented reality. Teleop. Virt. Environ. **6**(4), 355–385 (1997)
3. Neumann, U., You, S., Cho, Y.: Augmented reality tracking in natural environments. In: Proceedings of the IEEE International Symposium on Mixed Realities, Tokyo, Japan, January 1999
4. De Paiva Guimaraes, M., Farinazzo Martins, V.: A checklist to evaluate augmented reality applications. In: 2014 XVI Symposium on Virtual and Augmented Reality (SVR), pp. 45–52 (2014)
5. Jing, C., Yong-tian, W., Jun-wei, G., Wei, L.: Augmented reality technology applied on mobile phone platform. J. Univ. Electron. Sci. Technol. Chin. **39**, 80–84 (2010)
6. Gherghina, A., Olteanu, A.-C., Tapus, N.: A marker-based augmented reality system for mobile devices. In: 11th Roedunet International Conference (RoEduNet), pp. 1–6 (2013)
7. Youqun, H., Yongchen, J., Dan, L.: Research on interactive operations on augmented reality based on ARToolKit. JISUANJI YU XIANDAIHUA **9**, 97–100 (2008)
8. Shaoqing, H.J.H.: An implementation of augmented reality system based on ARToolkit plus and MVC framework. Comput. Digit. Eng. **40**(4), 1102–1104 (2012)
9. Augmented reality [EB/OL]. http://baike.baidu.com/view/104668.htm
10. Augmented reality technology [EB/OL]. http://news.xinhuanet.com/world/2009-08/04/content_11821085.htm
11. Augmented reality [EB/OL]. http://en.wikipedia.org/wiki/Augmented_reality
12. The future of mobile augmented reality [EB/OL]. http://www.mobilemarketingwatch.com/the-future-of-mobile-augmented-reality-4612/
13. Lucas, B.D., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: Processing of the International Joint Conference on Artificial Intelligence, Canada, pp. 674–679 (1981)
14. Horn, B.K.P., Brian G.S.: Determining optical flow. Artificial Intelligence Laboratory, Massachusetts Institute of Technology (1980)
15. Bradski, G., Kaehler, A.: Learning OpenCV: computer vision with the OpenCV library (2009)
16. Anderson, C.: Change detection and tracking using pyramids transformation techniques. In: 1985 Proceedings of SPIE-Intelligent Robots and Computer Vision, Cambridge, MA, vol. 579, pp. 72–78 (1985)

17. Stauffer, C., Grimson, W.E.L.: Adaptive background mixture models for real-time tracking. In: Proceedings of 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, p. 2. Fort Collins, CO, USA (1999)
18. Li, L.Y., Huang, W.M., Gu, Y.H.I., et al.: Foreground object detection from videos containing complex background. In: Proceedings of the Eleventh ACM International Conference on Multimedia 2003, pp. 2–10. Berkeley, California, USA (2003)
19. Zheng, F., Schubert, R., Weich, G., Martín-Gutiérrez, J., Saorín, J.L., Contero, M., Alcañiz, M.: A general approach for closed-loop registration in AR. In: IEEE Virtual Reality, 2013 IEEE Virtual Reality (VR) Conference, pp. 47–50 (2013)

A Study of Digital Games as a New Media of Cultural Transmission

Chenyang Cui[✉]

China Academy of Art, Hangzhou, Zhejiang, China
`cuicy@caa.edu.cn`

Abstract. By the 20th century, with the swift development in digital technology such as computer, network and multimedia, digital games have created an enormous social impact on all aspects of the community. Evolving from the initial simple means of entertainment to an interactive media visual culture, it would have more significant impact on human culture. Given its new form of technology-based representation, we need to examine it from a new horizon accordingly. This paper aims to articulate the cultural characteristics of digital games from three aspects, mass media, mass culture, international communication.

Keywords: Digital games · Mass media · Mass culture · Social context · International communication

1 Introduction

Games industry is drawing more and more attention with it is huge commercial success and development potential. According to DiGRA (Digital Games Research Association), the Digital Games is defined as digital technology as a means to digital devices designed to support the development and implementation of a variety of games, its type covering single games, online games, mobile terminals games and host games. Different from its increasingly commercial recognition and market demand, there is a lack of understanding on digital game's characteristics, role and implications in the academic field. It has long been debated whether digital games can be seen as an art form, or not. With the development of digital technology, digital games is getting more and more complex, the ability to express the world is becoming more and more powerful. As seen in the past, none of any new art forms was recognized and accepted, when it were emerged. Similar to the emergence of film art, the role and significance of film art has not been recognized, except a vaudeville peep show and circus side show merely, until a new subfield of contemporary philosophy of art was gradually established in the early decades of the twentieth century. Currently, there are a few academic studies on digital games found in both Western countries and China. However, it is noted they are merely initial investigation from theoretical perspective. For instance, from a rhetorical sense, Jane explored the structure of digital games and the consequent immersion brought to people and society around from the virtual worlds. From the aesthetic base, Graeme Kirkpatrick studied the common points sharing in the video game and traditional arts in terms of drawing painting and music aesthetics etc. Jenova Chen proposed a "Active

Dynamic Difficulty Adjustment" theory, pointed out that the game itself should be based on the elements which make corresponding self-adjustment, breaking game typical either alive or dead, either win or lose routine, trying to arouse people's emotional experience from the level of consciousness. Le Blanc proposes a MDA model and game mechanism is the important to aesthetic experience in the game. Chen Xiao-fei discussed aesthetics factor in the application of game design techniques to provide effective design ideas and reasonable technical support for the game art and design. Zeng Xiaodong analyzed stripping game and traditional narrative to learn, summarized game aesthetics research classification and development direction of context.

The above researches mere studied digital games from a specific perceive: such as aesthetic point of view, the structure of the game, aesthetic features, narrative methods etc. However, they ignored the important role played by digital games now as cultural media in the social-cultural content. Accordingly, this paper will elaborate cultural characteristics of digital games from mass media and popular culture, international communication and others.

2 Digital Games in Social Context

Kant, German classical aesthetics researcher, explicitly studied the relationship between the game and the art, and Freud also once said: "Art as an imaginative creation, just like dream, is the extension and the alternative of childhood games." Thus, the game as a kind of art expression form has already been done. Digital games, as a special game form in modern science and technology society, attracts a huge number of people or players, and the number of players is rapidly increasing. At present, there are mainly two kinds of views on digital games in society. One is that digital games is regarded as a harmful poison, which makes people addicted to the virtual world, and separated from the realistic society, and leads to a series of social problems. Media often report the digital games' negative influence on society. Especially, some students frequently truant in order to play the digital games, and finally which leads not to finish the normal study, and even some teenagers commit crimes as addicted to the games. The other view is that digital games is considered as a new cultural phenomenon under the development of modern science and technology, which are involved in all aspects of social life. Digital games can bring enormous wealth and social stability, and provide many employment opportunities, even some wonderful games can be used as a course in the class so that the students get the relevant knowledge from them. Whether there is controversy about digital games in society or not, the idea is widely recognized that digital games can be regard as a comprehensive art form integrated with music, painting, literature and other art elements. Although digital games bring the various positive and negative influence on the society, huge commercial development prospect is still attracting capital investment, and many colleges and universities worldwide have also created game design discipline and conveyed a variety of game design talents to the society. At present, digital games and various views on the games exist simultaneously in society, which is the real and main ecological environment of digital games in the social context.

3   Mass Communication Media

Digital games as a mass art has more influence on the player's thoughts and life than any other art forms. The contents of the games show the game designer's thought, emotion and values, the current society can be observed through the certain games and games industry phenomenon. In China, so many copied games disclose that all the society must be anxious to achieve quick success and get instant benefits. From the game itself, every game has its own world view. The contents, images and game mechanics in the game "journey" created by Chen, show the game designer's love of the game itself. Obviously, digital games can be regarded as a medium for the spread of culture.

The history of human information dissemination can be roughly divided into six stages. The first phase is the posture and the signal era, which is also known as the spread of the beginning. In this phase, human mainly use the voice, the expression and body movements as a basic mean for communication. The second stage is the production of language, which is also called the era of speech and language, language of this kind of symbol system plays an important role in greatly promoting the development of the whole human thinking ability and the progress of human society, and also makes the human culture accumulate, quickens the speed of the development and spread. Different from the previous two stages needed face-to-face communication, the third stage is text era. The emergence of the text makes the human information spread more widely, and at the same time the texts have the function of information recording and transmission. The fourth stage is printing. The large-scale use of printing reflects the evolution of the replication technology, greatly accelerates the spread of culture and the development process. The newspapers, books and magazines rapidly expand in the society. The fifth stage is for the electronic age, which is also true era of mass communication. At the beginning of the 20th century, electronic media such as film, radio, television and other media forms mark a specific era's beginning in which the mass share information. Compared with the previous several stages, the ways of transmission become more diverse, the speed of transmission is greatly accelerated, and the information transmission is greatly beyond the limitation of the time and space. And the electric media plays a more and more important role in the process of the historical development of human society, the people all over the world can get information through the electric media. The sixth phase is the internet. The internet brings the thorough changes from one-way transmission to a two-way interactive communication, the internet connects all worldwide user who exchange the information including science and technology, literature, art, and life from the realistic society to the virtual world, the internet has the colossal impact on all aspects of the whole society.

In the internet era, digital games are regarded as the important media for the transmission of the culture and digital games is involved in digital technology and internet. It is a fact that American blizzard's world of Warcraft (wow) as an online world classics and encyclopedia in game field, is a massively multiplayer online role-playing game (MMORPG). The game has a complete set of the world, an epic game background, the fascinating story with many loyal players all over the world. The game contains hundreds of delicate, majestic and magnificent scenarios and luxurious music background, brings players perfect audio-visual art and promotes more humanity. In today's society,

although the world is becoming more and more crowded, people is becoming more and more indifferent. And Warcraft provides such a virtual world, such relatively easy audio-visual environment makes people more easily than in the real world for emotional resonance and exchanges.

Thus, digital games have four basic functions such as spreading information, guiding consensus, education, and entertainment of the masses.

4   Mass Culture

In the 19th century, culture is only understood to be just listen to the elegant music, paintings, or the opera, and is still a way to escape from the reality. By postmodernism phase, art, politics, religion and philosophy are integrated, which forms the social integration of culture, mass culture becomes popular, the distance between high culture and popular culture is disappearing. The concept mass culture early appears in the book "Civil disobedience" written by American philosopher Ortega. Mass culture is often through the mass media (film, television, radio, newspapers, magazines, etc.) to convey and performance, although mass culture temporarily overcome the vacant feeling, loneliness and the real sense of crisis in the realistic world, but it is also likely to significantly reduce the real standard of human culture, even deepen people's alienation in the history of the long term. Therefore, some scholars think of mass culture as a result of the contemporary society nationalization, and mass culture marks the historical progress in social stability, cultural equality and disintegration of discourse hegemony. While the other critics regard mass culture as the decline of human civilization, the challenge of the capital, tradition, belief and values. Some people even consider mass culture as the human spiritual opium. Mass culture represented by digital games brings many new topics to the aesthetic culture, especially in the popular art and the serious art. There are apparent differences between the popular art and the serious art in the content, forms and function. At the same time, it is not denied that digital games are not only modern mass communication media but also the mass art with huge influence, and digital games integrate communication, art, culture and other elements into a special commodity form. So, digital games should have more social value and aesthetic value, and in this way, can undertake the aesthetic education and cognitive and aesthetic entertainment and many other functions, at the same time as a medium of transmission. Sometimes, driven by commercial interests, digital games is involved in some violence, pornographic contents for the general public, especially teenagers and children, and cause extremely harmful effects. Therefore, to prevent the digital games from the violence and the pornographic is the important and permanent task in the future of digital games industry. Digital games is a kind of unity of opposites such as industry versus art and commodity values versus aesthetics values. For digital games industry, the unity is based on the economic interests. For digital games art, the unity is based on the artistic value. Only in such a tension between a digital games industry and art of the digital games, digital games can bring a change as a vulgar entertainment in people's minds.

5   International Communication

Whether the digital games have the nationality or not lies in whether it really shows the life of certain nation. National life includes two aspects such as material and spiritual, which form a specific cultural environment and a specific cultural atmosphere. Digital games, therefore, often through the scene, roles, costumes, props, weapons, game play, and many other elements, convey some national properties and values and world views, the players in the game explore the unknown world by interaction with the various elements, and think and reflect about a known world.

The game Assassin's Creed is about the story in European medieval dark crusade period, and conveys the European political, economic and cultural values, European military, the people's livelihood, and so forth. Chinese Paladin is based on the Chinese Fairy lich ghosts-gods and the games convey the Chinese nation ancient culture and values. Both the above mentioned games attract a huge number of players all over the world, through the games, the European culture and Chinese ancient fairy folks are spread widely, the players can get to another thorough different world and know a different national culture instantly. The international communication of the digital games has reached an astonishing degree, and also brings a strong impact and influence of different culture. Digital games as a mass media are exercising the social functions of international cultural exchange.

6   Conclusion

Compared with other mature art form, digital games need to promote the aesthetic values and avoid vulgar content. Although driven by commercial interests, some unhealthy contents appear in the digital games, which is not the mainstream. Players itself will identify the quality of the game and improve the aesthetic taste of digital games in contact with different games. Regarded as a kind of mass culture and mass media, digital games will have healthy development environment and help people have the right attitude towards it as a culture art goods. Digital games is become increasingly perfect and mature.

References

1. Peng, J.: Movies and TV Aesthetics. Peking University Press, China (2010)
2. Ye, L.: Foundations of Aesthetics. Peking University Press, China (2011)
3. Chen, J.: Flow in games (and everything else). Commun. ACM **50**(4), 31–34 (2007)
4. Simon, N.: What we talk about when we talk about game aesthetics. In: Proceedings of the 2009 DiGRA International Conference (2009)
5. Hunicke, R., Le Blanc, M., Zubek, R.: MDA: a formal approach to game design and game research. In: Proceedings of the AAAI Workshop on Challenges in Game AI (2004)
6. Zeng, X.: Aesthetic Thinking on Electronic Games. Hunan Normal University, China (2004)
7. Dong, C.: Games and Arts. People's Publishing House, New Delhi (2004)

Research on Initialization of 3D Hand Pose Based on User and Computer Interaction

Shichang Feng[1], Zhiquan Feng[2(✉)], and Xiaohui Yang[2]

[1] School of Information Science and Engineering,
Shandong University of Science and Technology, Qingdao 266590, China
[2] Provincial Key Laboratory for Network Based Intelligent Computing,
Jinan 250022, China
`fzqwww@263.net`

Abstract. Finding 3D hand models corresponding to the user's 3D hand pose in the initial frames makes the initialization for the 3D hand model much more significant in 3D human hand tracking. Blending computer interaction techniques and cognition theories, a novel initialization approach for 3D hand model is put forward in the present paper to make the initialization process more human-oriented. The proposed initialization process is primarily divided into three steps. The first step is the approximate classification of the user's poses as dominated by a computer. The second step is to adjust by freehand as dominated by the user. The third step is to modify the 3D hand model as dominated by the computer. The present study attempts to describe and shape the user's behavioral model, upon the initialization algorithm is designed and optimized. To improve the performance of the initialization algorithm, User experience, time cost, and accuracy are fused into an evaluation criterion for the optimization of the proposed algorithm. The main contributions of the present work consists of modeling the operator's cognitive behavior, attempting to answer why and how the cognitive behavioral model guides the proposed algorithm in assigning tasks for the initialization between human and computer. The experimental results demonstrate good performance by the proposed method and its potential applications. In addition, the proposed approach could provide the user with an easier, more pleasurable, and more satisfactory experience. The developed initialization system is successfully applied to several application systems.

Keywords: 3D human hand gesture model · Features extraction · Initialization · 3D hand tracking · Human–computer interaction

1 Introduction

The reconstruction of a 3D hand pose aims to reconstruct the initial 3D hand models, in accordance with the users' poses in the first frames of an online hand video, from which recursive hand trackers can work [1–4].

Most of the prevailing tracking approaches, such as Kalman filter (KF), extended Kalman filter (EKF), unscented Kalman filter (UKF), and particle filtering (PF) [2], are featured with recursions. The next state of the tracked freehand is calculated based on the last state. These approaches do not work without the initial states. As a result,

research on the general approach to initializing the states of tracked freehand has become significant.

Although initialization of the visual tracking system is critical in the performance of freehand tracking systems, not much is known about its process. Most of the currently available algorithms assume that the initialization is done manually.

Additionally, initialization for 3D hand models is a very complicated issue. First, the recovery of a 3D hand structure from a single 2D hand image remains to be a challenge. Second, the human hand is a typical articulated and elastic object with high dimensionality; finding the real 3D hand model from nearly unlimited hand poses is almost impossible. The present study attempts to solve these initialization issues by the interaction between human and computer.

Motivated by the human–computer interface in 3D application systems, such as the drag-and-drop system, 3D virtual assembly system, and pointing devices, freehand tracking requires a reliable initial pose in the first frame with an easier, more pleasurable, and more satisfactory user experience.

2 Related Work

A person-independent recognition method for hand postures against complex backgrounds is proposed in [5] by combining different feature types at the graph nodes. To estimate the arbitrary 3D human hand postures, Shimada [6] accepts not only pre-determined hand signs, but also arbitrary postures in a monocular camera environment. The estimation is based on 2D image retrieval. More than 16,000 possible hand appearances are originated from a given 3D shape model by rotating the model's joints. The images are then stored in an appearance dataset. Rosales [7] proposed the specialized mappings architecture (SMA) approach to map image features to likely 3D hand poses using a machine learning architecture. The hand is tracked and its 3D configuration on every frame is tracked. No any restrictions are imposed on the hand shape and no manual initialization is required. The chamfer distance, edge orientation histogram, and moment are used in [8] to estimate the 3D hand shape and orientation by retrieving appearance-based matches from a large dataset of synthetic views, which are rendered by 26 predefined prototype shapes. The hand shape in the input image is assumed to be close to one of the 26 predefined shapes. A tree-based representation [9], in which the leaves define a partition of the state space with piecewise constant density, can be applied effectively to track 3D-articulated and non-rigid motion.

The single frame pose estimation approach [10] is based on a local search and keeps track of only the best estimation at each frame. This type of tracker is expected to work well at the initialization phase, because no previous data is used. One of the distinct features of the single-frame pose estimation approach is the retrieval of hand poses from a hand image dataset. The advantage of using appearance-based matching for 3D parameter estimation is that the estimation is done indirectly, by looking up the ground truth labels of the retrieved synthetic views. This method avoids the ill-posed problem of recovering the depth information directly from the input image.

Particle filtering is a well-known technique for implementing the recursive Bayesian filters using Monte Carlo simulations. The basic idea of particle filtering is to

represent an arbitrary probability density using weighted samples drawn from another easy-to-sample density called the importance density. The weights represent the probability of occurrence of each sample and the weighted samples are usually called particles. In case of tracking, the particles of the hand configuration distribution are to be updated at each frame. In [11], the Stochastic Meta-Descent (SMD) algorithm was employed and resulted in an eight-particle tracker. This tracker tracks high-dimensional articulated structures using far fewer samples than the previous methods. Additionally, it can handle multiple hypotheses, clutter, and occlusion, with which pure optimization approaches have problems.

The current authors proposed a new method to initialize the 3D pose and position of the freehand by fusing the three techniques—interaction between human and computer, modeling cognitive behaviors for operators, and visualizing information in the initialization process [12]. Focus was placed on the development of a method for 3D hand models. A model behavior of the operator's hand was created, upon which the design of the present initialization algorithm is based. The previous research shows that behavioral models are not only beneficial to the reduction of the cognitive burden of operators, because it allows computers to cater to changes in the operators' hand poses, but also helpful in addressing the high dimensionality of an articulated 3D hand model. The experimental results also show that the method provides an easier, more pleasurable, and satisfactory experience for operators. Currently, the developed initialization system has been successfully applied to the proposed 3D freehand tracking system. A distinctive disadvantage of this method is that it does not present a way to determine the best temporal points for the interaction between human and computer, which affects the interaction's degree of harmoniousness. In fact, the boundary point between human and computer is set by trial and error. The present paper deepens and expands the previous method [12]. The main contributions of the current work lies in modeling the operator's behavior, illustrating why and how the Cognitive Behavioral Model (CBM) guides the developed algorithm in assigning tasks for the initialization between human and computer.

3 Overview

3.1 Modeling Problem

The present research attempts to find a solution of S, V, and R for the following problem (See Fig. 1):

$$\underset{V_S, R_S}{Min}\{H(V, R)\} \tag{1}$$

where R is a real hand pose and V is a virtual synthesized 3D human hand model (or 3D hand model). In expression (1), $H(.)$, denoted as the undirected Hausdorff ditance, is the cost function used to evaluate the similarity of V and R. Formula (1) means that the user's real 3D hand pose should be fitted into the 3D hand model at frame 0, starting from the frame $-S + 1$.
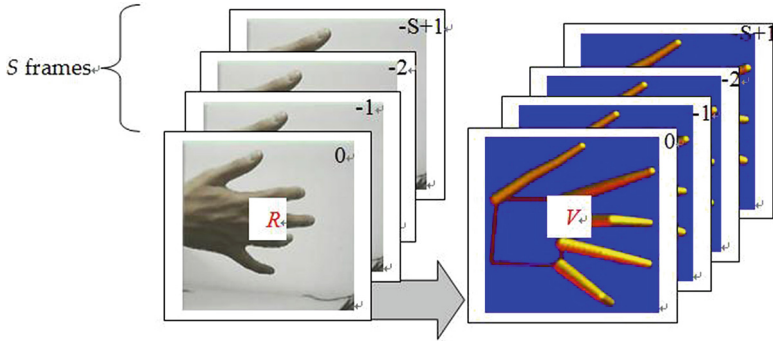
Fig. 1. This figure shows how to acquire the 3D hand model during the initial S frame images in an online video with an easy, pleasurable, and satisfactory user experience, which is the main objective of the present paper. The frame images **R** are to the left and the right images are the corresponding recovered 3D hand model **V**.

3.2   Overview

Freehand tracking using a PF tracker [2] is implemented in the present study. As soon as the initialization of hand pose is performed, the recursive system begins to work automatically, frame by frame (See Fig. 2).
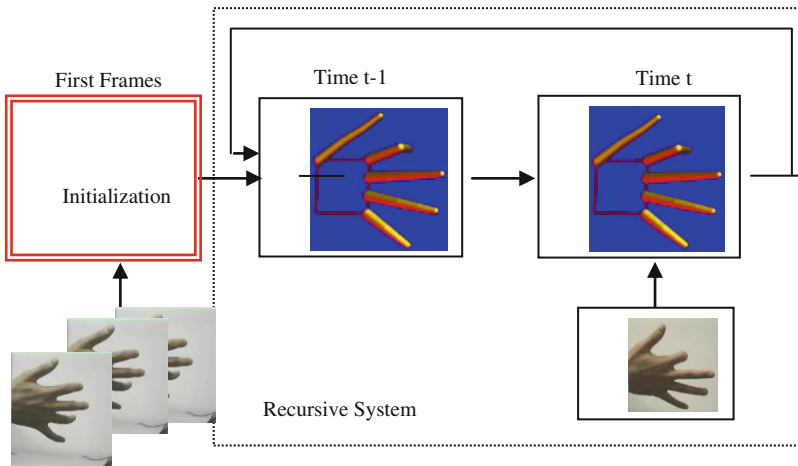
Fig. 2. Global framework of freehand tracking system, a typical recursive system. Acquisition of hand state during first frames is important. Focus in placed on initialization.

The objective of initialization for the 3D hand model is to determine a 3D hand model corresponding to a 3D hand posture during the first frame images.

As shown in related works, most of the known initialization approaches require the users' initial hand poses to be the same as those in a hand pose dataset, which means that users are asked to reconstruct the same predefined poses.
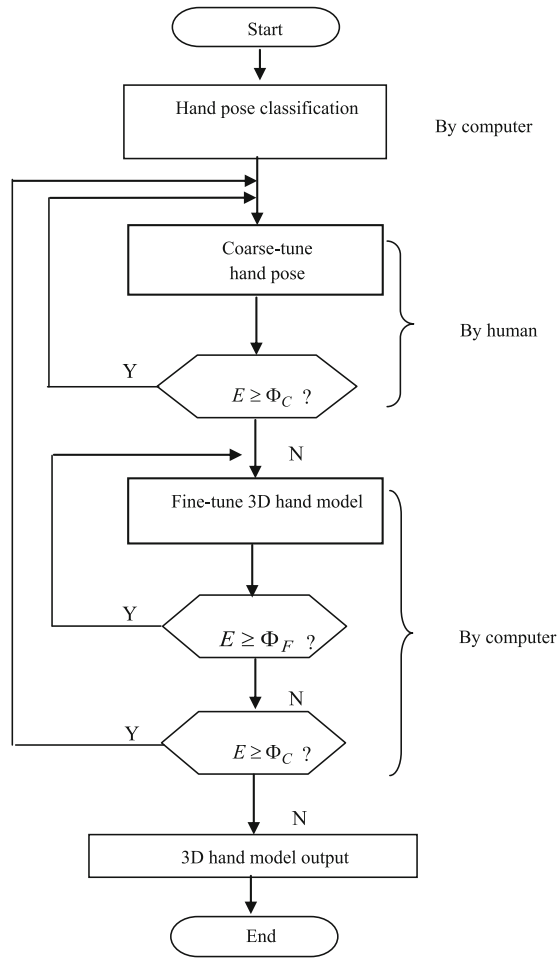
(a)
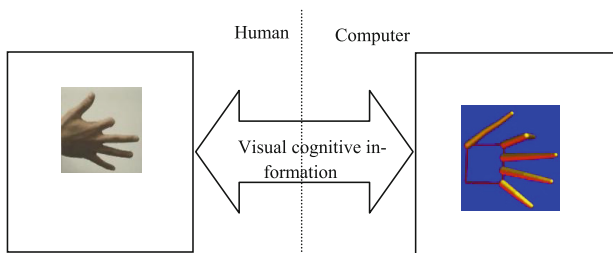
(b)

Fig. 3. Overview of OM. (a) Φ_C and Φ_F are threshold values used as temporal points between user and computer. After a hand pose is classified from the rough 3D hand model, OM performs the interaction between human and computer. (b) Framework for how human, computer, and cognitive information are integrated.

The algorithm framework of the proposed initialization method (referred to as OM) is depicted in Fig. 3. OM is composed of three phases, namely, hand pose classification, coarse-tuning the user's hand pose, and fine-tuning the 3D hand model. Figure 3 also shows how the 3D hand model is shaped by the three phases and how an initialization task is approximately assigned between user and computer for effective interaction.

The objective of hand pose classification is to determine a rough 3D hand model similar to the frame image by the retrieval from a hand posture dataset. Given a dataset of many poses, a pose that best matches the object in the input image is identified. However, the rough 3D hand model is only an initial value, $V - S + 1$, of V in Eq. (1). The objective of coarse-tuning the hand pose is to bring the user's hand image close to the projection of the 3D hand model while keeping the hand model fixed. The objective of fine-tuning the hand model is to making the hand model and 3D hand pose similar, while the hand pose is kept fixed.

To provide a new approach to initializing the 3D hand pose, the operation of which is human-oriented and convenient for online use, the three core techniques, namely, HCI in the initialization process, the visualization of cognitive information, and modeling cognitive behaviors to reduce the cognitive burden, are blended together in the proposed method. The states of temporary 3D hand models and image features from videos are fed back onto screen in the form of interactive graphics and imaging. The states are used as cognitive information upon which users adjust their behaviors. On the other hand, the computer fine-tunes the 3D hand models according to users' responses.

In Fig. 3a, both Φ_C and Φ_F are the Hausdorff distances between the projections of the 3D hand models onto the frame image and the frame image features. Φ_C determines the accuracy of coarse-tuned human hand pose and Φ_F determines the accuracy of the fine-tuned 3D hand model. In most cases, $\Phi_C \geq \Phi_F$.

To determine Φ_C and Φ_F, a behavioral model is built for users.

4 Modeling Users' Behavior

Research on cognitive behavioral model is helpful in exploring the cognitive mechanism and ways for the interaction between users and computer.

4.1 Experiment for CBM

The participators (users) were equipped with a data glove and position tracker on their hands. They would be requested to perform some initialization tasks. A synthesized 3D hand model based on data from the data glove and position tracker was displayed in the scene and each participator was requested to adjust his/her hand until it fit into the 3D hand model. In Fig. 4a, a user is doing an experiment for CBM and further analysis of the experiment is shown in Fig. 4b.

According to Fig. 4, the joint angles acutely change in one period of time, and placidly change in another period of time. The former is regarded as the coarse-tuning process, the latter as the fine-tuning process. This observation guides the introduction

(a)   (b)
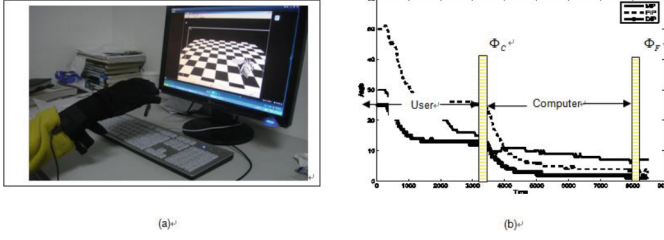
Fig. 4. An investigation of CBM for the user. (a) To reveal the CBM, the user was equipped with a data glove and a position tracker on the right hand to fulfill initialization for the 3D hand model. (b) The angle curves over the frames for the little finger in the process of grasping a virtual object; it describes the three angles per frame; the measurement unit in the y axis is the degree and the time unit is in ms. The time determined by Φ_C and Φ_F are the temporal points between user and computer; in most cases, $\Phi_C \infty \geq \Phi_F$ is satisfied. Φ_C and Φ_F are the Hausdorff distances.

of the technique of interaction between human and computer, and provides cues to assign interactive and cooperative tasks between user and computer. The CBM models are stated in the form of CBM features.

4.2 CBM Features

The main findings, also called CBM features, in the present study are presented as follows:

- CBM-0: The process of adjusting the pose can be divided into different sections, in which one features a large range regulation in the frontal period of time and another features a small range regulation in the back period of time.
- CBM-1: There are two paths that keep the hand pose consistent with the 3D hand model to maintain low cognitive burdens.
  Path-1: Initial pose → adjust the hand pose orientation → translate the hand pose → 3D hand model is slightly tuned, or
  Path-2: Initial pose → translate the hand pose → adjust the hand pose orientation → 3D hand model is slightly tuned.
- CBM-2: The variables set, x, in hand model X can be described with the unchanging variables as

$$X = U \cup (I - U) \tag{2}$$

where U refers to the set composed of unchangeable variables, I refers to the set composed of all variables in hand model X. Taking the fist-style pose as an example, in the whole process of initialization, all local angles are kept fixed. Set U can be described manually in the hand pose dataset.

- CBM-3: Compared with the approach asking users to imagine and shape their required hands according to required regulations or constraints, the approach allowing users to actively adjust their hand poses based on visual cognitive information onscreen is more concurrent with the users' cognitive customs and costs less cognitive burdens [12].

The CBM features lead to the best way to blend both HCI and visualization into the initialization process, with the purpose of maintaining low cognitive burdens for users.

5 Tune Users' Hand Poses

Based on CBM-0, the initial hand pose may be far from the 3D hand model, even if the rough pose is acquired from the retrieval hand pose dataset in the classification process. That is, finding the solution for formula (1) may require a computationally expensive search. Accordingly, the very interesting issue of how to use the CBM Features to guide the interaction between human and computer is discussed to adjust the user's hand pose for low cognitive burden and design the 3D hand model using cheap computation.

The superiority of the user over the computer is that he/she can flexibly make the decisions for all situations in completing a cognitive task, which is the reason why the HCI technique is introduced into the developed initialization system. The frontal period of time is assigned to a user. That is, the user adjusts his/her hand's position and pose in agreement with the CBM-1 Feature to superpose the online hand images onto the projection of the given initial 3D hand model while keeping the 3D hand model fixed.

The effectiveness of the HCI, to some extent, depends on the path of feedback and the visualization style of the hand image and the 3D hand model. The present study uses many approaches for visualization. For example, the 3D hand model is rendered and outputted by OpenGL, and the new hand images are displayed in real time with a visual style.

Clearly, the developed initialization process starts from an initial pose similar to that in a dataset. However, the final 3D hand model may not be limited to the initial pose.

6 Tune 3D Hand Model

The present study does not rely only on the user to align his/her hand to the hand shape displayed. In fact, the back period of time is assigned to the computer. Once the user's hand is close enough to the 3D hand model, the computer provides the user with feedback, flickering and highlighting, and begins to fine-tune the 3D hand model using the approach similar to PF [2]. Taking Fig. 5b as an example, the duration from 0 to 3700 ms is called the frontal period of time, and the duration from 3700 ms to 8500 ms is called the back period of time. Φ_C is the Hausdorff distance at time 3700 ms and Φ_F is the Hausdorff distance at time 8500 ms.

In this process, the CBM-2 feature follows. Although a coarse pose is obtained by the coarse-tuning process, because of high dimensionality problem, the computer

would have difficulty further fine-tuning the 3D hand model in rapid speed until the latter is the same as the user's 3D hand pose. Fortunately, the CBM-2 feature can help to alleviate this problem. For example, if the computer knows or predicts that some part of the variables in a hand model vector would change over time, it will focus on determining the values of these variables and pay little attention to the unchanged variables. This approach is equivalent to reducing the dimensionality of the hand pose vector.

Suppose X0 is the retrieved virtual initial 3D hand model from hand pose dataset, I is the hand frame image at the current frame, Ω is the feature set extracted from image I, which is composed of the contour, fingertips, roots of fingers, joints, and the intersection of the knuckle on the fingers. According to the proposed method, the feature extraction process is composed of two steps, namely, the coarse location phase (CLP) and the refined location phase (RLP), from coarseness to refinement. In the CLP phase, the hand contour is approximately described by a polygon with concave and convex. An approach to obtaining the hand shape polygon using locating points and locating lines is meticulously discussed. Subsequently, using a coarse location algorithm, the contour, fingertips, roots of fingers, joints and the intersection of the knuckle on different fingers can be extracted. In the RLP phase, a multi-scale approach is applied to the extracted features in the CLP phase by defining the response strength of different types of features; the accurate features can then be obtained.

The algorithm for adjusting the 3D hand model based on the CBM is presented as follows.

(A) The empirical values of Φ_C are determined by trial and error.

(B) For the user: Change hand shapes and hand positions.

(B.1) The user moves his/her hand along path-1 or path-2, as stated in CBM-1, while the 3D hand model is unchanged. The 3D hand model and hand image features are synchronously visualized onscreen.

(B.2) The user adjusts his/her hand postures according to the visualized cognitive information, such as the projection of the 3D hand model, hand image features, and other feedback.

(B.3) E is evaluated by

$$E = Hausdorff(X_0, \Omega) \tag{3}$$

where Hausdorff (X_0, Ω) is the Hausdorff distance [13] between the 3D hand model X_0 projection onto the hand image and the hand image features.

(B.4) If $(E \geq \Phi C)$ the computer feedbacks with a flicker and highlight, and goes to step (B) else X1 ← X0.

(C) For the computer: Fine-tune the 3D hand model with PF tracker.

(C.1) Superpose the 3D hand model projection onto the hand image.

(C.2) Sample

Generate N particles X_1^i of X_1, i = 1, 2, …, N, using the Gaussian distribution in agreement with CBM-2.

(C.3) Compute weight ω for each particle

(C.4) State Updating

$$X = \sum_{i=1}^{N} \omega_i X_1^{(i)} \tag{4}$$

(C.5) Evaluate E by

$$E = Hausdorff(X, \Omega) \tag{5}$$

If ($E \geq \Phi F$), X1 ← X and go to step (C).
If ($E \geq \Phi C$), X0 ← X and go to step (B).
Output X.

By looping between steps (B) and (C), the optimization function (1) could be satisfied approximately. Φ_C and Φ_F are threshold values used to control the accuracy of the coarse-tuning and fine-tuning processes, respectively. One of the hardest part in the pose recovery process is tracking under the pose, which is tackles by PF tracking. The role of CBM-1 in the OM is to reduce the cognitive burdens for the user. CBM-2 is used into the OM to avoid sampling for the unchanging parts in the 3D hand model.

7 Optimization of OM

One of the important issues in the OM is the way to choose the parameters Φ_C and Φ_F, by which the OM will be optimized in this section.

Suppose L algorithms, M1, M2, …, ML, are applied to the same initialization task. Ta is the average time cost of the L algorithms; Ak and Tk are the accuracy and time cost of the kth algorithm, where $1 \leq k \leq L$; α_k is the degree of harmoniousness (DOH) determined by the users, which reflects the users' cognitive burden.

Cognitive burden is evaluated by four factors in the present study: tiredness, joviality, freedom, and workability. Tiredness describes the extent of toil a user feels in the initialization process; joviality describes the degree of amusement the user feels; convenience describes the suitability to the user's purposes; and workability describes the extent to which the initialization approach is feasible. The four factors have a scale between 0 and 100, and are scored by the users.

The performance, as one of the evaluation criterions, is defined as formula (6):

$$\lambda_k = \alpha_k T_a \frac{A_k}{T_k} \tag{6}$$

where α_k is the average of J, the joviality, convenience, and workability of the algorithm k. Here, J is the difference between 1 and degree of tiredness.

One of the key issues in evaluating system performance is the availability of ground-truth data. Obtaining ground truth data for the 3D hand pose estimation is a difficult problem [4]. The widely used method to evaluate the 3D hand model is to project the hand model onto the input image to show how well the projection matches the image data, which can be measured by Hausdorff distance.

Accuracy is defined as

$$A_k = e^{-\beta H_k} \tag{7}$$

where Hk is the Hausdorff distance of the algorithm k and β is an experiential constant, which is set at 0.01 in the present experiments.

According to the definition of λ in formula (7), if the time cost of an algorithm is equal to the average time cost, the algorithm is evaluated by accuracy if the parameter α is ignored. On the whole, a large λ_K means high accuracy and low time cost.

The objective of the OM's optimization is to find the threshold values Φ_C and Φ_F to maximize λ_{OM} or

$$\arg(\max(\lambda_{OM})). \tag{8}$$

8 Experimental Results

8.1 Experimental Settings

The present study uses a color CCD camera ZT-QCO12 with a 4 mm lens that captures a 640×480 video at 30 Hz. The computer used has an Intel(R) Core(TM) 2 Quad CPU with a 2.66 GHz processor and 3.25 GB memory. A 26 DOF 3D hand model is used, with 6 DOFs for the global transformation and four DOFs for each finger. The length of each knuckle on the fingers and the size of the palm are fixed.

8.2 Experimental Procedures and Results

The experimental process is composed of three phases, namely, hand pose recognition, coarse-tuning, and fine-tuning. There are many hand pose recognition approaches, but most of them have difficulty effectively differentiating between two similar hand poses with the same protrudent fingers. Recognition involves finding the location in the input image that best match occurs. Density distribution features (DDF) [14] are used to describe the hand image features upon which the initial pose is retrieved from the dataset. DDF is defined by the following formula

$$DDF = (\rho_1, \rho_2, \ldots, \rho_N; \delta_1, \delta_2, \ldots, \delta_N) \tag{9}$$

in which the first feature vector represents the relative density of object pixels within each sub-image and the second represents the difference of relative density in the direction of radial coordinates. DDF is invariant to translation, scale, and rotation. In formula (9),

$$\rho_i = n_i/n \tag{10}$$

$$\delta_i = \begin{cases} |\rho_2 - \rho_1| & i = 1 \\ |2\rho_i - \rho_{i-1} - \rho_{i+1}| & 1 < i < N \\ |\rho_N - \rho_{N-1}| & i = N \end{cases} \qquad (11)$$

where n_i, $(i = 1, 2, \dots, N)$ is the total number of pixels of the hand skin in the ith circumcircle and n is the maximum of all n_i.

The OM is compared with two widely used methods, namely, the single frame (hereafter simplified to SF) pose estimation approach [10] and the SMD approach [11]. The first experiment is done using the OM. The number of particles used is 50. The proposed algorithm is evaluated using the same user, and different users. An experiment suing an optimized OM then follows.

Figure 5 is the $\lambda_{OM} - \Phi_C$ curve of the OM algorithm. Based on Fig. 5, the threshold value Φ_C has an effect on performance λ of the OM, as defined in formula (8). If Φ_C is too small or too big, the performance of the OM will decreases, and a maximum exists on the $\lambda_{OM} - \Phi_C$ curve. In the current experiment, when $\Phi_C = 36.96$, λ_{OM} reaches its maximum.

The reason for the above fact is that if Φ_C becomes smaller, the cognitive burdens become bigger; as a result, the parameter α_k becomes smaller. If Φ_C becomes bigger, the accuracy will become smaller; as a result, the parameter Ak becomes smaller. Thus, the performance parameter λ_k becomes smaller under either of the two situations.

Thus, the OM can be optimized by the optimized threshold value Φ_C. The performances of the SMD, SF, and optimized OM by Φ_C are shown in Fig. 6.

Clearly, of the three algorithms, the optimized OM by Φ_C has the best performance.
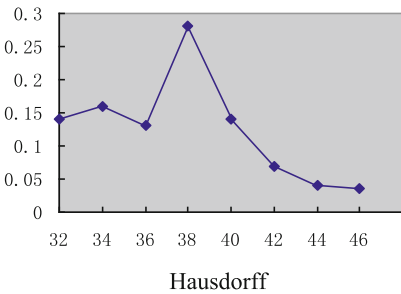
Fig. 5. Hausdorff-λ curve of OM. Here, the Hausdorff distance is the threshold value Φ_C used in OM. When $\Phi_C = 36.96$, λ_{OM} reaches its maximum.
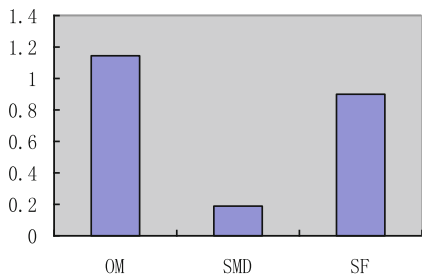
Fig. 6. Performance of three algorithms after OM is optimized by Φ_C.

For further comparison of the OM and optimized OM by Φ_C, the two algorithms are run 50 times each with the same user. The average results are shown in Table 1. The bigger the λ, the smaller the cognitive burden.

For 50 different users, the performance comparison between the OM and optimized OM by Φ_C are shown in Table 2.

Table 1. Performance comparisons of OM and optimized OM by Φ_C. The results are an average of 50 times with the same user.

Performance / Algorithm	Time Cost (ms)	Accuracy	λ
OM	1265.953	0.698	1.361
OM and optimized OM by Φ_C	955.200	0.715	1.430

Table 2. Performance comparisons of OM and optimized OM by Φ_C. The results are an average of 50 times with the same user.

Performance / Algorithm	Time Cost (ms)	Accuracy	λ
OM	1915.3	0.6896	0.7052
OM and optimized OM by Φ_C	2023.59	0.7146	0.80316

Table 3. Comparison of OM and optimized OM by Φ_C and Φ_F

Performance / Algorithm	Time Cost (ms)	Accuracy	λ
OM	1915.30	0.720	0.7368
OM and optimized OM by Φ_C	2023.59	0.744	0.8364

According to Tables 1 and 2, the optimized OM by Φ_C has better performance than the OM.

After the Φ_C is fixed, Φ_F can be determined by a similar method as Φ_C.

The comparisons for the time cost, accuracy, and λ of the OM and optimized OM by both Φ_C and Φ_F are shown in Table 3. The DOHs of the different methods are shown in Table 4.

Table 3 shows that the time cost by the optimized OM by Φ_C and Φ_F increased, and its global performance also increased. Table 4 shows that the optimized OM by Φ_C and Φ_F imposes the least cognitive burden for users.

Table 4. Comparison of DOH of the several algorithms

Algorithm	Optimized OM by Φ_C and Φ_F	OM	SMD	SF
α(DOH)	0.85	0.732	0.624	0.647

At the end of this section for the OM, the dependency of Φ_C and Φ_F on the participants and on the complexity of the pose are discussed. Φ_C and Φ_F are determined by formula (9). For different participants or different poses, the parameters Ta, Ak, Tk, and α_k are different. Thus, Φ_C and Φ_F are dependent on the participants and the complexity of the pose. In fact, let 50 different participants do the same or not the same poses for initialization, the variances of Φ_C and Φ_F are large and change in a large scale. However, for the same participant, the variances of Φ_C and Φ_F are small.

The OM and optimized OM by Φ_C and Φ_F expands the idea proposed in the present paper [12]. In the same experimental conditions, the performance is listed in Table 5.

Table 5. Comparison OM with the approach presented in paper [12]

Performance / Algorithm	Time Cost (ms)	Accuracy	λ
OM	1915.3	0.6896	0.7052
OM and optimized OM by Φ_C and Φ_F	1823.59	0.7146	0.80316
Previous Method [32]	2535.3	0.6392	0.6988

The OM and optimized OM by Φ_C and Φ_F perform the same initialization task with less time, higher accuracy, and higher λ, compared with the previous work [12].

9   Discussion and Conclusions

Compared with the previous related work [12], the main contributions of the present study involves the attempt to theoretically explore the fundamental "why, how, when (WHW) problems" in the process of fusing the three core techniques, namely, the HCI for initializing, visualization of the 3D hand model, and modeling the operator's cognitive behavior. Thus, the present study shows why and how the CBM guides the

proposed algorithm in assigning tasks for the initialization between human and computer, and informs when the temporal points between users and computer occur. These benefits are summarized as follows.

References

1. Erol, A., et al.: Vision-based hand pose estimation: a review. Comput. Vis. Image Underst. **108**, 52–73 (2007)
2. Julier, S.J., Uhlmann, J.K.: A new extension of the Kalman filter to nonlinear systems. In: Proceedings of of AeroSense: The 11th International Symposium on Aerospace/Defence Sensing, Simulation and Controls, pp. 182–193, SPIE, Orlando, Florida, USA (1997)
3. Salih, Y., Malik, A.S.: Comparison of stochastic filtering methods for 3D tracking. Pattern Recogn. **44**(10–11), 2711–2737 (2011)
4. Erol, A., Bebis, G., Nicolescu, M., Boyle, R., Twombly, X.: A review on vision-based full DOF hand motion estimation. In: Proceedings of the IEEE Workshop on Vision for Human-Computer Interaction (V4HCI), pp. 15–22, San Diego, California (2005)
5. Triesch, J., von der Malsurg, C.: A system for person-independent hand posture recognition against complex background. IEEE Trans. Pattern Anal. Mach. Intell. **23**(12), 1449–1453 (2001)
6. Shimada, N., Kimura, K., Shirai, Y.: Real-time 3-D hand posture estimation based on 2-D appearance retrieval using monocular camera. In: Proceedings of International Workshop RATFG-RTS, pp. 23–30 (2001)
7. Rosales, R.: The specialized mappings architecture with applications to vision-based estimation of articulated body pose. Ph.D. thesis, BOSTON University Graduate School of Arts and Sciences (2002)
8. Athitsos, V., Sclaroff, S.: Estimating 3D hand pose from a cluttered image. In: Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 432–439 (2003)
9. Stenger, B., Thayananthan, A., Torr, P.H.S., Cipolla, R.: Filtering using a tree-based estimator. In: Proceedings of IEEE International Conference on Computer Vision, vol. 2, pp. 1063–1070 (2003)
10. Tomasi, C., Petrov, S., Sastry, A.: 3D Tracking = Classification + Interpolation. In: Ninth IEEE International Conference on Computer Vision, pp. 1441–1448 (2003)
11. Bray, M., Koller-Meier, E., Gool, L.V.: Smart particle filtering for 3D hand tracking. In: Sixth IEEE International Conference on Automatic Face and Pose Recognition, pp. 675–680. IEEE Computer Society, Los Alamitos, CA, USA (2004)
12. Feng, Z., Zhang, M., Pan, Z., Yang, B., Xu, T., Tang, H., Li, Y.: 3D-freehand-pose initialization based on operator's cognitive behavior models. Vis. Comput. **26**(6–8), 607–617 (2010)
13. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Trans. Pattern Anal. Mach. Intell. **15**(9), 850–863 (1993)
14. Liu, H., Feng, S., Zha, H.: Document image retrieval based on density distribution feature and key block feature. In: Proceedings of 8th International Conference on Document Analysis and Recognition, pp. 1040–1044, Seoul, Korea (2005)

The Recognition of Human Daily Actions with Wearable Motion Sensor System

Benyue Su[1,2], Qingfeng Tang[1,2], Guangjun Wang[1,2], and Min Sheng[2,3(✉)]

[1] School of Computer and Information, Anqing Normal University, Anqing, China
[2] The University Key Laboratory of Intelligent Perception and Computing of Anhui Province, Anqing Normal University, Anqing, China
msheng0125@aliyun.com
[3] School of Mathematics and Computational Science, Anqing Normal University, Anqing, China

Abstract. This paper develops a method for recognition of human daily actions by using wearable motion sensor system. It can recognize 13 daily actions with the data in WARD1.0 efficiently. We just extract 11 features including the means and variances of vertical acceleration data of five sensors and the mean of horizontal angular speeds of the waist sensor. Then we randomly select 80 % of the samples as the training set, and the remaining samples as the test set. By removing the abnormal samples based on the confidence interval of the distance among the same type samples and using the SVM as the classifier, we present a new method for recognition of the human daily actions. Moreover, we optimize the parameters of SVM with K-CV (K-fold Cross Validation) method. The results of the experiments show that the proposed method can efficiently identify the 13 kinds of daily actions. The rate average recognition can approach to 98.5 %.

Keywords: Wearable motion sensor system · Recognition of actions · Feature extraction · Abnormal sample · Parameter optimization · SVM

1 Introduction

From a macro perspective, the human behavior can include the poses and movement processes of head, limbs and body, and it is an interactive form of expression between human and the surrounding environment. Human behavior is researched in many areas including psychology, medicine, sport science and so on. In recent years, with the development of science and technology, motion capture technology is increasingly mature, which can be more convenient to get the data of human behavior and provide a good platform for researching human behavior. Both the motion capture system based on optical principle in the early stage and the device based on MEMS in recent years have been widely used in: movies, cartoons, medical rehabilitation, auxiliary movement, etc. [1]. But compared to the motion capture system based on optical principle, the data of wearable motion sensor system can basically reflect the characteristics of the human movement. That offers a possibility to analyze the nature of human behaviors.

Generally speaking, the algorithm of actions recognition mainly includes two parts: feature extraction and classifier recognition. The feature extraction module is the process of reprocessing for the collected data. It can excavate the deep characteristics of human movement. The commonly characteristics cover: mean [1, 2, 4, 5], variance [3–6], coefficient of association [7–9], energy [10], the coefficient of FFT [1, 7], etc. The function of classifier recognition module is building a classifier by training set and using test set to examine the efficiency of the classifier. The usual classifiers contain: decision tree classifier [4], Naive Bayes classifier [8], SVM classifier [1, 4, 9, 11, 12], etc.

In recent years, a lot of works have been done to the recognition research of human daily actions by the scholars. Yang et al. [13] proposed a distributed recognition framework to classify continuous human actions, called distributed sparsity classifier (DSC) using sparse representation. Bao and Intille [8] used decision tree classifier to recognize 20 kinds of daily behaviors. The RVM as the classifier and the characteristics of mean, skewness, peaks of FFT and some other characteristics were extracted by He [1]. Lu et al. [14] puts the new and traditional features together, and builds a new SVM classifiers. But these works usually have two shortages: at first, the selecting characteristics are too many in some papers, even it can reach to several hundreds. Although the researchers have adopted PCA [1, 15], LDA [1] and other ways to reduce dimensions, but it has increased the complexity of the algorithm and reduced the efficiency of the classifier. Secondly, for the random of choosing the training set, it will affect the capacity of the classifier with doing not remove the bad samples.

Table 1. The description of WARD 1.0

Name of action	Description of action
1. Rest at standing	The subject stands still for more than 10 s
2. Rest at sitting	The subject sits still for more than 10 s
3. Rest at lying	The subject lies still for more than 10 s
4. Walk forward	The subject walks straight forward for more than 10 s
5. Walk forward left-circle	The subject walks in counter-clock circle for more than 10 s
6.Walk forward right-circle	The subject walks in clock circle for more than 10 s
7. Turn left	The subject stays at the same position and turn left for more than 10 s
8. Turn right	The subject stays at the same position and turn right for more than 10 s
9. Go upstairs	The subject goes up one flight
10. Go downstairs	The subject goes down one flight
11. Jog	The subject jogs straight forward for more than 10 s
12. Jump	The subject stays at the same position and jump for more than 5 times
13. Push wheelchair	The subject pushes a wheelchair/walker for more than 10 s

In this paper, we adopts the WARD1.0 behavior database in the experiment. WARD1.0 was constructed with using a wearable motion sensor network by Yang et al. [13] from University of California at Berkeley in 2008. They used five terminals, locating left forearm, right forearm, waist, left ankle and right ankle to capture the behavior of human. Each terminal has a triaxial accelerometer and a biaxial gyroscope. Each axis is reported as a 12bit value to the node, indicating values in the range of $\pm 2g$ and $\pm 500°/s$ for the accelerometer and gyroscope. So each sensor terminal will produce a data vector with 5 dimensionalities at each time point. The database took 7 females and 13 males with ages ranging from 19 to 75. Everyone has been collected 13 daily actions, each action runs 5 times. The specific contents are shown in Table 1.

2   Description of Algorithm

WARD1.0 has 1300 samples and contains 13 classes of actions. We extract 11 features from each sample. Then partial samples are chose as the training set, and the rest will be as the test set. We use the K-CV (K-fold Cross Validation) [16] method to optimize the parameters of SVM in the MATLAB environment. Then we take training set to build the classification model with SVM. At last, test set will be used to check the efficiency of the classifier. The detailed process will be introduced in the following. The rough process is shown in Fig. 1.
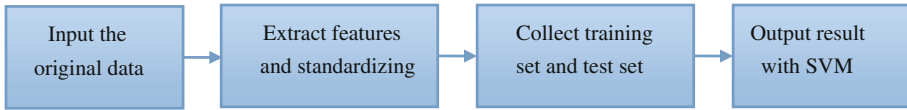
Fig. 1. The rough process of the algorithm

2.1   Extract Features of Behaviors

This paper extracts 11 kinds of features: the mean and variance values of vertical acceleration data of 2 wrist sensors; the mean and variance values of vertical acceleration data and mean values of horizontal angular speed data of waist sensor; the mean and variance values of vertical acceleration data of 2 ankle sensors.

Variance value reflects the dispersion degree of data, and the mean value is an indicator of central tendency. So variance and mean values can express the range and acuteness of actions to a certain degree. In these 13 kinds of actions, there are 3 static behavior, the others are dynamic behaviors. The variance values of each behavior are different, and the acuteness of the first three static behaviors is much smaller than others'. Meanwhile, the movement ranges of the 11th and 12th action are larger than the others'. Waist is the location of centre of gravity which can reflect the characteristics of the nature of human movement. Figures 2 and 3 show the variance and mean values of waist sensor:
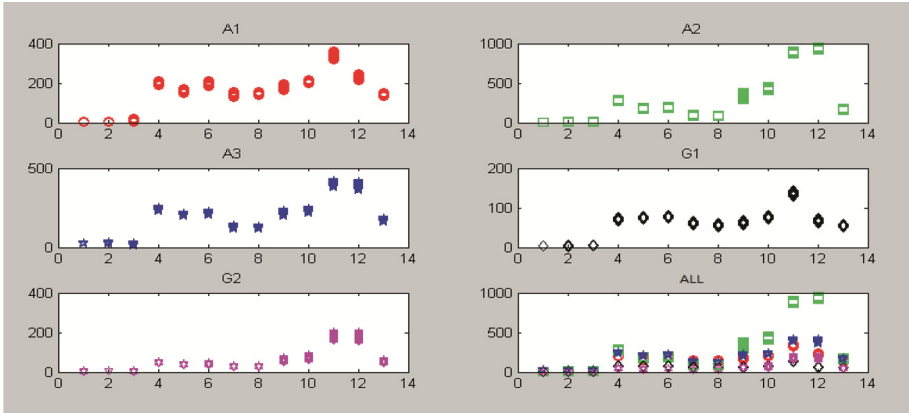
Fig. 2. The variance values of waist sensor data
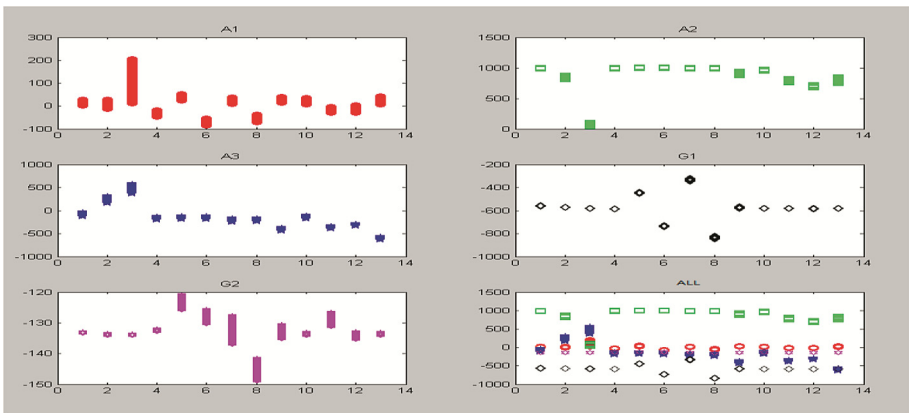
Fig. 3. The mean values of waist sensor data

When body turns from standing to lie down, the vertical direction of the waist sensor will change to horizontal direction almost. So the means values of vertical acceleration data approach about 1000 (1 g), but the 3rd' will close to 0. (It can be judged that A2 represents vertical acceleration data).

Then the mean values of A2 in Fig. 2 express the deflection range of sensors relative to the vertical direction. For what has been discussed above, firstly the mean and variance values of vertical acceleration data of 5 sensors will be extracted to express the range of body relative to the deflection of the vertical direction. Furthermore, the deflection of the horizontal direction should not be neglected. Specially, it should not be neglected in four behaviors with the 5th (Walk forward left-circle), the 6th (Walk forward right-circle), the 7th (Turn left) and the 8th (Turn right). Here, the 5th and the 7th behaviors are similar, the 6th (Walk forward right-circle) and the 8th (Turn right) are similar. We find that the horizontal angular speed values of the 5th and the 6th are clearly smaller

Table 2. The result of 10 characteristics

	1	2	3	4	5	6	7	8	9	10	11	12	13	result
1	190	10	0	0	0	0	0	0	0	0	0	0	0	0.95
2	50	150	0	0	0	0	0	0	0	0	0	0	0	0.75
3	0	6	194	0	0	0	0	0	0	0	0	0	0	0.97
4	0	0	0	178	16	0	0	0	0	3	1	0	2	0.89
5	2	0	0	11	176	0	8	0	0	2	0	0	1	0.88
6	1	3	0	26	0	146	0	5	0	1	4	0	14	0.73
7	41	0	0	0	11	0	148	0	0	0	0	0	0	0.74
8	33	0	0	0	0	0	0	167	0	0	0	0	0	0.835
9	0	0	0	90	22	1	0	0	40	20	0	14	13	0.2
10	0	0	0	62	5	0	0	0	9	108	2	9	5	0.54
11	0	0	0	0	0	0	0	0	0	0	193	7	0	0.965
12	0	0	0	0	0	0	0	0	0	4	0	196	0	0.98
13	20	0	0	0	0	0	0	0	0	0	0	0	180	0.9
overall	190	150	194	178	176	146	148	167	40	108	193	196	180	0.795

Table 3. The result of 11 characteristics

	1	2	3	4	5	6	7	8	9	10	11	12	13	Result
1	182	18	0	0	0	0	0	0	0	0	0	0	0	0.91
2	43	157	0	0	0	0	0	0	0	0	0	0	0	0.785
3	0	9	191	0	0	0	0	0	0	0	0	0	0	0.955
4	0	0	0	169	21	4	0	5	0	1	0	0	0	0.845
5	0	0	0	67	108	3	4	16	0	2	0	0	0	0.54
6	0	5	0	93	69	19	5	4	0	0	1	0	4	0.095
7	94	0	0	4	2	0	75	23	0	2	0	0	0	0.375
8	71	0	0	0	1	0	2	126	0	0	0	0	0	0.63
9	0	0	0	78	21	1	0	3	57	17	0	13	10	0.285
10	0	0	0	48	14	1	0	0	7	125	0	5	0	0.625
11	0	0	0	0	0	0	0	0	0	0	196	4	0	0.98
12	0	0	0	0	0	0	0	0	0	8	0	192	0	0.96
13	7	0	0	13	19	2	2	15	0	0	0	0	142	0.71
overall	182	157	191	169	108	19	75	126	57	125	196	192	142	0.669

than the 7th and the 8th. So we use the mean values of horizontal angular speed data of waist sensor (G1 shown in Fig. 3) to express the range of body relative to the deflection of the horizontal direction. The recognition results using the features of the 10 and 11 dimensions data respectively are shown in Tables 2 and 3 with the only difference just in containing or not the mean values of horizontal angular speed data of waist sensor.

In Table 2, we can find that the recognition results of the 5th (Walk forward left-circle), the 6th (Walk forward right-circle), the 7th (Turn left) and the 8th (Turn right) behaviors are dissatisfactory. The 5th behavior is often mistakenly judged into the 4th behavior, the 6th behavior is often mistakenly judged into 4th or the 5th behavior, the 7th behavior is often mistakenly judged into the first or the 8th behavior, and the 8th behavior is often mistakenly judged into the first behavior. After the mean values of G1 are added to the features vectors, the recognition results of the 5th, the 6th, the 7th and the 8th behaviors are improved obviously (shown in Table 3).

2.2 Choose Training Set and Test Set

When constructing the classification model, we need to choose the appropriate training samples. The requirement of training set is as far as possible full uniform containing various kinds of behaviors. Now, it has 1300 samples in WARD1.0 database, and each kind of action has 100 samples, with running 5 times for each action of 20 persons. When

we choose the training set, it should include each type of action of each person. So the proportion of training set and test set often can be identified as: 1:4, 2:3, 3:2 and 4:1. Then we contrast the results of these four circumstances in the experiment with choosing samples randomly (shown in Table 4).

Table 4. The results of different proportion

Proportion	1:4	2:3	3:2	4:1
Result	0.68538	0.70949	0.73077	0.79154

As can be seen in Table 4, when the proportion of training set and test set is 4:1, it can achieve the best results.

2.3 Optimize the Parameters of SVM

In the over section, we use the SVM as the classifier, but the result is ordinary case. We know that the parameters of SVM should be optimized with penalty parameter c and kernel function parameter g. In general, there is no best way to choose the optimal parameters of SVM. This paper uses K-CV method to train and acquire the best parameters c and g. According to the K-CV method, we separate all samples into five groups on average. Each group treated as the test set respectively, the rest of the four groups treated as the training set. So we can get 5 models, and use the average of the result of the 5 models as the standard of the K-CV. All in all K-CV can avoid over-learning and less-learning, thus the result is convictive.

At first, we let the parameters c and g change in a certain range. It will produce several pairs of point set based on the specific rule. Each pair of point set will induce the classifier index by K-CV method, and then we save the parameters c and g of the best classifier index as the optimal parameters of SVM. The rough process is shown in Fig. 4.

Fig. 4. The process of K-CV

2.4 Denoise the Sample Data

In the process of data collection, the collected data have certain errors and the mistakes due to various factors, so the samples are not entirely accurate. Chen and Li [17] denoises the sample data for large-scale sample set by defining the importance of samples; Liu et al. [18] proves the importance of boundary samples by analyzing the influence of different training sample set for recognition system.

The classifier is affected by the existent of noisy samples. So when choosing the training set and test set, we should get rid of the noisy samples as many as possible. The next will introduce the sample de-noising.

Defining a matrix A for the sample set and including k classes:

$$A = [A_1, A_2, \cdots, A_k] \tag{1}$$

If the ith class samples contain n samples, then:

$$A_i = [a_{i1}, a_{i2}, \cdots, a_{in}] \tag{2}$$

Extracting 11 features for each sample, the ith class samples can be expressed:

$$A_i = \begin{bmatrix} T1_{i1} & T2_{i1} & \cdots & T11_{i1} \\ T1_{i2} & T2_{i2} & \cdots & T11_{i2} \\ \vdots & \vdots & \ddots & \vdots \\ T1_{in} & T2_{in} & \cdots & T11_{in} \end{bmatrix} \in R^{n \times 11} \tag{3}$$

Defining the distance between a_{i1} and a_{i2} as $d(a_{i2}, a_{i2})$. Defining the among-class distance of all samples on the ith class samples:

$$D = \begin{bmatrix} d(a_{i1}, a_{i1}) & d(a_{i1}, a_{i2}) & \cdots & d(a_{i1}, a_{in}) \\ d(a_{i2}, a_{i1}) & d(a_{i2}, a_{i2}) & \cdots & d(a_{i2}, a_{in}) \\ \vdots & \vdots & \ddots & \vdots \\ d(a_{in}, a_{i1}) & d(a_{in}, a_{i2}) & \cdots & d(a_{in}, a_{in}) \end{bmatrix} \in R^{n \times n} \tag{4}$$

In the ith class samples, defining the distance between the mth ($m = 1, 2, 3 \cdots, n$) sample and the rest of the $(n - 1)$ samples:

$$D_m = \frac{\sum\limits_{k=1}^{n} d(a_{im}, a_{ik})}{(n - 1)} \tag{5}$$

If D_m is larger, then the distance between the mth sample and the rest of the $(n\text{-}1)$ samples is farther.

Defining D_d as the among-class distance matrix on the ith class samples:

$$D_d = [D_1, D_2, \cdots D_n] \tag{6}$$

Then we must remove the outliers of D_d. The dispersion of among-class distance is different on each action. Meanwhile, too much noise will affect the result of the experiment. Figure 5 shows the histogram of among-class distance in D_d of the 13 classes of action samples (Abscissa represents the distance, Ordinate represents action number).

If we get the among-class distance matrix D_d on the ith class samples, we also can calculate the confidence interval of D_d for each class. The confidence level can be early designated. For each D_d, if the data of D_d is larger than the upper limit of confidence interval, it will be judged as the noises and should be removed. The relation between confidence level and the proportion of noisy is shown in Fig. 6.
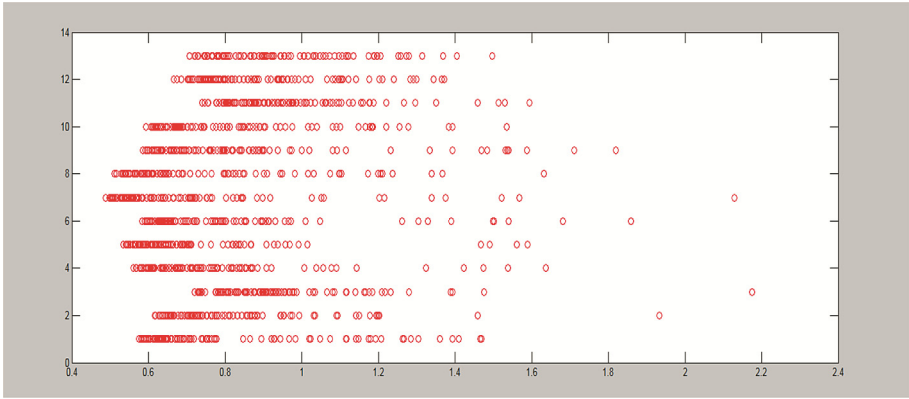
Fig. 5. The histogram of among-class distance of the 13 classes of action samples
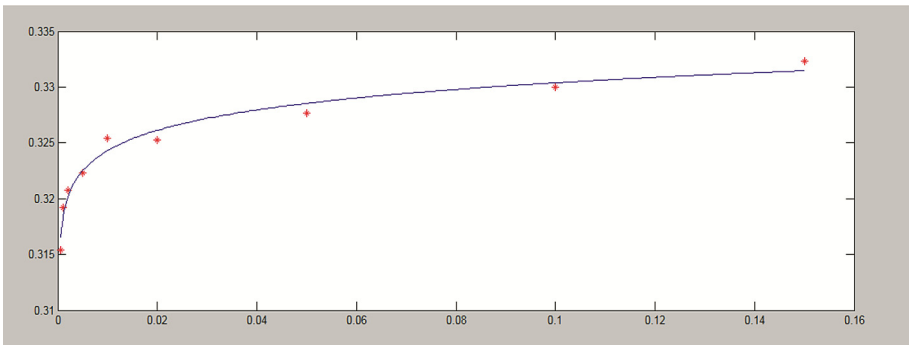
Fig. 6. The relation between confidence level and the proportion of noisy

We find that when the confidence level is smaller than 0.02, the slope of the curve is larger in Fig. 6. And the corresponding model will be unsteady. So 0.05 is chosen as the confidence level finally in this paper.

3   Result of Experiments

This paper extracts 11 features at last: the mean and variance values of vertical acceleration data of 5 sensors and the mean value of horizontal angular speed data of waist

sensor. At last, we optimize the parameters of SVM and get rid of the noisy, the corresponding results are shown in Tables 5 and 6.

Table 5. The result of the parameters optimization

	1	2	3	4	5	6	7	8	9	10	11	12	13	result
1	132	2	0	0	0	0	0	0	0	0	0	0	0	0.985
2	7	131	0	0	0	0	0	0	0	0	0	0	0	0.949
3	0	0	139	0	0	0	0	0	0	0	0	0	0	1
4	0	0	0	140	0	0	0	0	0	0	0	0	2	0.986
5	0	0	0	0	141	0	0	0	0	0	0	0	0	1
6	0	0	0	0	0	143	0	0	0	0	0	0	0	1
7	0	0	0	0	0	0	146	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	140	0	0	0	0	0	1
9	0	0	0	0	0	0	0	0	131	7	0	0	0	0.949
10	0	0	0	0	0	0	0	0	9	129	0	0	0	0.935
11	0	0	0	0	0	0	0	0	0	0	133	0	0	1
12	0	0	0	0	0	0	0	0	0	0	0	137	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	130	1
overall	132	131	139	140	141	143	146	140	131	129	133	137	130	0.985

Table 6. The result of the samples denoising

	1	2	3	4	5	6	7	8	9	10	11	12	13	result
1	183	17	0	0	0	0	0	0	0	0	0	0	0	0.915
2	12	185	3	0	0	0	0	0	0	0	0	0	0	0.925
3	0	3	197	0	0	0	0	0	0	0	0	0	0	0.985
4	0	0	0	180	3	0	0	0	0	1	2	0	14	0.9
5	0	0	0	1	189	0	1	0	1	1	0	0	7	0.945
6	0	0	0	0	0	197	0	2	0	0	0	0	1	0.985
7	0	0	0	0	0	0	200	0	0	0	0	0	0	1
8	0	0	0	0	0	0	0	200	0	0	0	0	0	1
9	0	0	0	0	1	0	0	0	178	21	0	0	0	0.89
10	0	0	0	0	0	0	0	0	26	173	1	0	0	0.865
11	0	0	0	0	0	0	0	0	0	0	199	1	0	0.995
12	0	0	0	0	0	0	0	0	0	0	0	200	0	1
13	0	0	0	0	0	0	0	0	0	0	0	0	200	1
overall	183	185	197	180	189	197	200	200	178	173	199	200	200	0.954

We contrast the Tables 3 and 5 and find that the recognition efficiency has significantly improved. Thus the method of parameters optimization of SVM is useful and important. In Table 6, the result of the 9th and the 10th behaviors shows that the recognition efficiency is obviously better than the case in Table 5. The reason is that the difference of the within-class of the 9th and the 10th behaviors is larger than others. So the samples de-noising can distinctly improve recognition efficiency.

4   Conclusion

This paper extracts 11 features based on the essence of human behavior, which can express the range and acuteness of body relative to the deflection of the vertical and horizontal direction. After optimizing the parameters of SVM and de-noising the 13 daily actions on the data in WARD1.0 database, human daily behaviors can be recognized effectively. It can also provide a thinking path of pattern recognition for other behavior recognition and health cares. But there are some deficiencies in this paper. For example, we do not consider the front or back direction characters of behaviors. How to excavate the specialty of human behavior is a worthy problem which should be studied for a long time.

References

1. He, W.: Study on the key technology for human activity recognition. Doctor thesis, Chongqing University (2012)
2. Shoaib, M., Bosch, S., Incel, O.D., Scholten, H., Havinga, P.J.M.: A survey of online activity recognition using mobile phones. Sensors **15**, 2059–2085 (2015)
3. Fahim, M., Fatima, I., Lee, S., Park, Y.-T.: EFM: evolutionary fuzzy model for dynamic activities recognition using a smartphone accelerometer. Appl. Intell. **39**, 475–488 (2013)
4. Yang, J., Wang, S.Q., Chen, N.J., et al.: Wearable accelerometer based extendable activity recognition system. In: IEEE International Conference on Robotics and Automation (ICRA), pp. 3641–3647 (2010)
5. Kwapisz, J.R., Weiss, G.M., Moore, S.A.: Activity recognition using cell phone accelerometers. ACM SIGKDD Explor. **12**, 74–82 (2010)
6. Xu, C., Gu, Q., Yao, M.: Activity recognition method based on three-dimensional accelerometer. Comput. Syst. Appl. **22**, 132–135 (2013)
7. Preece, S.J., Goulermas, J.Y., Kenney, L.P.J., et al.: A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. IEEE Trans. Biomed. Eng. **56**, 871–879 (2009)
8. Bao, L., Intille, S.: Activity recognition from user-annotated acceleration data. In: Ferscha, A., Mattern, F. (eds.) PERVASIVE 2004. LNCS, vol. 3001, pp. 1–17. Springer, Heidelberg (2004)
9. Wu, J., Pan, G., Zhang, D., Qi, G., Li, S.: Gesture recognition with a 3-D accelerometer. In: Zhang, D., Portmann, M., Tan, A.-H., Indulska, J. (eds.) UIC 2009. LNCS, vol. 5585, pp. 25–38. Springer, Heidelberg (2009)
10. Morillo, L.M.S., Gonzalez-Abril, L., Ramirez, J.A.O., Concepcion, M.A.A.: Low energy physical activity recognition system on smartphones. Sensors **15**, 5163–5196 (2015)
11. Gayathri, K.S., Elias, S.: Hierarchical activity recognition for dementia care using Markov Logic Network. J. Pers. Ubiquitous Comput. **19**, 271–285 (2015)
12. Fahad, L.G., Khan, A., Rajarajan, M.: Activity recognition in smart homes with self verification of assignments. Neurocomputing **149**, 1286–1298 (2015)
13. Yang, Y., Jafari, R., Sastry, S.S., Bajcsy, R.: Distributed recognition of human actions using wearable motion sensor networks. J. Ambient Intell. Smart Environ. **1**, 103–115 (2009)
14. Lu, X., Wang, H., Wang, Y., Xu, X.: Application research on acceleration data features in human behavior recognition. Comput. Eng. **40**, 178–182 (2014)
15. Zhang, J.: Research on human behavior identification technology. Master thesis, Chongqing University (2011)
16. MATLAB Chinese Forum: MATLAB Neural Network 30 Case Analysis. Beijing University of Aeronautics and Astronautics Press, Beijing (2010)
17. Chen, S., Li, L.: Denoising and sample reduction for large-scale sample set based on distance of nearest neighbors. Comput. Eng. **37**, 184–186 (2011)
18. Liu, G., Zhang, H., Guo, J.: The influence of different training samples to recognition system. Chin. J. Comput. **28**, 1923–1928 (2005)

A Rehabilitation Planning Based on Kinect Somatosensory Recognition and Cloud Computing

Hongpu Liu[1], Heyun Ma[1], Junhua Gu[1], Feng Wu[1], and Junjie Lv[2(✉)]

[1] School of Computer Science and Engineering, Hebei University of Technology,
Tianjin 300401, China
[2] Changsha Changjun Xiangfu School, Changsha 410114, Hunan, China
191383363@qq.com

Abstract. In order to solve the physical rehabilitation of stroke patients with apoplectic hemiplegia, an intelligent rehabilitation supporting system based on Kinect device is designed. The system uses Kinect device to capture the user's limb joints point information. DTW algorithm is combined with RANSAC algorithm to match the action in standard action library, thus the system can obtain the evaluation results. Then the patients' rehabilitation data is sent to the doctors with new treatment options intelligently. As a result, the experimental results show that this physical rehabilitation method is more convenient and efficient than the traditional training method.

Keywords: Kinect rehabilitation · DTW algorithm · RANSAC algorithm

1 Introduction

Nowadays, stroke has become a high incidence of disease. Along with the advance of medical level, the mortality rate of the disease has been greatly reduced. However, the disease disabled rate is as high as 75 %. Rehabilitation training is still a looming problem. The current rehabilitation training, basically depends on the artificial guidance. Only part of the rehabilitation hospital or rehabilitation agency introduces some orthoses training and limb rehabilitation exercise machine to assist the rehabilitation training. And because of various factors, the majority of patients due to the lack of professional rehabilitation of the correct guidance, causing that rehabilitation training can be not in place. It is difficult to restore a very good state. So, it is very necessary to design a set of convenient and fast rehabilitation training system which can be carried out in the community.

Based on Kinect somatosensory technology which is called is motion-sensing technology is one of the most original used in game industry, and the development of the technology infiltrates gradually other areas as medicine, architecture [1] and network [2, 3]. In the medical field, motion-sensing technology at present is mainly used in medical rehabilitation, medical imaging [3, 4], etc. Especially in the field of medical rehabilitation, motion-sensing technology development is very rapid.

Chang, for example, developed a Kinempt system [5] based on Kinect used for patients with cognitive impairment. In order to exercise rehabilitation of youth movement disorder guidance, the Kinerehab [6] system based on Kinect is developed by

Chang too. The system uses the image processing technology of Kinect sensor to detect the movement information of patients with movement disorders.

At present, in terms of rehabilitation training system research, our country mainly concentrated in a few engineering comprehensive university, and the research content mainly focus on the development of rehabilitation training equipment. Tsinghua University develops the human lower limb rehabilitation training vehicle, the nerve injury rehabilitation training device, and the upper limb complex motor rehabilitation training robot [7], etc..

2 System Architecture

The system is divided into three modules as the Kinect-based somatosensory recognition module, the cloud platform delivery module and the big data processing module (Fig. 1).
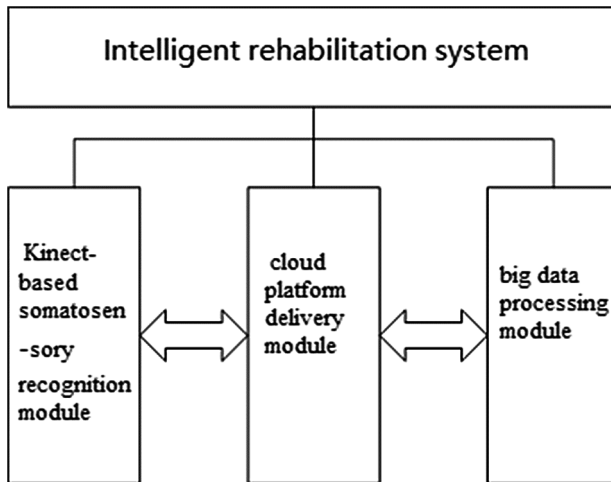
Fig. 1. The structure of the system

3 Somatosensory Recognition Algorithm

3.1 DTW Algorithm

Dynamic Time Warping (DTW) algorithm was proposed by Sakoe and Chiba [8] in 1978 with a combination of distance measure and time warping algorithm, which was first applied to speech recognition.

DTW algorithm, which is a mainstream algorithm, has been used in the field of speech recognition. When making the speech recognition, even though the same person said the same words, each pronunciation may not have exactly the same length of time. So when the template matches with the stored, you need to deal with the timeline of the new words, so that it corresponds to the time of the template. Dynamic time warping is a typical effective measure to solve the minimum distance between the two templates.

It uses the function of a certain condition to deal with the time correspondence between the under-recognition template and the sample reference template. In this system, there is a similar problem of movement recognition, the time for everyone to do the same action will not be exactly the same. For example, we see the coach's standard action as template A and patient's action as template B to determine the accuracy of B on the basis of A. As the Fig. 2 shows, everyone do the same operation which have the same start time and end time of action, in this case A and B to the similarity of the comparison, the similarity obtained may be 0 %, but the fact is that patients may not totally do incorrect action. So prior to action recognition, we must be on the standard action templates and test action template dynamic time warping.
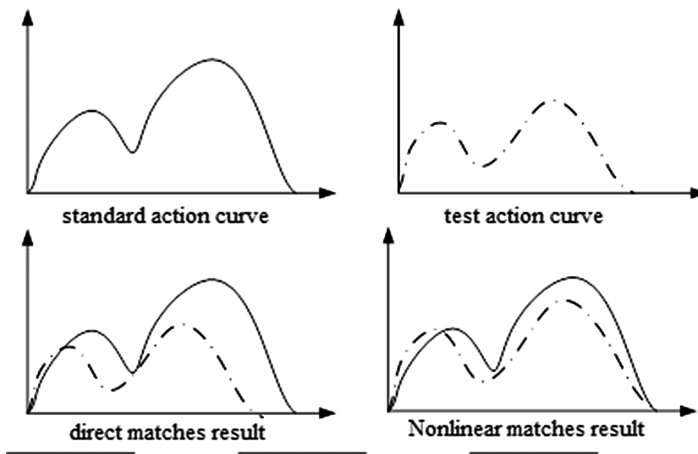

Fig. 2. Match result

DTW (dynamic time warping) is a nonlinear dynamic time warping algorithm which combines distance calculation and time warping. The main idea is to use dynamic programming, searching matching path for two different length sequences. Then we can obtain the path made cumulative amount of distortion minimal when the two template sequences are matching. For two sequences of different time axis, the time axis using the technique of dynamic structured template sequence wherein one stretching or structured to eliminate the difference between the degree of distortion to a minimum. For two sequences of different time axis, the time axis using the technique of dynamic structured template sequence wherein one stretching or structured to eliminate the difference between the degree of distortion to a minimum.

Suppose two templates for standard sample template R and test sample template T are the length of m as $R = \{r_1, r_2, \ldots r_i, \ldots r_m\}$ and a length of n as $T = \{t_1, t_2, \ldots t_j, \ldots t_n\}$. We construct a M × N matrix whose element represents the corresponding distance of vector in R and T. DTW find an optimal path that satisfies some constraints and the minimum amount accumulative distortion from point (1,1) to point (M, N) through the local path constraints. Constraint conditions include continuity constraints, monotonic constraints and boundary constraints. It means that point on the path to match and adjacent points must be aligned and

must not skip a point to match. It must be monotonous over time from one point to another point on the path with start pointing (1,1) and ending point (M, N). As shown is the process to search the matrix elements created by the optimal matching path (Fig. 3).
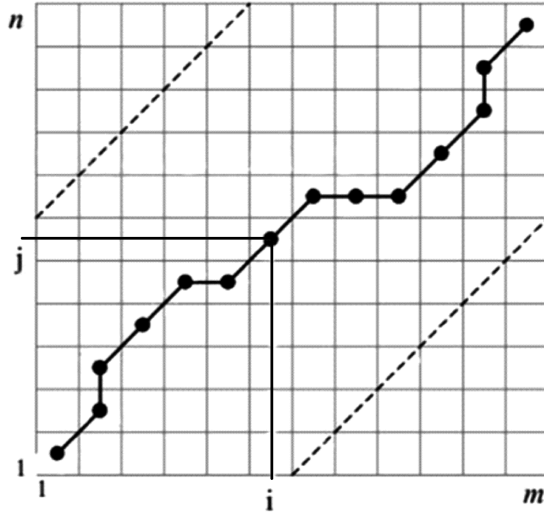
Fig. 3. DTW algorithm optimal path schematic

According to DTW algorithm to search the optimal path on the template sequence R and T, points on the optimal path respectively are (m1,n1), (m2,n2),…,(mi,ni),…,(mk,nk), where k satisfies $max(m, n) \leq k \leq m + n - 1$. Matrix element (i,j) which represents feature vector similarity $d(r_i, t_j)$ between r_i and t_j, can be obtained by Euclidean distance, Mahalanobis distance, etc. Searching the optimal path is to find the points of the minimum amount accumulative distortion of the adjacent elements between the current path and the surrounding matrix element. The current cumulative distance $l(i, j)$ is:

$$l(i,j) = d(r_i, t_j) + min\{l(i-1, j-1), l(i-1, j), l(i, j-1)\} \tag{1}$$

The optimal path of two templates is the Point of collection that we constantly calculating the minimum cumulative amount of distortion between the template.

Wanting to apply DTW motion recognition algorithm the action recognition problem, we need to determine the computational method of feature vector similarity of two template's points and standard action template. In this paper, the feature vector similarity calculation method uses chi-square test. Under normal circumstances, the Euclidean distance, Mahalanobis distance method can calculate the similarity feature vectors. To simplify the design process, while avoiding the complex calculation of the Euclidean distance, the similarity between the two templates are also calculated using the chi-square test method.

$$d(R, T) = \sum_{i=1}^{k} \frac{(r_i - t_j)^2}{r_i + t_j} \tag{2}$$

In order to improve the reliability of the standard action samples, and to improve the accuracy of the algorithm DTW action recognition, we collect multiple sets of data for a same standard action. We need take similar action samples from multiple sets of standard data, and calculate with similarity and threshold with other movement data. Then we can get a standard motion data template of the action by getting the sample average within the threshold range.

3.2 RANSAC Algorithm

RANSAC [9] is "the Random Sample Consensus (Random sampling Consensus)". It can be from a group of observation data set which contains "outsiders", through the iterative way to estimate the parameters of the mathematical model. It is an uncertain algorithm - it has a certain probability to arrive at a reasonable result; In order to increase the probability we need to improve the number of iterations. The algorithm was first proposed by Fischler and Bolles in 1981.

Input parameters of RANSAC algorithm are a set of observed data, a parametric model can be explained or adapt to the observed data and a number of credible arguments. RANSAC achieves their goals by repeatedly selecting data in a random subset. Selected subset point is assumed to be interior point, and verified by the following methods:

1. There is a model adapted to the assumptions of office interior point, that all points of the unknown parameters can be calculated from the assumption of interior point.
2. Model from 1 was used to test all of the other data, if a point is suitable for the estimated model, we call it office interior point.
3. If you have enough points to be classified as a hypothetical office interior point, it is estimated that the model is reasonable enough.
4. Then, we use all assumptions office interior point to re-estimate the model, because it is just estimated by initial assumptions innings.
5. Finally, we can through the estimated office interior point and error rate of the model to evaluate the model.

This process is repeated a fixed number of times. Each model either because office interior point too few to be discarded, either because better than the existing model is chosen.

3.3 Patient Rehabilitation Exercises Assessment

Although this method requires two assessments, it will slow down the speed of assessment. But DTW and RANSAC algorithm combine, it can make the results more objective and accurate, can rely.

With the DTW algorithm and RANSAC algorithm combining, the accuracy of patient rehabilitation action can be assessed. The patient motion data A and standard action data B will be sent to the DTW algorithm to get similarity evaluation results a,

and A and B will be sent to RANSAC to get algorithm similarity assessment b. This article sets confidence of DTW algorithm and RANSAC algorithm as M and N. In the process of similarity evaluation, the doctor can manually select a or b to trust and involve in the evaluation. Then the selected confidence of the credible algorithm plus 1. If the doctor does not participate in the evaluation process, the system will calculate the similarity combining confidence with assessments based on method of the formula, the following formula:

$$D = \frac{M}{M+N}a + \frac{N}{M+N}b \tag{3}$$

D represents the final similarity of A and B.

We can combine DTW and RANSAC algorithm, through the doctor's judgment, improving the accuracy and usability of evaluation results,and providing a more reliable basis for physicians to develop training programs.

4 System Implementation

4.1 Kinect-Based Somatosensory Recognition Module

Kinect-based rehabilitation aid system mainly uses theory and technology of computer graphics, kinematics, topography, rehabilitation medicine and other fields. It uses XBOX equipment with corrective action and rehabilitation aid as core. Through the Kinect device to collect data, it is a physical rehabilitation assistance system for stroke patients with hemiplegia. System main interface is shown in Fig. 4.

Rehabilitation aid system mainly consists of three sub-modules: standard position generating module, rehabilitation operation correction module, feedback and evaluation module.

Fig. 4. Somatosensory recognition module interface

1. Position generating module records the standard action that rehabilitation experts design for stroke hemiplegia. The module by recording rehabilitation specialists' bones node data, generates standard data used for rehabilitation.
2. Rehabilitation operation correction module can directly provide the patient with assisted rehabilitation and movement assisted rehabilitation and operation. The module, by reading the patient's bone node data in real time, compares with the standard action and gives the appropriate action guide.
3. Feedback and evaluation module mainly collects patient data for all actions to analysis and evaluate the rehabilitation training data. These results are provided directly to the patient or treatment personnel, and as a basis to determine the next treatment.

In this part, we mainly use the core technology of Kinect skeleton tracking. As shown in Fig. 5, through the camera, we can intelligently identify the body of the 20 key points. Each node has X, Y, Z three coordinate values. Video image data defaults to 640 * 480, RGB32 format. At the same time database also saves the time difference which is current skeleton node relative to the video recording start time, for simultaneous localization of bones node data and color image data. For action recognition, we use the algorithm above.
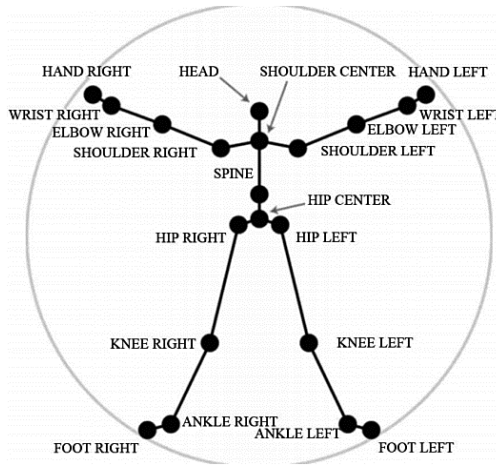

Fig. 5. Kinect skeletal point example

4.2 "Rehabilitation-Cloud"—— Rehabilitation Tracking Platform Based on Big Data and Cloud Computing

"Rehabilitation-Cloud" is divided into a cloud platform delivering module and big data processing module.

For patients, "Rehabilitation-Cloud" is a push platform for the movement database. The doctor will provide patients with a treatment plan. The patient carry out rehabilitation training. After the training, all the feedback data, such as the time of each step, the matching degree of each action, etc., will be back to the doctor. The data will be observed and studied by doctor, so that the doctor can develop the next step of treatment.

For doctors, "Rehabilitation-Cloud" is a big platform. Doctors can view the patient list through the platform, and observe the patient recovery data, to understand the progress of rehabilitation. Doctors can communicate with other physicians through "Rehabilitation-Cloud", to achieve resource sharing.

When a new patient appears, the doctor entered the patient's disease information, to develop rehabilitation programs for patients. And these rehabilitation programs will be imported into "Rehabilitation-Cloud" platform database.

Later, when another new patients with similar information appeared, the system, thorough data processing, recommended the development of several other doctors had developed the best rehabilitation treatment to the doctor. In this case, the doctor can adopt these different treatment options, and give a more accurate treatment of patients according to the actual situation (Fig. 6).

Fig. 6. Patients rehabilitation data display interface

5   Conclusion

This system mainly has some advantages as follows. First, former system needed professional artificial guidance, and the equipment was also more complex. While our system is put forward to make up for the defects. Rehabilitation training can be realized by a computer network and a common Kinect device machine. It saves a lot of time cost and artificial cost. Second, the doctor can communicate quickly through this way. It can improve levels of doctors from small rehabilitation institutions. It is also very beneficial for resource sharing.

To apply this topic research in clinical rehabilitation practice such as in hospitals, community medical institutions and family. It plays a very important role for patients with early recover. It will create conditions for patients to return to the society, improve the quality of life, reduce or even eliminate both material and spiritual burden of family and society.

References

1. Brian, M., Zak, M., Caroline, A.: Utilizing explorable visual environments for experiential applications. Procedia Comput. Sci. **8**, 261–266 (2012)
2. Boulos, M.N., Blanchard, B.J., Walker, C., et al.: Web GIS in practice X: a Microsoft Kinect natural user interface for Google Earth navigation. Int. J. Health Geogr. **10**, 45–59 (2011)
3. Michael, Z., Michael, M., Günther, G., et al.: Automatic reconstruction of personalized avatars. Comput. Anim. Virtual Worlds **22**(1), 195–202 (2011)
4. Luigi, G., Alessio, P.P., Mario, C.: Controller-free exploration of medical image data: experiencing the Kinect. In: 24th International Symposium on Computer-Based Medical Systems (CBMS), pp. 1–6. IEEE, Bristol, Britain (2011)
5. Chang, C.-Y., Lange, B., Zhang, M., et al.: Towards pervasive physical rehabilitation using Microsoft Kinect. In: International Conference on Pervasive Computing Technologies for Healthcare and Workshops, pp. 159–162 (2012)
6. Chang, Y.J., Chen, S.F., Chuang, A.F.: A gesture recognition system to transition autonomously through vocational tasks for individuals with cognitive impairments. Res. Dev. Disabil. **32**, 2064–2068 (2011)
7. Jiang, H., Xiang, J.: Kinect rehabilitation assistance system research. Artif. Intell. Robot. Res. **2**, 79–82 (2013)
8. Sakoe, H., Chiba, S.: Dynamic programming algorithm optimization for spoken word recognition. IEEE Trans. Acoust. Speech Signal Process. **26**, 43–49 (1978)
9. Chen, F., Wang, R.: Based on the consistency and rapid sampling algorithm. J. Softw. **16**(8), 1431–1437 (2005)
10. Sempena, S., Maulidevi, N.U., Aryan, P.R.: Human action recognition using dynamic time warping. In: Proceedings of 2011 International Conference on Electrical Engineering and Informatics (ICEEI), pp. 1–5 (2011)
11. Yamato, J., Ohya, J., Ishii, K.: Recognizing human action in time sequential images using hidden Markov model. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, pp. 379–385. IEEE (1992)
12. Yu, X.Y., Wu, L.Y., Liu, Q.Y., et al.: Children tantrum behaviour analysis based on Kinect sensor. In: Third Chinese Intelligent Visual Surveillance (IVS) Conference, Beijing, China, pp. 49–52 (2011)

Virtual Interface and Its Application in Natural Interaction

Hui Liu[1,2], Zhiquan Feng[1,2(✉)], Liwei Liang[1,2], and Zhipeng Xu[1,2]

[1] Department of Information Science and Engineering,
University of Jinan, Jinan 250022, China
`fengzq@ujn.edu.cn`
[2] Shandong Provincial Key Laboratory of Network Based Intelligent
Computing, Jinan 250022, China

Abstract. To gesture interaction based on computer vision, the effective area of hand activity is often the whole reach captured by cameras, so some subconscious action of users may be interpreted as effective computer command. In order to address this problem, we put forward the concept of virtual interface, which makes the effective area of hand activity narrow to a specific region. As the gesture in the virtual interface is regarded as effective, while the gesture outside it is regarded as invalid, it is possible to solve the "Midas Touch Problem". To begin with, we identify the position and size of virtual interface by utilizing the least square method through learning process. Then users interact with computer by virtual interface, during which the gesture commands are released through it. The experimental results show that the proposed virtual interface can efficiently solve the "Midas Touch Problem" and has a good user experience.

Keywords: Virtual interface · Midas touch problem · Gesture interaction

1   Introduction

Gesture interaction is a hot topic in the field of human-computer interaction and attracts more and more people's attention as a natural way to interact [1, 2]. Compared with traditional interaction style based on graphical user interface, VBGI (Vision-Based Gesture Interface), which has become a natural and unconstrained interaction, allows users to get rid of keyboard and mouse. The gesture interaction interface based on computer vision has been widely used in smart space, augmented reality, pervasive computing and gradually becomes a hot topic at home and abroad. However, due to the non-contact and ambiguity of VBGI, there remains a problem namely the "Midas Touch Problem". Since the effective range of hand activity is often the whole area captured by cameras, the system cannot distinguish between real intention and subconscious action, namely the effective and ineffective gesture.

The "Midas Touch Problem" is a widespread problem in VBGI. Paper [3] summarized that the solution to this problem can be divided into three categories which consisted of the strategy based on the time delay, the principle based on spatial proximity and the method of providing interactive context based on widget. The strategy

based on the time delay was proposed by Jacob [4] whose main idea was to calculate how long the cursor on gesture control interface stay in the interface objects to determine whether the object was selected. Obviously, there are some deficiencies in this method. The user must concentrate on the time staying on the object when in operation, which increases the cognitive and operational burden for the user. Kato et al. presented a strategy based on spatial proximity principle [5] whose main idea was to complete a series of interactive tasks by determining the distance between the hand-held paddle and virtual objects. The drawback of this method is that the distance between the paddle and virtual object is not easy to control, so it can also cause false choices. Kjeldsen proposed to establish certain mapping relation between gestures and visual interface widget [6]. Although this method solves the problem of false elections to some extent, it is also not natural. Because when the system needs to generate a lot of widget, it will undoubtedly increase the cognitive load of users. Paper [3] proposed an extensible visual gesture interaction model which divided gesture interaction into three phases, namely selection processing, distribution processing and centralization processing. Then he put forward a gesture recognition framework on the basis of this model. Even though the algorithm can solve the "Midas Touch Problem" to some degree, the recognition rate is not high for the training that has nothing to do with the user.

Based on previous studies, this paper puts forward a new solution that aims to solve the "Midas Touch Problem" by means of virtual interface. Firstly, we determine the size and position of virtual interface by way of learning process. Then the user interacts with 3D scene through partitioned virtual interface, during which the virtual interface is used as an input interface. In addition, when the body moves some distance to left or right during the process of interaction, the virtual interface will move towards the same direction. Therefore, the user can continue to operate in the original virtual interface.

2   Construct the Virtual Interface

2.1   Gesture Segmentation Based on Kinect

Gesture segmentation [7, 8] is to separate gesture image from complex background and only gesture is kept in the image. This paper segments the gesture based on depth data obtained by Microsoft's kinect sensor, which avoids the influence of light, background and other factors and improves robustness of gesture segmentation. The thoughts of gesture segmentation algorithm is to use the OpenNI to track hand to get the position of hand, and then extract portion of hand by using the depth information of hand [9]. Since the extracted hand may contain wrist, color models have been used to deal with the segmented hand. Finally the segmentation results need other process, such as denoising, corrosion, expansion [10–12]. Gesture segmentation results are as follows (Fig. 1):

2.2   The Specific Method of Constructing a Virtual Interface

One difference between the ordinary camera and kinect is that the kinect can get the depth information of hand, so spatial coordinate can be obtained by kinect. Spatial
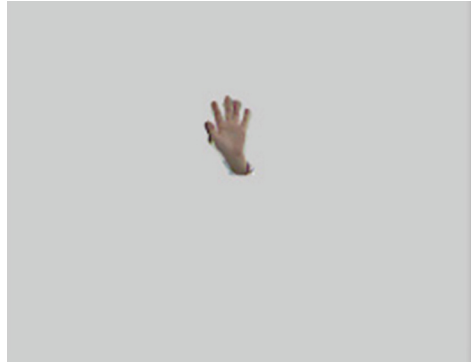
Fig. 1. Result of gesture segmentation

coordinate of hand can be converted into screen coordinate in the projected coordinate system by coordinate conversion function and the screen coordinate is namely the position of hand in color image captured by kinect. The process of constructing the virtual interface is also called the learning process whose learning form is fixed. It just needs operator to draw a rectangle track with depth information, which is namely the virtual interface. The size and location of virtual interface consists of two parts. The first part is the scope of determined by projection point of hand on the screen in the learning process. The planar rectangle, which is the projection of rectangle track, is represented by collection Range1. The other part, which is represented by collection Range2, is the hand depth range in the learning process. The former is realized by fitting each side of the projection rectangle linearly by using least square method and the latter is realized by getting the maximum and minimum of depth information.

When learning virtual interface, the operator needs to draw rectangle track in the order of from left to right and from bottom to top in space. When evaluating **$Range_1$**, we apply to the projection coordinates of hand spatial point on the screen. Firstly, we need to make sure the four corner points of the rectangle and then fit the four sides of the rectangle respectively. The four sides of the rectangle are separately named as $x = X_{min}$, $x = X_{max}$, $y = Y_{min}$, $y = Y_{max}$. The line $x = X_{min}$ represents the side located on the left; the line $x = X_{max}$ represents the side located on the right; the line $y = Y_{min}$ represents the side located on the top; the line $y = Y_{max}$ represents the side located on the bottom. Since this paper sets the four sides of the rectangle is vertical or parallel with the screen axis, the fitted line is special. The method of fitting $x = X_{min}$ is as follows:

Step 1: To save the initial point of the line $x = X_{min}$.
Step 2: To determine whether the learning process of virtual interface is completed. If the hand remains still within a few frames, it shows that the learning process is completed and the program will turn to Step 7. And if the hand is not still, the program will perform Step 3.

Step 3: To save the hand projection coordinates of current frame in collection C_1 and the hand depth value in collection C_2. n is the number of elements in collection C_1. If the following condition is satisfied, the program will save the projection point of the current frame in the intermediate variable *Temp* and then perform Step 2. If the following condition is not satisfied, the program will turn to Step 4.

$$Cur\_pointx - Sta\_pointx < b \tag{1}$$

Where *Cur_pointx* is x coordinate of hand spatial point of current frame; *Sta_-pointx* is x coordinate of hand spatial point of initial state; b is a constant value defined in advance whose recommended value is 30.

Step 4: To determine the lower-left corner point. *Temp* obtained through Step 3 is namely the lower-left corner point and is also the initial point of straight line $y = Y_{max}$.

Step 5: To fit the line $x = X_{min}$. X_{min} is obtained by means of least square method which uses C_1 and n.

Step 6: Search the lower-right corner and begin to fit the straight line $y = Y_{max}$.

Step 7: Determine the scope of depth of the virtual interface in learning process. Calculate the maximum HD_{max} and minimum HD_{min} of C_2.

Fit the straight line $y = Y_{max}$, $x = X_{max}$, $y = Y_{min}$ in the same way that fit the line $x = X_{min}$ and get Y_{max}, X_{max}, Y_{min}. In the end, we get **Range₁** = {X_{min}, X_{max}, Y_{min}, Y_{max}}, **Range₂** = {HD_{min}, HD_{max}}. So the scope of virtual interface **Range** obtained by learning process is **Range** = {X_{min}, X_{max}, Y_{min}, Y_{max}, HD_{min}, HD_{max}}.

In the following contents, *ProX* represents the projection coordinate x of hand, *ProY* represents the projection coordinate y of hand, *HandZ* represents the depth information of hand. Only when satisfying $ProX \in [X_{min}, X_{max}]$, $ProY \in [Y_{min}, Ymax]$, $HandZ \in [HD_{min}, HD_{max}]$, the gesture command is valid. The projection of constructed virtual interface on the screen as shown in Fig. 2 below and the blue rectangle section is namely it.

Fig. 2. Projection of virtual interface (Colour figure online)

3   Gesture Interaction Based on Virtual Interface

3.1   Virtual Interface Partition

After completing the learning process, the size and location of virtual interface are determined. The user is to interact with 3D scene through virtual interface. The main interaction task in this paper is to enlarge, shrink and rotate the object. In order to realize the three kinds of commonly used functions, this paper puts forward partition which parts the virtual interface into three sections and each section has different function. Therefore only a kind of gesture can complete all the functions. The partition is aimed at Range1. The projection of virtual interface namely the rectangle is divided into three parts on average according to the order of from left to right. They are respectively named as the first partition whose function is to enlarge the 3D objects; the second partition whose function is to shrink the 3D objects; the third partition whose function is to rotate the 3D objects. Specific partition algorithm is as follows:

Step 1: According to the following formula, partition the virtual interface.

$$xeach = (X_{max} - X_{min})/3 \qquad (2)$$

$$Value_1 = X_{min} + xeach \qquad (3)$$

$$Value_2 = Value_1 + xeach \qquad (4)$$

Where *xeach* represents the length of segmented rectangle when the rectangle is divided into three parts following the direction of horizon; X_{min} and X_{max} are the edge points value of the rectangle; $Valuex_1$ and $Value_2$ are the value of segmentation point and $Value_2$ is greater than $Valuex_1$.

Step 2: To determine the position of hand is in which partition. Get real-time the position of hand per frame and perform the following judgments:

(a)   When satisfying $ProX \in [X_{min}, Valuex_1]$, $ProY \in [Y_{min}, Y_{max}]$, $HandZ \in [HD_{min}, HD_{max}]$, it indicates that the hand is in the first partition and the volume of object in 3D scene becomes larger.

(b)   When satisfying $ProX \in [Value_1, Valuex_2]$, $ProY \in [Y_{min}, Y_{max}]$, $HandZ \in [HD_{min}, HD_{max}]$, it indicates that the hand is in the second partition and the volume of object in 3D scene becomes smaller.

(c)   When satisfying $ProX \in [Valuex_2, X_{max}]$, $ProY \in [Y_{min}, Y_{max}]$, $HandZ \in [HD_{min}, HD_{max}]$, it indicates that the hand is in the third partition and the object in 3D scene continues to rotate towards a fixed direction.

(d)   If (a), (b) and (c) are not satisfied, it suggests that the hand is out of the virtual interface and at this time the gesture command is invalid.

3.2   The Real-Time Refresh Movement of Virtual Interface

When the body moves some distance to left or right, the generated virtual interface does not disappear, but moves towards the same direction accordingly. Therefore,

when the position of operator changes, the operator can continue to operate in the virtual interface learned before. Since kinect can obtain the space position of human body, whether the virtual interface needs to move can be determined by whether the gravity coordinate of human body changes. Specific algorithm is as follows:

Step 1: To save initial gravity coordinate P_1 of human body when the learning process is finished.

Step 2: To get gravity coordinates P_2 of human body of current frame and calculate the distance d that human body moves in space. d is the mold of vector $\boldsymbol{P_1P_2}$.

Step 3: Determine whether the virtual interface needs to move. Judge whether $d > s_2$ is satisfied. s_2 is a constant value defined in advance whose recommend value is 200. If the condition is satisfied, the virtual interface will move. On this base, judge the direction of movement, namely the right or left.

(a) When the human body moves some distance to left, the virtual interface will move to the left as well. Update the value of elements in the collection **Range** according to the following formula:

$$\begin{cases} X_{min} = X_{min} - d * scale \\ X_{max} = X_{max} - d * scale \end{cases} \tag{5}$$

(b) When the human body moves some distance to right, the virtual interface will move to the right as well. Update the value of elements in the collection **Range** according to the following formula:

$$\begin{cases} X_{min} = X_{min} + d * scale \\ X_{max} = X_{max} + d * scale \end{cases} \tag{6}$$

where the value of scale is 0.7. Update P_1 and make $P_1 = P_2$. If the condition is not satisfied, the program will turn to Step 2.

4   Experimental Results and Analysis

We have experimented with our algorithm implemented using C++ and OPENGL language on a PC with windows 7 system, Intel Xeon W3520, 2.67 GHz, 8 GB main memory.

The operational scene in experiment and 3D scene is shown in Fig. 3. The whole window is divided into three parts; the upper left shows the real-time operational scene captured by kinect; the lower left is the results of hand segmentation based on depth information obtained by kinect; the right part is the virtual scene in which there is a magic square. When the virtual interface is determined, the user can interact with the magic square based on it.

The Fig. 4 shows the process that the operator interacts with the magic square in the 3D scene. The Fig. 4(a) shows the initial state of magic square; the Fig. 4(b) shows that when the hand is in the first partition whose function is to zoom, the volume of magic
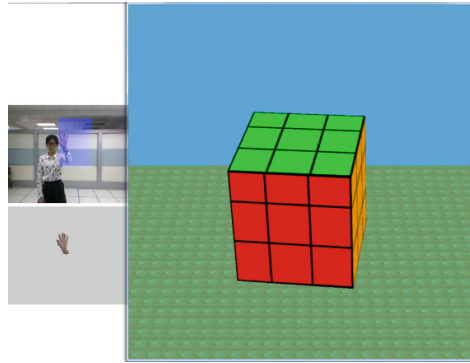
Fig. 3. Experimental scene

square will become larger on the base of initial state; the Fig. 4(c) suggests that when the hand is in the second partition whose function is to shrink, the volume of magic square will get smaller on the base of Fig. 4(b); the Fig. 4(d) indicates that when the hand is in the third partition that controls the rotation of object, the magic square will rotate a certain angle compared with the previous state; the Fig. 4(e) reveals that no change will occur when the hand is out of the virtual interface and the state of magic square is the same as the Fig. 4(d), which illustrates that the gesture command is invalid at this time.

The Fig. 5 describes that the virtual interface will move when the body moves some distance to left or right during the interaction. The Fig. 5(a) shows the initial projection position of virtual interface; the Fig. 5(b) represents that the virtual interface will move towards left when the human body moves some distance to left; the Fig. 5(c) explains that when the hand is in the first partition, the volume of magic square will become larger on condition that the projection of virtual interface moves towards left, which means that when the user moves, the constructed virtual interface is still usable; the Fig. 5(d) demonstrates that the virtual interface will move towards right when the human body moves some distance to right.

This paper just counts up the time and accuracy to evaluate the algorithm because the contents of this paper are novel and there are few things to compare with. The definition of time is how long it takes in the learning process of virtual interface and the accuracy is how many times the mistakes occur when the user interacts with 3D scene through virtual interface. It is necessary to set the task when calculate accuracy and the task is the user needs to interact with the 3D scene ten times each experiment. A graduate student of grade two conducts independent experiment for 50 times in the same condition to compute the time that the learning process costs and the number of mistakes that occur when the user interacts with 3D scene. The experimental results are as shown in Figs. 6 and 7. Figure 6 shows the time consumption in the process of learning from which we can see that the time is between 5000 ms and 5500 ms. We acquire the average time consumed in the learning process of virtual interface is 5294 ms by calculating the mean value of all the time data. Figure 7 displays the number of mistakes when the user interacts with 3D scene. As shown in Fig. 7, the number of mistakes is between 0 and 2.
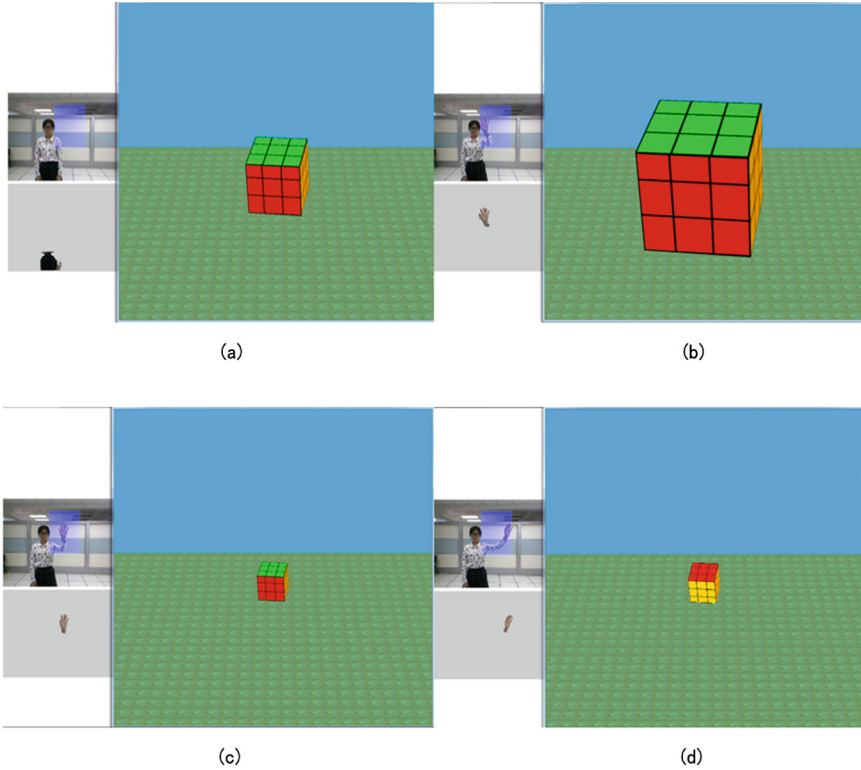
Fig. 4. The process of interaction

We get the error rate is five percent through that the sum number of errors in 50 experiments is divided by the total number of interaction times during the 50 experiments. It shows that the algorithm has a higher accuracy. It shows that the operator can finish interaction task well by using this system.

In addition, this paper also evaluates the algorithm in user interaction experience. The algorithm of this paper is named VIM (Virtual Interface Method), while the method that is not provided with virtual interface based on computer vision is called NVIM (None Virtual Interface Method). We select 100 undergraduates randomly as the experimenters for whom the experimental scene and interaction task are the same. We compare the system utilizing VIM with the system exploiting NVIM in the aspect of fatigue, effort, pleasure and feasibility. Fatigue is how tired the operator is in the process of operation; effort is how much work it takes to complete the interaction task; pleasure is how enjoyable the operator feels when manipulating; feasibility is the feasible degree to achieve the task. Specific grade standard is that the full mark is 100 points. The lower the score of fatigue and effort is, the better the performance of the system is; the higher the score of pleasure and feasibility is, the better the performance of the system is. By means of calculating the average of each performance, we acquire Fig. 8. We can see that in terms of fatigue and effort, the advantage of system based on
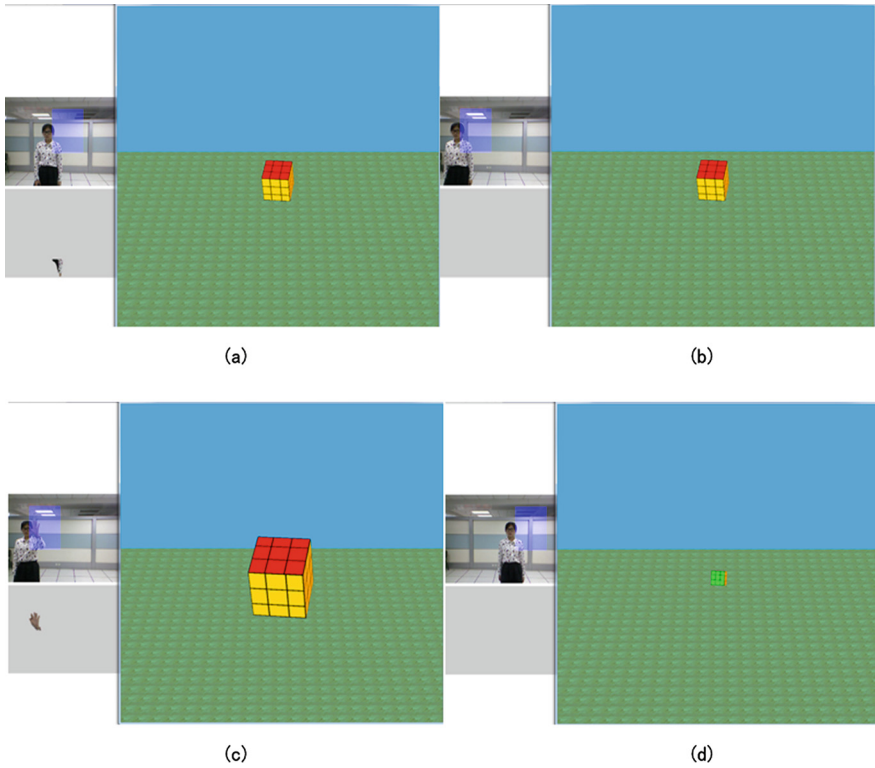
Fig. 5. The movement of virtual interface
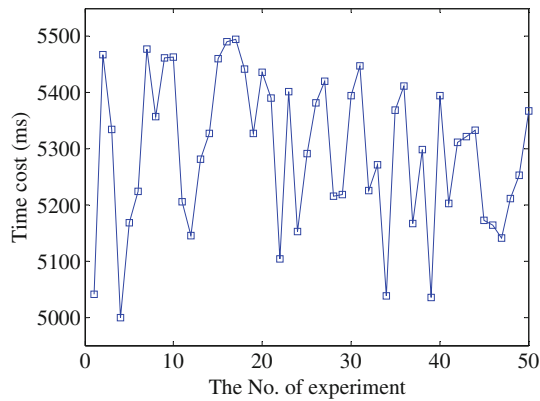
Fig. 6. Time cost in the learning process

VIM is especially obvious. That is because that compared with the system based on NVIM, the system based on VIM limits the operating space to a certain range. Consequently, the operator is able to complete the whole operation tasks in the limited
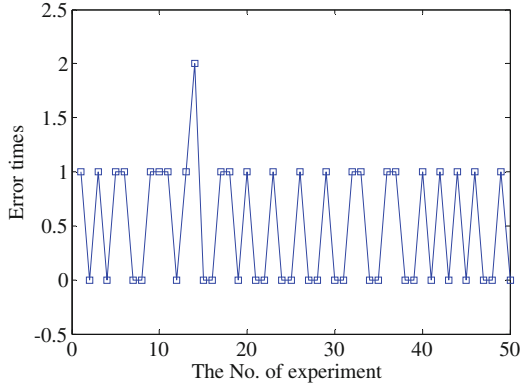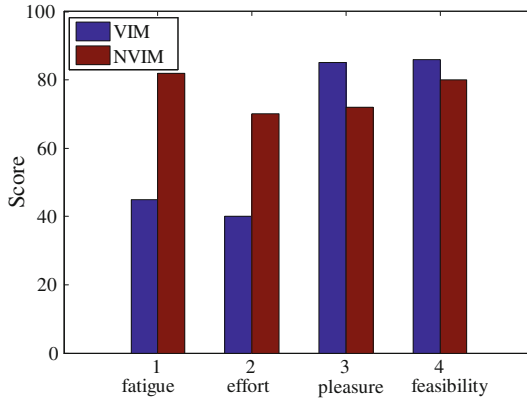
Fig. 7. Error times during interaction

Fig. 8. User experience (Colour figure online)

ranges without moving the body and stretching the arm. As the virtual interface can prevent the subconscious gesture from producing command, the system based on VIM has some advantage in aspects of pleasure and feasibility. It shows that the performances of system using algorithm proposed in this paper can be reduced or improved by 45 %, 43 %, 18 %, 8 % respectively.

5   Conclusions and Future Work

In this paper, we put forward the concept of virtual interface in order to solve the "Midas Touch Problem". The gesture in the virtual interface is regarded as an effective gesture, while the gesture outside the virtual interface is regarded as an invalid gesture and will not trigger any command. Firstly, the virtual interface is determined by the learning process and then the virtual interface is used to interact with the virtual scene. At the

same time, the virtual interface is able to keep pace with the human body, namely when the human body moves some distance towards left or right, the virtual interface will move towards the same direction. The research shows that the virtual interface can not only solve the "Midas Touch Problem" effectively, but also has a good user experience. However, there are also some drawbacks in this paper which should be amended. For example, the interactive tasks and operation scene is a bit single; For the condition that when the operators rotate or tilt their bodies how the virtual interface changes, this paper does not take into consideration; In the process of interaction, when the user moves a distance towards left or right, how to determine the change of virtual interface whether it needs to learned again or just the learned virtual interface moves with the user. These are all the problems that need to be solved in the future.

References

1. Wu, X., Zhang, Q., Xu, Y.X.: An overview of hand gesture recognition. Electron. Sci. Tech. **26**(6), 171–174 (2013)
2. Feng, Z.Q., Jiang, Y.: A survey of hand gesture recognition. J. Univ. Jinan (Sci. Tech.) **27**(4), 336–341 (2013)
3. Wu, Y.H., Zhang, F.J., Liu, Y.J., Dai, G.Z.: Research on key issue of vision-based gesture interfaces. Chin. J. Comput. 2030–2041 (2009)
4. Jacob, R.J.K.: Eye-movement-based human-computer interaction techniques: toward non-command interfaces. In: Proceedings of the Advances in Human-Computer Interaction, pp. 151–190. Ablex Publishing Corporation, Norwood (1993)
5. Kato, H., Billinghurst, M., Poupyrev, I.: Virtual object manipulation on a table-top AR environment. In: Proceedings of the ISAR 2000, Munich, pp. 111–119 (2000)
6. Kjeldsen, R., Levas, A., Pinhanez, C.: Dynamically reconfigurable vision-based user interfaces. Mach. Vis. Appl. **16**(1), 6–12 (2004)
7. Liu, S.T., Yin, F.L.: The Basic principle and its new advance of image segmentation methods. Acta Automatica Sinica **13**, 121–124 (2003)
8. Xu, X.Z., Ding, S.F., Shi, Z.Z., Jia, W.K.: New theories and methods of image segmentation. Acta Electronica Sinica **S1**, 76–82 (2010)
9. Min, F.W.: Research on Real-time Hand Tracking Algorithm Based on Cognitive-behavioral Model library and Kinect. University of Jinan (2014)
10. Feng, Z.Q., Yang, B., Zheng, Y.W., Xu, T., Li, Y., Zhu, D.L.: Gesture feature detection based on feature points distribution analysis. Comput. Integr. Manufact. Syst. **17**(11), 2333–2342 (2011)
11. Feng, Z.Q., Yang, B., Li, Y., Zheng, Y.W., Zhang, S.B.: Research on hand gesture tracking based on particle filtering aiming at optimizing time cost. Acta Electronica Sinica **37**(9), 1989–1995 (2009)
12. Yang, B., Song, X.N., Hao, X.Y.: Gesture recognition in complex background based on distribution feature of hand. J. Comput.-Aided Des. Comput. Graph. **22**(10), 1841–1848 (2010)

Building Targets Change Detection of SAR Images Based on Fuzzy Distances

Xiaoyan Li[1,2(✉)], Yun Sun[3(✉)], and Min Li[1(✉)]

[1] Xi'an Hi-Technology Institute, Xi'an, China
zgylixiaoyan@sina.com, clwn@l63.com
[2] Academy of Armored Force Engineering, Beijing, China
[3] The Logistics Academy, Beijing, China
Sunyun_hqxy@sina.com

Abstract. Propose a novel pixel-level change detection method based on the fuzzy distance for the interest building areas. Based on the imaging characteristic of Synthetic Aperture Radar (SAR) Images building areas, the variogram texture and gray feature are combined to be the feature vector in order to describe the image pixel feature clearly. And then the different image is computed using the fuzzy distance for detecting changes and obtaining quantitative detection results. Experimental results show that the proposed algorithm is feasible and effective to detect the change areas.

Keywords: SAR images · Building · Target detection · Fuzzy distance

1 Introduction

Target change detection based on SAR image has been a hot research problem in the field of graphic images. Due to speckle noise in SAR image and relatively bigger gray difference among multi-temporal SAR image, traditional difference method based on image gray information can't obtain good effect on change detection. In this paper, the traditional image difference method is improved by using the eigenvector combined gray features and texture features, and the fuzzy distance is used to detect the difference. The experiments show that the algorithm is simple, efficient and can achieve good effect on change detection.

2 Preprocessing of Change Detection

2.1 Image Registration

Image registration is the basis of target change detection and its purpose is to make the multi-temporal SAR images coincident with each other on position, image scale, etc. If change detection is made directly between two images without image registration preprocessing, a lot of error inspection area will be produced and the effect of the change detection cannot be assured.

2.2 Noise Suppression

Change detection Based on SAR image must meet high requirements in speckle noise suppression, and it is demanded that the noise reduction algorithm not only can effectively suppress speckle noise but also can keep the image details. Therefore, an adaptive spot reduction algorithm with good performance is required. Noise suppression filtering technology after image is commonly used, and among them the Multi-scale Geometric Analysis is mainly used in filtering method due to the better performance for preserving edges and smoothing noise.

3 Change Detection Method Based on Fuzzy Distance

3.1 Fuzzy Distance

Referring Shang's definition, the fuzzy likelihood function between point $\vec{a} = (a_1, a_2, \cdots a_n)$ and $\vec{b} = (b_1, b_2, \cdots b_n)$ in N-dimensional space can be defined as follow:

$$S_{\vec{a}\vec{b}} = \frac{1}{n} \sum_{i=1}^{n} \frac{a_i \wedge b_i}{a_i \vee b_i} \tag{1}$$

Among them $S_{\vec{a}\vec{b}} \in [0, 1]$, the bigger the $S_{\vec{a}\vec{b}}$, the smaller the difference between $\vec{a}$ and $\vec{b}$, that is to say, the shorter the distance between two points. Therefore, A new fuzzy distance between $\vec{a}$ and $\vec{b}$ can be defined as follow:

$$W_{\vec{a}\vec{b}} = 1 - S_{\vec{a}\vec{b}} \tag{2}$$

Calculation and comparison are done using the follow drafted points set:

$$X_1 = [1, 2, 3, 4], X_2 = [2, 3, 4, 5], X_3 = [2, 1, 4, 3], X_4 = [1, 2, 3, 8], X_5 = [3, 4, 5, 6]$$

And then the results are shown in Table 1.

Table 1. Euclidean distance and fuzzy distance calculations

	d_{12}	d_{13}	d_{14}	d_{15}
Euclidean distance	2	2	2	4
Fuzzy distance	0.321	0.375	0.125	0.475

The defect of Euclidean Distance can be seen intuitively from these above results. The d_{12}, same as d_{13} and d_{14}, is equal to 2. The Euclidean Distance can't distinguish the similarity among the previous four groups of sets, but Fuzzy Distance is suitable to compare it qualitatively, according to both equal degree of samples' distance and the resemblance of samples. Therefore, the Fuzzy Distance shows more the relationship among these sets than Euclidean Distance.

3.2   Change Detection Method Based on Fuzzy Distance

3.2.1   Feature Selection

Eigenvector, composition with the texture feature and gray feature of the SAR image, is introduced for change detection to solve the limit problem of those methods just using gray information. The texture feature is more stable than the gray feature. However the choice of texture features is key point of detection performance. The gray level co-occurrence matrix (GLCM), with high calculation complexity, was used in most of the previous literature. The variogram is used for texture feature extraction in this paper. The variation texture feature, with low calculation complexity, can reflect the dependence of gray among the pixel.

Calculating the unified window size of variation texture is consist of the following steps in this paper:

Step 1: Choose five pixels in an image randomly and obtain the window images with 25×25 pixels, whose center points are these five pixels.

Step 2: Calculate variogram curve of each window image. If the variable range a exists and $\left[\frac{a}{\sqrt{2}}\right]$ is odd, then the window size $w = \left[\frac{a}{\sqrt{2}}\right]$. If $\left[\frac{a}{\sqrt{2}}\right]$ is even, then $w = \left[\frac{a}{\sqrt{2}}\right] - 1$. If the variable range does not exist, then $w = 25$.

Step 3: Choose the Minimum size of five windows as the unified window size.

At the same time the local statistical characteristics is taken into account to compensate for the variation texture feature in the description of local image features. Local statistical features focus on the local characteristics of the image, mainly including each order moment of the image gray. The first order moment (standard) and the second order moment (deviation) of image gray are used to assist research.

Set the registered SAR image before change as $f(x, y)$, set the registered SAR image after change as $f'(x.y)$. Correspondingly (i, j) is the pixel in the $f(x, y)$ and (i', j') is the pixel in the $f'(x.y)$. Therefore, following these three steps, window size named by $W \times W$ can be obtained, window images named by $w(x, y)$ and $w'(x, y)$ can be constructed, the gray standard and deviation of window images can be calculated. At the same time, variation texture feature WR can be calculated through setting the lag distance $h = 1$. So, as to pixel (i, j) and (i', j'), the corresponding eigenvector can be obtained using the follow expression.

$$F_{ij} = \left[\mu_{ij}, \sigma_{ij}^2, WR_{ij}\right] \tag{3}$$

$$F'_{i'j'} = \left[\mu'_{i'j'}, \sigma'^2_{i'j'}, WR'_{i'j'}\right] \tag{4}$$

3.2.2   Fuzzy Difference Graph

In order to obtain the same proportion of influence on change detection results with each feature, fuzzy distance is adopted as a measurement of distance between the eigenvector in this paper. That is, fuzzy differences graph $D(x, y)$ can be defined as:

$$D(i,j) = W_{F_{ij}F'_{i'j'}} = 1 - \frac{1}{3}\sum_{k=1}^{3}\frac{F_{ij}^k \wedge F_{i'j'}^{\prime k}}{F_{ij}^k \vee F_{i'j'}^{\prime k}} \qquad (5)$$

3.2.3 Threshold Segmentation

A threshold is used to distinguish the regions with change or no change and this process is equivalent to the image segmentation. Therefore, the threshold selection affectes the change detection results directly. However, the threshold is usually obtained by experience or many times experiment. The automatic method of threshold selection includes iterative algorithm, the Ostu threshold method, etc.

(1) Iterative Algorithm

The thinking of the algorithm can be described as following steps:

Step1: select an initial threshold

Step2: segment an image into different regions with change or non-change

Step3: Calculate the standardes of gray level in the above region and select the average value as a new threshold

Step4: End the Iterative procedure until threshold change is less than the parameters set beforehand. Otherwise repeat iteration from step1.

(2) Ostu Algorithm

The Ostu threshold method uses the Maximum between-Cluster Variance of discrete probability density function to determine the threshold. A discrete probability density function is expressed as follows:

It is assumed that L grey levels would exist through applying standard grey level quantization to image. Set $i = 0, 1, \cdots, L - 1$, $N(i)$ is the pixel number marked with the i-th grey level, M is the total pixels number of the whole image, and then the appearance probability of the i-th grey can be expressed as:

$$P(i) = \frac{N(i)}{M} \qquad (6)$$

If choose a threshold K, the between class variance is defined as:

$$\sigma^2 = \omega_0(\mu_0 - \mu_T)^2 + \omega_1(\mu_1 - \mu_T)^2 \qquad (7)$$

The parameters of above formula can be obtained using the following expressions.

$$\omega_0 = \sum_{i=0}^{k-1} P(i) \qquad (8)$$

$$\omega_1 = \sum_{i=k}^{L-1} P(i) \qquad (9)$$

$$\mu_0 = \sum_{i=0}^{k-1} iP(i) \bigg/ \omega_0 \qquad (10)$$

$$\mu_1 = \sum_{i=k}^{L-1} iP(i) \bigg/ \omega_1 \qquad (11)$$

$$\mu_T = \sum_{i=0}^{L-1} iP(i) \qquad (12)$$

The threshold T maximizing the between class variance is the required threshold. Due to fuzzy distance reflects difference degree of two images, the smaller the distance, the smaller the difference and the lower the gray value in the fuzzy differences graph. Therefore, the change area can be obtained by making threshold segmentation on differential images. The T, classification threshold of the differential images, is automatically determined by the Ostu algorithm.

In addition, if the building target in the image was hit, the corresponding shadow region will also change. In order to eliminate the influence of shadow area, according to the gray value of shadow area, fuzzy differences graph can be defined as:

$$D(i,j) \begin{cases} \in & \text{changes} & if \quad D(i,j) > T \quad and f(i,j) > G \\ \in & \text{non - changes} & else \end{cases} \qquad (13)$$

Among them, G is a small positive number.

3.3   Change Detection Process of Building Targets

Algorithm steps in this Paper can be shown in Fig. 1.

4   Experimental Results and Analysis

4.1   Comparison of the Algorithm in This Paper and the Gray Level Difference Method

As shown in Figs. 2 and 3, there are two group of results contrast about building targets difference detection in SAR images respectively using traditional difference method based on image gray information (marked with c and d) and difference method in this paper (marked with e and f). It can be seen, difference detection can not only use gray information because there are a lot of clutter in the result figure, caused by radiation difference between images before and after hit, and changed region detection is not accurate.
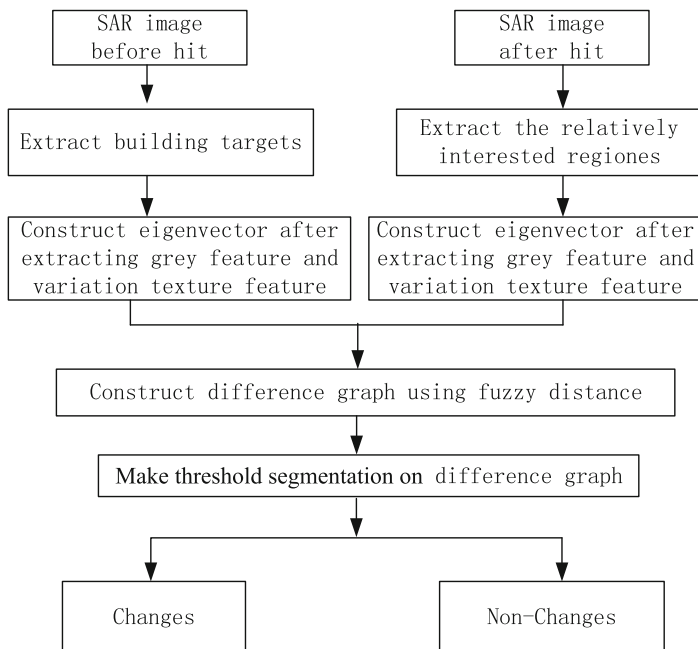
Fig. 1. Algorithm steps in this paper

From the pictures marked with f in Figs. 2 and 3, we can see that the building area with change in the image was successfully detected, and clutter was also reduced a lot, the detection precision is improved a lot. In the eigenvector calculation, window image size was set as 7 in Fig. 2 and window image size was set as 5 in Fig. 2.

## 4.2 Influence of the Window Size on the Change Detection Results

As to different types of images, it needs choose different size of window for feature calculation. The target size in the image is related to the window size. If the target area with changes is small, it uses a smaller window size. On the other hand, it needs a larger window for calculation.

As shown in Fig. 4, there are three group of results contrast about change detection respectively using three type size of windows. Among them, window size was set as 3×3 in c and d, window size was set as 5×5 in e and f, window size was set as 11×11 in g and h.

As can be seen from Fig. 4, 5×5 window is the most suitable for change detection experiment. For showing the characteristics of image pixels better, it needs comprehensive consideration of image local feature and dependence among pixels. It would ignore the dependence between the pixels and could not reflect the advantage of texture features when windows size is too small. Therefore change detection correct rate will be increased gradually by increasing window size. However, when the window size
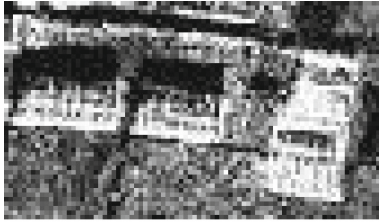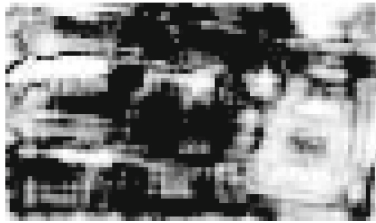
a. Image before hit
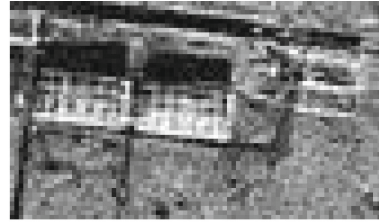
b. Image after hit
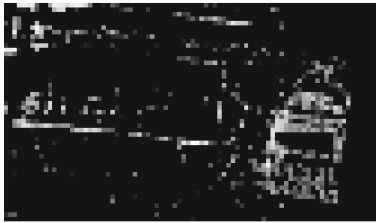
c. Traditional gray difference graph

d. Threshold segmentation graph of c

e. Difference graph using algorithm
in this paper

f. Threshold segmentation graph of e

Fig. 2. Result graph of building targets change detection in SAR images

a. Image before hit

b. Image after hit

c. Gray difference graph

d. Threshold segmentation graph of c

e. Difference graph using algorithm
in this paper

f. Threshold segmentation graph of e

Fig. 3. Result graph of building targets change detection in SAR images

increases to a certain degree, the image edge will be fuzzy, thus the precision degree of local image characteristics will be reduced, and then the correct rate of change detection will decrease.

The experiment results show that the proposed algorithm is feasible and effective to detect the change areas, and can solves the problem within a certain range, that it is difficult to extract key change from the detection information with radiation difference and image registration error.

a. Image before hit   b. Image after hit

Fig. 4. The detection results graph of building region with changes in SAR image

c. Region with change obtained
(window size is 3×3)
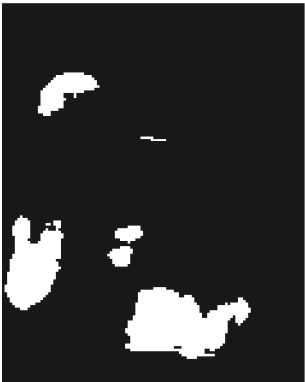
d. Mark c on the original drawing

e. Region with change obtained
(window size is 5×5)

f. Mark e on the original drawing

g. Region with change obtained
(window size is 11×11)

h. Mark g on the original drawing

Fig. 4.   (continued)

References

1. Po, D.D.Y., Do, M.N.: Directional multiscale modeling of images using the contourlet transform. IEEE Trans. Image Process. **15**(6), 1610–1620 (2006)
2. Marques, R.C.P., Sombra de Medeiros, F.N., Ushizima, D.M.: Target detection in SAR images based on a level set approach. IEEE Trans. Syst. Man Cybern. **39**(2), 214–222 (2009)
3. Grandi, G.D., Lee, J.S., Schuler, D.: Target detection and texture segmentation in polarimetric SAR images using a wavelet frame: theoretical aspects. IEEE Trans. Geosci. Remote Sens. **45**(11), 3437–3453 (2007)
4. Li, M., Liu, B.: Research and implementation of new image. Segmentation method. JCIT **7**(8), 110–119 (2012)
5. Chen, Z.: Fuzzy theory for the P2P subject trust evaluation model. IJACT **4**(8), 67–74 (2012)
6. Xu, X., Xu, S., Jin, L., et al.: Characteristic analysis of Otsu threshold and its applications. Pattern Recogn. Lett. **32**(7), 956–961 (2011)
7. Jia, T.-Q.: Adaptive threshold in segmentation algorithm based on fuzzy distances. Comput. Eng. Des. **35**(3), 856–860 (2014)

3D Plant Modeling Based on BP Neural Network

Jia Liu[1,2(✉)], Zhiguo Jiang[3], Hongjun Li[2,4], Weilong Ding[5], and Xiaopeng Zhang[2]

[1] School of Automation, Beijing Information Science and Technology University, Beijing, China
liujiahh@126.com
[2] LIAMA-NLPR, Institute of Automation, Chinese Academy of Sciences, Beijing, China
[3] School of Astronautics, Beihang University, Beijing, China
[4] College of Science, Beijing Forestry University, Beijing, China
[5] College of Computer Science & Technology, Zhejiang University of Technology,
Hangzhou, China

Abstract. We propose a fast plant modeling method based on BP neural network. The input is a plant image. Users can sketch the main branches and crown silhouettes on the image. Through branch copying, rotation and adjustment, the 3D main branches are obtained. A BP neural network is built and trained by analyzing the parameters of main branches. This network is used to forecast the parameters of small branches generated based on self-similarity. Finally, leaves are added and a 3D plant model resembling the input image is built. This method is based on one image and sketch. 2D sketch information is fully used for training the BP neural network and then forecasting the small branch parameters. This method relies on no database, and can be applied to many plant species. The experiments show that it runs fast and can build realistic plant models.

Keywords: BP neural network · Plant modeling · Image · Sketch

1 Introduction

In computer graphics, plant modeling is important and complex. Virtual plants are necessary in scene simulation, and can be applied to film making, game designing, etc. However, complex structures and diversiform species of plants bring difficulties for modeling. From 1960s, various methods have been developed, including procedural models (e.g. L-systems), image-based methods, interactive systems (e.g. sketch-based systems), and scanner based ones. They are applicable for a certain range of species, based on model database or parameterized rules, while their limitations exist in the aspects of diversity, convenient usability, and realistic effect.

We present a new image and sketch based plant modeling method with fast speed and realistic results. The system can be used as a convenient modeling tool for novices.

The motivation of this work is to model plants in our life as easy as possible. Through unskillful sketching, image processing is simplified. By fully utilizing 2D input information, no database is needed. By BP neural network training and forecasting, the complex problem of parameter adjustment in universal procedural models is solved. Its main contributions include:

- A novel 3D main branch modeling algorithm. The main branches are recovered from 2D sketches on one image by copying, rotating and adjusting. It can get even branch distribution and plausible branch structure.
- Parameters of small branches are obtained through BP neural network forecasting. The BP network is trained by parameters of main branches. It makes the small branches more reasonable.
- An efficient system for interactive modeling and designing 3D models of trees. The runtime of the system is less than 1 min on a common computer.

Experiments demonstrate that, various plants faithful to the 2D input sketches can be modeled by our system with no need for image registration and parameter adjustments.

2   Previous Work

Procedural models are the earliest developed plant models. Procedural methods generate contents by a procedure with the function of database amplification, and can be used to model plants, buildings, urban, texture, etc. In the aspect of plant modeling, they have been used to simulate botanical organs, growing process, and various plant structures.

L-system, a parallel string rewriting system, was first named by Lindenmayer in 1968 [1] to simulate cell division, and later extended to plant modeling by Prusinkiewicz and his collaborative. Now L-system has become a framework integrated with environment effects (such as Open L-systems [2]), interaction [3], physical quantities [4] and vision reconstruction [5], and various feasible plant models are obtained.

AMAP, another well known system for plant modeling, is based on biological knowledge and has become a commercial product. AMAP was first described in 1988 [6]. Every module of a plant model, such as buds, leaves, and roots have their own functions and fates. The growth of a plant is the evolution results of the corresponding modules [7].

Other procedural methods such as fractal-based methods [8] and particle systems [9, 10] are also used for plant modeling. Particle systems have been used to create a lot of natural objects such as plants, lightning, raining and cloud. Rodkaew et al. [11] and Runions et al. [12, 13] perform particle simulation for the creation of leaf vessels and trees. Beneš and Millán [14] model plants as systems of oriented particles that are able to sense their environment. Climbing plant structure is constructed by gradual growth of modules using directed random walking of particles. Neubert et al. [15] put forward a particle-flow method for fast tree modeling from images, as will be discussed in the image-based part below. More recently, Zhang et al. [16] use a hierarchical particle flow technique to model non-visible branches from scan points.

Although plausible models are obtained in the above methods, the final shape of the plant is not easy to control, and a lot of parameters are complex for users to adjust. Beneš et al. [17] propose a guided procedural modeling method to solve the problem of low controllability of procedural models. It divides each model into several parts generated by L-systems, and controls the final shape of the plant model by changing the silhouettes of the parts interactively, however, in the method, only 2D shapes can be modeled other

than 3D. Talton et al. [18] generate production that conforms to the specification by optimizing over the space of possible productions from the grammar. Pirk et al. [19] generate analogous models from an original one by approximating the tree's reverse natural growth. [20] gets appropriate tree models adapting environments by simulating a tree's proximity to solid obstacles and other trees. Wang et al. [21] present an variational computing approach for generating realistic trees in specific shapes.

Image-based methods are newly developed and become popular in recent years. Calibrated images are normally used as inputs in classical works, such as Shlyakhter et al. [22], Pan et al. [23], Quan et al. [24], Tan et al. [25], Teng and Chen [26], and Lopez et al. [27]. Shlyakhter et al. [22] construct a visual hull from 4–15 calibrated images with certain overlaps, and use its medial axis as skeleton. Details are added using L-systems. Pan et al. [23] and Teng and Chen [26] reconstruct plants from two pictures taken almost in the same direction by image matching. Quan et al. [24], Tan et al. [25] and Lopez et al. [27] recover the 3D point cloud or 3D skeleton of a plant based on structure from motion algorithm. In these methods, camera calibration and image segmentation are necessary steps.

The method of Neubert et al. [15] is an image-based method for approximate tree modeling using particle flows. It is based on the assumption that the camera models of input images are parallel projections and the input images are at an equal angle. A 3D direction field is built from input images, and the particles flow downward in the field to form a 3D tree skeleton shape. Initial positions of these particles are determined by calculating the volume density values through a rendering equation. Realistic results are achieved, however, as particle movement cannot be accurately controlled, branch models are greatly different from their input images. In our method, we realize the 3D construction by propagating and converting a 2D skeleton into 3D, and compared with [15], more reliable main branch structures can be obtained. Tan et al. [28] use a single image with several strokes as inputs and model trees by synthesizing branches from a growth engine with a small library of elementary, based on global position optimization. The strokes are used as guidance for image segmentation. In our method, we use strokes as inputs while the tree image provides sketching guidance for novices. In addition, parameter forecasting by BP neural network rather than global optimization for branch parameters makes our method run much faster than [28].

Scanner-based methods use 3D point clouds from scanners for plant reconstruction, such as [29–32]. Xu et al. [29] calculate shortest paths in a graph and cluster points to reconstruct the main skeleton. Cheng et al. [30] construct faithful 3D model of the trunk and main branches of a real tree from a single range image by shape decomposition and shape fitting based on incomplete data. Zhu et al. [31] reconstruct tree crown based on alpha shape modeling. Livny et al. [32] reconstruct main tree structures through 3D point processing, and enrich the model using L-system generated branchlets by lobe texturing. Yin et al. [33] present an interactive technique for surface reconstruction from incomplete and sparse scans of 3D objects possessing sharp features, and it can be used for plant modeling. Zhang et al. [16] use a data-driven technique to model trees from a single laser scan, in which visible branches are constructed through cylinder marching in scan data, while non-visible branches are generated based on a hierarchical particle flow technique. Scanner-based methods can achieve measurable branch structures; however,

generally, the scanners are costly and not convenient for data collection, which blocks their applications in fast modeling.

Unlike sketch-based modeling methods for other things [34], sketch-based plant modeling mainly deal with the problem of branch optimization. Okabe et al. [35] convert a user's 2D strokes into a 3D plant by optimizing angles between branches as large as possible. Chen et al. [36] convert a user's freehand sketch into a full 3D tree model by probabilistic optimization based on a database of tree models. These methods are flexible, however, they need artistic 2D sketches as input, and artistic drawing is difficult for novices. We solve this problem by sketching on a tree photo, which makes sketching easier for users.

Wither et al. [37] designed trees by inferring branches from silhouettes at multiple scales. Longay et al. [38] simulates competition of branches for space and light as the tree develops from a seedling. The tree structure can be controlled by directing growth with a procedural brush. Liu et al. [39] model a tree by branch retrieval from two different sketches, and achieve higher diversity in models conforming to left and side sketch constraints. Li et al. [40] model trees from a frame of video. They copy a 2D skeleton to more planes, and spread the branches in all directions uniformly within both local and global constraints.

Since most of tree modeling methods contain procedural process, and they generally can be sorted into more than one class, the sorting of all these methods is not rigid, and is just based on their main features. Our system discussed in this paper is also a procedural model using sketches on an image as inputs.

3   The Modeling System

Image-based methods generally use several calibrated images with large overlaps to construct plant models base on computer vision algorithms. Although the constructed models are realistic, the process is complex and slow; Sketch based methods are fast and agile, but sketching skills or special model databases are needed, and most of the sketch-based methods are only fit for plant species of similar shapes. Our tree modeling method is based on sketches on a tree image, and is designed for easy use, fast speed and realistic results of wide species.

The process of modeling begins with users' sketching on a single image to mark up the main branches and crown silhouettes. After sketching, 2D skeleton is built from strokes (Sect. 3.1). The 3D main branches are modeled through branch copying, rotating and angle adjusting (Sect. 3.2). A BP neural network is built and trained by analyzing the parameters of main branches (Sect. 3.3). The BP neural network is used to forecast the parameters of small branches, which are generated based on self-similarity (Sect. 3.4). Finally, leaves are added and a 3D tree model resembling the input image is built (Sect. 3.5).

3.1   Sketching and 2D Skeleton Building

A plant image captured by hand or downloaded from internet can be used for sketching. As the image is just used as a sketching guidance, there are no special requests as long as the plant can be distinguished by eyes. The user draws several strokes on the image to present main branches and crown silhouettes. For a novice, sketching generally takes 2–3 min.

The algorithm for 2D skeleton building in [39] is used to keep individual sketch characteristics and it achieves high speed. The process includes two steps conducted at the same time: local pixel analysis and new skeleton point linking. The algorithm is designed for two cases, viz. single-pixel strokes without radius information and brush strokes with branch radius information.

For single-pixel strokes, there are pixel superpositions in the connection parts, and all pixels are individual points rather than a hierarchically dendriform structure, therefore some necessary points are extracted by pixel clustering and selection in neighbor areas and linked into a tree structure. Brushes can be used when a user wishes to express specific branch thickness in the sketch. In this case, brush widths are detected as prior constrains in branch radius calculation. The algorithm is conducted by horizontal and vertical scanning, center pixel judgment and simplification, and center point linking. Designed for dendriform and connected strokes, these skeleton building algorithms are available and easier to implement. Personal characters of sketches are preserved while high speed is achieved.

3.2   Main Branch Modeling

The 3D main branches of a plant are modeled by converting a 2D skeleton into 3D. In this process, the 2D skeleton should be propagated and expanded by branch copying, rotating and adjusting. The details of the 3D main branch modeling are discussed below.

Branch copying. As we know that, in most cases, users can only mark up the main branches that can be distinguished in the image. They are just a part of the plant branches. We propagate the branches of the 2D skeleton before converting them into 3D by branch copying.

The amounts of plant ramification are different for different plants. According to the most cases, we increase the amount of branches of 2D skeleton to its 2–3 times. The process is conducted in the 2D skeleton according to the sequence of branch hierarchy as the following steps.

(1) Read in a branch $l_\sigma(\sigma = 0, 1, 2 \dots Q$, where Q is the total number of original 2D branches) in the 2D skeleton in sequence, go to step 2;

(2) If all the branches points are in the range of the sketched crown silhouette go to step 3, or else go to step 4;

(3) Write a ratio of the points into a new branch, which can be used as a new branch and is linked to the tree, go to step 1;

(4) Search the points at the edges of the crown silhouettes, and find the part points in the silhouettes, go to step 5;

(5) If the number of branch points in the silhouette is too small (for example, less than five), go to step 1, or else go to step 6;

(6) Write a ratio of the points in the silhouettes into a new branch, which can be used as a new branch and is linked to the tree;

The above process goes on until the total number of branches is qualified.

Branch rotation. After branch copying, a propagated 2D skeleton with N $(N > Q)$ branches is obtained. In this section, we rotate all the 2D branches to get a 3D skeleton, aiming at even distribution and plausible structure to 2D inputs.

The branches in the 2D skeleton are divided into two parts, viz. original ones and copied ones. The original branches comes from sketches, and the copied ones comes from branch copying. Branch rotation is conducted by rotating all branches with a roughly even distributed angle while restraining the original ones with relatively small angles only around the upright direction, as makes all branches distribute uniformly and keeps the original 2D shape viewed from one side. The process is described below.

(1) Read in a 2D branch $l_\sigma(\sigma = 0, 1, 2 \ldots N)$. If it is an original branch, go to step 2, or else go to step 5;

(2) If l_σ is the branch of level 0 (the trunk of the plant), go to step 3, or else go to step 4;

(3) No rotation is conducted. Given the root point $P_0(x_0, y_0, z_0)$ (where $z_0 = 0$) on branch l_σ, P_0 is converted into a 3D point $P'0(x_0, y_0, z'0)$ (where $z'0 = x_0$). For any other point $P(x, y, z)$ (where $z = 0$) on branch l_σ, its 3D position is set to be $P'(x, y, z')$ (where $z' = z0$). go to step 1;

(4) Rotate branch l_σ by angle c_σ around Y-axis (In the reference frames shown in Fig. 1. The input image lies in its XY plane). c_σ is calculate by the formula: $c_\sigma = \sigma * 2 * \pi/N$ (N is the total number of 2D branches). go to step 1;

(5) Rotate branch l_σ by angle c_σ around Y-axis, and then rotate branch l_σ by angle d_σ around X-axis. c_σ is calculated by the formula: $c_\sigma = \sigma * 2 * \pi/N$, and $d_\sigma = \pi * (-1)^\sigma/M$, where M is a constant, here we use $M = 10$. to step 1.

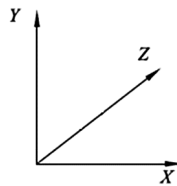The above process goes on until all branches are traversed.

Fig. 1. Reference frames

Branch adjustment. Now we have obtained a 3D skeleton, however, some 3D branches in the skeleton have abnormal angles with their neighbors, as making the structure unrealistic, for example, some angles are too big, and some are too small. The adjusting process contains coarse and accurate adjustments, as is described below (in the following steps, C, D, E are constant angle values, decided by experiments).

(1) Read in a 3D branch l_σ (σ = 0, 1, 2 ... N), whose farther branch is l_f (f = 0, 1, 2 ... N). If l_σ is an original branch, go to step 1, or else go to step 2;

(2) Suppose that the corresponding original branch of l_σ is l_k(viz. branch l_σ is copied from branch l_k). The coarse adjustments: If l_f is almost horizontal and the angle $\theta_{\sigma f}$ between l_σ and l_f is an obtuse angle, rotate l_σ around Y-axis by C degree to decrease the angle. If the angle $\varphi_{\sigma k}$ between l_σ and l_k is too small, rotate l_σ around Y-axis reversely by B degree to increase the angle. Go to step 3;

(3) The accurate adjustments: If l_f is almost perpendicular to the horizontal plane and the angle $\theta_{\sigma f}$ between l_σ and l_f is too small, rotate l_σ away from l_f by D degree on the plane made up of l_σ and l_f to increase the angle, and go to step 1, or else go to step 4;

(4) If the angle $\theta_{\sigma f}$ between l_σ and lf is too big, rotate l_σ towards l_f by E degree on the plane made up of l_σ and l_f to decrease the angle. Go to step 1.

As coarse adjustments are just rotation around Y-axis, which cost less time than accurate adjustments, we first do coarse adjustments for the majority and do accurate adjustments for the remainders.

Through branch copying, rotation and adjustment, a 3D skeleton with uniformly distributed main branches are obtained. In the next section, we build a BP neural network for parameter forecasting in the process of small branch propagation.

3.3 BP Neural Network Building

Artificial neural networks have been developed for dozens of years from 1940s. BP (Back Propagation) network is one of the most developed artificial neural networks, and has been applied in machine learning, pattern recognition, and automaton control.

A BP neural network is a multilayer feed forward network made up of an input layer, one or several hidden layers and an output layer, and trained by error back propagation algorithm. It can learn and store large numbers of relationships between the inputs and outputs, and be used to approach a mapping equation. Based on steepest descent method, weights are adjusted through error back propagation to minimize the error square sum. In our system, a BP network noted as *Net* is built and trained for parameter adjustment in small branch modeling.

Construction. The function of the BP neural network *Net* is to forecast the direction parameters of a new small branch based on the parameters of the main branches especially its father branch. Supposing that l_σ is a random branch grown from the node P_x (x = 0, 1, 2 ..., T) on the father branch l_f, the inputs of *Net* are parameters of the father branch l_f, including the normalized direction of l_f (viz. $v_f.x$, $v_f.y$, $v_f.z$), the ratio position of node P_x on l_f(viz. $r = length(P_0 \rightarrow P_x)/length(l_f)$, and the sibling number at node P_x(viz. N_s). The output parameters about the direction of branch l_σ are the cosine values of angle $\theta_{\sigma f}$ between l_σ and lf (viz.$\cos \theta_{\sigma f}$) and angle $\varphi_{\sigma k}$ between l_σ and its sibling branch lk(viz.$\cos \varphi_{\sigma k}$). In this way, we build a BP neural network of three layers, with node numbers 5-10-2, shown in Fig. 2.
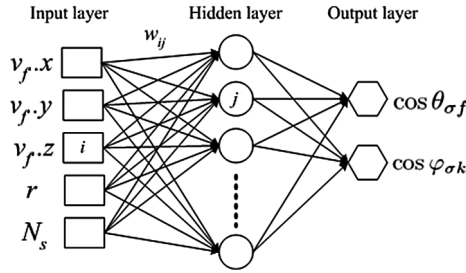
Fig. 2. BP neural network Net

Sigmoid function $f(x) = 1/[1 + exp(-x)]$ is used in *Net*. The weights between neighboring layers can be adjusted by training. Given that j is a node at the hidden layer or output layer, the input of node j is s_j, which is the weighted sum of the former layer, viz.

$$s_j = \sum_i w_{ij} y_i \tag{1}$$

In the formula above, w_{ij} is the weight between node i in the former layer and node j in the current lay. y_i is the output of node i in the former lay. Considering the threshold θ_j in node j, the actual input for node j is u_j shown below.

$$u_j = \sum_i w_{ij} y_i - \theta_j \tag{2}$$

The output of node j is y_j

$$y_j = f(u_j) = f(s_j - \theta_j) \tag{3}$$

In the formula above, $f(x)$ stands for sigmoid function.

Training. Training data for the BP neural network are calculated from the 3D skeleton of main branches. The BP training contains two processes, forward calculation and backward spread of errors, as are discussed below.

(1) Calculate parameters throughout all branches in the 3D skeleton built in Sect. 3.2, as the inputs and outputs of *Net*;

(2) Set initial values of weights and thresholds of *Net*;

(3) Forward calculation: the output of the node in the input layer is equal to its input x_i, while the input u_j and output y_j of node j in the hidden lay or output layer can be calculated as follows.

$$u_j = \sum_i w_{ij} y_i - \theta_j \tag{4}$$

$$y_j = f(u_j) = 1/[1 + \exp(-u_j)] \tag{5}$$

(4) Backward spread of errors: the remedy of weights is calculated in (6).

$$\triangle w_{ij} = w_{ij}(n+1) - w_{ij}(n) = \eta \delta_j y_i \qquad (6)$$

where η stands for the learning rate, y_i is the output of the node i in the former layer. n is the cycle time of this process. If j is a node of the output layer,

$$\delta_j = f'(u_j)(d_j - y_j) = y_j(1 - y_j)(d_j - y_j) \qquad (7)$$

where d_j stands for the target output of node j in the output layer. If j is a node of the hidden layer,

$$\delta_j = f'(u_j) \sum_i \delta_i w_{ij} = y_j(1 - y_j) \sum_i \delta_i w_{ij} \qquad (8)$$

(5) The forward and backward processes go on circularly until the error square sum become less than a certain value or the maximal cycle number is reached.

After the BP neural network *Net* is constructed and trained, we use it for parameter forecasting in small branch propagation.

3.4   Small Branch Propagation

For automatic parameter adjustment, we use the trained BP neural network to forecast parameters in small branch modeling. The small branches are copied from the main branches with angle parameters obtained from BP NN *Net*, while sketched 2D crown silhouettes are utilized for crown shaping.

Based on the principle of self-similarity, small branches are inferred from the structure of main branches and connected in a plausible way to the main structure. In this way, new generations keep the shape features of the main branches.

(1) For a main branch l_f, copy part of it, rotating, translating, and adding this part to certain nodes of branch l_f, and in this way a new branch l_σ is modeled. In this process, the angle $\theta_{\sigma f}$ between l_σ and l_f and the angle $\varphi_{\sigma k}$ between l_σ and its sibling branch l_k should be in certain ranges (viz.$Range(\theta_{\sigma f})$ and $Range(\varphi_{\sigma k})$). Details are illustrated in step 2 and 3.

(2) The angel ranges are calculated by *Net* forecasting in the following way. The parameters of l_f are given and they are used as the inputs of *Net*, and l_σ's angles with l_f (represented by $\cos \theta_{\sigma f}$) and with l_k (represented by $\cos \varphi_{\sigma k}$), can be obtained by *Net*'s forward calculation (by Eq. (5)). Then the range values can be calculated:

$$Range(\theta_{\sigma f}) = [\cos \theta_{\sigma f} - 0.2, \cos \theta_{\sigma f} + 0.2] \qquad (9)$$

$$Range(\varphi_{\sigma f}) = [\cos \varphi_{\sigma f} - 0.2, \cos \varphi_{\sigma f} + 0.2] \qquad (10)$$

(3) After step 2, search a direction v_σ, whose angles with l_f and l_k are in $Range(\boldsymbol{\theta}_{\sigma f})$ and $Range(\boldsymbol{\varphi}_{\sigma k})$ at the same time. v_σ is the final direction of branch l_σ.

In the process of branch modeling, 2D crown silhouette from the input image is used for branch pruning. Finally, by converting the 3D skeleton into generalized cylinders, all 3D branches with similar shapes to the 2D input are obtained.

3.5   Leaf Modeling

Each leaf or fruit is represented by a quadrilateral textured with a 4-channels image. Leaves and fruits are added by placing a phyllotaxis of quadrilateral leaf mesh models to each small branch. When leaves or fruits are added, random factors are employed to leaf position, direction and angles between leaf and branch. The average angle between leave and branch can be adjusted according to structural knowledge or actual data from measurement. Finally, the textured mesh model of a plant is produced. However, the density of virtual leaves are only decided by branch positions. Its consistency with real tree image has not been considered, as can be improved in the future work.

4   Results

Four tree models of different structures are presented in this paper. The process of 2D skeleton building and 3D main branch construction takes only a few seconds. Small branch growth takes about 1 s for creating 200 branches. After sketching, the total time for the tree modeling is typically less than one minute on a PC with a 2 GHz CPU.

Using the same input image with Neubert et al. [15], the tree model built by our method has a more plausible trunk to the input (Fig. 3). Compared with Tan et al. [28] (Fig. 4), our method runs much faster, after sketching, tree skeleton construction takes less than 1 min on a PC with a 2 GHz CPU, while [28] costs about 20 min on a PC with 2.4 GPU. Although we may take more time for sketching, we could control the branch structure in crown, and our method can be used for various plant species of different characters. A Chinese gooseberry with bine is modeled from strokes on an image downloaded from internet, shown in Fig. 5, and a pine with verticillate branches is modeled in Fig. 6. In Fig. 7, three tomato plants are modeled using images downloaded from internet.

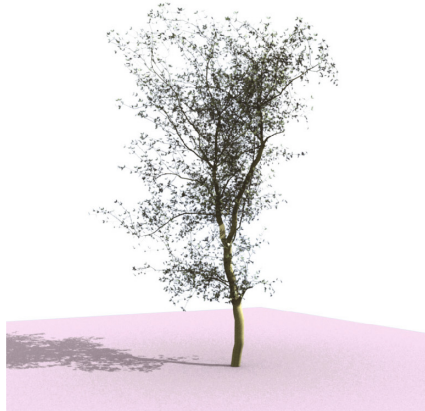(a) Input image [15]   (b) Our sketch
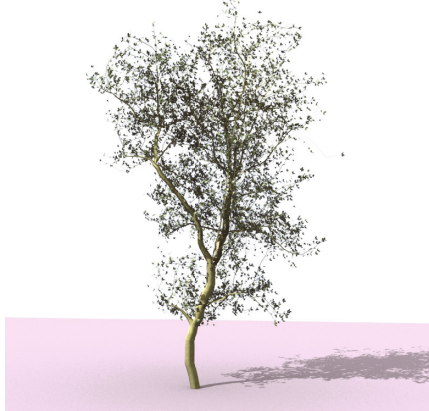
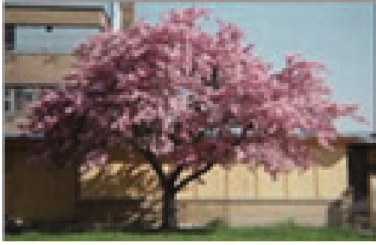(c) Tree model in [15]   (d) Our tree model

(e) Another view of our model

Fig. 3. Comparison with Neubert et al. [15]
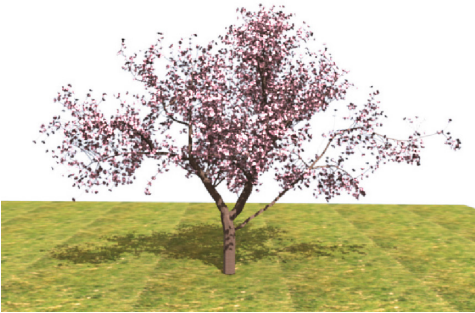
(a) Input image [28]

(b) Our sketch

(c) Branches in [28]

(d) Our branches

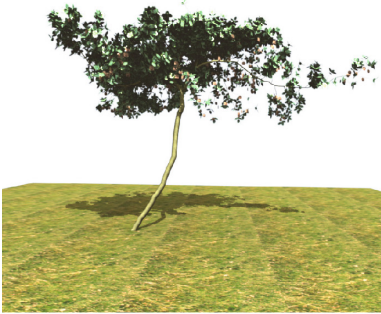(e) One view of our model

(f) Another view of our model

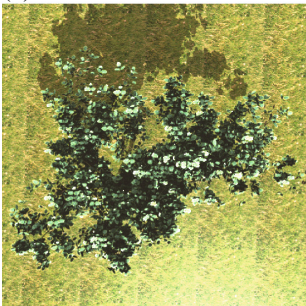**Fig. 4.** Comparison with Tan et al. [28]

(a) Input image   (b) Sketch
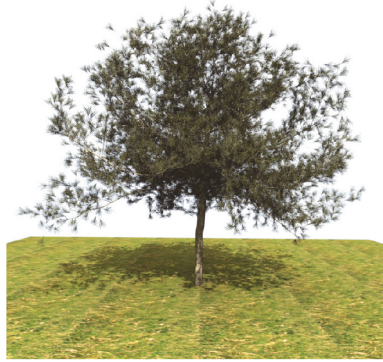

(c) Front side of our model   (d) Another side of our model

(e) Top of our model

Fig. 5. A Chinese gooseberry is modeled from strokes on an image downloaded from internet (bbs.gx118114.net)
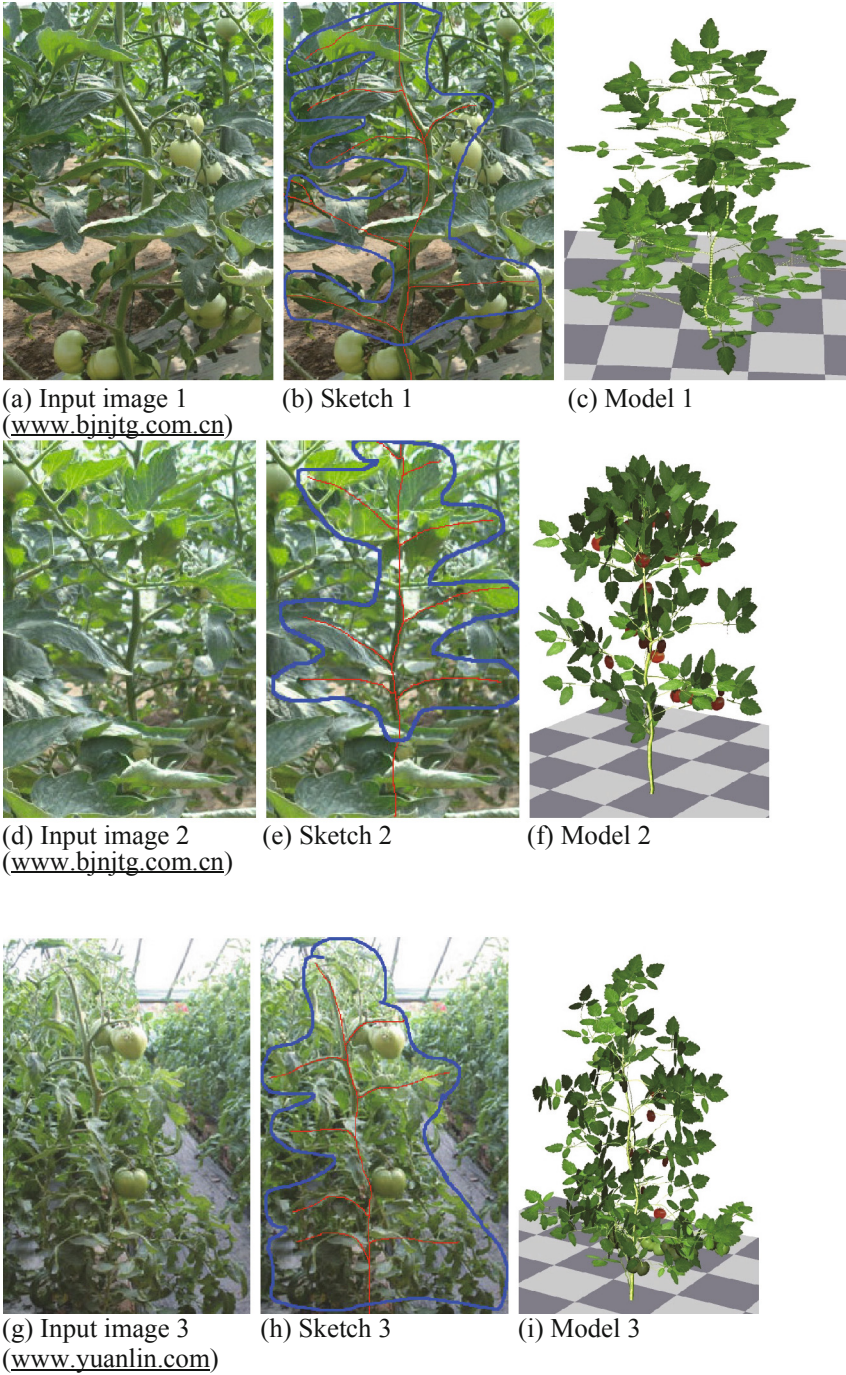
(a) Input image with strokes

(b) Front side of our model
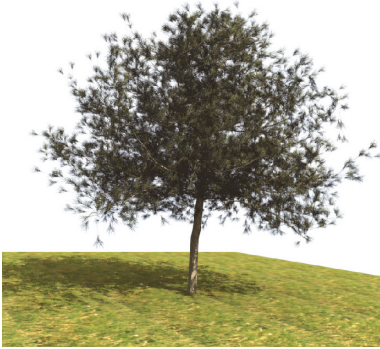
(c) Another side of our model
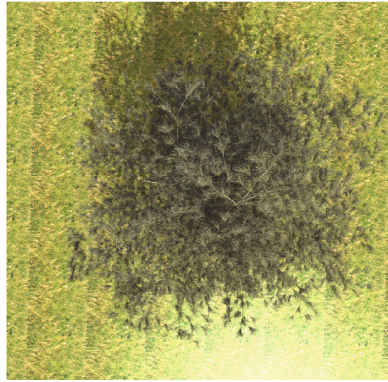
(d) Top of our model

Fig. 6. A pine is modeled from strokes on an image

(a) Input image 1   (b) Sketch 1   (c) Model 1
(www.bjnjtg.com.cn)

(d) Input image 2   (e) Sketch 2   (f) Model 2
(www.bjnjtg.com.cn)

(g) Input image 3   (h) Sketch 3   (i) Model 3
(www.yuanlin.com)

Fig. 7. Three tomato plants are modeled from strokes on images downloaded from internet

5   Conclusion

A fast method for plant modeling from sketches on a single image is proposed. The system can be used as a convenient tool for modeling various plant species with fast speed, realistic results and no parameter adjustment. In the further, local branch shapes should be optimized to avoid rigid corner in certain branch, and more biological information should be integrated to improve the visual effect.

References

1. Lindenmayer, A.: Mathematical models for cellular interaction in development. Parts I and II. J. Theoret. Biol. **18**, 280–315 (1968)
2. Měch, R., Prusinkiewicz, P.: Visual models of plants interacting with their environment. In: Proceedings of the 23rd Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1996, New York, NY, USA, pp. 397–410. ACM (1996)
3. Prusinkiewicz, P., Mundermann, L., Karwowski, R., Lane, B.: The use of positional information in the modeling of plants. In: Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 2001, pp. 289–300. ACM (2001)
4. Power, J.L., Brush, A.J.B., Prusinkiewicz, P., Salesin, D.H.: Interactive arrangement of botanical l-system models. In: Proceedings of the 1999 Symposium on Interactive 3D Graphics. I3D 1999, New York, NY, USA, pp. 175–182. ACM (1999)
5. Teng, C., Chen, Y., Hsu, W.: Constructing a 3D trunk model from two images. Graph. Models **69**(1), 33–56 (2007)
6. de Reffye, P., Edelin, C., Françon, J., Jaeger, M., Puech, C.: Plant models faithful to botanical structure and development. Comput. Graph. **22**(4), 151–158 (1988)
7. Lecoustre, R., de Reffye, P., Dinouard, P.: Controlling the architectural geometry of plant's growth - application to the begonia genus. In: Magnenat Thalmann, N., Thalmann, D. (eds.) Creating and Animating the Virtual World. CAS, pp. 191–214. Springer, Tokyo (1992)
8. Bloomenthal, J.: Modeling the mighty maple. In: Proceedings of the 12th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1985, New York, NY, USA, pp. 305–311. ACM (1985)
9. Arvo, J., Kirk, D.: Modeling plants with environment-sensitive automata. In: Proceedings of Ausgraph 1988, pp. 27–33 (1988)
10. Greene, N.: Voxel space automata: modeling with stochastic growth processes in voxel space. In: Proceedings of the 16th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1989, New York, NY, USA, pp. 175–184 ACM (1989)
11. Rodkaew, Y., Chongstitvatana, P., Siripant, S., Lursinsap, C.: Particle systems for plant modeling. In: Plant Growth Modeling and Applications, pp. 210–217 (2003)
12. Runions, A., Fuhrer, M., Lane, B., Federl, P., Rolland-Lagan, A.G., Prusinkiewicz, P.: Modeling and visualization of leaf venation patterns. ACM Trans. Graph. **24**(3), 702–711 (2005)

13. Runions, A., Lane, B., Prusinkiewicz, P.: Modeling trees with a space colonization algorithm. In: Proceedings of the Third Eurographics Conference on Natural Phenomena, NPH 2007, Aire-la-Ville, Switzerland, pp. 63–70. Eurographics Association (2007)
14. Beneš, B., Millán, E.U.: Virtual climbing plants competing for space. In: Proceedings of the Computer Animation. CA 2002, Washington, DC, USA. IEEE Computer Society (2002)
15. Neubert, B., Franken, T., Deussen, O.: Approximate image-based tree-modeling using particle flows. ACM Trans. Graph. (TOG) **26**(3), 88 (2007)
16. Zhang, X., Li, H., Dai, M., Ma, W., Quan, L.: Data-driven synthetic modeling of trees. IEEE Trans. Visual Comput. Graph. **20**(9), 1214–1226 (2014)
17. Beneš, B., št'ava, O., Měch, R., Miller, G., Quan, L.: Guided procedural modeling. Comput. Graph. Forum **30**(2), 325–334 (2011)
18. Talton, J.O., Lou, Y., Lesser, S., Duke, J., Měch, R., Koltun, V.: Metropolis procedural modeling. ACM Trans. Graph. **30**(2), 11:1–11:14 (2011)
19. Pirk, S., Niese, T., Deussen, O., Neubert, B.: Capturing and animating the morphogenesis of polygonal tree models. ACM Trans. Graph. **31**(6), 1–10 (2012)
20. Pirk, S., Stava, O., Kratt, J., Said, M.A.M., Neubert, B., Měch, R., Benes, B., Deussen, O.: Plastic trees: interactive self-adapting botanical tree models. ACM Trans. Graph. **31**(4), 1–10 (2012)
21. Wang, R., Yang, Y., Bao, H., Hongxin, Z.: Variational tree synthesis. Comput. Graph. Forum **33**(8), 82–94 (2014)
22. Shlyakhter, I., Rozenoer, M., Dorsey, J., Teller, S.: Reconstructing 3D tree models from instrumented photographs. IEEE Comput. Graphics Appl. **21**(3), 53–61 (2001)
23. Pan, Z., Hu, W., Guo, X., Zhao, C.: An efficient image-based 3D reconstruction algorithm for plants. In: Laganá, A., Gavrilova, M.L., Kumar, V., Mun, Y., Tan, C., Gervasi, O. (eds.) ICCSA 2004. LNCS, vol. 3044, pp. 751–760. Springer, Heidelberg (2004)
24. Quan, L., Tan, P., Zeng, G., Yuan, L., Wang, J., Kang, S.: Image-based plant modeling. ACM Trans. Graph. (TOG) **25**(3), 599–604 (2006)
25. Tan, P., Zeng, G., Wang, J., Kang, S., Quan, L.: Image-based tree modeling. ACM Trans. Graph. (TOG) **26**(3), 87 (2007)
26. Teng, C., Chen, Y.: Image-based tree modeling from a few images with very narrow viewing range. Vis. Comput. **25**, 297–307 (2009)
27. Lopez, L., Ding, Y., Yu, J.: Modeling complex unfoliaged trees from a sparse set of images. Comput. Graph. Forum **29**(7), 2075–2082 (2010)
28. Tan, P., Fang, T., Xiao, J., Zhao, P., Quan, L.: Single image tree modeling. ACM Trans. Graph. **27**(5), 108 (2008)
29. Xu, H., Gossett, N., Chen, B.: Knowledge and heuristic-based modeling of laser-scanned trees. ACM Trans. Graph. **26**(4), 19 (2007)
30. Cheng, Z., Zhang, X., Chen, B.: Simple reconstruction of tree branches from a single range image. J. Comput. Sci. Technol. **22**(6), 846–858 (2007)
31. Zhu, C., Zhang, X., Hu, B., Jaeger, M.: Reconstruction of tree crown shape from scanned data. In: Pan, Z., Zhang, X., El Rhalibi, A., Woo, W., Li, Y. (eds.) Edutainment 2008. LNCS, vol. 5093, pp. 745–756. Springer, Heidelberg (2008)
32. Livny, Y., Pirk, S., Cheng, Z., Yan, F., Deussen, O., Cohen-Or, D., Chen, B.: Texture-lobes for tree modelling. ACM Trans. Graph. **30**(4), 53 (2011)
33. Yin, K., Huang, H., Zhang, H., Gong, M., Cohen-Or, D., Chen, B.: MORFIT: Interactive surface reconstruction from incomplete point clouds with curve-driven topology and geometry control. ACM Trans. Graph. **33**(6), 1–12 (2014). (Proceedings of SIGGRAPH Asia 2014)

34. Xiao, D., Pan, Z., Zhou, R.: Sketch-based instancing of parameterized 3d models. In: Zhang, X., Zhong, S., Pan, Z., Wong, K., Yun, R. (eds.) Edutainment 2010. LNCS, vol. 6249, pp. 550–561. Springer, Heidelberg (2010)
35. Okabe, M., Owada, S., Igarash, T.: Interactive design of botanical trees using freehand sketches and example-based editing. Comput. Graph. Forum 24(3), 487–496 (2005)
36. Chen, X., Neubert, B., Xu, Y., Deussen, O., Kang, S.: Sketch-based tree modeling usingmarkov random field. ACM Trans. Graph. **27**(5), 109:1–109:9 (2008)
37. Wither, J., Boudon, F., Cani, M., Godin, C.: Structure from silhouettes: a new paradigm for fast sketch-based design of trees. Comput. Graph. Forum **28**(2), 541–550 (2009)
38. Longay, S., Runions, A., Boudon, F., Prusinkiewicz, P.: Treesketch: interactive procedural modeling of trees on a tablet. In: Proceedings of the International Symposium on Sketch-Based Interfaces and Modeling, SBIM 2012, Aire-la-Ville, Switzerland, Switzerland, pp. 107–120. Eurographics Association (2012)
39. Liu, J., Jiang, Z., Li, H., Zhang, X.: Easy modeling of realistic trees from freehand sketches. Front. Comput. Sci. **6**(6), 756–768 (2012)
40. Li, C., Deussen, O., Song, Y., Willis, P., Hall, P.: Modeling and generating moving trees from video. ACM Trans. Graph. (TOG) **30**(6), 127 (2011)

The Seamless Integration Achievement
of the Actual Situation of the Scene

Jinhui Huang and Haichao Shi[✉]

School of Computer and Information Engineering,
Beijing Technology and Business University, Beijing 102488, China
shihaichao1995@l26.com

Abstract. In the augmented reality systems, in order to give users results of visual consistency, the key is to achieve a virtual object and the real scene seamless overlay, combined with complex scenes direction of the light source to achieve a true desktop environment based on augmented reality systems, and in the system implements the seamless integration of the actual situation, including static and dynamic two parts. You first need to get the real scene of three-dimensional computer data in conjunction with the relevant algorithm to achieve, then the virtual object should be consistent with the real scene of human action, that is, action by the user to control the virtual object moves, mainly using Kinect tracking and bones action recognition, combined with interactive 3D engine to achieve.

Keywords: The actual situation · Seamless integration · Kinect · Interactive 3D · Point cloud data · Depth

1   Introduction

The augmented reality technology is developed in the virtual reality technology based on a new technology, a new expansion of virtual reality technology. Virtual objects and real scene superimposed seamless visual consistency can bring results, which is crucial to enhance reality technology [10], three-dimensional computer-generated virtual objects with real objects in the scene seamless integration, and thus play on real information to complement and enhance the role of the World. The seamless integration of static shielding effect is mainly a virtual object and the real scene, and seamless integration is a dynamic attitude to real-time control by the body movement of the virtual object, and add the collision detection between virtual objects and the real scene.

To achieve the seamless integration of the actual situation of the scene, you first need to complete a three-dimensional reconstruction of the real scene, through Kinect real scene reconstruction, three-dimensional scene with a virtual object superimposed and then rebuilt, and add the appropriate lighting effects and masking effects to achieve Augmented Reality system and the reality seamless integration. Virtual objects and real scene can not be a simple superposition, which requires visibility fusion system to study the actual situation, the real object and the virtual object if there is a depth of occlusion, occlusion effect establish and actual situation of the scene can be 3D interactive simulation engine Virtools achieve. Depth camera for rapid reconstruction

of the real scene and gesture detection and identification of the user, allowing users to achieve real-time control of the virtual object, the two achieve a depth camera-based augmented reality system, the actual situation of the seamless integration of interactive design.

2   Based on the Kinect's Reconstruction of Real Scene

The three-dimensional reconstruction [12] refers to digital information to represent the real scene, so that it can be displayed on the computer, which is based augmented reality scene of the key technologies in the computer.

In computer vision, 3D reconstruction refers to the process according to the image reconstructed single view or multiple views of information technology in various fields that have more and more widely used, also got the computer graphics, computer vision researchers attention. How the object in the computer for fast, effective reconstruction, has been the focus and difficulty of the study. Since Kinect appeared, for various reasons of its portability, low-cost, etc., a large number of domestic and international aid agencies are beginning to do research on Kinect, particularly in the three-dimensional reconstruction of the scene. In this paper, there would be analysis and comparison about the achieve results of three-dimensional reconstruction of different principles based on the Kinect, the following are the two methods commonly used in the reconstruction.

2.1   Based on the Point Cloud Data [11]'s Three Dimensional Reconstruction

Kinect color camera can capture the scene of material information, and infrared projector combined with infrared camera can get the depth information of the scene, but they did not merge, that is the real scene depth information and color information are separately stored, shown in Fig. 1:
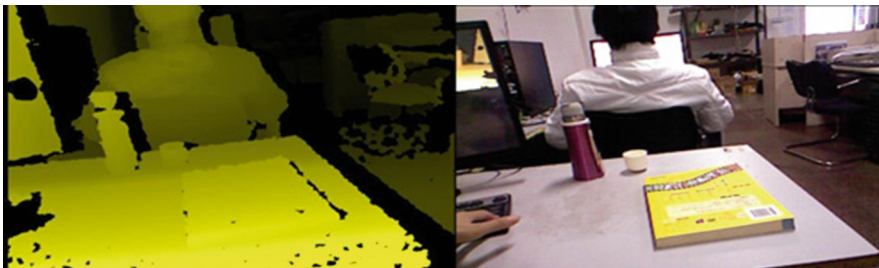
Fig. 1.   Kinect generated depth map and color map (Color figure online)

By default, the color image data obtained by Kinect [3] data is a const pointer to a one-dimensional matrix. The matrix size is of $640 \times 480 \times 3$, each of three groups representing the RGB values of the image.

The depth information is imaged by the realization of the principle of people's eyes, focusing on the need to acquire the baseline and focus, as well as parallax data between two cameras. By constructing a matrix Q, the parallax data can recover dimensional position information [15].

Since the Kinect depth camera and color camera [1] in the device is split into ends, but not exactly the same parameters of the camera lens itself, the two camera images collected will be slight different, from Fig. 1 you can see the left and right of the depth map display color graphics display is a difference, to complete a three-dimensional reconstruction of the scene you need the depth map and color images to achieve exactly the same, so the need to write code to a depth camera [5] perspective correction, this mainly using OpenCV interface to achieve. Skanect is such a reconstruction system based on point cloud data, when the system through the Kinect movement in the scene, it moves to the viewpoint of capturing images consecutive multiple scenes, including the scene in different locations and material information, real-time automatic the integration of depth information and generate three-dimensional scene of material information, the resulting point cloud data to be saved in .ply format. Ply file format is developed by Stanford University, a three-dimensional mesh model data format, the final reconstructed 3D point cloud data is all from the composition of the coordinates of a three-dimensional point.

The point cloud data can be opened for further editing in Meshlab or in Blender. Its reconstruction out of the scene contains material information, which is closer to the real scene. However, due to the resulting point cloud information only records the scene coordinate a three-dimensional point, and did not form a three-dimensional grid, as is shown in Figs. 2 and 3, so this is difficult to achieve virtual object and real scene cover.

Fig. 2. Reconstruction scene depth information and color information combined (Color figure online)

Fig. 3. Larger point cloud data in Meshlab

2.2   Based on the Grid Data's Three Dimensional Reconstruction [2]

This reconstruction of the above point cloud data based on the existence of difficulties in the practical application, so he appeared to point cloud data generated grid computing data reconstruction techniques, it is the emergence of a major breakthrough in the field of three-dimensional reconstruction.

This reconstruction method is based on the point cloud data to achieve, so in the implementation process need to get reconstructed point cloud data, due to the reconstruction of the point cloud data point will be noise, on the other hand the resulting point cloud file is too large for an object the description does not require much point. So the need for 3D point cloud data records generated noise streamline data after the vertex filter out key points and so on.

But screening is still not out of the point topology, as Fig. 4(a), the establishment of topology is an extension of the neighborhood's main data for each sampling point. P represents a point cloud with a collection of sample points for pi ∈ p, which can be defined as a neighborhood, it means that the k-th point in the region is the nearest point

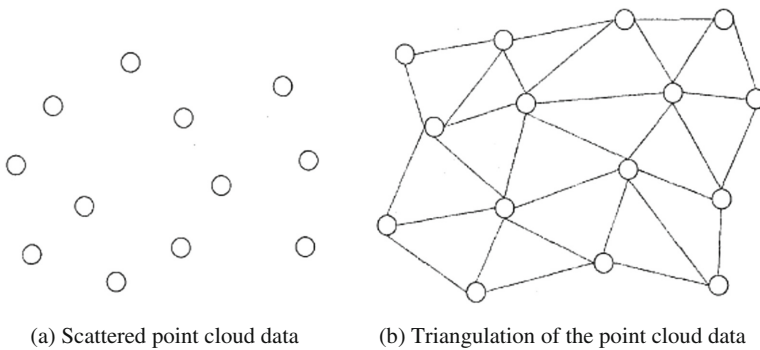(a) Scattered point cloud data   (b) Triangulation of the point cloud data

Fig. 4. Figure of point cloud data

from the point pi, these points form a collection K-nearest neighbors. To establish topology disordered point cloud midpoint between points by neighborhood expanding points. Neighborhood point search methods are octree, spatial bounding box method and k-d tree method.

With the topology of point cloud data to be converted to mesh a variety of methods, generally using triangulation of methods to achieve, Fig. 4(b), the most famous is the Delaunay triangulation as shown. It all points to the plane in cloud data, there is only a triangulation method, so to generate triangular patches meet split 'maximum - minimum angle' principle.

Delaunay triangulation has two important features:

(1) The minimum angle is the maximum: In the method of producing triangle subdivision, the triangle mesh Delaunay triangulation to obtain the minimum angles of maximum.
(2) Empty circumcircle: Any Delaunay triangulation of a triangle circumscribed circle containing the triangle formed only three points, and not include any other points.

Therefore, the formation of triangular mesh Delaunay triangulation is unique, the best grid.

Based on the above method, ReconstructMe reconstructs mesh three-dimensional scene, rebuilding Mesh data out of the scene, and can be stored as OBJ or STL format, into the mainstream of modeling software for scene changes, as Fig. 5 shows:
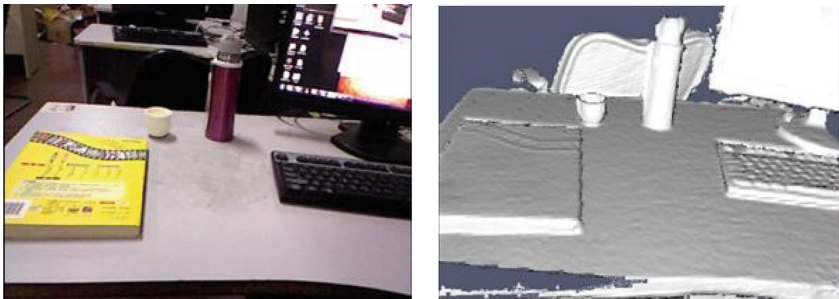

Fig. 5. ReconstructMe's reconstruction effect

3 Enhancing Reality System Integration of the Actual Situation of the Scene

After completing the three-dimensional reconstruction [8] of the real scene, and then reconstructed three-dimensional scene with a virtual object superimposed, and add the appropriate lighting effects and masking effect, in order to achieve the reality augmented reality system seamless integration.

In augmented reality system virtual object and the real scene of the most straightforward overlay effect, both before and after that will be the simple sum: Before

the three-dimensional virtual objects in the real scene of the video image, the super-imposition process without considering occlusion between them. But fusion accomplished not true, because the scene real objects and virtual objects in depth will certainly keep the situation obscured. Therefore, the study of the actual situation in the visibility of the fusion system is very necessary.

There are many ways to deal with visibility between objects, according to the realization of the principle can be divided into three categories, one is the object of an accurate method, it is an object considered as a whole, between the objects are compared with each other to determine each object the visible part of the image. The image is an accurate method for image processing of discrete representation, the whole idea is a solution by determining the image of each pixel in the image visible object for resolution. The Z-buffer algorithm [7] is based on a typical example of this type. A third approach is to target the exact method and the exact method by mixing images. In this context, it is often used interactive 3D engine Virtools, and occlusion of relations between the virtual object and the lighting effects and virtual objects in the real scene between the environment, as well as adding the collision detection in real-time interaction with virtual objects and interaction.

3.1   Actual Situation and Blocked the Scene to Establish the Effect of

Virtools Dev that contains a wealth of interactive virtual reality interactive module behavior editing software, which integrates three-dimensional models, pictures and particle effects, and other digital information, from which it can make different use of multimedia interactive products, such as computer games, city planning, education and entertainment, among other products.

Virtools Dev 'behavior and interaction module' can be freely combined to form different interaction effect, by simply dragging and dropping the behavior actions given on a specific object or character, and gradually formed edit interactive virtual world [4]. Virtools Dev in addition to good behavior and interaction module contains the set, but also includes application programming interfaces, designed by the programmer or generate new interactive module executable files.

Virtools operating interface consists of following parts (see Fig. 6):

(1) 3D Layout window: Displays the current scene in real-time 3D environment, and provides all the components necessary for operation of the 3D tools and navigation tools, the last generation of interactive screen is viewed in this window.
(2) Building Block window: Contains Virtools all comes with a 'behavior module', according to the category to find the behavior module belongs.
(3) Level View window: a clear tree structure diagram to find all the objects contained in the project editor and can be of different classes of objects in the data set in the window.
(4) Schematic window: mainly for different objects flowchart editor, to achieve the appropriate interactive action.

The front of the three-dimensional reconstruction using Kinect Grid [13] into the scene in 3D modeling software 3dsmax, and its make appropriate changes, and then
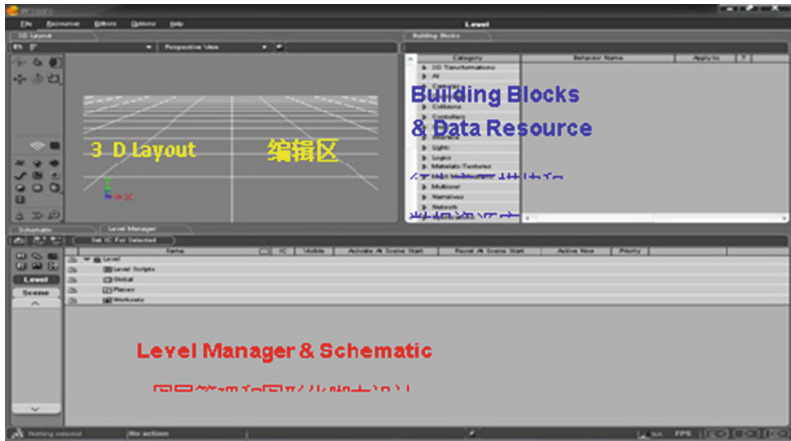
Fig. 6. Virtools interface

plug the three-dimensional scene using Virtools_3DSMax_Exporter guided into Virtools own format (.vmo or .nmo), and import it into Virtools in. Then the real scene contains a perspective view of adding cameras, adjust the camera's view, to make the picture of the object model and the object image coincide on the real scene. As is shown in Fig. 8(a) below.

In order to get better and the reality seamless fusion, light source information obtained by the coordinate conversion to convert it to 3D interactive engine Virtools world coordinate system, and set the appropriate properties of the virtual objects cast shadows, re-use interactive module ShadowCaster achieve shadow effect, as shown in Fig. 8(b). Then use the Z-Buffer principle to achieve real and virtual objects obscured scene, first introduced the principle Z_Buffer following algorithm:

Z-Buffer algorithm is the most famous image space visible algorithm, based on the fundamental idea is an associative array frame buffer, one that is associated with the pixel array elements, during the execution of the algorithm, keeping the pixels corresponding to the "depth" of information, but also is the Z coordinate value of each pixel, the initial value of 1.0, the algorithm for each object in the scene processing as shown:

(1) Each point (x, y, z) on the surface of an object corresponds to one pixel on the display screen (xs, ys), the coordinates are transformed through pan and zoom the display from standard, set I (x, y, z). It is the color intensity in the appropriate shading model scene in this point.

(2) If z < Z [xs, ys], then set the pixel (xs, ys) of the intensity I (x, y, z), and the Z value is assigned to the Z [xs, ys]; otherwise, do nothing.

So for now the pixel being processed, if the color needs to be rewritten, it must be represented by the new value of the point in the scene represented by the color compared to the previous point from the point of view closer to the scene, the angle Z buffer is recorded in the frame buffer color intensity corresponding to the currently set distance. The algorithm is the pixel level to replace close-range vision, object order of

appearance on the screen is irrelevant. Such substitution method to implement flexible and simple than the general sort, is conducive to hardware implementation [18]. So 3D interactive engine you need to set up a three-dimensional model of the real scene of Z buffer information, then the depth of information compared with the virtual object, to determine whether, and show which part of the scene, there are two general occlusion relationship, One is the real object occlusion virtual object, the other is a virtual object obscured the true object, in this article have done both cases verification, this part of the flow chart shown in Fig. 7:
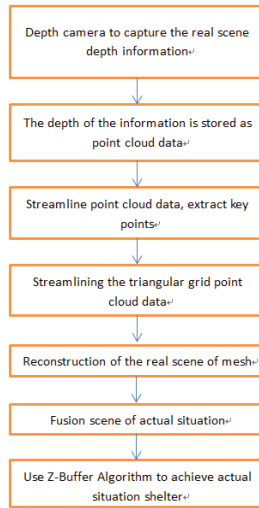
Fig. 7. The Flowchart about block of implementation of the actual situation

The ultimately results are in Fig. 8(c) and (d) below.

4   Based on Depth Camera Real-Time Interaction

In this part of the completion of two aspects, one is based on the depth camera to capture the body posture combined with interactive 3D engine Virtools [17] identify the user's action, on the other hand will recognize the movement into action command allows the user to manipulate virtual objects in real time, including control of virtual objects to move forward or backwards, and left and right to control objects with the steering, and collision detection combined with virtual objects or bypass forward promptly turned when confronted real objects in the scene. This part of the realization of the principle is shown in Fig. 9.

First modeled on the human body in the process, simplify the main bone joints, and body posture represents each named after the action, since the body's different actions have different amplitudes, such as hands can lift 30 cm, can be move higher, so be on
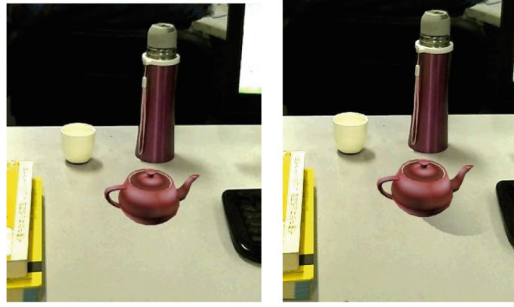
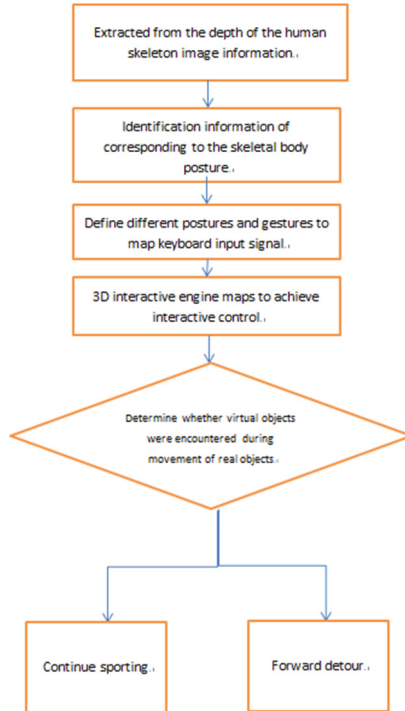(a) Actual situation fusion scene (b) Shadow effect of the virtual object

(c) The effect of real object blocking the virtual object

(d) The effect of the real object closuring virtual object

Fig. 8. The result of the actual situation of the scene

Fig. 9. The flowchart of real-time interactive control

the condition value different action to limit, thus avoiding misuse. In addition to the motion capture can be divided into static and dynamic, such as hand raised this action can be understood as a hand-held static after a good action, action would only use static analysis of the current condition of the target state value can be, but sometimes dynamic action needs to be analyzed, such as wave action or a circle, etc. These actions need to speed restrictions. Here the different attitude corresponding actions and restrictions set forth in Table 1.

In practice, the appropriate action needs to be added to the setting value and keyboard mapping [6], such as lean_forwards 15 key w represents the body forward 15 degrees mapping w key on the keyboard, and then set the w key corresponding to the engine through interactive 3D actions to complete the real-time virtual object manipulation. But it should be noted that due to some action represents an event might be triggered simultaneously, such as walk action due to heels on foot_up may trigger events, so the application should be taken to avoid setting the similarity larger movements. In this system, only set up four large discrimination action [16], respectively, lean forward 15 degrees forward, backward on behalf of the back, lift the left hand on behalf of the left, move to the right on behalf of right-hand turn. In the lab environment used to simulate the operation of a virtual human character models are provided by 3D interactive engine Virtools resource libraries, model skeleton, has pre-configured with different movements, such as forward and backward, the following

Table 1. Body posture represents the action set list and the corresponding threshold

| Types	Action name	Action description	Action threshold (unit)
Static action	lean_left	Body left	Body left-leaning angle (Degree)
	lean_right	Body right	After the body tilt angle (Degree)
	lean_forwards	Lean forward	Leaning forward angle (Degree)
	lean_backwards	Body backward	Leaning back angle (Degree)
	left_arm_forwards	Left hand reach	Distance from the shoulder to the hand (in)
	left_arm_down	Left hand down	Distance from the shoulder to the hand (in)
	left_arm_up	Left hand raised	Distance from the shoulder to the hand (in)
	left_arm_out	Left hand left stretch	Distance from the shoulder to the hand (in)
	left_arm_across	Left hand right stretch	Hand staggered distance from the shoulder into the body (in)
	right_arm_ forwards	Right hand reach	Distance from the shoulder to the hand (in)
	right_arm_down	Right hand down	Distance from the shoulder to the hand (in)
	right_arm_up	Right hand raised	Distance from the shoulder to the hand (in)
	right_arm_out	Right hand stretched	Distance from the shoulder to the hand (in)
	right_arm_across	Left hand stretched	Hand staggered distance from the shoulder into the body (in)
	left_foot_forwards	Left foot forward	Distance from the feet to the left hip (in)
	left_foot_sideways	Left foot	Distance from the feet to the left hip (in)
	left_foot_backwards	Left back	Distance from the feet to the left hip (in)
	left_foot_up	Left foot lift	Height from left foot to ground (in)
	right_foot_forwards	Right foot forward	Distance from the feet to the right hip (in)
	right_foot_sideways	Right foot right	Distance from the feet to the right hip (in)
	right_foot_backwards	Right foot back	Distance from the feet to the right hip (in)
	right_foot_up	Right foot lift	Height from right foot to ground (in)
	jump	Jump	

(*Continued*)

Table 1. (*Continued*)

Types	Action name	Action description	Action threshold (unit)
			Height from the two feet to ground (in)
	walk	Walk	Height from the feet to ground while walking (in)
Dynamic action	push	Push	Speed (in/s)
	swipe_up	Play up	Speed (in/s)
	swipe_down	Play down	Speed (in/s)
	swipe_left	Left swing	Speed (in/s)
	swipe_right	Right swing	Speed (in/s)
	circle	Circle	Radius (in)
	wave	Waving	Default 0

also need to set up operation and control of the avatar with the keyboard mapped to the corresponding key. Use role in the experiment control and keyboard mapping these two interactive modules combined action recognition based on depth camera enables real-time interactivity, integration of the virtual object scene graph and real-time interaction effects shown in Figs. 10 and 11:

Fig. 10. Containing a virtual object augmented reality scene
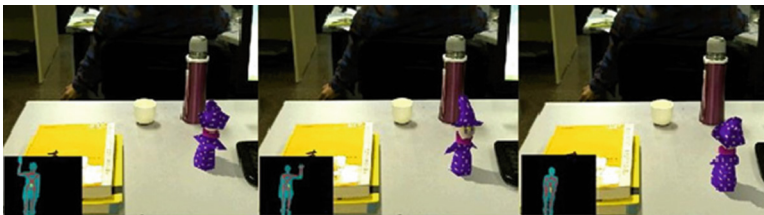
Fig. 11. The virtual object controlled by a different attitude

After the completion of the real-time control of the virtual object, virtual objects found occurrence "through walls" [14] in the face of real objects in the scene, shown in Fig. 12(a), it is because in the system is not set collision detection. Collision detection is very important in augmented reality systems, good collision detection requirements in the case of a variety of people blocked the forward direction as much as possible so that people have to move forward along the direction of rational rather than through the wall or through internal objects.

(a) No collision detection scene   (b) Virtual objects and collision cups, lids, books
                                              happened detection

Fig. 12. The contrast of before and after the collision detection

Several obvious obstacles are set up by the system in the scene, including a cup, a lid and two books, the virtual object in the face of these objects should stop conversion direction. In order to achieve the following results collision detection [19], the need for two aspects of the operation.

(1) Select the obstacles in its properties panel to add collision Manager of Fixed obstacle property, in the interactive process, Virtools collision detection engine detects whether the virtual object and set the corresponding attribute of an object collision.

(2) Add the virtual object itself the Prevent collision behavior module to prevent the virtual object through real objects at run time.

After finish the above operation, you can see the virtual objects while walking, collision detection engine detects whether the scene will be real-time detection of the collision, a collision occurs if the virtual object can not pass through glass, can only walk from the side. As shown in Fig. 12(b) below. Finally verify the occlusion effect in real-time interactive process between the virtual object and the real scene, the shelter was set lid, shown in Fig. 13:

Fig. 13. Real-time interaction during the occlusion effect

5   Conclusion

Augmented reality systems need to integrate the reality scenes to give users the experience of realism, but more importantly is controlled by the user scene virtual object interaction occurs between real objects and by this real-time interaction to expand awareness of the real world perception, but also enhance the realism of the system, but also improve an important means to enhance visual consistency reality system. At present, this aspect of the research work for a relatively small, but the emergence of Kinect depth camera for the augmented reality system provides a lot of interactivity possible, the paper realized the depth of the camera based on the actual situation of augmented reality system seamless integration of interactive design mainly it includes two aspects, one is by using a depth camera for quick reconstruction to achieve real scene actual situation of each scene occlusion effect, compared with the previous manual modeling improve work efficiency, and ensure accurate position of objects in the scene between sex. The other is by the depth camera gesture detection and identification of the user enables the user to control virtual objects in real-time operation, realized the function of real-time interaction, and add collision detection between virtual objects and real objects to make the effect more realistic.

References

1. Ju, R.: Joint scaled depth and color camera and in augmented reality application. Zhejiang University (2014)
2. Zhang, C.: Kinect depth camera indoor three-dimensional scene reconstruction. Dalian University of Technology (2013)
3. Peng, G., Xiangning, C., Bin, L.: Kinect sensor calibration color and depth camera. China Image Graph. **11**, 1584–1590 (2014)
4. Li, X., Guo, W., Li, S., Chen, C., Sun,. L.: A closed algorithm pose accuracy of the depth camera. Estimate Robot, February 2014
5. Rong, Z.: Get technology research depth information based on light field camera. North University (2014)

6. Ho, P., Gui, J., Lin, X.: Combining fast video matting kinect depth map algorithm. Tsinghua University (Natural Science), April 2012
7. Luo, Y., Hsieh, Y.: Chang Design and realization of the system kinect sensor-based intelligent wheelchair gesture controlled robot, January 2012
8. Quan, H., Chen, W.G., Zheng, B., Xu, Z.: Kinect application in a video conference system. Guangxi University (Natural Science) (S1) (2011)
9. Yu, X., Yang, X., Likai, Y., He, H., Zheng, X., Li, M.J., Yuan, J.J., Huhong, Y., Wu, T., Shi, K., Wang, R., Zhang, Y.-G.: Research and Implementation of B/S mode of the PACS Based on. J. Biomed. Eng. (3) (2004)
10. Dong, S.: Progress in human-computer interaction and challenges. Comput. Aided Des. Comput. Graph. (1) (2004)
11. Tong, J., To, X., Tian, H., Pan, Z., Cheung, M.: Use flight time three-dimensional shape of the camera's non-rigid three-dimensional reconstruction. Comput. Aided Des. Comput. Graph. (3) (2011)
12. Application of the depth camera on Computer Vision and Graphics (English). Comput. Sci. Explor. (6) (2011) to Xueqin, Pan Zhigeng, child crystals
13. Yang, C., Qi, Y., Shen, X.-K., Zhao, Q.-P.: A fast 3D scan data automatic registration method. J. Softw. **21**(6), 1438–1450 (2010)
14. Pan, H., Wang, Q., Xie, B., Xu, S.: ACTUATORS. TOF method data processing method for three-dimensional imaging camera. Zhejiang Univ. (Eng. Sci.) (6) (2010)
15. Zhuyan, J., Zhou, R., Zhang, L.: Scattered Data matching algorithm. Comput. Aided Des. Comput. Graph. (4) (2006)
16. Giles, J.: Inside the race to hack the kinect. New Sci. **208**, 22–23 (2010)
17. Chang, Y.J., Chen, S.F., Huang, J.D.: A kinect-based system for physical rehabilitation: a pilot study for young adults with motor disabilities. Res. Dev. Disabil. **32**, 2566–2570 (2011)
18. Zöllner, M., Huber, S., Jetter, H.-C., Reiterer, H.: NAVI – a proof-of-concept of a mobile navigational aid for visually impaired based on the Microsoft Kinect. In: Campos, P., Graham, N., Jorge, J., Nunes, N., Palanque, P., Winckler, M. (eds.) INTERACT 2011, Part IV. LNCS, vol. 6949, pp. 584–587. Springer, Heidelberg (2011)
19. Michael, N., Fink, J., Kumar, V.: Cooperative manipulation and transportation with aerial robots. Auton. Robots **30**, 73–86 (2011)

An Improved Edge Detection Method Using Adaptive Threshold

Xiangjiu Che, Li Wang[(✉)], and Xiaoxin Guo

Key Laboratory of Symbol Computation and Knowledge Engineering, Ministry of Education, College of Computer Science and Technology, Jilin University, No. 2699 Qianjin Street, Changchun 130012, China
liwangl99l@outlook.com

Abstract. Edge detection is an important step for extracting interesting feature information in image processing and computer vision. Although ant colony optimization (ACO) has been improved by using distributed adaptive threshold strategy (DATS), the artificial ants of this approach still easily ignore weak edges with lower edge gradient which results in detecting discontinuous edges of interesting features. To detect more continuous edges of features by using ACO in color and gray scale images, this work proposes an image pre-processing for ACO with DATS. The result of image pre-processing, which is the image after binarization processing by using adaptive threshold generated form Otsu's method, is taken as input for ACO. The purpose of image pre-processing is to provide salient changes of image gradient that original images couldn't provide for artificial ants. By quantitative analysis and subjective comparison of images in different sizes and types used as benchmarks for edge detection, our method extracts more continuous edges and provides more accuracy of interesting feature information than original ACO with DATS does. What's more, our approach detects all positive edge points of ground truth in our experiments.

Keywords: Edge detection · Adaptive global threshold · Image pre-processing · Ant colony optimization

1 Introduction

Important information contained in the edges of objects in digital images could be utilized as low level features in various image analysis and computer vision systems (Umbaugh 2005) such as object detection (Li et al. 2013), object recognition (Olson and Huttenlocher 1997), object tracking (Sullivan and Carlsson 2002) and image retrieval (Kokare et al. 2003). Edge detection is an essential step in image processing, especially in the areas of feature detection and feature extraction (Gonzalez 2002; Pratt 1991). In ideal cases, we hope the results of edge detection algorithms could provide continuous and smooth boundaries of objects focused. In practice, however, it is very difficult to obtain an accurate detection, especially when there exists noise in the image (Gonzalez and Woods 2007).

Inspired by the nature of swarm intelligence, such as the distributed and adaptive specialization observed in social insects and ants, scientists utilized such nature approaches into further increasing efficiency of self-organizing process and improving performance of solving problem. To date, Ant Colony Optimization (ACO) has been used as swarm intelligence into the application of image edge detection and feature extraction. The ACO is modeled based on a population of artificial ants to simulate the process of finding food source, which is completed by natural ants. However, a general ACO always extracts jagged edges from original digital images due to the amount of pheromone deposition and the selection rules from current location to a next pixel. Therefore, many researchers proposed various improved methods to enhance detection performance. For example, hybrid edge detection method(HM), proposed by (Manish et al. 2012), combined canny filter with scale multiplication (CSM) (Bao et al. 2005) and ACO with a static global threshold. The proposed method used the optimized canny filter outputs of different images as the inputs (pheromone map) for ACO. The amount of discontinuous edges reduces, but accuracy could be quite low when CSM can not detect weak edges (Manish et al. 2012); An distributed adaptive threshold strategy (DATS) (Mullen et al. 2013) is proposed to improve the low performance due to one global threshold, which results in weak edges with lower edge gradient being ignored by artificial ants. This approach allows each ant to carry their own local thresholds, then local thresholds will be dynamically adjusted when their owners move around the image. However, the proposed method provides lower detection performance when regions of interest or changes of image gradient could not be salient for ant agents in color images and gray scale images.

Static global threshold by user-set approaches usually extract discontinuous edges when its value is over high or too many useless edges when its value is over low. Choosing a suitable threshold value for detecting continuous edges of interesting features in different images is a difficult and time-consuming task. In order to solve the problem of selecting a suitable global threshold value to provide salient changes of image gradient for detecting continuous edges in both color and gray scale images, this paper proposes a novel method based on ACO with two adaptive thresholds: one is for image pre-processing and the other is for swarm self-organizing process.

2　Background

This section will provide a brief description of ant colony optimization (ACO), the difficult choice of threshold value T in ACO and distribute adaptive threshold strategy.

2.1　ACO for Image Edge Detection

This algorithm employs a certain number of artificial ants as computational agents. N ants in the process of initialization occupy N pixels within the image, where pixels in the image represent states in the search environment. Locating and mapping out boundaries of objects within the image is the purpose of ants moving around the image. In order to achieve this goal, solving the problem how to help each ant in current position choose one pixel from its allowed surrounding pixels as its next pixel is a

primary key. Thus, heuristic information is introduced. Heuristic information is an additional weight value on each pixel in the image map and help ants to choose a new pixel to move in. Pheromone deposited by ants in a pixel is along with the movement in ACO, which could record how many ants move through this pixel. And the higher amount of pheromone in a pixel shows a higher number of ants walk through, which means a higher probability for guiding future ants to move in. The amount deposited could be a function about change in image gradient. There is a form of forgetting, evaporation of pheromone, avoiding the algorithm converging too rapidly in order to favor the exploration of strange areas in search environment. Updating the pheromone map occurs at each iteration, along with the pheromone deposition and a constant evaporation. The transition rule is then defined as a function about the heuristic information and pheromone map.

The heuristic information could be defined as the visibility, Vxy, which represents the largest changes of local image gradient around current pixel location (x,y).

The quantity of pheromone deposited by the ith ant at the pixel location(x,y) could be given by:

$$\Delta\tau_{xy}^i = \begin{cases} V_{xy}/255 & \text{if the } i\text{th ant at pixel } (x,y) \text{ and } V_{xy} > T \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

Where T is static global threshold value defined by user that only allows ants to deposit pheromone at the pixels of edges or boundaries when their heuristic information is above a certain 'strength'.

2.2 Difficult Selection for Single Static Global Threshold in ACO

Ant agents are allowed to deposit pheromone if they are following a probable edge of gradient greater than static global threshold value T. Thus value T directly affects the pheromone map, and moreover, affects the performance of the algorithm. A low threshold value T allows ants to deposit amount of pheromone in the large regions of image which contains very weak edges and extract out lots of unwanted background information in details. However, a higher threshold value T restricts ant agents to make pheromone deposition in a small area of the image where there are only strong edges present, which results in missing weak edges of objects and extracts out discontinuous edges of interesting features with many breakpoints. Therefore, choosing the threshold parameter is critical to the performance of ACO. Besides, manual selecting a satisfactory static global threshold for every different image may need prior knowledge and time consuming. Thus it's a tough task as well.

2.3 Distribute Adaptive Threshold Strategy (DATS)

Instead of requirement of a global user-set threshold, an adaptive threshold strategy for all different images that each ant could maintain their own local thresholds' value and automatically change it as they move around the image (Mullen et al. 2013). The update rule is given as:

$$T_i^n = \begin{cases} T_i^{n-1} + 1 & \text{if } V_i(x, y) > T_i^{n-1} \\ T_i^{n-1} - 1 & \text{if } V_i(x, y) < T_i^{n-1} \\ \bar{T} & \text{if } T_i^{n-1} < \bar{T} \end{cases} \tag{2}$$

where $V_i(x,y)$ is the visibility of ant i at pixel location(x,y), and $\bar{T}$ is the computed average threshold over all ant after each iteration. Only one of the first two rules would happen at each step of one ant, and the third one occurs at the end of nth iteration to ensure that all T_i^n is latest for computing $\bar{T}$.

This update rule can allow each ant to increase or decrease their own local threshold values depending on their individual search route of the image, while the threshold maintains a value above the level of unwanted background noise computed from the global perception of the entire swarm. This method effectively deals with the situation of extracting interesting features which require various threshold values.

3 Otsu's Method for an Adaptive Global Threshold in Image Pre-processing

ACO with DATS has one prior assumption that every pixels in a given image has only one image intensity value, which means that this image need to be a gray scale or binary image. Thus color images must be converted into one of two types before feeding to ACO. In this research, we discuss about how to convert color images into better input for ACO instead of directly using gray scale images. Usually, the projective method converts a pixel from three-dimensional space (R, G, B) to one-dimensional space, which projects a point in a R, G, B for the axis of three-dimensional space onto straight line R = G = B. The computational formula is Gray value = 0.29900 * R + 0.58700 * G + 0.11400 * B. The gray scale image by using this method still contains lots of background information which will mislead ant agents in ACO. In order to further reduce unwanted background information and distinguish majority of interested edges from gray scale images, this work plans to use Otsu's method to generate adaptive global threshold for extracting interested features and binarization of the resulting image as pre-processing. Otsu's method (Otsu 1976) is widely used to determine a static global threshold value for binarization process. It divides all pixels of an image into two subsets (edge and non-edges) where it maximizes the discriminating criteria of interset variance between the pixel intensities in these subsets. The image pre-processing of this paper is that using the projective method converts color images into a gray scale image and binary image through Otsu's method. Then the result image is taken as input for ACO with distributed adaptive threshold strategy.

4 Experiments and Analysis

In this section, we provide two different image sets for subjective comparison and quantitative analysis respectively to ensure that our method will detect more continuous edges by using image pre-processing than original ACO with DATS in our experiments.

The first image sets includes three color images and three gray scale images which are commonly used as benchmarks for edge detection. The images, Brick House, Girl (Tiffany), Mandrill are downloaded from the USC-SIPI Image Database and McGuire Graphics Data. The second image sets including their ground truth are provided from Segmentation Evaluation Database which contains color images along with gray scale images and ground truth segmentations. To test our method, we use this approach to test different sizes of images and three color images and gray scale images for subjective comparison respectively.
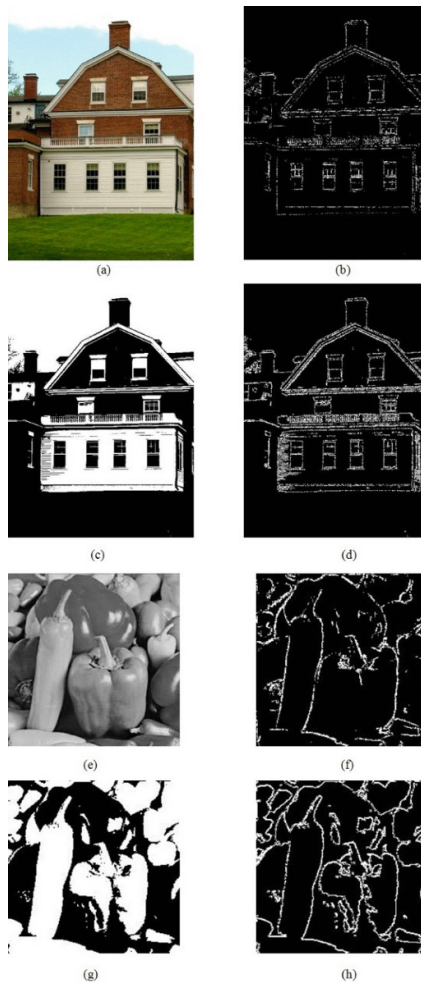
Fig. 1. Example image (a) and image (b) are original color images and original gray scale images. Images (b) and (f) are results obtained from ACO using DATS without proposed image pre-processing. Images (c) and (g) are results after image pre-processing. Image (d) and (h) are our results. Please refer to a digital copy for clear images. (Color figure online)

4.1   Subjective Comparison

In Fig. 1. Two images are used for subjective comparison between ACO using distributed adaptive threshold strategy (DATS) without image pre-processing and our method. It's noticeable that image (d) extracts more continuous edges of windows and eaves than image (b) and provides more clear structure under left chimney because of image (c) shows more salient changes of image gradient than original images. Even in gray scale images, our result, image (h), provides more details in edges of peppers than image (f).

In Fig. 2. Result images by using Otsu's method for image pre-processing all detect many more continuous edges and display more interesting feature information than ones without image pre-processing.
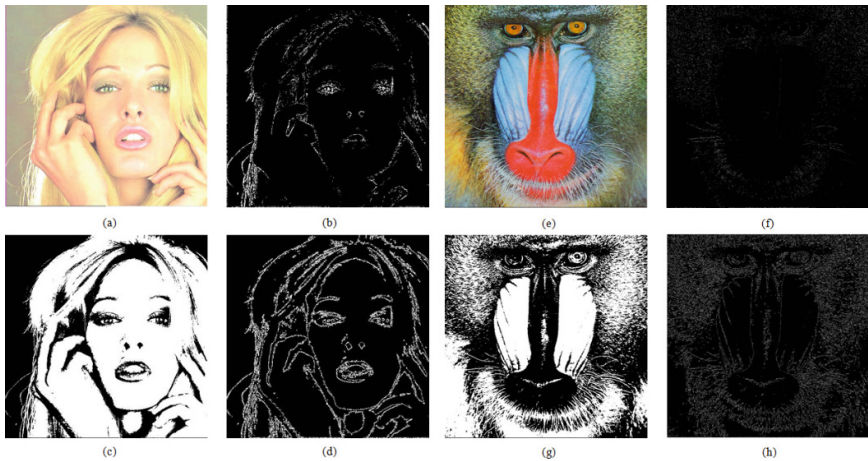
Fig. 2. Example images (a) and (e) are original color images. Images (b) and (f) are results obtained from ACO using DATS without proposed image pre-processing. Images (c) and (g) are results after image pre-processing. Image (d) and (h) are our results. (Color figure online)

In Fig. 3. In order to test further whether our method could still do better with gray scale images as inputs than original ACO with DATS, two gray scale images, cameraman and lena, were used to make the subjective comparison. As can be seen, our method still outperforms excellent results in providing salient changes of image gradient even if the input images are in different sizes (512*512 and 128*128 respectively).

4.2   Quantitative Analysis

We compare our results and results obtained from original ACO using DATS with ground truth for quantitative analysis. The comparison of edge maps extracted from two edge detection methods and ground truth is inferred by three parameters: proportion of correctly detected pixels (Pc), proportion of not detected pixels (Pn) and proportion of erroneously detected pixels (Pe).

$$Pc = TP/Max(N_g, N_d)$$

$$Pn = FN/Max(N_g, N_d)$$

$$Pe = PF/Max(N_g, N_d)$$
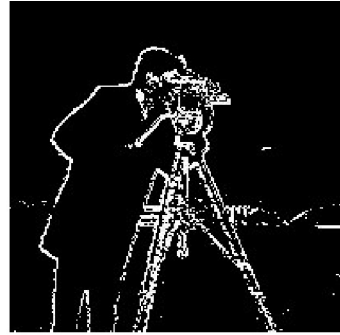
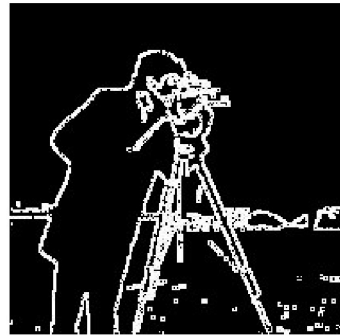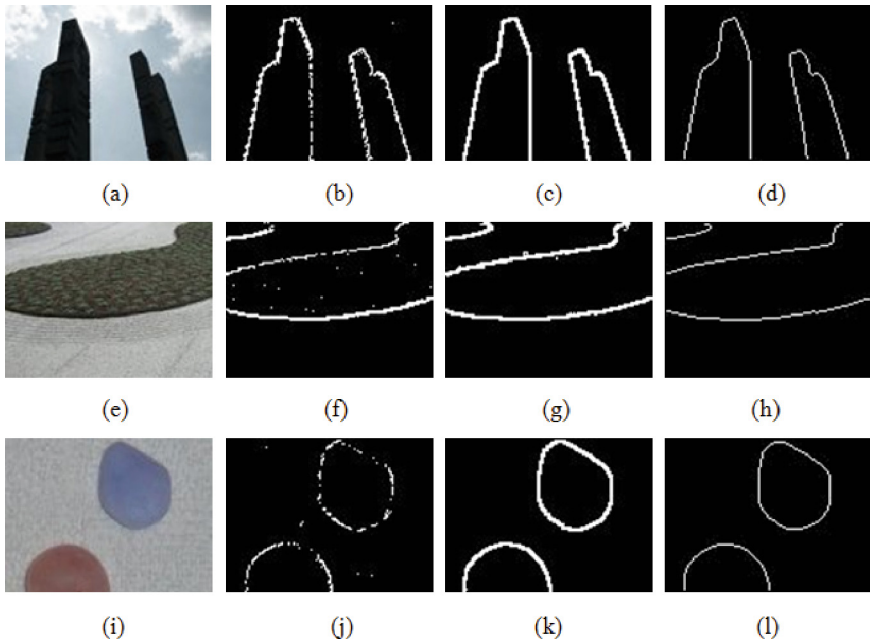Fig. 3. Example images (a) and (e) are original color images. Images (b) and (f) are results obtained from ACO using DATS without proposed image pre-processing. Images (c) and (g) are results after image pre-processing. Image (d) and (h) are our results.

where True Positive (*TP*) represents the number of correctly detected pixels, False Negative (*FN*) represents the number of not detected edge pixels (Abdou and Pratt 1979), False Positive (*FP*) represents the number of erroneously detected pixels, Ng is the number of edge pixels in ground truth and N_d is the max number of edge point detected by two edge detection methods.

Table 1. Measurements for three images and three measures

Images	Image 1		Image 2		Image 3	
Measures	ACO with DATS	Our method	ACO with DATS	Our method	ACO with DATS	Our method
Pc	0.774834	0.893491	0.543860	0.753894	0.388350	0.844262
Pn	0.119205	0.106509	0.456140	0.246106	0.203883	0.155738
Pe	0.225166	0.000000	0.087719	0.000000	0.611650	0.000000

In Table 1. Our method detects about 89 %, 75 %, 84 % correct edge pixels in image 1, image 2 and image 3 respectively, whose accuracy are 1.153, 1.386 and 2.173 times of original ACO with DATS respectively. Also, our approach extracts out all edge pixels in ground truth (Pe = 0). Figure 4 shows results from ACO with DATS, ours and ground truth. It is noticeable that our method does detect more correct edge

(a)   (b)   (c)   (d)

(e)   (f)   (g)   (h)

(i)   (j)   (k)   (l)

Fig. 4. Example images (a), (e), (i) are original color images. Images (b), (f), (j) are results obtained from ACO using DATS without proposed image pre-processing. Images (c), (g), (k) are results from our method with proposed pre-processing. Image (d), (h), (l) are ground truth. (Color figure online)

pixels and less false ones. The shortcoming is that ACO often extracts thick edges, which eventually reduces the global accuracy of detection in the image.

5   Conclusion

In this research, we found that unobvious edge gradient in color image and gray scale image made artificial ants ignore weak edges of interesting features in ACO with distributed adaptive threshold strategy (DATS). And an image pre-processing by using Otsu's method for binarization image is proposed to combine ACO with DATS to perform better results eventually. Two different types of performance evaluation measures, subjective comparison and quantitative analysis, are utilized to evaluate the performance of two edge detection methods and show that our method detects more continuous edges and interesting feature information by using adaptive global threshold for salient changes of image gradient in color images and gray scale images. In future, we will pay more attention on increasing the accuracy by detecting more thin edges.

References

Abdou, I.A., Pratt, W.: Quantitative design and evaluation of enhancement/thresholding edge detectors. Proc. IEEE **67**(5), 753–766 (1979)

Bao, P., Zhang, L., Wu, X.L.: Canny edge detection enhancement by scale multiplication. IEEE Trans. Pattern Anal. Mach. Intell. **27**(9), 1485–1490 (2005)

Gonzalez, R.: Digital Image Processing. Prentice Hall, Upper Saddle River (2002)

Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice Hall, Upper Saddle River (2007)

Kokare, M., Biswas, P.K., Chatterji, B.N.: Edge based features for content based image retrieval. Pattern Recogn. **36**(11), 2649–2661 (2003)

Li, H., Wang, Y.J., Liu, W.F., Wang, X.M.: Detection of static salient objects based on visual attetion and edge features. In: Proceedings of the Fifth International Conference on Internet Multimedia Computing and Service, pp. 252–255 (2013)

Manish, T.I., Murugan, D., Kumar, T.G.: Edge detection by combined Canny filter with scale multiplication & ant colony optimization. In: Proceedings of the Second International Conference on Computational Science, Engineering and Information Technology, pp. 497–500 (2012)

McGuire Graphics Data. http://graphics.cs.williams.edu/data/images.xml

Mullen, R.J., Monekosso, D.N., Remagnino, P.: Ant algorithms for image feature extraction. Expert Syst. Appl. **40**(11), 4315–4332 (2013)

Olson, C., Huttenlocher, D.: Automatic targetrecognition by matching oriented edge pixels. IEEE Trans. Image Process. **6**(1), 103–113 (1997)

Otsu, N.: A threshold selection method for gray level histograms. IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1976)

Pratt, W.K.: Digital Image Processing, 2nd edn. Wiley, New York (1991)

Segmentation Evaluation Database. http://www.wisdom.weizmann.ac.il/~vision/Seg_Evaluation_DB/index.html

Sullivan, J., Carlsson, S.: Recognizing and tracking human action. In: Proceedings of the 7th European Conference on Computer Vision-Part I, pp. 629–644 (2002)

The USC-SIPI Image Database. http://sipi.usc.edu/database/

Umbaugh, S.E.: Computer Imaging: Digital Image Analysis and Processing. CRC Press, Boca Raton (2005)

Vertex Mesh Simplification Algorithm Based on Curvature and Distance Metric

Yunliang Jiang[1,3(✉)], Wuyang Nie[3], Liang Tang[4], Yong Liu[2],
Ronghua Liang[5], and Xiulan Hao[1]

[1] School of Information Engineering, Huzhou University, Huzhou 313000, China
`jylsy@hutc.zj.cn`
[2] National Laboratory of Industrial Control Technology, Zhejiang University,
Hangzhou 310027, China
[3] School of Computer Science, Hangzhou Dianzi University, Hangzhou 310018, China
[4] China Ship Development and Design Center, Wuhan 430064, China
[5] College of Computer Science, Zhejiang University of Technology, Hangzhou 310023, China

Abstract. In traditional edge collapse simplification algorithms are mostly using the Quadric Error Metrics (QEM) as the cost to carry out edge collapse simplification, these methods will lose the original model shape feature; the simplified algorithm based on Vertex Curvature can preserve model shape feature, but the error between its folded model and original model can't be controlled well. This article plans to combine the advantages of these two methods, propose a novel vertex mesh simplification algorithm based on curvature and distance metric, using vertex curvature on the boundary type classification, get the alternative vertex of the folded edge through the LOOP subsidiary patterns, According to folding cost which is the distance metric from alternative vertex to adjacent triangle to simplify the complex mesh. In our method, those edges whose two vertex are all in the curvature threshold are marked foldable, and the less the cost, the first to fold. Compared with the simplification algorithm based on butterfly pattern, the experimental results show that this algorithm can preserve the model shape feature better.

Keywords: Curvature · Edge folding · Mesh simplification · Cost function · Distance metric

1 Introduction

The processing of large complex mesh model makes heavy demands on the processing and storage capabilities, large number of vertex and patch data calculation requires a lot of memory storage, and the time of processing and display of such grid model is geometric growth. At this time, vertex and patch simplification of large complex mesh model is very important, especially in some case the detail requirements of model is not high, do not need too many vertices and patch. The premise to simplify the mesh model is to maintain the details of original shape, the difference between simplified model and original model is not too big. The existing mesh model simplification algorithms can be divided into merging and deleting the geometric elements and resample vertex according

to the different approximate error of fitting surface, the direct method is to delete the geometric elements (point, edge, patch), including the deletion of vertices and triangle, edge collapse, triangle collapse and so on [1–4], in which the deletion of vertices and triangles will produce hole which need triangulation again, the two methods are seldom used because they need a large amount of calculation and time-consuming. Edge collapse is a mesh optimization algorithm offered by Hoppe in 1993, which fold a edge into a new point which connect the adjacent vertices of the edge, this method can avoid hole. The similarity between simplified model and original model in topological structure, boundary and shape feature of edge is determined by the way to produce new vertex. New vertex resample methods mainly have the following several kinds: Using Quadratic error matrix to calculate the sum of square of distance between the new point and the adjacent surface, which offered by Garland [5], and the principle of minimum volume change offered by Guezier [6], the two methods can ensure the spatial error minimum after the mesh be simplified, but it's difficult to ensure shape detail of the model when there is a large curvature variation. Sun-jeong put forward the mesh curvature which take the minimum change of mesh curvature as simplification criteria, the disadvantage of this method is that curvature have related to the length of edge, so the edge will be simplified when the edge is large and the change of vertex curvature is little. To solve the problems existing in these methods, this paper presents a method which take the vertex discrete curvature as the simplified criteria, calculate the curvature of each vertex of edge, the edge will be simplified only if each vertex curvature meet the simplified condition, at the same time, whether an edge can be simplified also needs to consider the boundary streamline costs. Using the similar Guezier distance metric to calculate the distance between the new vertex (we can get it by LOOP sub-model) and all adjacent surface, and we can simplify the edge whose distance metric is least priority.

2 Vertex Mesh Simplification Algorithm Based on Curvature and Distance Metric

2.1 Algorithm Thought and Process

This paper simplify mesh model through two simplified guideline: the vertex distance metric and vertex curvature, under conditions that the greatest degree of retaining original mesh boundary and shape features to simplify mesh vertex and triangle number. Through the calculation of vertex discrete curvature to screen prominent or obvious area and simplify the flat and smooth area, only the edge whose two vertices both meet the simplified condition will be simplified. We sort each edge by simplified cost, then simplify the edge that its cost is smallest, which can ensure the error between original mesh and simplified mesh is smallest.

Algorithm steps are as follows:

- Read the original model, extract the vertex information and triangular mesh topology information, establish data structure of each vertex adjacency relationship and adjacent triangles, and get all edge information.
- Set the simplified classification value of vertex discrete curvature.

- Calculate each vertex discrete curvature to determine whether each edge can be folded to simplify.
- Produce new vertex from the edge that can be folded, calculate edge cost and sort the edge order by the cost.
- Fold the edge that its cost is smallest and can be folded, if not, then algorithm end.
- If there is a edge that meet the folding requirement, update the adjacent vertex information and adjacent triangle information, to step 3 (in the step 3, only calculate the updated vertex curvature and edge cost, the unchanged vertex and edge do not need to update.)

Algorithm flowchart as shown in Fig. 1.
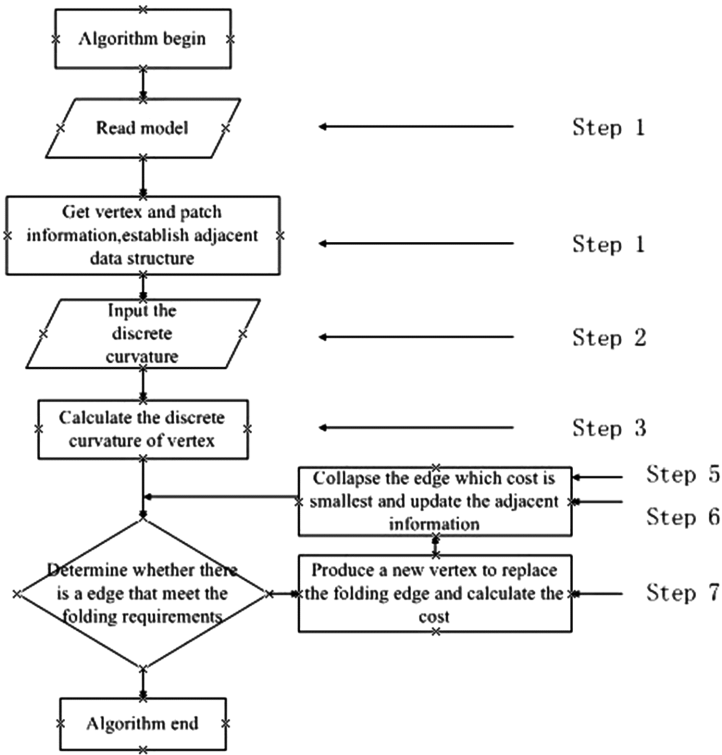
Fig. 1. Algorithm flowchart.

2.2 Edge Collapse

The algorithm in this paper is similar to the edge collapse algorithms of Hoppe [7]. First, fold the edge P_1P_2 to a new vertex P_{new}, delete the edges and triangles which are adjacent with the new vertex and add new edges and triangles, edge collapse method can avoid the hole after deleting the vertex and triangle and do not need triangulation. In specific process can directly use P_{new} replace the vertex P_1 and vertex P_2 (Fig. 2).
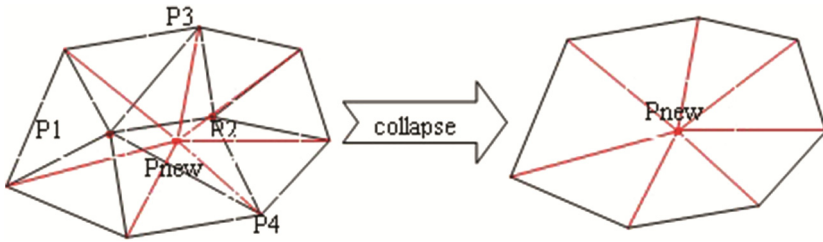
Fig. 2. Edge collapse.

2.3 Calculate Discrete Curvature

In those places where the mesh shape feature is particularly prominent, for example, boundary, shape edge and so on, the vertex curvature is larger than those smooth and flat places. So we must preserve the vertex which curvature is large and delete the vertex that curvature is small. This method not only preserve the shape feature of original mesh but also simplify the mesh.

This paper use discrete approximation algorithms based on mesh vertex mean curvature offered by Desbrun [8, 9]. The vertex curvature at P is:

$$H_n = \frac{1}{4A} \sum_{i \subseteq N(P)}^{n} \left(\cos \alpha_i + sos\beta_i \right) \left(Q_i - P \right) \tag{1}$$

In which, A is the sum of the area of triangular patch which is adjacent with point P, N(P) is the set of vertices that adjacent with P, α β is the angle corresponding to the edge QP in the triangle (Fig. 3).
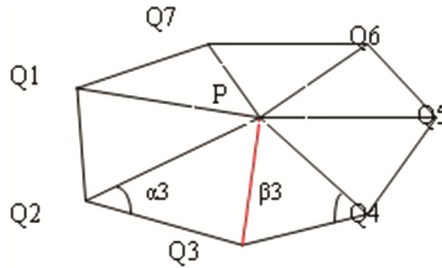
Fig. 3. Vertex curvature calculation schematic diagram.

A_3, β_3 are the angles corresponding to the edge PQ_3 in the triangle Q_2PQ_3 and Q_3Q_4P, With the coordinates of point P, $Q_i - P$ can be expressed as a vector, $H_n(P)$ is also a vector, in this paper, the vertex discrete curvature is expressed with mold of vector.

2.4   Distance Cost Function

Edge collapse, according to the fold cost of each edge, priority to collapse the edge which cost is smallest. The cost of edge, which is similar to the cost function offered by Garland, is a method based on distance metric. The cost of the collapsed edge is expressed with the sum of distance between the new point and the adjacent triangular patch. For example, the new point $P_{new}(X_{new}, Y_{new}, Z_{new})$, the adjacent patch can be expressed as $A_i*x + B_i*y + C_i*z + D_i = 0$, the normal vector is $N_i(A_i, B_i, C_i, D_i)$, the distance between P_{new} and flat is:

$$\text{Distance}\left(P_{new}, N_i\right) = \frac{P_{new} * N_i}{\|N_i\|} \tag{2}$$

If triangle V have three points, they are $P_1 (x_1, y_1, z_1)$, $P_2 (x_2, y_2, z_2)$, $P_3 (x_3, y_3, z_3)$, then the normal vector N is:

$$N = \begin{bmatrix} i & j & k \\ x_2 - x_1 & y_2 - y & z_2 - z_1 \\ x_3 - x_1 & y_2 - y_1 & z_2 - z_1 \end{bmatrix} \tag{3}$$

The cost of collapsed edge is:

$$\text{Distance}\left(P_{new}, N_i\right) = \sum_{i=1}^{n} \frac{P_{new} * N_i}{\|N_i\|} \tag{4}$$

$\|*\|$ is the mold of $*$, Distance(P_{new}, e) is the cost of collapsed edge e, calculate the cost of all edges according to the new point, sort the cost and determine whether this edge meet the folding criteria.

2.5   LOOP Sub-patterns

We get a new mesh topological structure after folding the edge into a new point, the way producing the new point directly determine the primitive geometric feature of the simplified mesh. The most commonly used method is that the two endpoints and the midpoint of the collapsed edge are considered as the candidate point, we calculate their folding cost, and then the point which the cost is smallest is used as the new point. The defect of this method is imprecise and will lose the geometric feature of original mesh. The second method is to produce a quadric surface using triangular mesh approximation algorithm, then get a new point on the surface, the defect of this method is numerical instability and complex in calculation. The method of this paper is to subdivide the surface into sub-surface step by step using subdivision method, then get a new point on the sub-surface. Subdivision method commonly have Catmull-Clark pattern [10, 11], Loop pattern [12], butterfly pattern and so on. This paper use Loop pattern, approximation static subdivision pattern of split surface, this pattern can get a more smooth surface than butterfly pattern. Loop pattern is offered by Charles Loop [13], the main idea is that

insert a new point into each edge of each triangle and connect with each other, so get four triangles from a triangle; then calculate the position of all vertex in the next subdivision. After several subdivision, the original surface divide into a limit surface and the vertex are all on the limit surface.

Calculation of new point as follow:

New edge vertex (E-vertex): the two endpoint V_0 and V_1 of internal edge E, the two adjacent triangles $T_1(V_0, V_1, V_2)$ and $T_2(V_0, V_1, V_3)$, then E-vertex as follow:

$$V_E = \frac{3}{8}\left(V_0 + V_1\right) + \frac{1}{8}\left(V_2 + V_3\right) \tag{5}$$

New vertex (V-vertex): the adjacent vertex of internal vertex V is V_0, V_1, V_2, V_3,…, V_{n-1}, the n is the number of the adjacent edge of vertex V, then V-vertex as follow:

$$V_n = \left(1 - n\alpha_n\right)V + \alpha_n \sum_{i=0}^{n-1} V_2 \tag{6}$$

$$\alpha_n = \frac{1}{n}\left(\frac{5}{8} - \left(\frac{3}{8} + \frac{1}{4}\cos\frac{2\pi}{n}\right)^2\right) \tag{7}$$

a_n is weights of the adjacent vertex, $1 - a_n$ is the weights of vertex V, when n is equal to 3, a_n is equal to 3/16, when n is greater than 3, a_n is equal to 3/8n.

New vertex of boundary edge: the midpoint of two endpoints; new vertex of boundary vertex: the sum of three-eighths of adjacent vertex and three-fourths of itself.

In Fig. 4, (a) E-vertex, (b) V-vertex, (c) vertex of boundary edge, (d) boundary vertex.
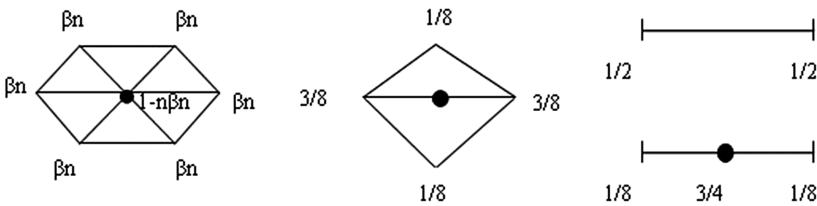
Fig. 4. The algorithm of vertex calculation in Loop patterns.

The edge E will produce a E-vertex after one subdividing, and then subdivide the E-vertex for three times will get the new vertex, we consider the sum of the distance between new vertex and adjacent triangle as the cost of edge E.
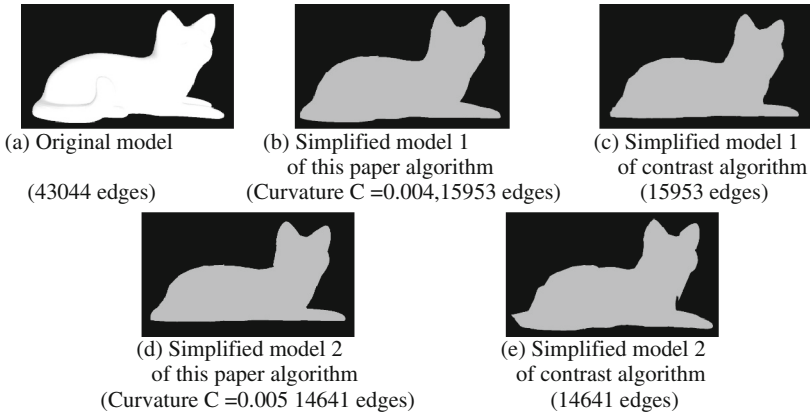
3 The Result of Experiment

Figure 5(a) is original mesh model, 12004 edges, (b) and (c) are the simplified model when the curvature value are 0.2 and 0.3 respectively. From the Fig. 5, we can see that the model is not changed too much, the number of edge are 11280 and 8970. (d) and (e) are the simplified model when curvature value is 0.04 and 0.05, in which the number of edge is 5881 and 3409 respectively. We can see the shape feature of the ear, head, and

(a) Original model
(12004 edges)

(b) Curvature C=0.02
(11280 edges)

(c) Curvature C=0.03
(8970 vertex)

(d) Curvature C=0.04
(5881 edges)

(e) Curvature C=0.05
(3409 edges)
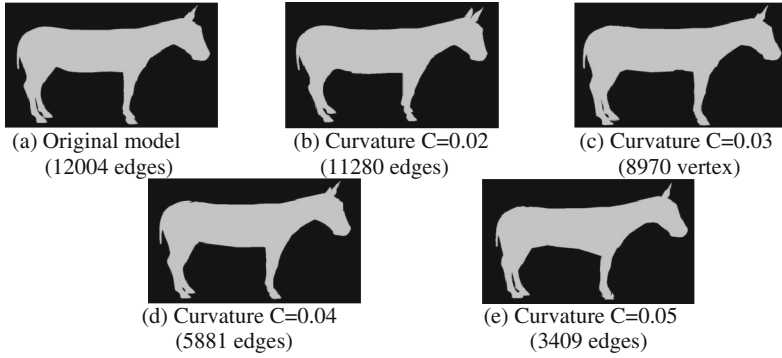
Fig. 5. The result of experiment.

feet of model is obvious, the ear and tail is very sharp, the contour of feet is prominent. In the process of simplification, the place in which the shape feature is obvious not be simplified basically. While the abdomen is simplified obviously because it is flat and smooth, in the flat and smooth place, the curvature is small, so can be simplified, at the same time, the algorithm can preserve the shape feature.

(a) Original model

(43044 edges)

(b) Simplified model 1
of this paper algorithm
(Curvature C =0.004,15953 edges)

(c) Simplified model 1
of contrast algorithm
(15953 edges)

(d) Simplified model 2
of this paper algorithm
(Curvature C =0.005 14641 edges)

(e) Simplified model 2
of contrast algorithm
(14641 edges)

Fig. 6. Contrast experiment 2.

The following are the simplified example using this paper algorithm and butterfly patterns algorithm based on the quadratic error. In Fig. 6, (a) is a cat model which have 43044 edges and 7340 vertex, (b) and (d) are simplified model of this paper algorithm, (c) and (e) are simplified model of contrast algorithm, (b) and (c) have 15953 edges, (d) and (e) have 14641 edges. From the result of experiment, we can see that this paper algorithm can preserve the shape feature of the tail and back, while the contrast algorithm do not preserve it, the tail grow sharp and the back grow coarse; the simplified model of this paper algorithm have a little deformation on the back, but it preserve the shape feature of original model overall, the contrast algorithm have a obvious deformation

when the edge reach a certain level. This is due to simplify the place where the curvature is big and the shape feature is obvious.

4   Conclusions

This paper algorithm consider the vertex curvature as the simplified restrictions, only if the two endpoints of the edge meet the restrictions, the edge can be simplified, so it can preserve the original shape feature, and the error is smallest when we give priority in collapsing the edge which the cost is smallest, what is more, the Loop sub-patterns can ensure that the error between the cost of collapsed edge and original model is smallest and get a best result.

References

1. Zhou, K., Pan, Z.G., Shi, J.Y.: Mesh simplification algorithm based on triangle collapse. Chin. J. Comput. **21**(6), 506–513 (1998)
2. Zhu, R., Shen, W.: Edge collapse considering triangular mesh for model simplification. In: Yin, Z., Pan, L., Fang, X. (eds.) Proceedings of the Eighth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA 2013), vol. 212, pp. 1175–1181. Springer, Heidelberg (2013)
3. Li, W., Chen, Y., Wang, Z., Zhao, W., Chen, L.: An improved decimation of triangle meshes based on curvature. In: Miao, D., Pedrycz, W., Slezak, D., Peters, G., Hu, Q., Wang, R. (eds.) RSKT 2014. LNCS, vol. 8818, pp. 260–271. Springer, Heidelberg (2014)
4. Vidal, V., Wolf, C., Dupont, F.: Combinatorial mesh optimization. Vis. Comput. **28**(5), 511–525 (2012)
5. Garland, M., Heckbert, P.S.: Surface simplification using quadric error metrics. Comput. Graph. **31**(3), 209–216 (1997)
6. Guezier: Surface Simplification Inside a Tolerance Volume. IBM Research Report Computer Science/Mathematics, 1-56, RC 20440 (1996)
7. Hoppe, H.: Mesh optimization. Comput. Graph. **27**(1), 19–26 (1993)
8. IsLer, V., Lau, R.W.H., Mark, G.: Real-time multi-resolution modeling for complex virtual environments. In: Symposium on Virtual Reality Software and Technology, Hong Kong, pp. 11–19. ACM Press (1996)
9. Lu, S.M., Yang, X.N., Wang, G.Z.: Edge collapse mesh simplification algorithm based on discrete curvature. J. ZheJiang Univ. **31**(3), 297–299 (2004)
10. Schroeder, W.J., Zarge, J.A., Lorensen, W.E.: Decimation of triangle mesh. ACM Comput. Graph. **26**(2), 65–70 (1992)
11. Hamann, B.: A data reduction scheme for triangulated surfaces. Comput. Aided Geometric Des. **11**(2), 197–214 (1994)

12. Kim, S.J., Kim, C.H., Levin, D.: Surface simplification using a discrete curvature norm. Comput. Graph. **26**, 657–663 (2002)
13. Loop, C.T.: Smooth Subdivision Surfaces Based on Triangles. Department of Mathematics, The University of Utah (1987)

Visualization of the Color Family Order System

Jianwen Song[1(✉)], Jianming Song[1], Guosheng Hu[1],
and Zhigeng Pan[2]

[1] China Academy of Art, Hangzhou, China
`songjw888@163.com`
[2] DMI Research Center, Hangzhou Normal University, Hangzhou, China

Abstract. Color perception always depends on subjective factors. When it comes to a color design project, different designers will persist in different opinions. Although, color designing tools are handy to access nowadays, it takes long-term accumulation of color knowledge and abundant experiences of color designing for designers to master color principles and strengthen color perception. Based on objective color analyzing, this study established a visualization system of color harmony theory of family order and its digital model, on which a method of color designing was also developed. This method bridged the gap between objective colors and subjective perception in virtue of three-dimensional visualizing technology. This paper presented the color analyzing and designing toolkits of the project product. The color visualizing and designing system not only accords with designer's intelligence and experience, but also proposes a new method for color analyzing and teaching.

Keywords: Objective color · Subjective color · Color space · Colors family order · Visualization

1 Introduction

1.1 Color Space

People have experienced a long history of curiosity and researching for color vision. Since Issac Newton refracted white light into a spectrum of colors as red, orange, yellow, green, cyan, blue, purple with a prism and discovered the attributes of visible lights in 1660s (Fig. 1), color exploring involved in science domain. From then on, J.H. Lambert proposed the pyramid color map concept, P.O. Runge drafted the color sphere then, a hemisphere by M.E. Chevreul, a cone by H. von Helmholtz and a slanted double cone by A. Kirschmann as well as a symmetric color wheel by J.W. von Goethe, colorists represented and analyzed color from linear wheel of hue to different color space in three-dimension. Color analyzing theory was profoundly accomplished via over 300 years' research and accumulation.

Color space with three dimensions of attributes is commonly used for color representation and analyzing nowadays. Usually, the attributes of dimensions follows in hue, saturation, and value. There are several color spaces in common use:

Fig. 1. Newton visible light spectrum (Color figure online)

(1) Munsell color system is the first color space that specifies colors based on three color dimensions: hue, value (lightness), and chroma (color purity). It is the only well-rounded color space that is based on subtractive color mixing theory contrary to the later color spaces based on additive lights color. It is still widely used in color design and teaching today.

(2) CIE XYZ color space was created by the International Commission on Illumination (CIE) in 1931 (Fig. 2) [1, 3]. In this model, X, Y, and Z roughly mean hue, luminance, and saturation. The data embodied the experimental color measurement results of human vision between 1° to 4° standard observer. Later, CIE released the CIE 1964, which is the 10° CIE standard observer.

(3) RGB color space, is a converted color model with three additive primary colors (R for red, G for green, and B for blue). It is widely used in computer graphics (CG). CIE RGB and CIE XYZ were both defined by CIE 1931 color space standard.

(4) Other color spaces like HSI color cylinder, HSV/HSL color space, L*a*b* color space, etc., are mainly used in color management and design.
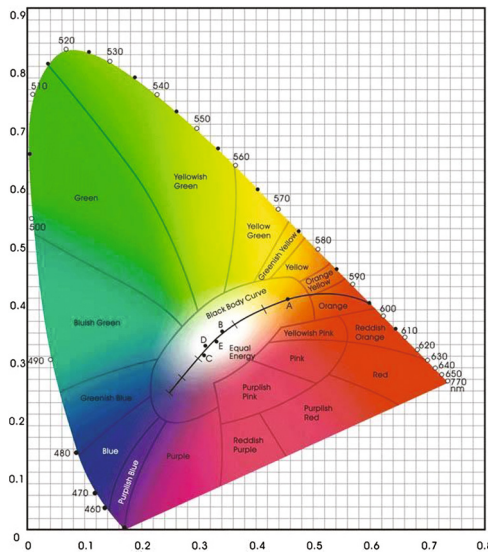
Fig. 2. CIE 1931color space (Color figure online)

1.2   Color Standard and Its Digitized Research

The famous enterprises in research and application of color are committed to color digitized, such as American's Munsell and Pantone, Germany's RAL, Sweden's NCS, and Japan's DIC, etc. They occupy the international market of color digital tools in art design. The international color digital research mostly concentrated in the color equipment manufacturing, especially in the fashion industry. Establishing color standards are viewed as great important, and seen it as an far-reaching information resource, that will pay a remarkably strategic role in business competition.

Currently the most widely used color standards consists of: (1) Munsell color system developed by the U.S. National Bureau of Standards, (2) The Natural Colour System®© was published by Scandinavian Colour Institute AB, Sweden, and (3) RAL (color space system) was published Imperial Commission for Delivery Terms and Quality Assurance, German.

In early 1990s, a project on The Chinese Color System [4] conducted by Daheng Wang et al. from Institute of psychology, China Academy of Sciences. The project focusing on color visual experience of Chinese people established The theoretical models of Chinese Color System and the Chinese Color System Sample Volumes.

Based on these studies, "CNCSCOLOR (China National Color System Color) System for FASHION" [5, 6] was established by China Textile Information Center and China Fashion Color Association in 2008. The system is designed as standard platform and system tools in fashion color application, so that it helps enterprises to improve service and management.

Meanwhile, progresses have already achieved in the research field of color digitalizing and visualization. Shigenobu's color image scale was devised by the use of an original color-projection technique, analysis of variance, cluster analysis, factor analysis, and the semantic differential method. It can be used for representing and evaluating color images considering perceptual factors [7]. Yamazaki and Kondo introduced a method of editing color schemes in Kansei scales, which adopted human perception, e.g. "warm to cool," "soft to hard," and "natural to artificial", to represent emotional feeling [8]. Meier et al. developed a set of interactive color palette tools that provide eight visualized methods for color mixing: palette browser, image and composition tools, gradient mixer, dial-a-color tool, frequency visualizer, palette breeder, name IPT, and grouper [9]. Hu et al. presented an interactive method for generating harmonious color schemes in the visualized HSV/HSL color space, in which user can obtain harmonious color schemes by simply sitting a few personal parameters [10].

2   Objective Color Analyses and the Presentation of Color Order Relation

2.1   Image Analysis of Objective Color

In the process of product design, it is an important step for designers to determine the color scheme, on which the core values is to establish a proper color system and harmonious relationship. These skills dominated by subjective impressions are accumulated and well-shaped in the long-term design experience for designers.

In the example shown in Fig. 3, the system analyses a photo with a few colors in order to access the basic structural relation of colors. By analyzing intuitively the colors of the tulips in the photo, we found that the image is mainly composed of: light blue, gold yellow, red and dark green ▨▨▨▨.

Fig. 3. Photo example for color analyses (Color figure online)

In practice, however, if designers want to determine precisely the four colors and use them in design, their works may show a more inflexible, because our vision is vague and varies with the environmental impact of uncertainty distinguish colors.

By image analyzing, we found that the four colors presented in human vision are composed of four similar color systems:

Light blue series ▨▨▨▨▨
Gold yellow series ▨▨▨▨▨
Red color series ▨▨▨▨▨
Green color series ▨▨▨▨▨

2.2 Ordering of Color

The sense of order about color is from nature. The perception of human vision to color varies according to the distance between the color and the eye location. Shown in Fig. 4, as the color block goes away from the eye, the color chroma becomes lower and grayer until it vanishes. Therefore, a specific ordering in color space is created.

The ordering of color is to classify the color analyzed by image processing method according to the color attributes of lightness, hue or chroma, to arrange them in some ordered style, and to visualize the location of the colors in the cubic color space.
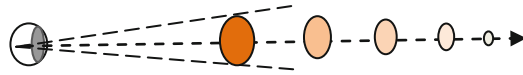
Fig. 4. Color varies as distance (Color figure online)

3   Family of Color and Color Family Order

3.1   Family of Color

The concept of Family of Color was proposed by Professor J.P. Lenclos, a France color scholars, to describe the characteristics of color harmony. Lenclos claimed that harmony of color likes kinship, having some kind of same identity of genes [11]. Harmony gene among colors comes from the similar identity of colors (i.e. similar parameters of lightness, hue or saturation), where color components having such characteristics can be viewed as a color family. As shown in Fig. 5, the choice of color family members is a sequence of color samples left by filtering in the same or very similar color photo screening in accordance with certain rules.
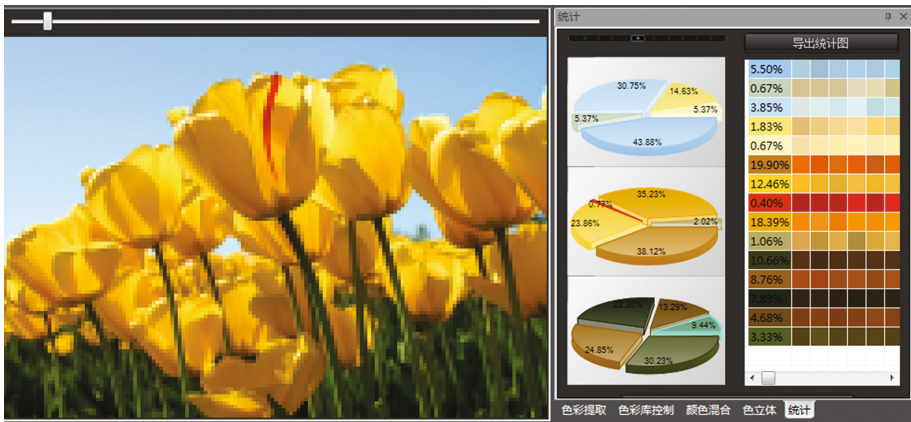
Fig. 5. 90 % proximity with the source image (Color figure online)

For the same photo, the color family members also changes, accompanied by changing the proximity in the image processing. With pie data analyzing, shown in Fig. 5, it divides the colors, which share the same saturation, in five-element classification-identity measurement, into categories of High Lightness, Mid Lightness and Low Lightness respectively.

The data revealed in the pie chart represents the scaling relations that the proportion of the pie for various elements in be determined by the same proximity and lightness.

The *Exported Statistical Chart* shows the color samples that share similar color tendency and distributed in uniform steps.

From Table 1, the image proximity choice depends on designer's sitting. By reducing the image proximity, the objective colors from the source image are gradually

Fig. 6. 70 % proximity with the source image (Color figure online)

Table 1. Families of color analysis table

Fig	proximity	The color families element sequence and the proportion of the pie chart						Conclusion
6	90%	High Lightness	43.88%	30.75%	5.37%	14.63%	5.37%	Closer to the source image color
		Mid Lightness	35.23%	35.23%	38.12%	2.02%	0.77%	
		Low Lightness	9.44%	30.23%	22.20%	24.85%	13.29%	
7	70%	High Lightness	29.10%	50.94%	7.35%	5.88%	7.35%	Deviating Image color from the source
		Mid Lightness	14.31%	22.0%	29.69%	29.54%	4.46%	
		Low Lightness	22.69%	19.33%	20.8%	11.53%	25.63%	
8	50%	High Lightness	19.72%	12.68%	30.99%	28.17%	8.45%	Subjective color
		Mid Lightness	9.15%	19.01%	40.14%	30.28%	7.41%	
		Low Lightness	17.65%	15.97%	10.92%	34.45%	7.53%	

changed into the subjective color families sequence. The objective color shown in Fig. 4 gradually changes to the family color order system that match designers thinking habits.

3.2   The Order of the Color Family

A coloring domain order exists in families of color. It makes the three-dimensional distribution legible in the color solid space.

Figure 7 indicates that the source image with 90 % proximity in HSL color of three-dimension. Figures 8 and 9 represent the different color distributions in RGB and HSL color space in the same proximity of the source image. It is important to reveal the inherent law of color system by studying the color family color order and analyzing the
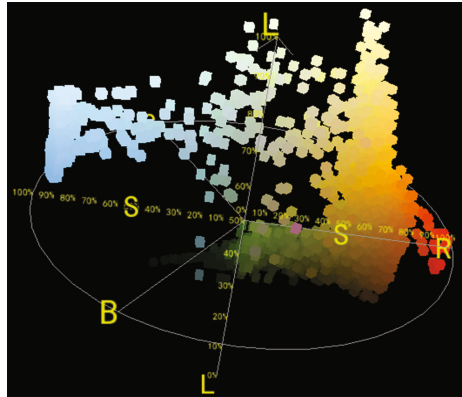
Fig. 7. 90 % degree of proximity with the source image color family order (Color figure online)
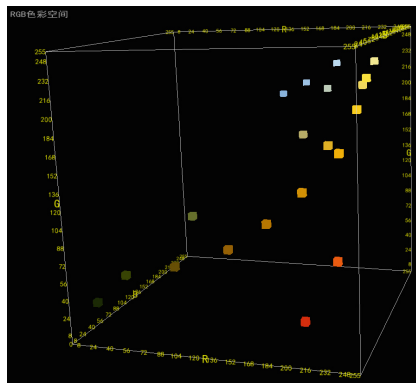
Fig. 8. Order chain in RGB color space about Fig. 6 (Color figure online)

color using habits of designers or painters. The theory of color family order provides visualized ideas and methods for studying the color using experiences of artists.

The harmonious key colors can be accessed in the color space area along the relation track. This track can be represented as the Harmonious Color Order Chain. Essentially, the color order chain is the enveloped boundary line of the color distribution area.

4 Design of the Color Family Order System

The color family order system (CFOS) consists of two sub-systems; (1) Color Family Digitalization Basic System (CFDBS), (2) Color Family Order Analysis System (CFOAS). Figure 10 describes the two connected sub-systems sharing the data base.
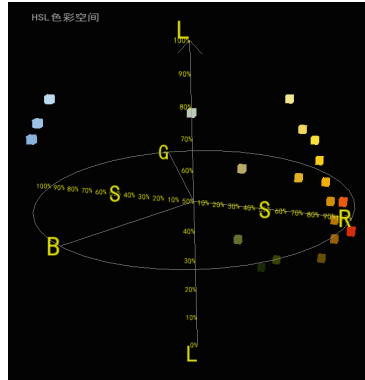
Fig. 9. Order chain in HSL color space about Fig. 6 (Color figure online)
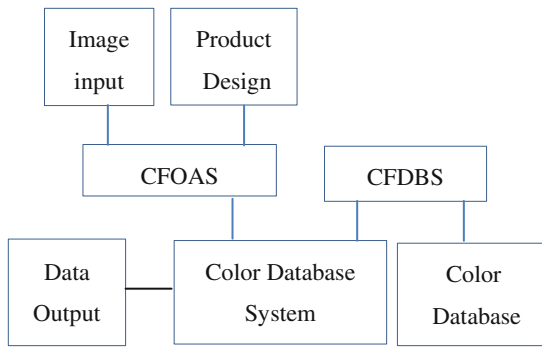
Fig. 10. The design diagram of CFOS

4.1 Design of Color Family Order Analysis System (CFOAS)

Figure 11 is the module diagram of CFOAS, which is constructed by adopting the thinking habits of designers. The design is based on the color family order theory introduced in Sects. 3.1 and 3.2, and the system is implemented with the support of image processing, color analyses, color order, and color order visualization. The system can be thought as a practical and efficient tool for color digital model.

4.2 Design of the CFDBS

In color management domain, there are widely applied color spaces like HSI color cylinder, Lab color space, Munsell color system, HSV color cylinder and RGB color cube space. The transforming algorithms among these color spaces consummated the color family order theory. In the CFDBS, a standard color space is established with OpenGL, performing quantizing control to chrom values. In the following we illustrate the digitalized transforming model of the main color standards;
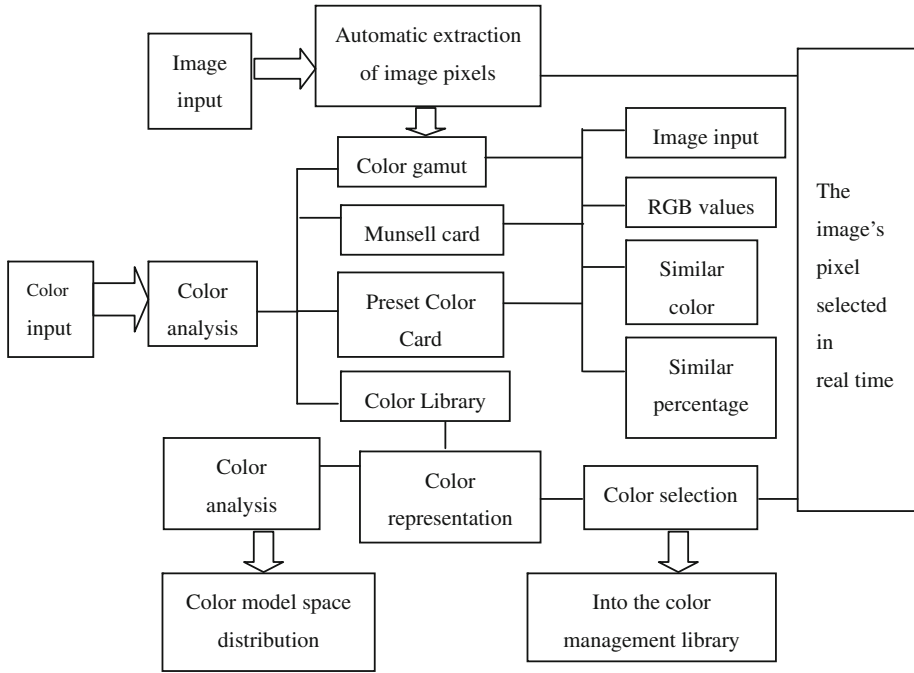
Fig. 11. The process of the color design about CFOAS

(1) *RGB color cube space construction*

In RGB (r,g,b) color cube model (Fig. 12), a color can be represented as the additivity ratio of red, green and blue. The common formats are: (1.0, 0.0, 0.0) digital algorithm, percentage algorithm (100 %, 0 %, 0 %), RGB 8 bit component (255, 0, 0) algorithm and 16 bit component algorithm.

(2) *HSL color cube space construction*

In our system, the HSL color space model is represented by h, s, and l, value to define colors, in which h (hue) value is between 0 and 360, the value of s and l is between 0 and 1, while in RGB space, (r, g, b), the value of r, g and b is between 0 and 1 (Fig. 13).

(1) *Lab color cube space construction*

Lab color space is designed to approach the human vision, which is described in L, a, and b values. L component can be used to adjust brightness compare (Fig. 14). There are two common Lab color cube spaces: L,a,b and L*a*b:

(a) L*a*b color space (CIELAB): L represents lightness, a and b represent opposite dimension; −a is green value, +a is red value, −b is blue value, +b is yellow value.
(b) Hunter 1948 L, a, b color space [12, 13]: it was created by R. S. Hunter, the components are L, a and b. It should be noted that in practice Lab is usually be
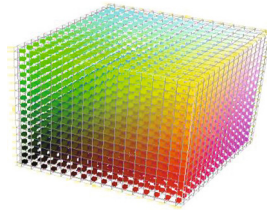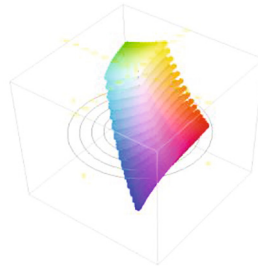
Fig. 12. RGB color cube space (Color figure online)

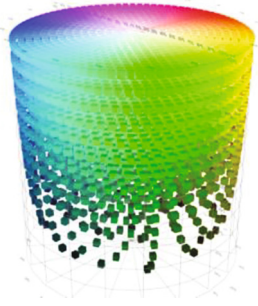**Fig. 13.** HSL color cube space (Color figure online)

Fig. 14. Lab color cube space (Color figure online)

used to represent CIE 1976 (L*, a*, b*) color space with components in L*, a* and b*.

The two schemes are mutually connected in color space, but with different methods; CIELAB employs cubic root calculation, while Hunter Lab employs square root calculation [14].

(4) *HSI color space construction*

In HSI color space, color is described with Hue, Saturation, Chroma and Intensity or Brightness (See Fig. 15). Human vision system usually adopts HSI color space [15], which is more suitable than RGB color space for human's vision attribute. In image processing and visualization, many algorithms can be used in HSI color space.

Fig. 15. HSI color cube space (Color figure online)

(5) *Munsell color cube space construction*

Munsell color system, created by A. H. Munsell, is also one of the widely used color standards (Fig. 16). In the digitalizing construction, since Munsell color system and RGB color cube represents two separate systems for describing colors (material color, light simulating color), transforming between the two color spaces remains difficult.

In the CFDBS system, the modules can be used individually for color teaching or color design (Fig. 17), or can be used jointly with the CFOAS (Fig. 18) by sharing color data information.

As shown in Fig. 18, after the photo is loaded into the system, we can get the masked picture which is in 80 % of proximity to the original photo. In the Color Management toolbar, we can get color order link and color order distribution by entering the toolbar Color Cube.
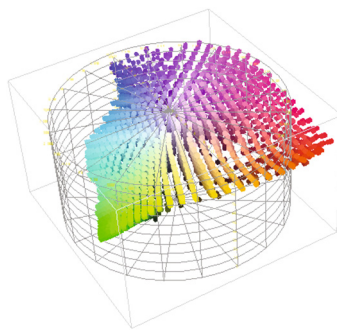
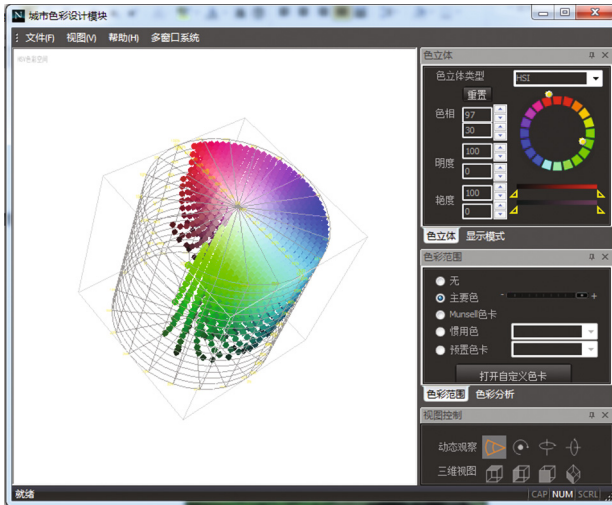**Fig. 16.** Munsell color cube space (Color figure online)

Fig. 17. The process of the city color design module of the CFDBS (Color figure online)
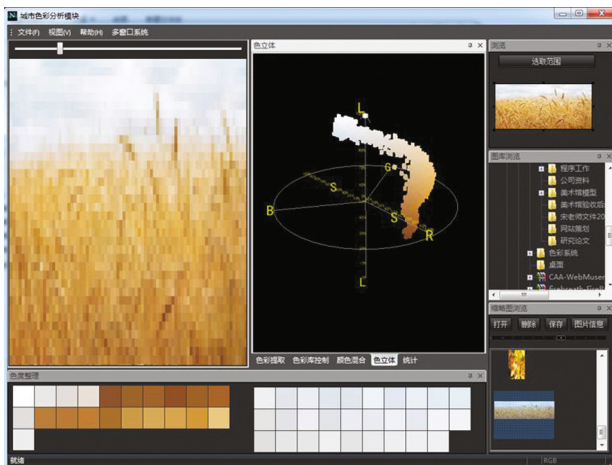

Fig. 18. Implementation results of the CFOAS (Color figure online)

5   Implementation and Conclusion

The implemented system has been employed in teaching of color design at China Academy of Art. The Color Design Professional Teaching Software implemented based on our theory has won the China Color S&T Award in 2011/2012 from China Fashion & Color Association. The Institute of Color in China Academy of Art has applied the CFOS system in multiple projects for city color design in China [17, 18],

and excellent feedback has been obtained. The projects will be introduced in another paper.

In summary, based on objective color analyzing, we present a visualization system of the color harmony theory of family order and its digital model, on which a method of color designing [18] is also developed. In addition, this paper presented a color analyzing and designing toolkits which can be used for city color design project and color teaching. Experimental results illustrate the efficiency of our method.

References

1. http://www.cie.co.at/main/freepubs.html
2. http://www.iso.org/iso/home.html
3. http://en.wikipedia.org/wiki/Category:Color_space
4. People's Republic of China: The Chinese Color System (GB\T15608-2006). China Standard Press (2006). ISBN 066128554
5. Shenzhen HaiChuan Color Technology, Ltd.: Chinese color system standard sample volumes, GSB16-2062-2007
6. http://www.cncscolor.com/en/Products.asp
7. Shigenobu, K.: Color Image Scale. Kodansha Ltd., Tokyo (2001)
8. Yamazaki, H., Kondo, K.: A method of changing a color scheme with Kansei scales. J Geom. Graph **3**, 77–84 (1999)
9. Meier, B.J., Spalter, A.M., Karelitz, D.B.: Interactive color palette tools. IEEE Comput. Graph Appl. **24**, 64–72 (2004)
10. Hu, G.S., Pan, Z.G., Zhang, M.M., Chen, D., Yang, W.Z., Chen, J.: An Interactive method for generating harmonious color schemes. COLOR Res. Appl. **39**, 70–78 (2012)
11. Lenclos, D., Lenclos, J.-P.: Colors of the World: The Geography of Color. W. W. Norton & Co., New York (2004)
12. Hunter L,a,b Versus CIE 1976 L*a*b*. HunterLab Applications Note **13**(2)
13. http://www.hunterlab.com/
14. Hunter, R.S.: Photoelectric color-difference meter. J. Optic. Soc. Am. **38**, 661 (1948)
15. Pascale, D.: A Review of RGB Color Spaces: From xyY to R'G'B'. The BabelColor Company, Montreal (2003)
16. Song, J.M.: Graceful Showing of the Giant Golden Building: Architecture Color Design Notes of Nanjing South Station. Architecture Design (Chinese Edition), vol. 154, pp. 72–76
17. Song, J.M.: The Color Expression of Hangzhou: Experience of Color Planning for Hangzhou City. Public Art (Chinese edn.), vol. 5, pp. 28–34
18. Song, J.M.: Reading the Colors of Macao. New Geographies: Urbanisms of Color v.3, Harvard Graduate School of Design, pp. 125–141

An Affine Transformation Based Image Shuffling and Watermarking Scheme

Ruisong Ye[1], Haiying Zhao[2(✉)], and Sun Chang Ping[3]

[1] Department of Mathematics, Shantou University,
Shantou 515063, Guangdong, China
rsye@stu.edu.cn
[2] Mobile Media and Cultural Calculation Key Laboratory of Beijing,
Century College, BUPT, Beijing 102613, China
zhy.yn@163.com
[3] School of Electrical Engineering, Tianjin University of Technology,
Tianjin, China

Abstract. Thanks to the exceptionally good properties, such as ergodicity, pseudo-randomness, sensitivity to initial conditions and control parameters, chaotic maps have been widely studied and applied to devise image shuffling, encryption and watermarking schemes. In this paper, the chaotic nature of affine transformation is studied. The affine transformation is then utilized to generate pseudo-random number sequence and applied to shuffle the host image during the preprocessing of watermarking. The watermarking is performed in the spatial domain, where the watermark bits are also encrypted by the skew tent map and then embedded in the shuffled host image. The proposed shuffling scheme yields good scrambling effects compared with the conventional scheme. The watermarked images are robust against various attacks, such as cropping, JPEG compression, noise, etc.

Keywords: Affine transformation · Skew tent map · Chaotic map · Shuffling · Watermarking

1 Introduction

Spurred by the fast developments of multimedia and network technology, digital images, video, and audio have revolutionized in the way of largely stored, manipulated, and transmitted over the Internet and wireless networks in the past decades. Creative ways of storing, accessing and distributing data have generated lots of benefits into the digital multimedia field. However, these benefits also bring with a serious effect on copyright encroachment. Digital multimedia data can be replicated, edited easily and spread openly through the Internet, which causes a serious problem that it is difficult

even impossible to give proof of copyright once pirated. As a result, the copyright isn't protected legally, and the authors of digital multimedia may feel economically deprived of their reputations by fake copies. To provide copyright protection of digital data, two complementary techniques have been developed: encryption and watermarking [1–5]. Encryption can be used to protect digital data during the transmission process. However, after the receiver has received and decrypted the data, the original data is obtained but is no longer protected. Watermarking is a compliment encryption by embedding a secret imperceptible signal, a watermark, into the original data in such a way that it always remains present. Digital watermarking provides an effective approach for copyright protection and content authorization. The watermark is imperceptible usually, but it can be detected or extracted by specific algorithm, even after some manipulations to the watermarked data. In contrast to encrypted data, watermarked data can still be used while remaining protected. It does not necessarily prevent the copying of digital data, but should rather identify the original data source, so that copyright violations can at least be detected. The hidden signal travels with the data, which thus remains "marked" and protected, until the intended receiver removes the watermark.

Since a watermark is used to prove ownership, there are several important issues in the watermarking system including transparency, robustness, security and blindness. Transparency is required so that a watermark may remain invisible in order to maintain its secrecy. Transparency is also needed to prevent the visual distortion of an original image to maintain the commercial value of the image. A watermarking technique is referred to as blind if the original image and watermark are not needed during extraction. The blindness is necessary if it is difficult for us to obtain the host image and watermark. Indeed, both security and robustness are also important because the watermark bits are expected to be irremovable and unperceivable. The watermark must be able to effectively resist common image attacks, such as nosing, cropping, JPEG compression, etc.

There are two methods of performing watermarking, one in spatial domain, and the other in frequency domain. Each technique has its own advantage and disadvantage. Watermarking in the spatial domain is easy to implement by embedding a watermark in selected areas on the texture of the host image by changing the grey levels of some pixels [6, 7]. The disadvantage of this kind of watermarking is that the inserted information may be easily detected using computer analysis and the watermark cannot effectively resist image processing attacks such as compression, noise, and filtering. In order to enhance the robustness against malicious attacks, a variety of improved watermarking schemes have been proposed recently, see for example, [8–11]. Regarding watermarking in the spatial domain, the common way is to perform pre-processing to the host image and/or the watermark by some shuffling algorithms, which makes the watermark more transparent, secure and strongly robust [12–14]. As an approach of image encryption, digital image shuffling technique is a research emphasis all through [2, 15–17]. It can be used directly in digital encryption, and can also be used as the preprocessing and post processing to enhance the robustness of digital image hiding and digital image watermarking. The shuffling technique has two advantages. On one hand, it makes the watermark bits spread uniformly all over the host image. Therefore, it can resist the attacks such as cropping, noise, compression effectively. On the other hand, it can also enhance the security of watermarking schemes by setting the

encryption keys. Only the authorized consumers know the secret keys to restore the watermark and the original host image. Many researchers have been focusing on security recently. The common way is design some secure shuffling schemes to perform the preprocessing in watermarking.

In this paper, an image shuffling scheme based on the ergodicity of affine transformation is proposed and applied to shuffle the host image in the watermarking preprocessing to enhance the security, the transparency and the watermarking robustness against attacks. The watermarking is performed in the spatial domain, where the watermark bits are first encrypted by the chaotic skew tent map and then embedded in the shuffled host image. Such a performance spreads the watermark bits in the host image more uniformly, and therefore increases the security and transparency of the watermarking scheme. Experimental results show that the watermarked images are robust against various attacks, such as cropping, JPEG compression, noising, etc.

The rest of the paper is organized as follows. In Sect. 2, affine transformation is introduced and its chaotic nature is studied numerically. The pseudo-random numbers generated by the affine transformation are then used to shuffle the host image in spatial domain. The image shuffling scheme is presented in Sect. 3. The key sensitivity and shuffling degree are performed to show that the proposed image shuffling scheme is efficient. In Sect. 4, a watermarking scheme based on the shuffling preprocessing of host image and watermark image encrypted by skew tent map is presented. In Sect. 5, attacking tests are performed to show that the image watermarking scheme is robust against attacks such as cropping, JPEG compression, noising.

2   2D Affine Transformation

The following 2D affine transformations are widely used to generate fractals [18]

$$T(x, y) = \begin{pmatrix} a & 0 \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} = A \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix}. \tag{1}$$

From the dynamical system's point of view, we can consider (1) as one discrete dynamical system and study its long term dynamical behaviour. The dynamical system of (1) may be written as

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = T(x_k, y_k), k = 0, 1, 2, \cdots. \tag{2}$$

In this paper we are interested in those ones whose orbits own chaotic nature, such as ergodicity, pseudo-randomness, sensitivity to initial conditions and control parameters. In order to make the dynamical system (2) possess chaotic properties, the control parameters a, b, c, d should be chosen appropriately. As a matter of fact, let the two eigenvalues of matrix A be λ_1, λ_2, if the determinant is greater than 1, then at least one of $|\lambda_1|, |\lambda_2|$ should be greater than 1, and therefore the dynamical system (2) is expanding at x and/or y direction. To make the dynamical system possess more chaotic sense, we choose the control parameters satisfying $|a| > 1, |d| > 1$ and c, e, f

arbitrarily, which make the dynamical system (2) expand in both x and y directions. In order to confine the orbits to the unit square $[0, 1] \times [0, 1]$, we add a modular function to (2), that is,

$$\begin{pmatrix} x_{k+1} \\ y_{k+1} \end{pmatrix} = A \begin{pmatrix} x_k \\ y_k \end{pmatrix} + \begin{pmatrix} e \\ f \end{pmatrix} \bmod 1, \; k = 0, 1, 2, \cdots. \tag{3}$$

The chaotic nature of (3) is demonstrated visually in Fig. 1, where the initial control parameters and initial conditions are set as follows.

$$a = 2.16, c = 2.31, d = 3.76,$$
$$e = 0.35, f = 0.51, x_0 = 0.46, y_0 = 0.63.$$

A typical orbit of $z_0 = (x_0, y_0)$ derived from the dynamical system is $\{z_k = T^k(z_0), k = 0, 1, \cdots\}$, which is shown in Fig. 1(a). The orbit shall fill the unit square $[0, 1] \times [0, 1]$ if the orbit evolves infinity, which indicates that the system is chaotic. The control parameters and the initial conditions can be regarded as encryption keys as the system is used to design image shuffling schemes. There also exist some good dynamical features, such as desirable auto-correlation and cross-correlation features indicating pseudo-randomness. The iterated trajectory is used to calculate the correlation coefficients, which are shown in Fig. 1(b)-(c) respectively.
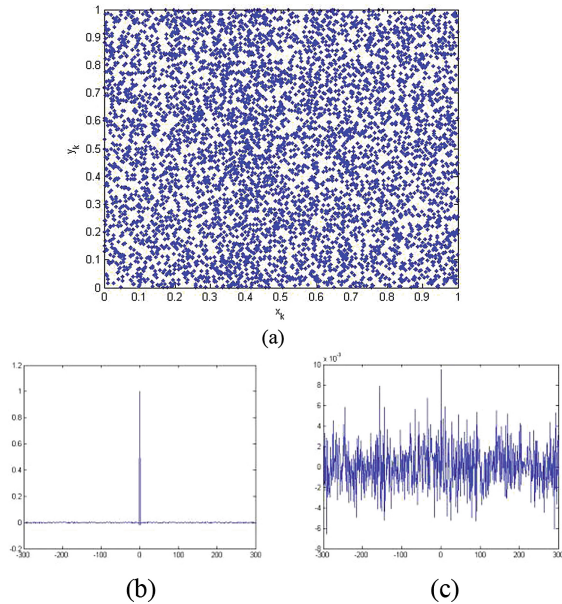
Fig. 1. (a) The chaotic orbit with 6000 points; (b) The auto-correlation of sequence $\{y_k\}$; (c) The cross-correlation between $\{x_k\}$ and $\{y_k\}$

3 Application of Affine Transformation in Image Shuffling

The ergodic property of chaotic system (3) means that for almost the initial values (x_0, y_0), one can get chaotic orbits of (3) filling the unit square, which implies that a sufficient long orbit will fill the square. For any pixel position $(i,j) \in [0, M-1] \times [0, N-1]$, there exists $(x_k, y_k) \in [0,1] \times [0,1]$ approximating $(i/M, j/N)$. This ergodic property can be used to get an order number for every point $(i/M, j/N)$, $i = 0, \cdots,$ $M-1; j = 0, \cdots, N-1$. The obtained order number for pixel (i,j), namely k, can be applied to change the pixel position (i,j) to the new position $([k/N], \mathrm{mod}(k,N))$, where $[x]$ refers to the largest integer not larger than x and $\mathrm{mod}(k,N)$ is $k - N * [k/N]$. If the same procedure is employed to all the pixels, one will get a shuffled digital image at last. The shuffled image owns a better shuffling result than that by conventional affine transformation shuffling method. Another advantage for this kind of schemes is that it owns more secret keys, that is the initial position (x_0, y_0). From the security point of view, the proposed scheme in this paper is superior. Suppose that the original image is A with size $M \times N$ and the shuffled image is B with the same size. The proposed shuffling scheme is outlined as follows.

Step 1. Set the values of $x_0, y_0, a, c, d, e, f, T, IterM$.

Step 2. Iterate (3) with initial condition (x_0, y_0) by T rounds to discard the transitional part of the orbit $\{(x_n, y_n) : n = 0, \cdots, T\}$.

Step 3. Set one index matrix to be zero matrix with size $M \times N$, and $k = 0$. Do the following loop.

```
For n=1: iterM   do
     Iterate (3) to get a new orbit point
     (z_n, w_n) = (x_{n+T}, y_{n+T});
          If  index([M * z_n], [N * w_n]) == 0 then
                index([M * z_n], [N * w_n]) = 1;
                B([k / N], mod(k, N)) = A([M * z_n], [N * w_n]);
                k = k + 1;
          Endif
     Endfor
```

Step 4. If there are pixels not ergodic, i.e., there are some elements in index matrix are still zero after Step 3, then do the following loop.

```
For i=1:M do
For j=1:N do
   If  index(i, j) == 0 then
        B([k / N], mod(k, N)) = A(i, j);
            k = k + 1;
       Endif
    Endfor
Endfor
```

We note that one can get the order numbers for all the pixel positions theoretically as long as the orbit is long enough. In practice, in order to save the computational time and the storage space, one can't get the order numbers for all the pixel positions. If there are pixel positions not ergodic, they can be put in the remainder part in the shuffled image orderly, that is what we show in Step 4. The experimental results are shown in Fig. 2(a), where we set $(x_0, y_0) = (0.46, 0.63)$, $a = 5.16$, $b = 0$, $c = 2.31$, $d = 3.76$, $e = 0.35$, $f = 0.51$, *IterM* = 300000, $T = 30$ and apply the scheme one round. We discard the transitional part of the orbit $\{(x_n, y_n) : n = 0, \cdots, T\}$ to get an orbit with more sensitivity to initial conditions and control parameters. One can observe that the shuffling effect is pretty good compared with the shuffled image performed by the traditional scheme proposed in [19], see Fig. 2(b) for shuffling image performed by the traditional shuffling scheme.

(a) original image Lena
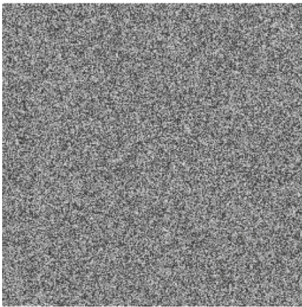

(b) shuffling scheme propsed here
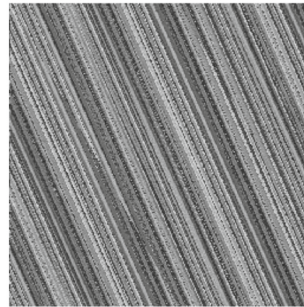
(c) traditional shuffling scheme

Fig. 2. Shuffling result

It is well known that a good encryption algorithm should be sensitive to the cipher keys. In order to test the sensitivity of the cipher keys, namely, x_0, a, e, some experiments are performed. In Fig. 3(a), we set $x_0 = 0.46 + 10^{-16}$ and the other keys unchanged and apply the scheme one round. The difference between Figs. 2(b) and 3(a) is 99.34 %. In Fig. 3(b), we set $a = 5.16 + 10^{-15}$ and the other keys unchanged and perform the scheme one round. There is a difference of 99.28 % between Figs. 2(b) and 3(b).

In Fig. 3(c), we set $e = 0.35 + 10^{-16}$ and the other keys unchanged. The difference between Figs. 2(b) and 3(c) is 99.27 % and the difference image is shown in Fig. 3(d). The obtained results imply that shuffling scheme is sensitive to the all the keys. The proposed shuffling scheme has a better disordering effect compared with the traditional shuffling scheme.

(a)   (b)

(c)   (d)
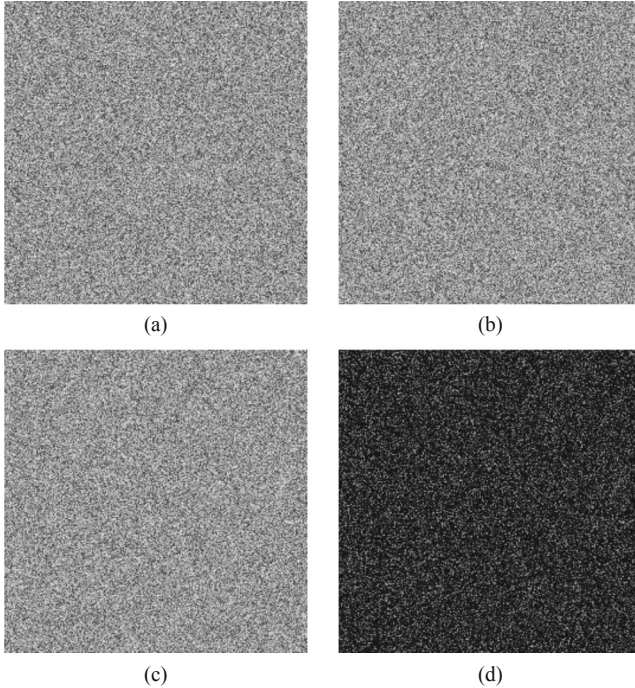
Fig. 3. Sensitivity test for the shuffling scheme

Shuffling degree is another numeric index to evaluate shuffling schemes [12]. For one digital image P sized $H \times W$, We first compute the gray difference between pixel (i,j) and its four neighboring pixels by

$$GD(i,j) = \frac{1}{4} \sum_{i' j'} [P(i,j) - P(i',j')]^2,$$

$$\text{where}(i',j') = \{(i-1,j), (i+1,j), (i,j-1), (i,j+1)\}.$$

Calculate all $GD(i,j)$ for the whole image except those pixels at four sides and take the mean

$$E(GD) = \frac{\sum\limits_{i=2}^{H-1} \sum\limits_{j=2}^{W-1} GD(i,j)}{(H-2) \times (W-2)}.$$

Let $E(GD), E'(GD)$ be the mean gray differences of the original image and the shuffled image respectively. The shuffling degree is defined as follows.

$$GDD = \frac{E'(GD) - E(GD)}{E'(GD) + E(GD)} \tag{4}$$

The value of GDD lies in [-1,1]. We note that if GDD is more near 1, the better shuffling effect is obtained. The conventional shuffling scheme proposed by [19] and the proposed shuffling scheme in this paper are compared at the same iteration round. The shuffling degree calculated by Fig. 2(a) and (b) is 0.9170. The shuffling degree yielded by Fig. 2(a) and (c) is 0.8744. From the result, one can conclude that the proposed scheme is superior in shuffling effect.

4   Watermarking Scheme

An image watermarking scheme is proposed in this section. We first use the shuffling scheme proposed in Sect. 3 to shuffle the host image and then imbed the watermark bits in the shuffled host image. The watermark image can be detected or extracted easily. The watermark embedding scheme is proposed as follows.

Step 1. Apply the shuffling scheme to shuffle the original host image A sized $M \times N$, we get a shuffled host image B.

Step 2. Iterate the skew tent map with control parameter $p = 0.29$ for $L = W \times H$ times to get the orbit $\{v_1, \cdots, v_L\}$ of initial conditions $v_0 = 0.36$:

$$v_{n+1} = f(v_n) = \begin{cases} v/p, & \text{if } v \in [0,p], \\ (1-v)/(1-p), & \text{if } v \in (p,1]. \end{cases}$$

Step 3. Sort $\{v_1, \cdots, v_L\}$ to get the new set $\{\bar{v}_1, \cdots, \bar{v}_L\}$ and the index set $\{i_1, \cdots, i_L\}$ such that $\bar{v}_n = v_{i_n}, n = 1, \cdots, L$. Reshape the original watermark W to be a vector V with length L. Then set $V1(k) = V(i_k), k = 1, \cdots, L$. Reshape $V1$ to be a 2D matrix, that is the encrypted watermark W_1.

Step 4. Take the left-top part B_1 of the shuffled image B with the same sizes as those of W:

$$B_1 = \{B(i,j)|0 \leq i < W - 1, 0 \leq j < H - 1\}.$$

The following formula is applied to imbed the watermark image into B_1:

$$B_s = (1 - t)W_1 + tB_1, \, t \in (0, 1).$$

Step 5. Put B_s back to the left-top part of B and apply the inverse shuffling scheme to get an image A_w imbedded watermark.

We note that $t \in (0, 1)$ should be chosen suitably in the imbedding scheme. If t is too small, the quality of the watermarked image will be influenced. If t is too large, the information of the watermark image becomes weak and it is difficult to extract the watermark image. The extraction scheme is just the inverse of the imbedding scheme and is easy to work. Figure 4 shows the image imbedded watermark with $t = 0.9$, the watermark image and the extracted watermark image.

5   Attacking Tests

In this section, experiments are performed to test the robustness of the proposed watermarking scheme. Attacks in the experiments are cropping, salt and pepper noising, Gaussian noising, speckle noising, JPEG compression, etc. Experimental results show the proposed scheme is robust against the mentioned kinds of attacks.

(a) Watermarked Lena image
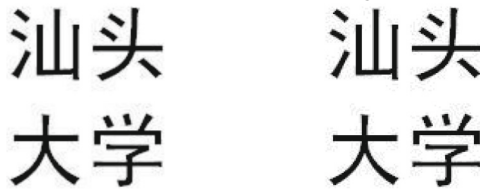
(b) Original watermark   (c) Extracted watermark

Fig. 4. The watermarking results

All the attacking tests are applied to the watermarked image performed with those keys set in the shuffling scheme in Sect. 3 and those set in the watermarking scheme. The experimental results are shown in Fig. 5.

(a) Cropping attack with cut-off 128X128 at the left-top corner
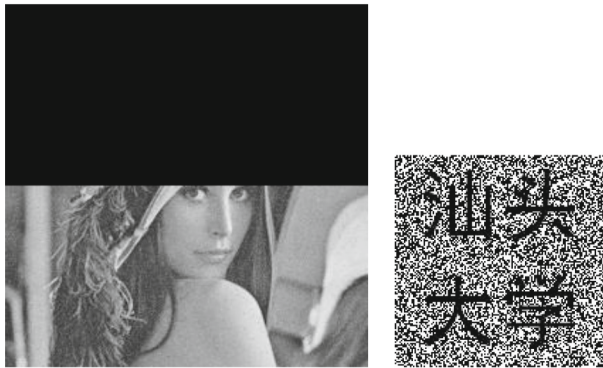
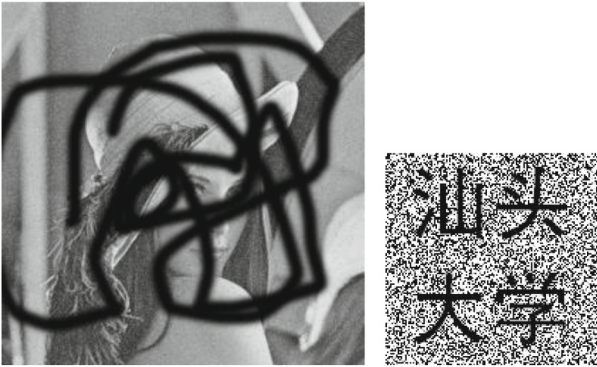(b) Cropping attack with cut-off 128X256 at the top half

(c) Salt & pepper noising attack with density 0.1

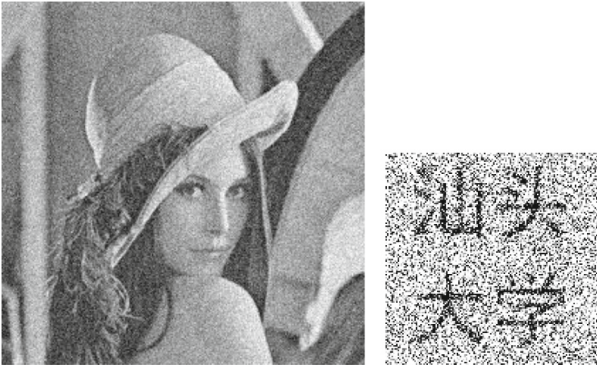**Fig. 5.** Results of attacking tests
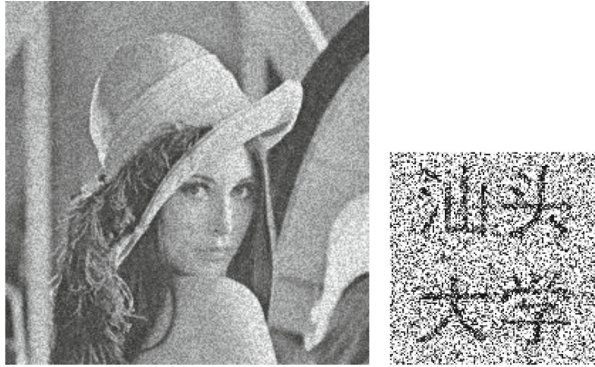
(d) JPEG compression attack with quality 50

(e) Daubing attack

(f) Gaussian noising attack with mean 0 and variance 0.005

Fig. 5. (continued)

(g) Speckle noising attack with mean 0 and variance 0.02

Fig. 5.   (continued)

References

1. Cox, I.J., Kilian, J., Leighton, F.T., Shamoon, T.: Secure spread spectrum watermarking for multimedia. IEEE Trans. Image Process. **6**, 1673–1687 (1997)
2. Fridrich, J.: Symmetric ciphers based on two-dimensional chaotic maps. Int. J. Bifurcat. Chaos **8**(6), 1259–1284 (1998)
3. Wang, Y., Doherty, J.F., Dyck, R.E.V.: A wavelet-based watermarking algorithm for ownership verification of digital images. IEEE Trans. Image Process. **11**, 77–88 (2002)
4. Wong, P.H.W., Au, O.-C., Yeung, Y.-M.: A novel blind multiple watermarking technique for images. IEEE Trans. Circ. Syst. Video Technol. **13**(8), 813–830 (2003)
5. Ye, R.: A novel chaos-based image encryption scheme with an efficient permutation-diffusion mechanism. Opt. Commun. **284**, 5290–5298 (2011)
6. Mukherjee, D.P., Maitra, S., Acton, S.T.: Spatial domain digital watermarking of multimedia objects for buyer authentication. IEEE Trans. Multimedia **6**(1), 1–15 (2004)
7. Nikolaidis, N., Pitas, I.: Robust image watermarking in the spatial domain. Sig. Process. **66**, 385–403 (1998)
8. Lu, C.-S., Huang, S.-K., Sze, C.-J., Liao, H.Y.M.: Cocktail watermarking for digital image protection. IEEE Trans. Multimedia **2**, 209–224 (2000)
9. Lee, Z.-J., Lin, S.-W., Shun-Feng, S., Lin, C.-Y.: A hybrid watermarking technique applied to digital images. Appl. Soft Comput. **8**, 798–808 (2008)
10. Chang, C.-C., Yih-Shin, H., Tzu-Chuen, L.: A watermarking-based image ownership and tampering authentication scheme. Pattern Recogn. Lett. **27**, 439–446 (2006)
11. Kima, B.-S., et al.: Robust digital image watermarking method against geometrical attacks. Real-Time Imag. **9**, 139–149 (2003)
12. Ye, R., Li, H.: A novel image scrambling and watermarking scheme based on cellular automata. In: Proceedings of the 2008 International Symposium on Electronic Commerce and Security, Guangzhou, pp. 938–941 (2008)
13. Ye, R.: A novel digital image scrambling and watermarking scheme based on orbits of Arnold transform. In: Proceedings of the 2009 Pacific-Asia Conference on Circuits, Communications and System, pp. 485–488 (2009)
14. Zhu, L., Li, W., Liao, L., Li, H.: Novel image scrambling algorithm for digital watermarking based on chaotic sequences. Int. J. Comput. Sci. Netw. Secur. **6**(8B), 125–130 (2006)

15. Qi, D.-x., Zou, J.-c., Han, X.-y.: A new class of scrambling transformation and its application in the image information covering. Sci. China (Ser. E) **43**(3), 304–312 (2000)
16. Chen, G., Miao, Y., Chui, C.K.: A symmetric image encryption scheme based on chaotic cat maps. Chaos, Solitons Fractals **21**, 749–761 (2004)
17. Mao, Y., Chen, G., Lian, S.: A novel fast image encryption scheme based on 3D chaotic baker maps. Int. J. Bifurcat. Chaos **14**(10), 3613–3624 (2004)
18. Barnsley, M.F.: Fractals Everywhere, Academic Press, 1993. Zhu, G., Cao, C., Hu, Z., He, S., Bai, S.: An image scrambling and encryption algorithm based on affine transformation. J. Comput.-Aided Des. Comput.- Graph. **15**, pp. 715–717 (2003)
19. Zhang, G.J., Liu, Q.: A novel image encryption method based on total shuffling scheme. Opt. Commun. **284**, 2775–2780 (2011)

A Novel Emergency Detection Approach Leveraging Spatiotemporal Behavior for Power System

Wanxing Sheng[1], Ke-yan Liu[1], Yixi Yu[2,3], Rungong An[2,3], Xin Zhou[2,3], and Xiao Zhang[2,3(✉)]

[1] China Electric Power Research Institute, Beijing 100192, China
{wxsheng, liukeyan}@epri.sgcc.com.cn
[2] Key Laboratory of Data Engineering and Knowledge Engineering,
Ministry of Education, Renmin University of China, Beijing 100872, China
[3] School of Information, Renmin University of China, Beijing 100872, China
{yuec, anrungong, zhouxin3l4l59, zhangxiao}@ruc.edu.cn

Abstract. Emergency detection is of significant value in preventing damages to power systems and even saving lives by alerting anomalous behaviors of electronic devices. However, existing works only identify manually encoded events or patterns on power supply and thus consume huge amount of manpower and cannot deal with undefined emergency. In this paper, we propose a novel method 3D-LRT (Three Dimensional Likelihood Ratio Test) to detect emergency in large scale power system. To the best of our knowledge, this is the first work that leverages spatiotemporal behavior characteristics to identify anomalous patterns in power system. For scalability of large scale power systems, we further optimize our 3D-LRT using pruning and parallelization methods to save time overhead. We conduct experiments on real-world synthetic data sets. The results demonstrate that our 3D-LRT method is both effective and efficient.

Keywords: Data mining · Anomaly detection · Power grid data · LRT · Optimization method

1 Introduction

Power system is prone to frequent faults, which may occur in any of its components, such as electric distribution network, transmission network and so on [1]. It is well known that faults can destabilize the entire system and cause injuries to personnel. Detection of faults is therefore of paramount importance from economic and operational viewpoints.

In this paper, we apply a statistical approach on a great quantity of power data, which are explored to extract the electrical anomaly pattern in power system. We know that an area can be plied by thousands of electrical monitoring equipments twenty four

hours per day. Most of these equipments remain their position unchanged and record data over time, so that spatiotemporal coordinates of power data are available. If an area accessible to electricity is partitioned into a grid, then the cell data will settle into a pattern and vary periodically. For example, in residential areas, the power consumption is higher from seven to eleven in the evening than other time periods. Thus, at a given time, we can estimate cell data in the grid, such as active power and reactive power. Occasionally, if the cell data is far departure from its periodical behavior, then some unforeseen events like an earthquake or a power failure may be happening.

Our objective is to discover regions and time intervals which have the greatest statistically significant imparity from their expected behavior. Once such regions and time intervals are recognized then experts can identify events which may induce unexpected disasters. This, in turn, can help make provisions to deal with electrical behaviors in the future.

So, what's the problem? (1) Most of the statistical anomaly detecting methods base on purely space data, without considering the time dimension; (2) Even when the time dimension is considered, most of existing works on spatiotemporal anomaly detection are simply used for some fixed statistical models that can only detect some dramatically increasing or decreasing regions and cannot let users customize the data model according to actual conditions; (3) Conventional statistical anomaly detection methods still have such high time and space complexity that is not applicable to today's massive power data.

In an article of particular relevance to our work, the LRT framework [2] states the LRT computation steps in spatial grid and develops methods to prune rectangles to avoid computing their associated LRT statistics. In this paper, we propose the 3D-LRT method inspired by Wu et al. [2] to detect anomalous electrical patterns. More specifically, the contributions are:

- We propose a general and efficient method named 3D-LRT for spatiotemporal anomaly detection in power system.
- We propose two methods to optimize the performance of 3D-LRT.
- We carry out experiments on the 3D-LRT method whereby several methods are compared and we verify the accuracy, pruning rate and performance of the optimization methods on large-scale real-world data sets.

The rest of this paper is organized as follow. Section 2 reviews related work. Section 3 proposes our 3D-LRT method. Section 4 illustrates our optimization methods of 3D-LRT. Section 5 shows the experiments and case studies. Section 6 ends the paper with some conclusions.

2   Related Work

2.1   Classification and Faults Detection in Power System

Faults detection and management has been a long-standing problem in power distribution networks. As our society is becoming increasingly dependent on electric power, the economic and societal cost due to loss of loads caused by distribution faults is severe.

To avoid the problem, many experts and scholars have proposed various anomaly data mining technology for power system. These methods are summarized as follows: expert system, artificial neural network, Petri network optimization method, rough set theory, fuzzy set theory, multi-agent system (MAS).

Expert systems have been proposed to identify, classify and diagnose power system events within a limited number of events [3]. If too many event types or features are to be analyzed, the expert system would become excessively complicated and thus increase the risk of losing selectivity. Another drawback is that expert systems are not always portable due to the settings that they depend mostly on the designer or operator for a particular set of events. ANNs have been applied for power system security, optimal static VAR compensation (SVC) [4], SMs [5], and power factor compensations [6]. One of the powerful tools for modeling parallel processing is Petri net [7]. There have been some proposed model-based systems that use Petri net or colored-and-timed Petri net for faster inference [8]. A multi-agent system (MAS), consisting of multiple distributed intelligent agents [9], is one of the promising approaches in order to deal with power system restoration tasks.

2.2   Anomaly Detection Methods

Anomaly detection has wide application prospect in power systems, such as detecting credit card fraud and network intrusion. Therefore, it has gained extensive attention from data mining, machine learning and statistics related fields. Anomaly detection methods can be broadly classified as: distance-based, classification-based, statistic-based, cluster-based and so on. In the paper [10], authors present some high-level overview and summary of anomaly detection methods and techniques.

In this paper, we focus on the statistic-based approach to detect spatiotemporal anomaly. In the domain of spatiotemporal applications, most statistical anomaly detection approaches proposed so far are purely about spatial searching [11]. Wu et al. [2] proposes a generic framework called LRT (Likelihood Ratio Test, LRT) for any underlying statistics model. It uses the classic likelihood ratio test statistic as a scoring function to evaluate the "anomalousness" of a given spatial region with respect to the rest of the spatial region. Moreover, a generic pruning strategy was proposed to greatly reduce the number of likelihood ratio tests. However, it is used only for spatial data without considering the temporal property. Although some of existing works take time aspect into consideration, they are simply applicable for some fixed statistical models, such as the LRT-based spatial and space-time scan statistic (SSS) framework [12]. Such models were originally designed for Poisson and Bernoulli distribution, but now many different variations of ordinal, exponential and normal distributions [13] have been implemented in many softwares such as SaTScan [14], which cannot be handled by existing works.

In this paper, we propose the 3D-LRT method to detect anomalies of 3D spatiotemporal data. Furthermore, we present the optimized methods to reduce the computation overhead of 3D-LRT.

3   The 3D-LRT Statistical Method

3D-LRT is a generic anomaly detection method based on likelihood ratio test for 3D spatio-temporal data. Our generic 3D-LRT method generalizes the set of statistical models that can be used for searching spatiotemporal anomalies. The core idea using LRT to find spatiotemporal anomalies is to test how many times more likely the spatiotemporal data under normal status than anomaly status.

The procedure of applying 3D-LRT to detect anomaly is as follows:

(1) **Model the research object and define the likelihood function**

When using likelihood ratio test to detect anomaly, the first thing that a user must do is to postulate an appropriate stochastic model to describe the data generating process within each spatiotemporal cell. The stochastic model for a spatiotemporal cell c is characterized by a probability density function (PDF), denoted by $f(X_c|\theta_c)$.

Suppose that there is a sample $X_R = \{c_1, c_2, \ldots, c_n\}$ of n independent and identically distributed observations in a spatiotemporal region R. For an IID sample, the joint density function is

$$f(c_1, c_2, \ldots c_n | \theta_R) = f(c_1|\theta_R) \times f(c_2|\theta_R) \times \cdots \times f(c_n|\theta_R)$$

Then the likelihood of region R is:

$$L(\theta_R|X_R) = f(X_R|\theta_R) = \prod_{i=1}^{n} f(c_i|\theta_R)$$

(2) **Specify two competing hypotheses and define the test set**

Informally, given a test area R, the two competing hypotheses take the form:

H_0: assumes that the process generating data of the cells inside R is not substantially different from the process generating data of the cells outside R.

H_1: assumes that the process generating data of the cells inside R is substantially different from the process generating data of the cells outside R.

The process generating the data is modeled via PDF, so the hypotheses are equivalent to determining whether the parameters to the process differ within and without R. So we should define the parameters in PDF that are tested to see whether they are same within a region R and outside region R. These parameters form the test set and are denoted by TP (Test Parameters, TP).

Example. We assume the data in power system obeys normal distribution, so the PDF for given cell c:

$$f(X_c|\theta_c) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(X_c-\mu)^2}{2\sigma^2}}$$

The test set is

$$TP = \{\sigma, \mu\}$$

In this case, the two competing hypotheses become:

H_0: $\forall \theta \in TP$, $\theta_R = \theta_{\bar{R}}$.

H_1: $\exists \theta \in TP$, $\theta_R \neq \theta_{\bar{R}}$, where $\bar{R}$ denoted the area outside R.

(3) Use the maximum likelihood estimation method to estimate parameters

The likelihood ratio defined as follows:

$$\Lambda(x) = \frac{\sup\{L(\theta|x)|H_0\}}{\sup\{L(\theta|x)|H\}}$$

In order to construct functions that obey the Chi-square distribution, we define the likelihood ratio of region R:

$$D(R) = -2\log(\Lambda(X_R))$$

So,

$$D(R) = -2\log\big(\sup\{L(\theta|x_R)|H_0\} + 2\log\big(\sup\{L(\theta|x_R)|H\}\big)\big)$$

Calculating $\sup\{L(\theta|X_R)|H_0\}$ is equivalent to maximize $L(\theta_R|X_R)|H_0$, the maximization is done under the constraint that for every two cells c_1 and c_2 in grid G, and for every test parameter tp, $tp_{c1} = tp_{c2}$. Then,

$$L(\theta|X_R)|H_0 = L(\theta_R|X_R) = L(\theta_G|X_G)$$

Similarly, calculating $\sup\{L(\theta|X)|H\}$ is equivalent to maximize $L(\theta|X)|H$, where H represents the complete parameter space. We should consider X_R and $X_{\bar{R}}$ and compute their likelihood respectively. We choose the values in θ_R and $\theta_{\bar{R}}$ that maximize the entire grid´s likelihood. So,

$$L(\theta|x)\big|H = L\big(\theta_R|X_R\big) \times L\big(\theta_{\bar{R}}|X_{\bar{R}}\big)$$

In statistics, maximum likelihood estimation (MLE) is a method of estimating the parameters of a statistical model. The procedure to perform MLE in the null parameter space for an region R is denoted by $\theta_G = MLE_0(f(R))$; the procedure to perform MLE in the complete parameter space for an region R is denoted by $MLE_1(R, f(G))$.

Example. Based on the maximum likelihood method, the MLE of region R is:

$$\widehat{\theta_R} = \left(\widehat{\mu_R}, \widehat{\sigma^2 R}\right) = \left(\bar{x}_R, \sum_{c_i \in R}(x_{c_i} - \bar{x}_R)^2 \Big/ n\right)$$

(4) Calculate 3D-LRT score and output top-k anomalous regions

After getting the parameter values using MLE, we calculate the likelihood ratio test for each region R in the whole spatiotemporal space G as follow:

$$D(R) = -2\log L(\theta_G|X_G) + 2\log L(\theta_R|X_R) + 2\log L(\theta_{\bar{R}}|X_{\bar{R}})$$

Here, $\theta_G = MLE_0(f(G))$, $(\theta_R, \theta_{\bar{R}}) = MLE_1(R, f(G))$.

We can get likelihood ratio test values of all the sub-regions. The greater $D(R)$, the anomalousness of region R is more obvious. We can sort $D(R)$ for all sub-regions, or set a threshold using the Chi-square distribution combined with the confidence level, to get top-k regions that are the most significantly anomalous. We present the formal process of 3D-LRT in Algorithm 1.

The 3D-LRT is quite general that it virtually works with any underlying spatio-temporal statistical model and only needs a few distribution functions to be instantiated. To improve the generalization and applicability of 3D-LRT, we propose the optimized method of 3D-LRT in Sect. 4 to reduce the time overhead and compute cost.

4   Optimization Methods of 3D-LRT

It is expensive to compute the likelihood ratio for each region. Besides, the time complexity to enumerate all sub-regions is $O(n^4 \times t^2)$. It is time-consuming to enumerate all sub-regions and compute likelihood ratio for each sub-region. We present two methods to optimize 3D-LRT. One method which reduces the times of computing likelihood ration named BP-3D-LRT, the other method which enumerates all sub-regions in parallel using multi-threads named MT-PRLL-3D-LRT. This paper introduces these two detail methods in Subsects. 4.1 and 4.2.

```
Algorithm 1 3D-LRT anomaly detection method
  Input: an n n t 3D spatiotemporal grid G
  Output: top-k most anomalous regions
  for each region R in the grid G do
```
    Let $\theta_G = MLE_0(f(G))$;

    Let $(\theta_R, \theta_{\bar{R}}) = MLE_1(R, F(G))$;

    Let $D(R) = -2\log L(\theta_G \mid X_G) + 2\log L(\theta_R \mid X_R) + 2\log L(\theta_{\bar{R}} \mid X_{\bar{R}})$;
```
    If D(R) is in the top-k anomaly found so far, then
remember R.
  end for
  return top-k most anomalous regions
```

4.1   Optimization Method Based on Block Pruning

The core idea of BP-3D-LRT is to avoid computing the likelihood ratio of sub-region which can't contribute to the final result. Before computing the likelihood ratio test of region R, BP-3D-LRT accesses the upper bound of its likelihood ratio, if the upper

bound does not exceed the anomaly threshold, region R and its sub-regions are not accessed.

At each iteration of 3D-LRT in Algorithm 1, it is easy to handle $MLE_0(f(G))$, since the value is the same no matter which region R represents. So we only need to compute $MLE_0(f(G))$ once and reuse the value for every iteration. However, it is more difficult to get the value of $MLE_1(R, f(G))$. Fortunately, we can get the upper bound of $L(\theta_R|X_R)$ and $L(\theta_{\bar{R}}|X_{\bar{R}})$.

According to the formula in paper Wu et al. [2], we can get $L(\theta_R|X_R) \leq L(\theta'_{R1}|X_{R1}) \times L(\theta'_{R2}|X_{R2})$.

Here,

$$\theta_R = MLE_1(R, f(G)), \theta'_{R1} = MLE_0(f(R_1)), \theta'_{R2} = MLE_0(f(R_2)), R = R_1 \bigcup R_2.$$

So, we can tile R to several sub-regions and get the upper bound of $L(\theta_R|X_R)$. Similarly, tile $\bar{R}$ can get the upper bound of $L(\theta_{\bar{R}}|X_{\bar{R}})$. This paper presents the detail process to tile region R and $\bar{R}$ as following.

Tiling region R: In BP-3D-LRT, we get the $n \times n \times t$ 3D grid from time dimension and divide it into some $n_i \times n_j \times t$ 3D sub-grids. We recursively split the $n_i \times n_j \times t$ 3D sub-grid block into two sub-blocks of the same size and calculate their MLE_0 value respectively. Figure 1 shows an example of the block pruning approach. Algorithm 2 shows the divide-and-conquer method in detail and return the upper bound of $L(\theta_R|X_R)$. This method always tiles R with the largest possible rectangle to guarantees the smallest number of tiles.

```
Algorithm 2 The upper bound of L(θ_R | X_R) based on block
pruning
  Procedure Tile_R(x1, y1, z1, x2, y2, z2, low, high)
  Let mid=(low+high)/2 //base case
  if (z1==low && z2==high) or (z1==low && z2==mid) or
    (z1==mid && z2==high) then return regionLRT(x1, y1,
    z1, x2, y2, z2) //recursive case
    //regionLRT() returns the stored grid log likelihood
  end if
```

Tiling region $\bar{R}$: Now we turn our attention to the tiling methods for $\bar{R}$ in parallel using block pruning. Figure 2(a) and (b) show region $\bar{R}$ can be regarded as solid time block and hollow time block. The former, composed by upside $\overline{Ru}$ and downside $\overline{Rd}$,
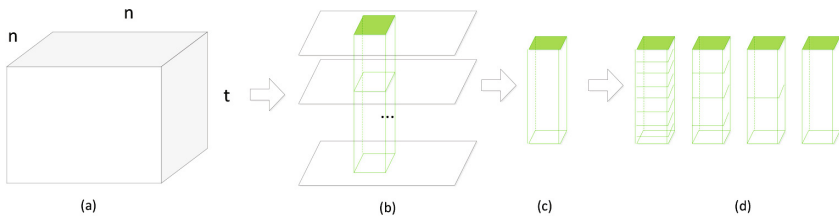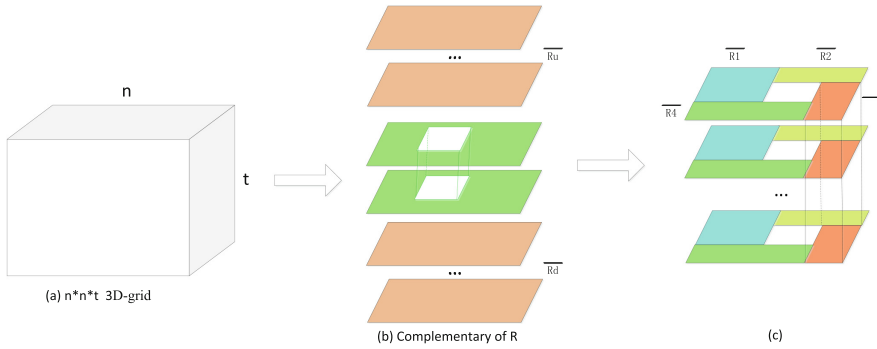

Fig. 1. Tiling region R based on block pruning

Fig. 2. Tiling region R using block pruning

can use the block's whole 2D grids' likelihood. The latter, can be cut into four parts:$\overline{R1}$, $\overline{R2}$, $\overline{R3}$ $\overline{R4}$, which is illustrated in Fig. 2(c).

Then, $L(\theta_R|X_R)$ is bounded by:

$$L(\theta_{\bar{R}}|X_{\bar{R}}) \leq L\left(\theta'_{\overline{RU}}|X_{\overline{RU}}\right) \times L\left(\theta'_{\overline{Rd}}|X_{\overline{Rd}}\right) \times L\left(\theta'_{\overline{R1}}|X_{\overline{R1}}\right) \times L\left(\theta'_{\overline{R2}}|X_{\overline{R2}}\right) \times L\left(\theta'_{\overline{R3}}|X_{\overline{R3}}\right)$$
$$\times L\left(\theta'_{\overline{R4}}|X_{\overline{R4}}\right)$$

4.2 Multi-threaded Parallel Method

MT-PRLL-3D-LRT divides the $n \times n \times t$ 3D spatiotemporal grid to smaller 3D blocks. For each small 3D block Ri, we regard it as a whole, as illustrated in Fig. 3. We distribute each 3D sub-grid Ri to a thread and calculate each block in parallel.

For a region Ri_j in Ri, we use the following formula to get its likelihood:

$$D\left(Ri_j\right) = -2\log L(\theta_G|X_G) + 2\log L\left(\theta_{Ri_j}\right) + 2\log L\left(\theta_{\overline{Ri_j}}|X_{\overline{Ri_j}}\right)$$
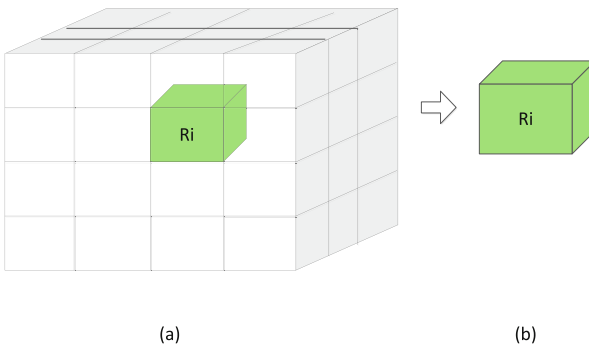

Fig. 3. Parallelization method of dividing 3D grid

When we finish calculating each sub-grid Ri, we can merge all output anomalies of each sub-grid and get the global top-k ones. Or, we can apply this parallel method recursively. It should be noted that this method cannot detect regions which across different sub-grids. If, in practice, we need to get more fine-grained anomalies, we can first filter out normal regions using the 3D-LRT score, and then handle the possibly anomalous regions to obtain all anomalies.

We propose a general 3D-LRT method (NV-3D-LRT) and two optimized methods (BP-3D-LRT and MT-PRLL-3D-LRT) of 3D-LRT. The procedure and summarization of all above-mentioned methods is described in Fig. 4. The effective and efficient of our methods are proved in Sect. 5.
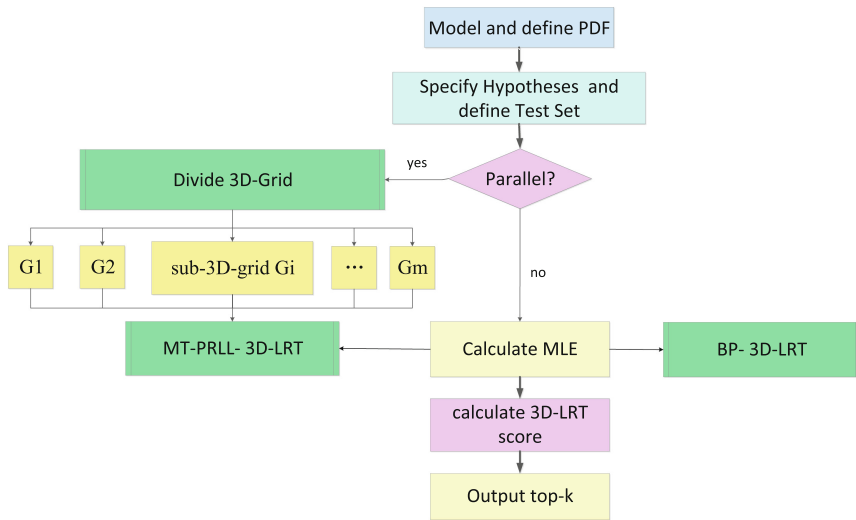

Fig. 4. The procedure and summarization of all mentioned 3D-LRT methods

5 Experiments

This paper presents synthetic data based on Poisson distribution to evaluate the effective and efficient of 3D-LRT. All experiments are finished on a server with 16 physical Intel Xeon E5-2670 dual-core CPU, 252 GB RAM. The Operating System on this server is Linux 2.6.32-431.el6.x86. The prototype is implemented by C language. The experimental setup and results are shown as follows.

Experimental Setup. For each cell c in the $n \times n \times t$ 3D grid, we generate a population size b_c uniformly by a normal distribution. Then the number of "successes" k_c in the cell is generated by sampling from a $P(b_c p)$ random variable of success rate p. Given a test region R, we denote the "success" rate within R as p and within R as q. The test parameter is p, while the null hypothesis assumes that $p = q$ and the alternative hypothesis assumes that $p \neq q$. The different parameter values of PDF are outlined in Table 1.

Table 1. The parameter values of PDF

Scenario	Hypothesis	μ	δ	Data distribution	Success rate p_c
Scenario 1	H_0 holds true	1×10^4	1×10^3	Uniform	0.001
Scenario 2	H_0 holds true	1×10^5	1×10^3	Uniform	0.001
Scenario 3	H_1 holds true	1×10^4	1×10^3	Nonuniform	0.003
Scenario 4	H_1 holds true	1×10^4	1×10^3	Nonuniform	0.01

Since the complete parameter space has only one more dimension than the null parameter space, the null distribution obeys Chi-squared with one degree of freedom. For our tests, we run the experiments on our 3D-LRT methods: NV-3D-LRT, BP-3D-LRT and MT-PRLL-3D-LRT. For each method, we run 50 trials on grids size from $4 \times 4 \times 4$ to $64 \times 64 \times 64$. We record the pruning rate (the number of regions pruned/the number of all anomalous regions), and check if there are false alarms. Our initial cutoff value for pruning is chosen to correspond to an overall false positive rate of 5 %. In all tests, we seek the top sub-region.

Experimental results and analysis.

1. Test of accuracy. The results of the four 3D-LRT methods are the same, so we give the results all together in Table 2.

Table 2. The accuracy of 3D-LRT methods

	$4 \times 4 \times 4$	$8 \times 8 \times 8$	$16 \times 16 \times 16$	$32 \times 32 \times 32$	$64 \times 64 \times 64$
Scenario 1	No false alarm	No false alarm	No false alarm	No false alarm	No false alarm
Scenario 2	No false alarm	No false alarm	No false alarm	No false alarm	No false alarm
Scenario 3	100 %	100 %	100 %	100 %	100 %
Scenario 4	100 %	100 %	100 %	100 %	100 %

2. Test of pruning rate. The average pruning rate of BP-3D-LRT on 4 scenarios are in Table 3.

Table 3. The pruning rate of BP-3D-LRT

	$4 \times 4 \times 4$	$8 \times 8 \times 8$	$16 \times 16 \times 16$	$32 \times 32 \times 32$	$64 \times 64 \times 64$
BP-DP-3D-LRT	99.9268 %	99.9958 %	99.9996 %	100 %	100 %

3. Test of performance.

(a) When the size of the 3D Grid is like $n \times n \times n$, the performance comparison of NV-3D-LRT, BP-3D-LRT and MT-PRLL-3D-LRT is shown in Fig. 5(a).

(b) Test the performance of MT-PRLL-3D-LRT. When n is 1000, 10000, 100000 and the number of threads is 2, 4, 8, 16, 32, 64, the performance of MT-PRLL-3D-LRT is shown in Fig. 5(b). It can be seen that the running time decreases dramatically as the number of threads increases.
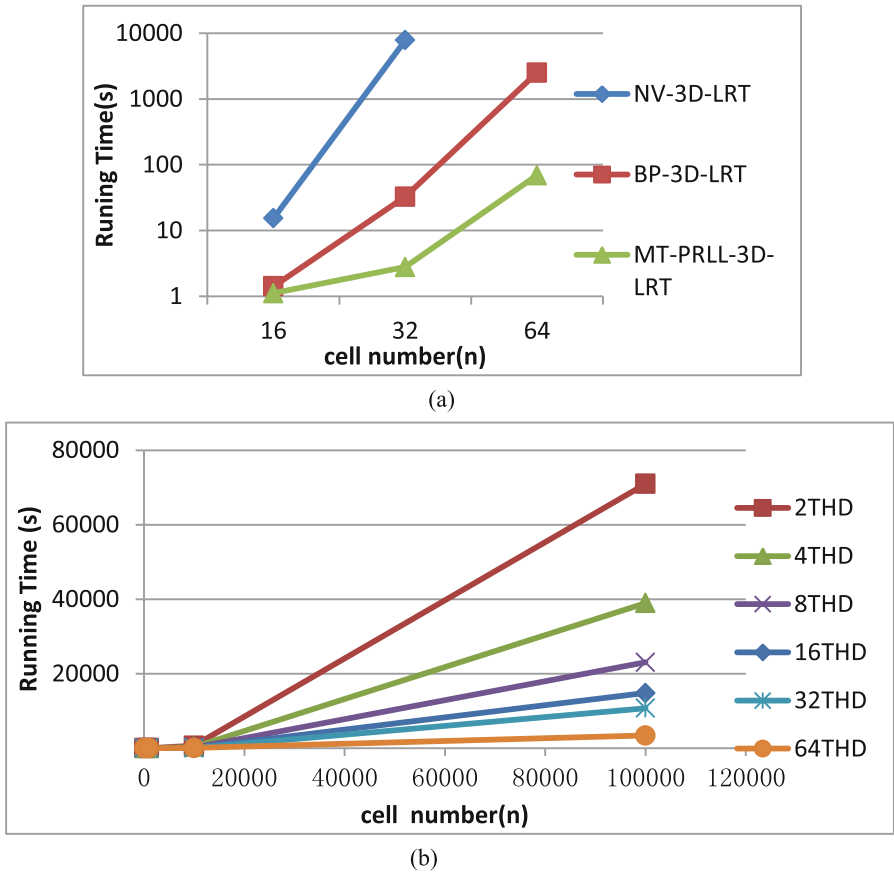
(a)

(b)

Fig. 5. Performance of 3D-LRT methods

Discussion

1. **Accuracy.** With an overall confidence score of 0.05, our 3D-LRT did not make any false alarms in tests where the null hypothesis held. The occurrence of false alarms may result from our application of a conservative Poisson distribution. Furthermore, in both the "subtly anomalous" and "extremely anomalous" cases, our 3D-LRT had 100 % detection accuracy.

2. **Pruning rate.** In general, the results show that our pruning methods have good effect. The larger the data size is, the more effective our pruning methods are, because under such circumstance most of the regions are normal. When the number of cells is close to *32 × 32 × 32*, the pruning rate almost reached 100 %.

3. **Performance.** The running time of NV-3D-LRT is up to two hours when tested on *32 × 32 × 32* 3D grid. The performance of NV-3D-LRT declines rapidly as n increase. The performance of two optimized methods DP-3D-LRT and MT-PRLL-3D-LRT improves significantly compared with NV-3D-LRT. By comparing the

performance of BP-3D-LRT and MT-PRLL-3D-LRT, we can see that on $n \times n \times n$ 3D grid, MT-PRLL-3D-LRT is 3 to 20 times faster than BP-3D-LRT.

These results show that the performance gain of MT-PRLL-3D-LRT is very impressive. The performance of MT-PRLL-3D-LRT is proportional to the number of threads when the number of threads is within the range from 2 to 256, which is because MT-PRLL-3D-LRT has very little data competition. MT-PRLL-3D-LRT is suitable for small and medium data sets and its output is identical to that of DP-3D-LRT.

Above all, our methods of 3D-LRT are effective and efficient.

6   Conclusions

In this paper, we divide the power system with time dimension into an $n \times n \times t$ 3D grid and put forward some methods to solve the problem of spatiotemporal anomaly detection. We propose the 3D-LRT and its optimization methods: the pruning strategies can minimize computation times of handling sub-regions; the multi-threaded parallelization of 3D-LRT can be used to deal with massive spatiotemporal data. Experiments on synthetic data and real-life applications show that our methods are tenable and effective.

References

1. Keegan, K.: Power system fault detection using the discrete wavelet transform and artificial neural networks (2014)
2. Wu, M., Song, X., Jermaine, C., Ranka, S., Gums, J.: A LRT framework for fast spatial anomaly detection. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD 2009), pp. 887–896 (2009)
3. Reaz, M., Choong, F., Sulaiman, M., Yasin, F., Kamada, M.: Expert system for power quality disturbance classifier. IEEE Trans. Power Deliv. **22**(3), 1979–1988 (2007)
4. Moghavvemi, M., Yang, S.S.: ANN application techniques for power system stability estimation. Electr. Mach. Power Syst. **28**(2), 167–178 (2000)
5. Senjyu, T., Shingaki, T., Uezato, K.: A novel high efficiency drive strategy for synchronous reluctance motors considering stator iron loss in transient conditions. In: Proceedings of the IEEE 32nd Annual Power Electronics Specialists Conference, PESC, vol. 3, pp. 1689–1694 (2001)
6. Sagiroglu, S., Colak, I., Bayindir, R.: Power factor correction technique based on artificial neural networks. Energy Convers. Manag. **47**(18–19), 3204–3215 (2006)
7. Murata, T.: Petri nets: properties, analysis and applications. Proc. IEEE **77**(4), 541–580 (1989)
8. Lo, K.L., Ng, H.S., Trecat, J.: Power System fault diagnosis using Petri nets. IEE Proc. Gener. Transm. Distrib. **144**(3), 231–236 (1997)
9. Khamphanchai, W., Pipattanasomporn, M, Rahman, S.: A multi-agent system for restoration of an electric power distribution network with local generation. In: 2012 IEEE Power and Energy Society General Meeting, pp. 1–8. IEEE (2012)

10. Chandola, V., Banerjee, A., Kumar, V.: Anomaly detection: a survey. ACM Comput. Surv. **41**, 1–58 (2009)
11. Ng, R., Clarans, J.H.: A method for clustering objects for spatial data mining. IEEE Trans. Knowl. Data Eng. **14**, 1003–1016 (2002)
12. Neill, D.B., Moore, A.W., Sabhnani, M., Daniel, K.: Detection of emerging space time clusters. In: Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining (KDD 2005), pp. 218–227 (2005)
13. Jung, I., Kulldorff, M., Richard, O.: A spatial scan statistic for multinomial data. Stat. Med. **29**(18), 1910–1918 (2010)
14. http://www.satscan.org (2008)

Computer Assisted Chinese Birthday Couplets Generation

Shunting Wang, Shicheng Zhang, and Zhigeng Pan[✉]

DMI Research Center, Hangzhou Normal University, Hangzhou, China
zgpan@cad.zju.edu.cn

Abstract. We propose a method of generating birthday couplets fitting to gender, age and birthday time. The work of this paper are as follows: Using memory mapping file to load large number of files which can significantly reduce the initializing time; Using the TFIDF method to train key words related age and gender; Using a dynamic Bi-gram graph model to generate the more semantically coherent left roll which can conversely alter the weights between words of the model.

Keywords: Name-embedded birthday couplet · TFIDF feature extraction · Memory mapped files · Dynamic Bi-gram graph model

1 Introduction

Chinese culture has a long and vast history of antithetical couplet of which birthday couplet is a wonderful work. Birthday couple [1] is a communicative couplet which is used to celebrate for the birthday. We create birthday couplet according to people's gender, age, identity, status class, experience, moral, behavior and so on. We aim to make them more appropriate and have their own characteristics.

The birthday couplets generated by this system can not only fit the main features of birthday couplets, but can embed our names in birthday couplet. The name-embedded sentences must have the same pattern, the same number of words and the relative meaning, apart from this, their rhyme must be harmonious. For birthday couplet, the sentences should not only highlight "happiness and longevity", but be appropriate and fit people's gender, age, season and so on. Although creating birthday couplet seems easy, it needs meet not only the requirements of the general couplet, but the specific requirements of name-embedded couplet and birthday couplet which increase the difficulties to create birthday couplet. So this system aims to generate name-embedded birthday couplet automatically to provide a reference for the creation of birthday couplet.

At present, a great number of couplet generation system only generate the left roll according to the given right roll, for example, Kaixu Zhang and Maosong Sun's ancient prose and couplets generation system based on statistics and rules [2], Yabin Zheng, Jiawei Cao and Zhiyuan Liu's couplet generation system based on maximum matching and hidden Markov model [3] and Microsoft Asia Research Institute's computer couplet [4]: generating the left roll according to the given right roll which is limited in ten Chinese characters. Similarly, Yue Fei in Institute of Automation of, Chinese Academy of Sciences [5] use neural network method to generate his chunlian generation system

which sentences are less than six words. Yong Yi in Chongqing University [6] consider the process of couplets generation as a supervised machine learning modeling which provides us a new idea and reference. He builds several models learning from error rate drive [7], a hidden Markov model (HMM) and probabilistic language model [8] and has a good effect. Jing Guo in Hangzhou Normal University [9] put forward the name-embedded couplets generation system for the first time in her master's thesis, but it takes a long time to load the couplet library and doesn't include other features of people except their name. This paper applies memory-mapping file method to solve the problem of loading large scale corpus. And the system in this paper can generate birthday couplets including people's name and further fitting age, gender, season and other features. This paper trains the initial corpus to extract keywords of different features and generate and evaluate the generated birthday couplets from the features.

This paper is organized as follows: Sect. 2 introduces the main technologies used to generate name-embedded birthday couplet, including the technology to establish the birthday couplet library, the technology to load a great deal of birthday couplets: memory-mapping file technology [10] and dynamic binary word graph technology to generate the left roll of name-embedded birthday couplet. Section 3 is mainly about our name-embedded birthday couplet system design: the process of birthday couplet generation according to the particularity of birthday couplet. Section 4 introduces the implementation of our couplet generation system and list some birthday couplets generated by our system.

2   Birthday Couplet Technologies

2.1   Building Birthday Couplets Library

Building birthday couplets is the cornerstone for automatic generation couplets system, while learning the rules of the couplets is our focus. In Ruiying Jiang's master thesis [11], he first extracts word list based on statistical Chinese word list extraction, after that he segments sentences into words using Forward and Backward Maximum Matching Combined with Naive Bayesian Model. After manually marking the segmented 20000 word and using hidden Markov model to learn the nominal of words, he sets up an automatic nominal tagging tool. At last he adds rhyme library to judge level and oblique corresponding in the tones. Besides, this paper use the algorithm TFIDF [12–14] to extract the key words of people's age, gender and season and other features as the extraction constraint.

The TFIDF algorithm gives each segmented word a weight to weigh the importance of describing the document based on the documents classified by age, gender, season and other features. It often need to consider three factors: term frequency tf indicating the frequency of the word appearing in the document; inverse document frequency idf indicating the distribution of the words in different category of birthday couplets which formulas is $\log(N/nk + 0.01)$ and N indicating the number of the documents, nk indicating the number of documents including the word; normalization factor, standardizing each component. We can get the formula for the weight:

$$w_{ik} = \frac{tf_{ik} \log \left(N / n_k + 0.01 \right)}{\sqrt{\sum_{k=1}^{t} \left(tf_{ik} \right)^2 \left[\log \left(N / nk + 0.01 \right) \right]^2}}$$

Through the formula for the weight, we can get key word for each feature for further evaluation.

2.2 Memory-Mapped Files

Win32 API and MFC have provided functions and classes supporting for file processing, but these are generally used to handle files within the 2 GB and are helpless for dozens of GB or even bigger files. For such large files, memory-mapping files is generally used.

Memory-mapping files maps a file to a block of memory. Through memory-mapping file, we can retain a region of an address and will submit the physical memory to this region. The physical memory of memory-mapping files is from one file already exists on the disk and you must first map this file before you use the file. So application can access the files on the disk through the memory pointer like accessing the memory having loaded files. So when we want to access files on the disk, we will no longer have to perform I/O operations on the files. Memory-mapping files play an important role in processing large scale files.

Hu et al. [15] applies memory-mapping files to process remote sensing images which capacity increases exponentially. And in this paper, memory-mapping files is applied to process about 10 GB corpus which can significantly reduce the time loading the corpus. When not using memory-mapping files, time of loading system is 42 s and using memory-mapping files, time of loading system is 18 s on the same machine with a condition of the same period.

2.3 Dynamic Binary Word Graph

Guo [9] used binary word graph to generate the right roll of the couplet. She establishes the binary word graph on the corpus through a single sentence. As shown in Fig. 1, there is an edge from the node "文明" to the node "盛世" which weight is the score calculated by binary language model to indicate the common degree of the two word nodes, while the direction indicates the order of the two word nodes. The reason why the weight is negative is that we take natural logarithm of the language model probability. But the dynamic binary word graph we used can dynamically adjusts the weight of the edge. For every new generated rolls, we first segment the words from the rolls and then adjust the weight of the edge according to the new frequency of the words from the rolls so that each iteration can ensure the common degree between the words is always the latest. Figure 1 is part of the binary word graph when you input the word "盛" while Fig. 2 is the new dynamic binary word graph.
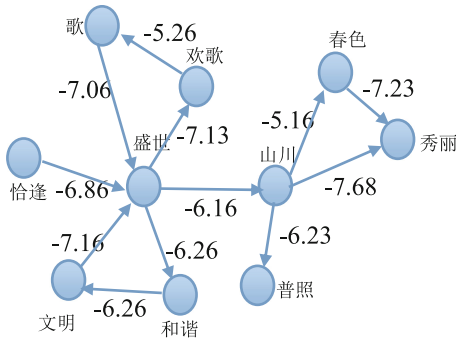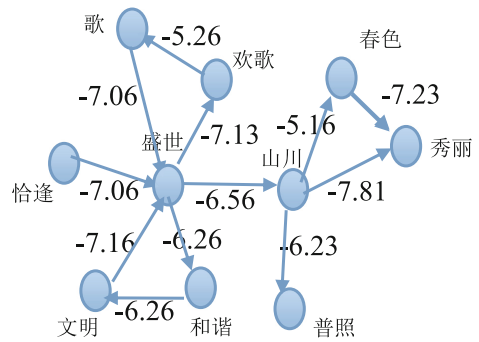
Fig. 1. Binary word graph

Fig. 2. Dynamic binary word graph

3 Name-Embedded Birthday Couplet System Design

The problem we must solve first is to load large scale corpus. Before solving this problem, we need load all the 10 GB corpus into the memory and it takes 42 s to initialize the system which brings bad user experience. This system solve the problem through memory-mapping files which don't need to load all the corpus to the memory. All we need to do is open up a region in memory to map the file to the region so that we can access the disk file directly avoiding loading them into memory. It can save time for loading files to memory and the I/O operation.

Due to the strong connection to age and gender, the generation of birthday couplet is significantly higher than the general couplet. Both the male and female birthday couplet have its representative words, for example, we usually use the word "大椿" or "松柏" to symbol the male and express the meaning by using chess, wine, poetry and other similar words. For the female birthday couplet, we usually use "母寿", "宝婆", "王母" and other words including female. The couple birthday couplet is to celebrate both the husband and wife birthday, it usually contains "共", "并", "双", "齐" and other words having double meaning. These words are appropriate and poetic used to congratulate the couple's birthday. Birthday couplet for different age often uses different words, for example, we can use the word "豆蔻年华" on a thirteen years old girl, use "半老徐娘" on a thirty years old woman, use "弱冠" on a twenty years old boy, use "而立" on a thirty years old man, use "不惑" on a forty years old man, use "知命" on a fifty years old man, use "花甲" on a sixty years old man and so on. Every age has its particular keywords, so you need to select the gender and age while generate the birthday couplet.

The birthday couplet we generated can not only fit gender and age, our system can also embed name. If the name has two words, we embed one word in the right roll and the other in the left roll. If the name has three words, we provides two ways: embedding two words of last name in the right roll and the left roll; embedding two words of the name in the right roll and embedding the other word in the left roll. If the name has four words, we also provides two ways: embedding the two words of last name in the right roll and the left roll; embedding one word of first name and one word of last name in the right roll and embedding the other word of first name and last name in the left roll.

For the couple birthday couplet, we need to have two names and embed the first name in the right roll and the other in the left roll. At last, we evaluate the rolls of couplet and select the best of them.

We add a few different kind of birthday couplet picture template to the existing couplet picture template due to the special meaning of birthday couplet. There are template of Chinese zodiac signs, template of "fu" and template of "Shou".

Both the birthday couplet and the couple birthday couplet can use idioms to generate. We collect the idioms library and get rid of negative words.

The generating candidate rolls need to be evaluated. The system apply statistical language model, statistics based grammar model and mutual information model to evaluate and select the top N couplets from the candidate couplets. The generation process of birthday couplets is shown in Fig. 3.
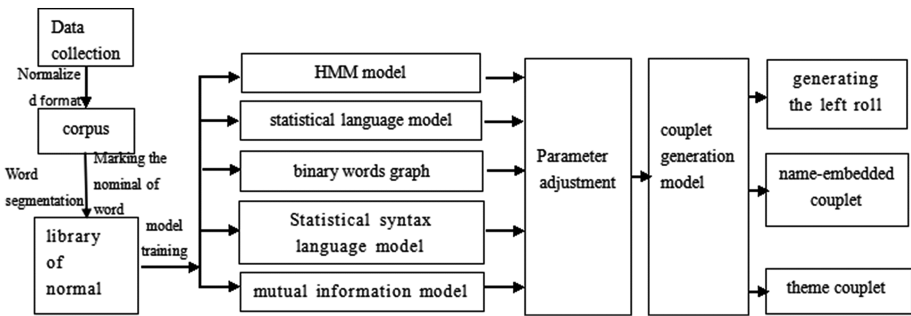

Fig. 3. Birthday couplet's generation process

4   Birthday Couplet Generation Model

In this paper, we mainly discuss two kinds of personalized birthday couplet: the customization of embedding single name fitting people's gender and age and the customization of embedding two names called couplet birthday couplet.

4.1   Generation for the Right Roll

The technique of dynamic binary word graph is based on the language model, that is, the appearance of a word is related to the N words in front of it. We build directed graph of words according to every segmented sentences, so that we can use it generate the short sentence. The process of generation as follows: (1) First search the directed words graph for the word nodes including the inputting embedded word; (2) Expand the word node forward and backward by the breadth-first searching strategy to stitch to short sentences; (3) Grade the sentences according to their language model score and get the top N sentence candidates; (4) If we have several words to be embedded, then splicing their short sentences in the word orders to form the candidate couplets; (5) Grade the candidate sentences according to their language model score and get the top ten sentences as the candidate right roll; (6) Segment every generation of the sentence,

rebuild the words graph. If the word has been in the words graph, then we only need to change its weight, otherwise we need to create the word node. For example, the embedded name is "王盛波", we need to first select the embedded word "盛" and then find the corresponding nodes including it from the word graph. Every word has its own corresponding word graph. Part of the binary word graph of the word "盛世" are shown in Fig. 2.

If the embedded word is "盛", we stitch the words searching from word graph to short sentences and part of the dynamic binary word graph is shown in Fig. 2. We will find many short sentences like this including the word "盛" and then we grade these candidate sentences according to their language model score and get the top N short sentences. If you need embed two words, you can stitch different short sentences of different word according to their order. At last, we also need to grade the candidate sentences and get the top ten sentences as the candidate right roll.

We also use Viterbi algorithm with embedded word to generate the left roll as the method used in Jing Guo's master thesis.

Every time when generated the couplet, we will segment the sentences of couplet to recalculate the word's frequency generate and also the edge's weight of binary word graph can be recalculated so that we can use the words with the highest frequency to generate sentences of couplet.

4.2 Generation for Couple Birthday Couplet

For couple couplet, we need to embed the first people's name in the right roll and embed the second one's name in the left roll. We also use dynamic binary word graph to generate short sentences and stitch them to form the right roll. After generating the left roll using Viterbi algorithm with embedded word, we need to evaluate the candidate right rolls and select the top N rolls. For the results, we also need segment them into words and recalculate the word frequency to update the edge's weight of dynamic binary word graph.

4.3 Generation and Storage for Couplet Pictures

After generating birthday couplets, you can choose one birthday couplet to generate kinds of couplet pictures based on different birthday picture template. It can dynamically adjust the size and position of words in the birthday couplet picture according to the couplet length. As shown in Fig. 4, the inputting name is "金星" and choose one of the generation couplets to generate couplet picture. As you can see from Fig. 4, we choose one template of "shou" and one template of "fu" and one template of Chinese zodiac signs. This system can not only generate different couplet pictures, but can also save the couplet pictures to the computer. You can share them to your friends or you can print them and present to your friends.

Fig. 4. Generation and storage for couplet pictures

5   Experimental Results and Conclusions

This section mainly introduces the experimental results of the system from three aspects: loading time of the system, the generation results with different gender and same name and the generation results with different age and same name.

Fig. 5. Generation results with different gender

The system has 10 GB, but only 4 GB are commonly used. Before using memory-mapping files, splash Screen starts from 60 s and displays loading time required every one second. After displaying "20 s required", the system displays "initialization comple-tion", that is, the system's loading time is 42 s. But after using memory-mapping files, the system's loading time is 18 s.

This system uses TFIDF algorithm to extract keywords for gender and age so that the birthday couplets have obvious characteristics of gender and age. In Fig. 5, the system generates different birthday couplets for different gender with the same name "林心如". As shown in Fig. 6, there will be different results for different ages. In Fig. 7, the system can generate good results for couple couplet with the name "潘志庚" and the name "张明敏".

Fig. 6. Generation results with different age

Fig. 7. Generation of couple couplet

This paper learn knowledge from a large scale corpus with a statistical language model, dynamic binary word graph, hidden Markov model. This system can use these

models to generate birthday couplet fitting to gender and age. Besides, we can generate good couple couplet either.

References

1. Gu, X.: The study of Chinese couplet. J. Peking Univ. **35**, 126–135 (1998)
2. Zhang, K., Sun, M.: An Chinese couplet generation model based on statistics and rules. Chin. Inf. Newspaper **23**, 100–105 (2009)
3. Zheng, Y., Cao, J., Liu, Z.: The paper collection of SCWL 2008, Shanxi (2008)
4. Zhou, M.: Microsoft couplet generation system [EB/OL]. Natural Language Computing (NLC) Group of Microsoft Research Asia (2006). http://duilian.msra.cn
5. Fei, Y.: Research on multi level integration of Chinese semantics–the design of chinese couplets art. Institute of Automation of Chinese Academy of Sciences, Beijing (1999)
6. Yi, Y.: A study on style identification and chinese couplet responses oriented computer aided poetry composing. Chongqing University, Chongqing (2005)
7. Bril, E.: Automatic grammar induction and parsing free text: a transformation-based approach. In: Proceedings of the ARPA Workshop on Human Language Technology, pp. 259–264 (1993)
8. Zhai, C.: Statistical language models for information retrieval a critical review. Found. Trends Inf. Retrieval **2**(3), 137–213 (2008)
9. Guo, J.: Study on generation of computer assisted customized couplet. Hangzhou Normal University, Hangzhou (2013)
10. Memory-Mapped Files [EB/OL]. MSDN. http://msdn.microsoft.com/en-us/library/dd997372.aspx
11. Jiang, R.: Computer assisted Chinese poetry and couplet generation. Zhejiang University, Hangzhou (2012)
12. Salton, G., McGill, M.J.: Introduction to Modern Information Retrieval. McGraw-Hill, New York (1983)
13. Salton, G., Fox, E.A., Wu, H.: Extended Boolean information retrieval. Commun. ACM **26**(11), 1022–1036 (1983)
14. Salton, G., Buckley, C.: Term-weighting approaches in automatic text retrieval. Inf. Process. Manage. **24**, 513–523 (1988)
15. Weizhong, H., Liu, N., Liu, R.: Quick-read means for large scale images based on storage mapping file technique. Appl. Res. Comput. **2**, 111–112 (2005)

Author Index