Andrzej Dziech
Andrzej Czyżewski (Eds.)

Communications in Computer and Information Science 149

Multimedia Communications, Services and Security

4th International Conference, MCSS 2011
Krakow, Poland, June 2011
Proceedings

Springer

Communications
in Computer and Information Science 149

Andrzej Dziech Andrzej Czyżewski (Eds.)

Multimedia Communications, Services and Security

4th International Conference, MCSS 2011
Krakow, Poland, June 2-3, 2011
Proceedings

Springer

Volume Editors

Andrzej Dziech
AGH University of Science and Technology
Department of Telecommunications
al. Mickiewicza 30
30-059 Krakow, Poland
E-mail: adzie@tlen.pl

Andrzej Czyżewski
Gdansk University of Technology
Multimedia Systems Department
Narutowicza 11/22
80-233 Gdansk, Poland
E-mail: indect@sound.eti.pg.gda.pl

ISSN 1865-0929 e-ISSN 1865-0937
ISBN 978-3-642-21511-7 e-ISBN 978-3-642-21512-4
DOI 10.1007/978-3-642-21512-4
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: Applied for

CR Subject Classification (1998): C.2, H.4, I.2, E.3, H.3, I.4, D.2

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

In recent years, multimedia communications, services and security have been contributing extensively to our life experience and are expected to be among the most important applications in the future. The objective of the Multimedia Communications, Services and Security Conference (MCSS 2011) is to present current research and developing activities contributing to theoretical and experimental aspects of multimedia systems, information processing and their applications for security research.

The conference objectives are in conformance with selected objectives of the European Union Seventh Framework Programme (FP7) Security Research. The main objectives of FP7 Security Research are as follows:

- To develop technologies for building capabilities needed to ensure the security of citizens from threats (terrorism, natural disasters, crime), while respecting human privacy
- To ensure optimal use of technologies to the benefit of civil European security
- To stimulate the cooperation for civil security solutions
- To improve the competitiveness of the European security industry
- To deliver mission-oriented research results to reduce security gaps

A parallel poster/demonstration session entitled "Demonstrations of Multimedia Communications, Services and Security" was organized along with the main conference. This event encouraged researchers to present and discuss the work-in-progress or experience-in-practice of their current studies/implementations and technology demonstrators covering the topics of MCSS 2011.

A session on intelligent monitoring, object tracking, and threat detection was also organized, as well as a special exhibition, where research groups had the opportunity to present their innovative achievements and prototypes in small booths or stands.

I firmly believe that this conference stimulated knowledge exchange among specialists involved in multimedia and security research.

June 2011 Andrzej Dziech

Organization

The International Conference on Multimedia Communications, Services and Security (MCSS 2011) was organized by the AGH University of Science and Technology within the scope of and under the auspices of the INDECT project.

Executive Committee

General Chair:	Andrzej Dziech AGH University of Science and Technology, Poland
Committee Chairs:	Andrzej Dziech AGH University of Science and Technology, Poland
	Andrzej Czyżewski Gdansk University of Technology, Poland

Scientific Committee

Emil Altimirski	Technical University of Sofia, Bulgaria
Fernando Boavida	University of Coimbra, Portugal
Ryszard Choras	University of Technology and Life Sciences, Poland
Marilia Curado	University of Coimbra, Portugal
Andrzej Czyżewski	Gdansk University of Technology, Poland
Andrzej Dziech	AGH University of Science and Technology, Poland
Marek Domański	Poznan University of Technology, Poland
Andrzej Duda	Grenoble Institute of Technology, France
Carolyn Ford	U.S. Department of Commerce, USA
Czesław Jędrzejek	Poznan University of Technology, Poland
Christian Kollmitzer	FHTW Wien, Austria
Bożena Kostek	Gdansk University of Technology, Poland
Zbigniew Kotulski	Warsaw University of Technology, Poland
Anton Kummert	University of Wuppertal, Germany
David Larrabeiti	Universidad Carlos III de Madrid, Spain
Antoni Ligęza	AGH University of Science and Technology, Poland
Józef Lubacz	Warsaw University of Technology, Poland
Wiesław Lubaszewski	AGH University of Science and Technology, Poland
Andreas Mauthe	Lancaster University, UK
Jaroslav Zdralek	Vysoka Skola Banska – Technicka Univerzita Ostrava, Czech Republic

Edward Nawarecki	AGH University of Science and Technology, Poland
Andrzej R. Pach	AGH University of Science and Technology, Poland
Zdzisław Papir	AGH University of Science and Technology, Poland
Thomas Plageman	University of Oslo, Norway
Władysław Skarbek	Warsaw University of Technology, Poland
Tomasz Szmuc	AGH University of Science and Technology, Poland
Ryszard Tadeusiewicz	AGH University of Science and Technology, Poland
Vladimir Vasinek	Vysoka Skola Banska – Technicka Univerzita Ostrava, Czech Republic
Jakob Wassermann	FHTW Wien, Austria
Jan Węglarz	Poznan University of Technology, Poland

Program Committee

Jacek Dańda	AGH University of Science and Technology, Poland
Jan Derkacz	AGH University of Science and Technology, Poland
Andrzej Głowacz	AGH University of Science and Technology, Poland
Nils Johanning	InnoTec Data, Germany
Jozef Juhar	Technical University of Kosice, Slovakia
Ioannis Klapatfis	University of York, UK
Mikołaj Leszczuk	AGH University of Science and Technology, Poland
Suresh Manandhar	University of York, UK
Tomasz Marciniak	Poznan University of Technology, Poland
Libor Michalek	Vysoka Skola Banska – Technicka Univerzita Ostrava, Czech Republic
Haluk Sarlan	PSI Transcom GmbH, Germany
Mikołaj Sobczak	Poznan University of Technology, Poland
Piotr Szczuko	Gdañsk University of Technology, Poland
Andrzej	Szwabe, Poznan University of Technology, Poland
Manuel Urueña	Universidad Carlos III de Madrid, Spain
Plamen Vichev	Technical University of Sofia, Bulgaria
Joerg Velten	University of Wuppertal, Germany

Eduardo Cerqueira	Universidade Federal do Pará, Brazil
Alexander Bekiarski	Technical University of Sofia , Bulgaria
Nitin Suresh	Georgia Institute of Technology, USA
Irena Stange	U.S. Department of Commerce, USA

Organizing Committee

Jacek Dańda	AGH University of Science and Technology, Poland
Jan Derkacz	AGH University of Science and Technology, Poland
Sabina Drzewicka	AGH University of Science and Technology, Poland
Andrzej Głowacz	AGH University of Science and Technology, Poland
Michał Grega	AGH University of Science and Technology, Poland
Piotr Guzik	AGH University of Science and Technology, Poland
Małgorzata Janiszewska	AGH University of Science and Technology, Poland
Paweł Korus	AGH University of Science and Technology, Poland
Mikołaj Leszczuk	AGH University of Science and Technology, Poland
Andrzej Matiolański	AGH University of Science and Technology, Poland
Piotr Romaniak	AGH University of Science and Technology, Poland
Szymon Szott	AGH University of Science and Technology, Poland

Sponsoring Institutions

- European Commission, Seventh Framework Programme (FP7)
- Institute of Electrical and Electronics Engineers (IEEE)
- Intelligent Information System Supporting Observation, Searching and Detection for Security of Citizens in Urban Environments (INDECT Project)
- AGH University of Science and Technology, Department of Telecommunications

Table of Contents

A High-Capacity Annotation Watermarking Scheme 1
Paweł Korus, Jarosław Białas, Piotr Olech, and Andrzej Dziech

Quality Assessment for a Licence Plate Recognition Task Based on a Video Streamed in Limited Networking Conditions 10
Mikołaj Leszczuk, Lucjan Janowski, Piotr Romaniak, Andrzej Głowacz, and Ryszard Mirek

Integrating Applications Developed for Heterogeneous Platforms: Building an Environment for Criminal Analysts 19
Arkadiusz Świerczek, Roman Dębski, Piotr Włodek, and Bartłomiej Śnieżyński

INACT — INDECT Advanced Image Cataloguing Tool 28
Michał Grega, Damian Bryk, Maciej Napora, and Marcin Gusta

Distributed Framework for Visual Event Detection in Parking Lot Area 37
Piotr Dalka, Grzegorz Szwoch, and Andrzej Ciarkowski

INCR — INDECT Multimedia Crawler 46
Michał Grega, Andrzej Głowacz, Wojciech Anzel, Seweryn Lach, and Jan Musiał

Detection and Localization of Selected Acoustic Events in 3D Acoustic Field for Smart Surveillance Applications 55
Józef Kotus, Kuba Łopatka, and Andrzej Czyżewski

Brightness Correction and Stereovision Impression Based Methods of Perceived Quality Improvement of CCTV Video Sequences 64
Julian Balcerek, Adam Konieczka, Adam Dąbrowski, Mateusz Stankiewicz, and Agnieszka Krzykowska

Semantic Structure Matching Recommendation Algorithm 73
Andrzej Szwabe, Arkadiusz Jachnik, Andrzej Figaj, and Michał Blinkiewicz

Hierarchical Estimation of Human Upper Body Based on 2D Observation Utilizing Evolutionary Programming and "Genetic Memory" 82
Piotr Szczuko

Assessing Task-Based Video Quality — A Journey from Subjective Psycho-Physical Experiments to Objective Quality Models 91
Mikołaj Leszczuk

Graph-Based Relation Mining 100
Ioannis P. Klapaftis, Suraj Pandey, and Suresh Manandhar

On Occlusion-Handling for People Detection Fusion in Multi-camera Networks 113
Anselm Haselhoff, Lars Hoehmann, Christian Nunn, Mirko Meuter, and Anton Kummert

An Informatics-Based Approach to Object Tracking for Distributed Live Video Computing 120
Alexander J. Aved, Kien A. Hua, and Varalakshmi Gurappa

M-JPEG Robust Video Watermarking Based on DPCM and Transform Coding 129
Jakob Wassermann

Assessing Quality of Experience for High Definition Video Streaming under Diverse Packet Loss Patterns 137
Lucjan Janowski, Piotr Romaniak, and Zdzisław Papir

LDA for Face Profile Detection 144
Krzysztof Rusek, Tomasz Orzechowski, and Andrzej Dziech

Performance Evaluation of the Parallel Codebook Algorithm for Background Subtraction in Video Stream 149
Grzegorz Szwoch

Automated Optimization of Object Detection Classifier Using Genetic Algorithm 158
Andrzej Matiolański and Piotr Guzik

Traffic Danger Ontology for Citizen Safety Web System 165
Jarosław Waliszko, Weronika T. Adrian, and Antoni Ligęza

A Multicriteria Model for Dynamic Route Planning 174
Wojciech Chmiel, Piotr Kadłuczka, and Sebastian Ernst

Extensible Web Crawler – Towards Multimedia Material Analysis 183
Wojciech Turek, Andrzej Opaliński, and Marek Kisiel-Dorohinicki

Acoustic Events Detection Using MFCC and MPEG-7 Descriptors 191
Eva Vozáriková, Jozef Juhár, and Anton Čižmár

Analysis of Particular Iris Recognition Stages 198
Tomasz Marciniak, Adam Dąbrowski, Agata Chmielewska, and Agnieszka Krzykowska

Analysis of Malware Network Activity 207
Gilles Berger-Sabbatel and Andrzej Duda

Software Implementation of New Symmetric Block Cipher 216
Jakub Dudek, Łukasz Machowski, Łukasz Romański, and Marcin Święty

Multicriteria Metadata Mechanisms for Fast and Reliable Searching of People Using Databases with Unreliable Records 225
Julian Balcerek and Paweł Pawłowski

Performance Measurements of Real Time Video Transmission from Car Patrol 233
Maciej Szczodrak, Andrzej Ciarkowski, Bartosz Głowacki, and Kamil Kacperski

Fast Face Localisation Using AdaBoost Algorithm and Identification with Matrix Decomposition Methods 242
Tomasz Marciniak, Szymon Drgas, and Damian Cetnarowicz

WSN-Based Fire Detection and Escape System with Multi-modal Feedback 251
Zahra Nauman, Sohaiba Iqbal, Majid Iqbal Khan, and Muhammad Tahir

Building Domain-Specific Architecture Framework for Critical Information Infrastructure 261
Norbert Rapacz, Piotr Pacyna, and Grzegorz Sowa

Prototypes of a Web System for Citizen Provided Information, Automatic Knowledge Extraction, Knowledge Management and GIS Integration 268
Antoni Ligęza, Weronika T. Adrian, Sebastian Ernst, Grzegorz J. Nalepa, Marcin Szpyrka, Michał Czapko, Paweł Grzesiak, and Marcin Krzych

Human Tracking in Multi-camera Visual Surveillance System 277
Piotr Marcinkowski, Adam Korzeniewski, and Andrzej Czyżewski

Quantum Cryptography Protocol Simulator 286
Marcin Niemiec, Łukasz Romański, and Marcin Święty

Human Re-identification System on Highly Parallel GPU and CPU Architectures 293
Sławomir Bąk, Krzysztof Kurowski, and Krystyna Napierała

Face Occurrence Verification Using Haar Cascades - Comparison of Two Approaches 301
Piotr Boryło, Andrzej Matiolański, and Tomasz M. Orzechowski

Implementation of the New Integration Model of Security and QoS for MANET to the OPNET 310
Ján Papaj, Anton Čižmár, and Ľubomír Dobož

One Approach of Using Key-Dependent S-BOXes in AES 317
Nikolai Stoianov

Scalability Study of Wireless Mesh Networks with Dynamic Stream Merging Capability 324
Jun Ye and Kien A. Hua

Overview of the Security Components of INDECT Project 331
Nikolai Stoianov, Manuel Urueña, Marcin Niemiec, Petr Machník, and Gema Maestro

Interactive VoiceXML Module into SIP-Based Warning Distribution System 338
Karel Tomala, Jan Rozhon, Filip Rezac, Jiri Vychodil, Miroslav Voznak, and Jaroslav Zdralek

A Solution for IPTV Services Using P2P 345
Le Bich Thuy and Yoshiyori Urano

Author Index 353

A High-Capacity Annotation Watermarking Scheme

Paweł Korus, Jarosław Białas, Piotr Olech, and Andrzej Dziech

AGH University of Science and Technology
al. Mickiewicza 30, 30-049 Kraków, Poland
pkorus@agh.edu.pl

Abstract. Annotation watermarking is a technique that allows to associate textual annotations with digital images in a format independent manner. The embedded annotations are robust against common image processing operations which makes it a convenient tool for knowledge transfer. In this paper, we present a new approach to annotation watermarking. Our scheme resembles a traditional packet network and adopts the fountain coding paradigm for encoding the watermark payload. In our study, we focus on high capacity annotations which is a challenging goal when robustness against popular image processing operations is required. The presented scheme is robust against cropping and lossy JPEG compression. The paper describes the principles of the proposed approach and presents the results of it's experimental evaluation. In particular, we assess the achievable capacity, robustness and the image quality impact.

Keywords: digital watermarking, annotations, watermark coding.

1 Introduction

Efficient searching and filtering of digital images requires easily accessible textual descriptions of their content. For this purpose annotations or tags are commonly used. These labels can be associated either with the whole images or with their selected fragments. Annotations often carry ground truth data for evaluation of object detection and recognition algorithms. Thus, carefully annotated image sets are highly valued in computer vision communities. In practical systems, annotations are often used to store the results of automatic object recognition.

One of the problems with traditional annotations is that they are stored separately from the images. Among the most common choices are plain text files and relational databases. As a result, this information is usually lost when transmitting the images without an explicit care for the annotations.

Annotation watermarking is a technique that allows to embed these descriptions into the images in a persistent and format-independent manner [2]. The annotations are tied to the images by means of imperceptible modifications of their appearance. The watermarks need to be robust against common image processing operations, e.g. a cropped fragment of the original image is expected to provide a valid description of that particular region.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 1–9, 2011.

Embedding annotations via watermarks is particularly common in medical applications [4,11]. The typical annotations include the identifiers of both the doctor and the patient as well as the most relevant extract of the medical history of the latter. The payload of such schemes is limited as one of the most important factors is the limited quality impact of the embedded information. In certain cases, it is required that the watermarking process should be fully reversible.

In this paper we present a new approach to annotation watermarking with the emphasis on the user data payload. We communicate the annotations through an image and the designed transmission architecture resembles a traditional packet network. We propose to adopt the fountain coding paradigm for the purpose of encoding the watermark's payload [8]. Our system allows for straightforward incorporation of content adaptivity and remains robust against cropping and lossy JPEG compression.

This paper is organized as follows. Section 2 describes the principles of the proposed approach. The results of experimental evaluation are shown in Section 3. Conclusions, supplemented with a comparison with existing schemes are presented Section 4.

2 Proposed Scheme

The proposed annotation scheme uses two independent watermarks. The first one allows for synchronization with the original block grid in case of cropping. The second one carries the payload of the annotations and the necessary headers. The general idea of the considered scheme is shown in Fig. 1.

The first step of the encoder is to divide the image into blocks. Successive steps of the algorithm require different block sizes, which imposes a three-level block hierarchy, described in more detail in Section 2.2.

The first of the watermarks is embedded in the spatial domain using the *additive spread-spectrum* technique [2]. A correlation detector implemented in the decoder recovers the shift between the original and the current block grid. The details of this synchronization procedure are presented in Section 2.2.

The next step is to perform the block-based forward Discrete Cosine Transform (DCT). This domain allows for straightforward selection of frequencies least affected by the JPEG compression. Due to high capacity requirements, we have used the Distortion Compensation Quantization Index Modulation technique for embedding the main watermark [1,3]. Each coefficient eligible for carrying the watermark is modified according to:

$$X^*_{i,j} = (1-\gamma)\cdot sign(X_{i,j})\cdot \Delta \cdot Q_m(\frac{|X_{i,j}|}{\Delta}) + \gamma \cdot X_{i,j} \tag{1}$$

where $X_{i,j}$ is the cover image coefficient, $X^*_{i,j}$ is the watermarked coefficient, Δ is the quantization step and γ is the distortion compensation parameter. $Q_m(\cdot)$ is a quantizer for message bit m, i.e. $Q_0(x) = 2\cdot \lfloor \frac{x}{2} + 0.5 \rfloor$ and $Q_1(x) = 2\cdot \lfloor \frac{x}{2} \rfloor + 1$.

For the purpose of generating the main watermark, the payload of each annotation is encoded by a fountain code [8] and supplemented with the necessary

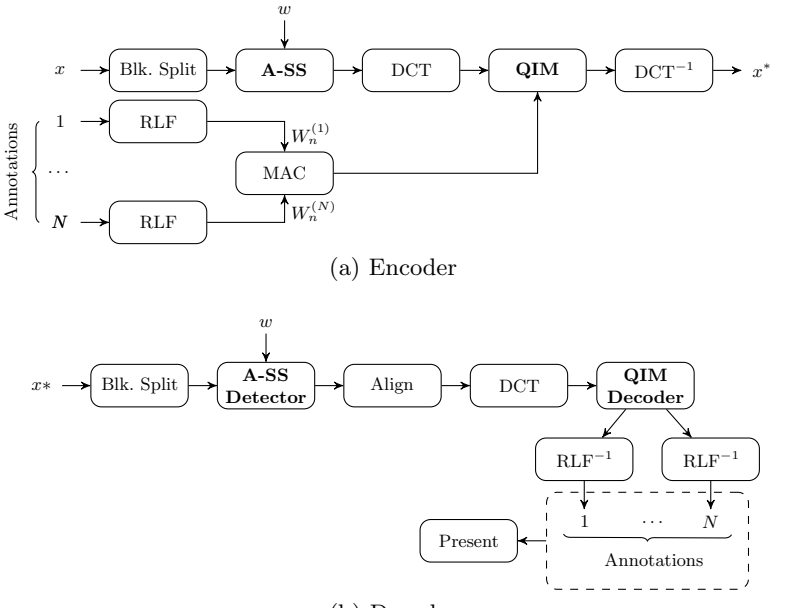


Fig. 1. Operation of the annotation encoder and decoder; x - cover image, $w \in \{-1, 1\}^{64x64}$ - synchronization watermark, additive spread spectrum, x^* - watermarked image

headers. The Medium Access Control (MAC) module assigns the available capacity and multiplexes the data streams from multiple annotations. In the last step, the modified spectrum is transformed back to the spatial domain using inverse DCT.

2.1 Annotation Transport Architecture

The principle of the proposed annotation watermarking scheme is to deliver a layered architecture analogous to traditional packet networks. The payload of each annotation is divided into constant length symbols, which are encoded in order to introduce the necessary redundancy. For this purpose, we propose to adopt the fountain coding paradigm [8]. The fundamental assumption of this code is that successful decoding is possible from arbitrary fragments of the symbol stream. The only requirements is that the decoder needs to receive a certain portion of the symbols. Let ρ denote the desired probability of decoding the message. Then, given the number of input symbols k, the number of symbols which are necessary for successful decoding is determined by (2).

$$s = k + \log_2(1/(1-\rho)) \tag{2}$$

A fountain code produces output symbols by calculating linear combinations of random input symbols. The code is capable of delivering a limitless symbol

stream. In the considered system, the output stream length is constant. We generate the output symbols for every macro block in the image. The MAC module decides which of them are actually going to be used. In practice, the unnecessary output symbols do not need to be calculated. The fountain code however, needs to explicitly generate null symbols in for the sake of the proper encoder-decoder synchronization.

The symbol length does not affect the performance of fountain codes. We have experimentally determined the symbol length for the application at hand by evaluating the impact of JPEG compression on the coefficients of the DCT spectrum. We use 60-bit symbols with additional 16-bit hash for error detection. Thus, the final watermark is divided into 76-bit symbols. The described transmission architecture is shown in Fig. 2.

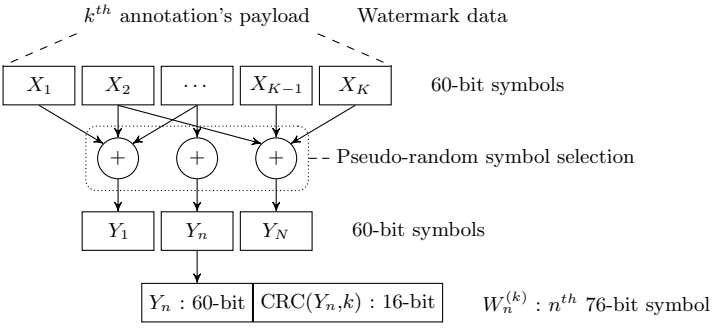


Fig. 2. Encoding and encapsulation of the payload of a single annotation (k^{th})

In this study, we have employed the Random Linear Fountain (RLF) code. Practical systems usually use the LT codes [7], which have more tractable computational complexity. For the purpose of presenting our concept, it is sufficient to base this study on the simpler RLF.

2.2 Block Hierarchy and Grid Synchronization

Each of the steps of the proposed scheme uses a different block size. The designed three-layer block hierarchy is shown in Fig. 3. The lowest-level division is based on the 8x8 px grid used by JPEG. Thus, we can directly assess the impact of lossy compression on individual frequencies of the spectrum.

The capacity of an individual lowest-layer block is insufficient for the discussed application. Thus, for the purpose of embedding the watermark symbols W_n, we group 4 lowest-layer blocks into 16x16 px macro blocks, which is capable of carrying one watermark symbol.

The highest layer of the designed hierarchy groups 16 macro blocks into 64x64 px synchronization blocks. The function of synchronization blocks is twofold. Firstly, they impose a strict organization of the macro blocks. The first macro block is referred to as the meta block and it encodes fundamental properties of

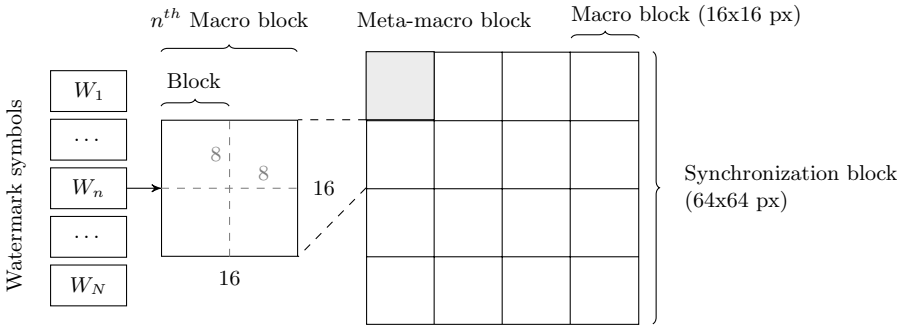


Fig. 3. The considered three-layer block hierarchy

the image. Specifically it contains the original dimensions of the annotated image, which are necessary for synchronization of the stream of watermark symbols and for translation of annotations' coordinates in case of cropping.

Secondly, the synchronization blocks are used during embedding of the auxiliary spread spectrum watermark. A uniform bipolar pseudo-random pattern $w \in \{-1, 1\}^{64\times64}$ is tiled to match the image size. We use the additive spread spectrum technique for embedding in the spatial domain (3), [2].

$$x^*_{i,j} = x_{i,j} + \alpha w_{i \ mod \ 64, j \ mod \ 64} \tag{3}$$

This auxiliary watermark allows for rapid resynchronization with the original blocking grid [6]. The detector calculates the correlation of an average synchronization block $\overline{x}$ with the known watermark pattern w. The location of the watermark detection peak corresponds to the grid misalignment vector. An exemplary detector output is shown in Fig. 4.

For the sake of computational efficiency, the decoder calculates the correlation in the Fourier domain. The magnitude of the spectrum is discarded to increase the detection performance [6]. The correlation matrix C is obtained using a coefficient-wise multiplication of the image and watermark spectra: $C = ft^{-1}(\Phi(ft(\overline{x}) \cdot \Phi(ft(w))$ where $ft(x)$ is the Fast Fourier Transform and $\Phi(x)$ is a magnitude discarding function [6]. This detector is equivalent to the Symmetric Phase Only Matched Filtering (SPOMF).

2.3 Medium Access Control

The MAC module is responsible for multiplexing data streams from multiple annotations, i.e. it assigns the macro blocks to the annotations. The embedded object descriptions need to be recoverable from cropped fragments of the original image. As a result, the MAC module attempts to allocate the surrounding macro blocks to each of the defined regions. The assignment is not known to the decoder, which attempts to validate the recovered watermark symbols for each of the annotations. The CRC of each symbol can be successfully validated only for

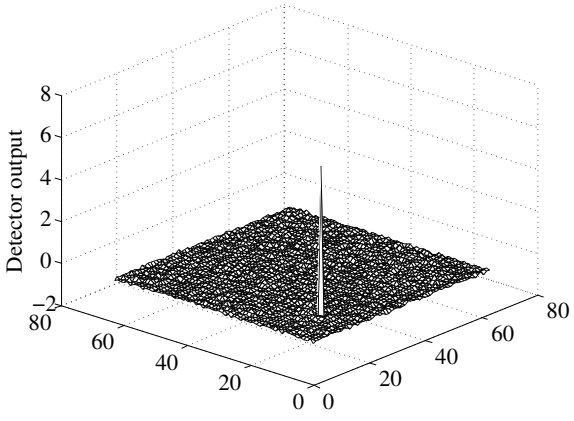


Fig. 4. An exemplary watermark detection result. The peak corresponds to the detected grid misalignment.

the correct annotation. Thus, the decoder quickly learns which macro blocks describe which stream and marks appropriate blocks as *erased* for the successive decoding step.

Analogously, the proposed system can be easily extended to support content adaptation. The utilized watermark embedding technique allows for straightforward incorporation of this feature by adapting the distortion compensation parameter γ. A Human Visual System model can be used to decide which macro blocks are more suitable for information embedding. This problem is one of the most challenging in digital information hiding [2]. In traditional approaches, the decoder needs to estimate the HVS model to recover the selection channel. This often leads to sub-optimal HVS models, which tend to be robust against prospective content modifications.

3 Experimental Evaluation

In this study, we focus on three main performance aspects of the proposed scheme. Firstly, we evaluate the quality impact of both of the watermarks. Secondly, we validate the robustness against lossy compression and cropping. We address this issue both from the grid synchronization and data retrieval perspectives. The last aspect is the capacity of the proposed scheme, i.e. the achievable payload of user data.

3.1 Quality Impact

The considered scheme uses two independent watermarks: a spatial domain auxiliary watermark and the main DCT domain watermark. The former is embedded using the additive spread spectrum technique (3) with constant embedding

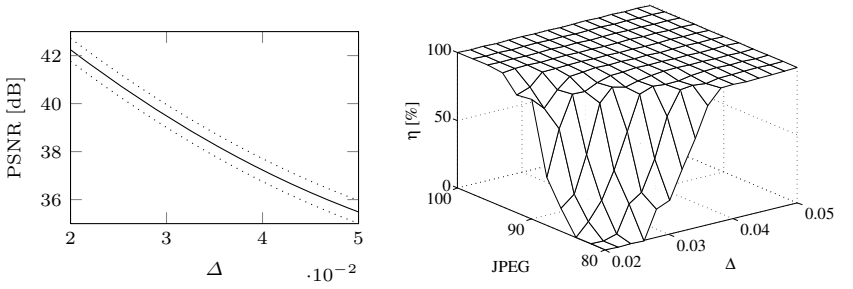


Fig. 5. (a) the joint quality impact for both of the embedded watermarks, (b) watermark recovery rate for different quantization steps and lossy compression settings

strength $\alpha = 1$ for all of the performed experiments. The average distortion introduced by the this watermark is 48 dB in terms of Peak Signal to Noise Ratio (PSNR).

The main watermark is embedded in the DCT domain (1). All of the macro blocks are used for watermark embedding. This simplifies the quality assessment as the only factor that influences the introduced distortion is the quantization step Δ, i.e. the strength of the watermark. Fig. 5 shows the average distortion vs. Δ along with 95% prediction intervals.

3.2 Robustness

We consider the robustness of the proposed scheme against cropping and lossy JPEG compression. Cropping causes grid misalignment and the decoder needs to be able to resynchronize prior to watermark recovery. The average success rate of $99.5 \pm 0.2\%$ with 95% confidence has been calculated from 7200 independent replications of the experiment on a test set of 90 images. The scenario included all possible grid misalignment vectors with step 2 and in combination with lossy compression from the range $[80; 100]$.

The most important performance criterion is the watermark recovery rate η. In this paper, we consider only flawlessly recovered symbols. Implementation of error correction mechanisms is expected to provide a considerable improvement. Due to a limited length of this paper, this issue is not discussed in more detail. The average recovery rate vs. the quantization step Δ and the JPEG quality setting is shown in Fig. 5b.

3.3 Capacity

The proposed scheme embeds 60 bits of watermark payload in each 256-bit macro block. Considering the overhead of necessary headers in the meta blocks, this results in the payload of 0.22 bpp. Not all of the capacity can be used to carry user data as the utilized fountain code introduces additional redundancy. The

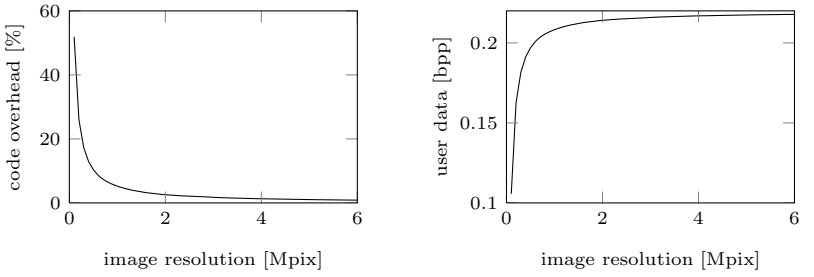


Fig. 6. The estimated pessimistic encoding overhead vs. image resolution

Table 1. A short comparison with existing annotation watermarking schemes

Ref.	Domain	Embedding	Robustness	PSNR [dB]	Payload [bpp]	ECC	Embedded Info.
[10]	DCT(ICA)	Custom	$\eta \approx 0.95$ for JPEG 50	45	0.004	N/A	Plain ACSII
[4]	Haar-DWT	Custom quantization	JPEG > 75	42	0.026	N/A	Medical annotations, robust ID
[9]	DFT	DDD	JPEG > 90, cropping	47	0.012	Reed-Solomon	Hierarchical annotations
[11]	Spatial	LSB	JPEG > 60	40	0.016	N/A	Medical annotations Authentication data
[5]	LBT	SS + regional statistical quantization	$\eta \approx 0.8$ for JPEG 10 $\eta \approx 1.0$ for JPEG > 50 cropping	46	32 bits = const	N/A	32-bit image ID
Our	DCT	QIM + A-SS	$\eta \approx 1.0$ for JPEG 80, cropping	38	0.22	RLF	Hierarchical annotations

annotations are encoded independently and the overhead of each individual annotation needs to be taken into account. The designed system allows to embed at most 32 annotations. From (2) for $\rho = 0.99$, the pessimistic final overhead equals 5.19% for a 1 Mpix image. This overhead drops asymptotically to 0 with the increase of the image size. Fig. 6 shows the behavior of the coding overhead and the resulting effective user data payload.

4 Conclusions

In this paper we have presented a novel annotation watermarking scheme. The scheme adopts the fountain coding paradigm to achieve straightforward multiplexing of independent data streams and overcome the problems with the non-shared selection channel. Our approach allows for straightforward incorporation of content adaptivity. A thorough evaluation of this property will be one of the main aspects of future research.

Our scheme is robust against cropping ans lossy JPEG compression while having significantly higher payload compared to existing systems at the cost of slightly lower image quality. We have distinguished a small set of parameters

that can be used to adjust the behavior and trade-off the robustness in case higher image quality is required. Table 1 contains a brief comparison of the main characteristics of existing annotation watermarking schemes.

Acknowledgment

The research leading to these results has received funding from the INDECT project funded by European Community's Seventh Framework Programme under grant agreement no. 218086 and from the European Regional Development Fund under INSIGMA project no. POIG.01.01.02-00-062/09. The latter has provided an implementation of the RLF and the SPOMF-based watermark detector.

References

1. Chen, B., Wornell, G.: Quantization index modulation: a class of provably good methods for digital watermarking and information embedding. IEEE Transactions on Information Theory 47(4), 1423–1443 (2001)
2. Cox, I.J.: Digital Watermarking and Steganography. Morgan Kaufmann Publishers, San Francisco (2008)
3. Eggers, J., Bauml, R., Tzschoppe, R., Girod, B.: Scalar Costa Scheme for Information Embedding. IEEE Transactions on Signal Processing 51(4), 1003–1019 (2003)
4. Giakoumaki, A., Pavlopoulos, S., Koutouris, D.: A medical image watermarking scheme based on wavelet transform. In: Proceedings of the 25th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 2003. , vol. 1, pp. 856–859 (2003)
5. He, S., Kirovski, D., Wu, M.: High-fidelity data embedding for image annotation. IEEE Transactions on Image Processing 18(2), 429–435 (2009)
6. Kalker, T., Depovere, G., Haitsma, J., Maes, M.: A video watermarking system for broadcast monitoring. In: SPIE: Security and Watermarking of Multimedia Contents, vol. 3657, pp. 103–112 (1999)
7. Luby, M.: Lt codes. In: Proc. 43rd Ann. IEEE Symp. on Foundations of Computer Science, vol. (16-19) (2002)
8. MacKay, D.: Fountain codes. In: IEE Proceedings Communication, vol. 152(6) (2005)
9. Schott, M., Dittmann, J., Vielhauer, C.: Annowano: An annotation watermarking framework. In: Proceedings of 6th International Symposium on Image and Signal Processing and Analysis, ISPA 2009, pp. 483–488 (2009)
10. Sun, G., Sun, H., Sun, X., Bai, S., Liu, J.: Combination independent content feature with watermarking annotation for medical image retrieval. In: Second International Conference on Innovative Computing, Information and Control, ICICIC 2007, p. 607 (2007)
11. Zain, J., Clarke, M.: Reversible watermarking surviving jpeg compression. In: 27th Annual International Conference of the Engineering in Medicine and Biology Society, IEEE-EMBS 2005, pp. 3759–3762 (2005)

Quality Assessment for a Licence Plate Recognition Task Based on a Video Streamed in Limited Networking Conditions

Mikołaj Leszczuk, Lucjan Janowski, Piotr Romaniak, Andrzej Głowacz, and Ryszard Mirek

AGH University of Science and Technology, Department of Telecommunications, al. Mickiewicza 30, PL-30059 Kraków, Poland
{leszczuk,janowski,romaniak,glowacz}@kt.agh.edu.pl, rysiekmirek@gmail.com
http://www.kt.agh.edu.pl/

Abstract. Video transmission and analysis is often utilised in applications outside of the entertainment sector, and generally speaking this class of video is used to perform a specific task. Examples of these applications are security and public safety. The Quality of Experience (QoE) concept for video content used for entertainment differs significantly from the QoE of surveillance video used for recognition tasks. This is because, in the latter case, the subjective satisfaction of the user depends on achieving a given functionality. Moreover, such sequences have to be compressed significantly because the monitored place has to be seen on-line and it can be connected by an error prone wireless connection. Recognising the growing importance of video in delivering a range of public safety services, we focused on developing critical quality thresholds in licence plate recognition tasks based on videos streamed in constrained networking conditions.

Keywords: quality of experience, content distribution, real time (live) content distribution.

1 Introduction

The transmission of video is often used for many applications outside of the entertainment sector, and generally this class of video is used to perform a specific task. Examples of these applications are security, public safety, remote command and control, and sign language. Monitoring of public urban areas (traffic, intersections, mass events, stations, airports, etc.) against safety threats using transmission of video content has became increasingly important because of a general increase in crime and acts of terrorism (attacks on the WTC and the public transportation system in London and Madrid). Nevertheless, it should be also mentioned that video surveillance is seen as an issue by numerous bodies aimed at protecting citizens against "permanent surveillance" in the Orwellian style. Among these, we should mention the Liberty Group (dedicated to human rights),

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 10–18, 2011.

an Open Europe organisation, the Electronic Frontier Foundation, and the Ethics Board of the FP7-SEC INDECT (Intelligent information system supporting observation, searching and detection for the security of citizens in an urban environment) [3]. This matter was also one of the main themes ("Citizens Security Needs Versus Citizens Integrity") of the Fourth Security Research Conference organised by the European Commission (September 2009) [5]. Despite this, many studies suggest that public opinion about CCTV is becoming more favourable [8]. This trend intensified after September 11, 2001. Furthermore, there do exist methods that partially protect privacy. They are based on the selective monitoring of figureheads, automatic erasing of faces/licence plates not related to the investigation or data hiding techniques using digital watermarking.

Anyone who has experienced artefacts or freezing play while watching an action movie on TV or live sporting event, knows the frustration accompanying sudden quality degradation of a key moment. However, for practitioners in the field of public safety the usage of video services with blurred images can entail much more severe consequences. The above-mentioned facts convince us that it is necessary to ensure adequate quality of the video. The term "adequate" quality means quality good enough to recognise objects such as faces or cars.

In this paper, recognising the growing importance of video in delivering a range of public safety services, we have attempted to develop critical quality thresholds in licence plate recognition tasks, based on video streamed in constrained networking conditions. The measures that we have been developing for this kind of task-based video provide specifications and standards that will assist users of task-based video to determine the technology that will successfully allow them to perform the function required.

Even if licence plate algorithms are well known, practitioners commonly require independent and stable evidence on accuracy of human or automatic recognition for a given circumstance. Re-utilisation of the concept of QoE (Quality of Experience) for video content used for entertainment is not an option as this QoE differs considerably from the QoE of surveillance video used for recognition tasks. This is because, in the latter case, the subjective satisfaction of the user depends on achieving a given functionality (event detection, object recognition). Additionally, the quality of surveillance video used by a human observer is considerably different from the objective video quality used in computer processing (Computer Vision). In the area of entertainment video, a great deal of research has been performed on the parameters of the contents that are the most effective for perceptual quality. These parameters form a framework in which predictors can be created, so objective measurements can be developed through the use of subjective testing. For task-based videos, we contribute to a different framework that must be created, appropriate to the function of the video — i.e. its use for recognition tasks, not entertainment. Our methods have been developed to measure the usefulness of degraded quality video, not its entertainment value. Assessment principles for task-based video quality are a relatively new field. Solutions developed so far have been limited mainly to optimising the network QoS parameters, which are an important factor for QoE of low quality video streaming

services (e.g. mobile streaming, Internet streaming) [11]. Usually, classical quality models, like the PSNR [4] or SSIM [13], were applied. Other approaches are just emerging [6,7,12].

The remainder of this paper is structured as follows. Section 2 presents a licence plate recognition test-plan. Section 3 describes source video sequences. In Section 4, we present the processing of video sequences and a Web test interface that was used for psycho-physical experiments. In Section 5, we analyse the results obtained. Section 6 draws conclusions and plans for further work, including standardisation.

2 Licence Plate Recognition Test-Plan

The purpose of the tests was to analyse the people's ability to recognise car registration numbers on video material recorded using a CCTV camera and compressed with the H.264/AVC codec. In order to perform the analysis, we carried out a subjective experiment. The intended outcome of this experiment was to gather the results of the human recognition capabilities. Non-expert testers rated video sequences influenced by different compression parameters. The video sequences used in the test were recorded at a parking lot using a CCTV camera. We adjusted the video compression parameters in order to cover the recognition ability threshold. We selected ITU's ACR (Absolute Category Rating, described in ITU-T P.910 [9]) as the applied subjective test methodology.

The recognition task was threefold: 1) type in the licence plate number, 2) select car colour, and 3) select car make. Testers were allowed to control playback and enter full screen mode. The experiment was performed using diverse display equipment in order to analyse the influence of display resolution on the recognition results.

We decided that each tester would score 32 video sequences. The idea was to show each source (SRC) sequence processed under different conditions (HRC) only once and then add two more sequences in order to find out whether testers would remember the registration numbers already viewed. The n-th tester was screened with two randomly selected sequences and 30 SRCs processed under the following HRCs:

$$\mathrm{HRC} = mod(n - 2 + \mathrm{SRC}, 30) + 1 \tag{1}$$

The tests were conducted using a web-based interface connected to a database. In the database both information about the video samples and the answers received from the testers were gathered. The interface is presented in Fig. 3.

3 Source Video Sequences and the Automatic Detection of Number Plates

Source video sequences were collected at AGH University of Science and Technology by filming a car parking lot during high traffic volume. In this scenario,

Fig. 1. Test interface

the camera was located 50 meters from the parking lot entrance in order to simulate typical video recordings. Using ten-fold optical zoom, 6m×3.5m field of view was obtained. The camera was placed statically without changing the zoom throughout the recording time, which reduced global movement to a minimum.

Acquisition of video sequences was conducted using a 2 mega-pixel camera with a CMOS sensor. The recorded material was stored on an SDHC memory card inside the camera.

All the video content collected in the camera was analysed and cut into 20 second shots including cars entering or leaving the car park. Statistically, licence plate was visible for average 17 seconds in each sequence. The length of the sequences was dictated mostly by the need to capture the vehicles not only when they were stopped by entrance barrier but in motion. The parameters of each source sequence are as follows:

- resolution: 1280×720 pixels (720p)
- frame rate: 25 frames/s
- average bit-rate: 5.6 - 10.0 Mb/s (depending on the local motion amount)
- video compression: H.264/AVC in Matroska Multimedia Container (MKV)

The owners of the vehicles filmed were asked for their written consent, which allowed the use of the video content for testing and publication purposes.

4 Processed Video Sequences

If picture quality is not acceptable, the question naturally arises of how it happens. The sources of potential problems are located in different parts of the end-to-end video delivery chain. The first group of distortions (1) can be introduced at the time of image acquisition. The most common problems are noise, lack of focus or improper exposure. Other distortions (2) appear as a result of further compression and processing. Problems can also arise when scaling video

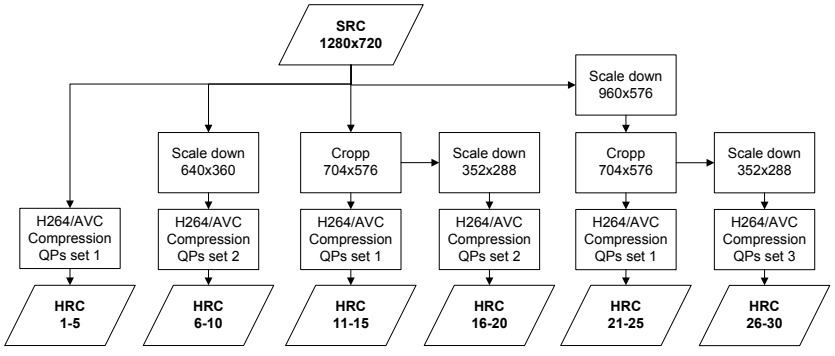


Fig. 2. Generation of HRCs

sequences in the quality, temporal and spatial domains, as well as, for example, the introduction of digital watermarks. Then (3), for transmission over the network, there may be some artefacts caused by packet loss. At the end of the transmission chain (4), problems may relate to the equipment used to present video sequences.

Considering this, all source video sequences (SRC) were encoded with a fixed quantisation parameter QP using the H.264/AVC video codec, x264 implementation. Prior to encoding, some modifications involving resolution change and crop were applied in order to obtain diverse aspect ratios between car plates and video size (see Fig. 2 for details related to processing). Each SRC was modified into 6 versions and each version was encoded with 5 different quantisation parameters (QP). Three sets of QPs were selected: 1) {43, 45, 47, 49, 51}, 2) {37, 39, 41, 43, 45}, and 3) {33, 35, 37, 39, 41}. Selected QP values were adjusted to different video processing paths in order to cover the plate number recognition ability threshold. Frame rates have been kept intact as, due to inter-frame coding, their deterioration does not necessarily result in bit-rates savings [10]. Furthermore, network streaming artefacts have been not considered as the authors believe in numerous cases they are related to excessive bit-streams, which had been already addressed by different QPs. Reliable video streaming solution should adjust video bit-stream according to the available network resources and prevent from packet loss. As a result, 30 different hypothetical reference circuits (HRC) were obtained. HRC is described by QP parameters nevertheless in the rest of the paper we use bit-rate as more network driven parameter. Note that the same QP for different views will result in different bit rates.

Based on the above parameters, it is easy to determine that the whole test set consists of 900 sequences (each SRC 1-30 encoded into each HRC 1-30).

5 Analysis of Results

Thirty non-expert testers participated in this study and provided a total of 960 answers. The subjects average age was 23 years old with maximum 29 and

minimum 17 years old. 11 female and 19 mails taken part in the study. We also asked about how well a subject know car marks, the obtained result shows that half of them know car marks. Each answer obtained can be interpreted as 1 or 0, i.e. correct or incorrect recognition. The goal of this analysis is to find the detection probability as a function of a certain parameter(s) i.e. the explanatory variables. The most obvious choice for the explanatory variable is bit-rate, which has two useful properties. The first property is a monotonically increasing amount of information, because higher bit-rates indicate that more information is being sent. The second advantage is that if a model predicts the needed bit-rate for a particular detection probability, it can be used to optimise the network utilisation.

Moreover, if the network link has limited bandwidth the detection probability as a function of a bit-rate computes the detection probability, what can be the key information which could be crucial for a practitioner to decide whether the system is sufficient or not.

The Detection Probability (DP) model should predict the DP i.e. the probability of obtaining 1 (correct recognition). In such cases, the correct model is logit [1]. The simplest logit model is given by the following equation:

$$p_d = \frac{1}{1 + \exp(a_0 + a_1 x)} \tag{2}$$

where x is an explanatory variable, a_0 and a_1 are the model parameters, and p_d is detection probability.

The logit model can be more complicated; we can add more explanatory variables, which may be either categorical or numerical. Nevertheless, the first model tested was the simplest one.

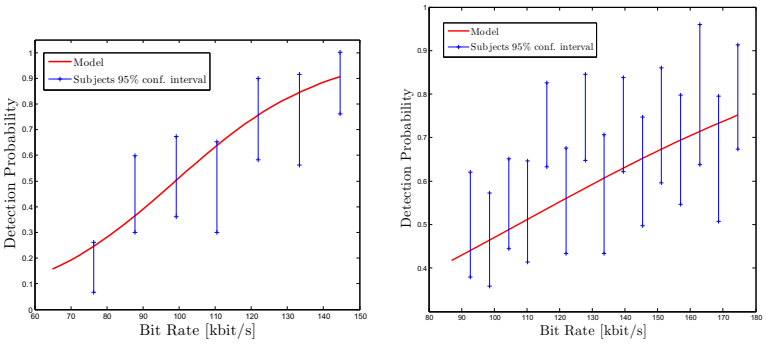
Building a detection probability model for all of the data is difficult, and so we considered a simpler case based on the HRCs groups (see section 4). Each five HRCs (1-5, 6-10, etc.) can be used to estimate the threshold for a particular HRCs group. For example, in Fig. 3(a) we show an example of the model and the results obtained for HRCs 20 to 25.

The obtained model it crosses all the confidence intervals for the observed bit-rates. The saturation levels on both sides of the plot are clearly visible. Such a model could successfully be used to investigate detection probability.

We present an extension of the sequences analysed to all HRCs results in the model drawn in Fig. 3(b).

The result obtained is less precise. Some of the points are strongly scattered (see results for bit-rate 110 to 130 kb/s). Moreover, comparing the models presented in Fig. 3(a) and Fig. 3(b) different conclusions can be drawn. For example, 150 kb/s results in around a 90% detection probability for HRCs 20 to 25 and less than 70% for all HRCs. It is therefore evident that the bit-rate itself cannot be used as the only explanatory variable. The question then is, what other explanatory variables can be used.

In Fig. 4(a) we show DP obtained for SRCs. The SRCs has a strong impact on the DP. It should be stressed that there is one SRC (number 26) which was not detected even once. The non-zero confidence interval comes from the corrected



(a) For HRC 20 to 25. (b) For all HRCs.

Fig. 3. Example of the logit model and the obtained detection probabilities

confidence interval computation explained in [2]. In contrast, SRC number 27 was almost always detected, i.e. even for very low bit-rates. A detailed investigation shows that the most important factors (in order of importance) are:

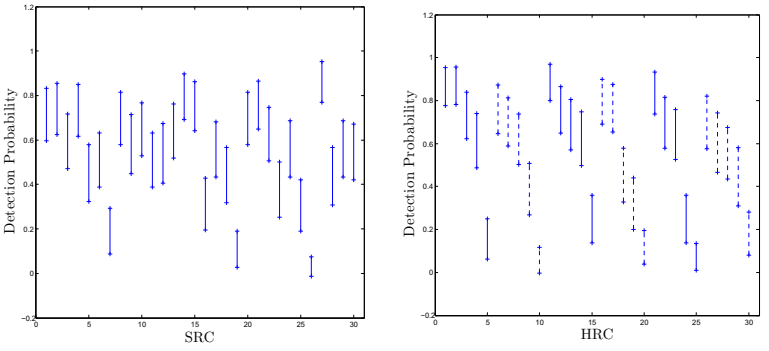
1. the contrast of the plate characters,
2. the characters, as some of them are more likely to be confused than others, as well as
3. the illumination, if part of the plate is illuminated by a strong light.

A better DP model has to include these factors. On the other hand, these factors cannot be fully controlled by the monitoring system, and therefore these parameters help to understand what kind of problems might influence DP in a working system. Factors which can be controlled are described by different HRCs. In Fig. 4(b) we show the DP obtained for different HRCs.

For each HRC, all SRCs were used, and therefore any differences observed in HRCs should be SRC independent. HRC behaviour is more stable because detection probability decreases for higher QP values. One interesting effect is the clear threshold in the DP. For all HRCs groups two consecutive HRCs for which the DPs are strongly different can be found. For example, HRC 4 and 5, HRC 17 and 18, and HRC 23 and 24. Another effect is that even for the same QP the detection probability obtained can be very different (for example HRC 4 and 24).

Different HRCs groups have different factors which can strongly influence the DP. The most important factors are differences in spatial and temporal activities and plate character size. The same scene (SRC) was cropped and/or re-sized resulting in a different output video sequence which had different spatial and temporal characteristics.

In order to build a precise DP model, differences resulting from SRCs and HRCs analysis have to be considered. In this experiment we found factors which influence the DP, but an insufficient number of different values for these factors



(a) For different SRCs with 90% confidence intervals

(b) For different HRCs. The solid lines corresponds to QP from 43 to 51, and the dashed lines corresponds to QP from 37 to 45.

Fig. 4. The detection probabilities obtained

was observed to build a correct model. Therefore, the lesson learned from this experiment is highly important and will help us to design better and more precise experiments in the future.

6 Conclusions and Further Work

The paper has answered the practical problem of a network link with a limited bandwidth and detection probability is an interesting parameter to find. We have presented the results of the development of critical quality thresholds in licence plate recognition by human subjects, based on a video streamed in constrained networking conditions. We have shown that, for a particular view, a model of detection probability based on bit rate can work well. Nevertheless, different views have very different effects on the results obtained. We have also learnt that for these kinds of psycho-physical experiments, licence plate characteristics (such as illumination) are of great importance, sometimes even prevailing over the distortions caused by bit-rate limitations and compression.

One important conclusion is that for a bit rate as low as 180 kbit/s the detection probability is over 80% even if the visual quality of the video is very low. Moreover, the detection probability depends strongly on the SRC (over all detection probability varies from 0 (sic!) to over 90%.)

The next research task will concern the development of thresholds for automatic detection tools, which will be useful in machine number plate recognition. Then, we anticipate the preparation of quality tests for other video surveillance scenarios.

The ultimate goal of this research is to create standards for surveillance video quality. This achievement will include the coordination of research work being conducted by a number of organisations.

Acknowledgements

The work presented was supported by the European Commission under Grant INDECT No. FP7-218086. Preparation of source video sequences and subjective tests was supported by European Regional Development Fund within INSIGMA project no. POIG.01.01.02-00-062/09.

References

1. Agresti, A.: Categorical Data Analysis, 2nd edn. Wiley, Chichester (2002)
2. Agresti, A., Coull, B.A.: Approximate is better than "exact" for interval estimation of binomial proportions. The American Statistician 52(2), 119–126 (1998); ISSN: 0003-1305
3. Dziech, A., Derkacz, J., Leszczuk, M.: Projekt INDECT (Intelligent Information System Supporting Observation, Searching and Detection for Security of Citizens in Urban Environment). Przeglad Telekomunikacyjny, Wiadomosci Telekomunikacyjne 8-9, 1417–1425 (2009)
4. Eskicioglu, A.M., Fisher, P.S.: Image quality measures and their performance. IEEE Transactions on Communications 43(12), 2959–2965 (1995), http://dx.doi.org/10.1109/26.477498
5. European Commission: European Security Research Conference, SRC 2009 (September 2009), http://www.src09.se/
6. Ford, C., Stange, I.: Framework for Generalizing Public Safety Video Applications to Determine Quality Requirements. In: 3rd INDECT/IEEE International Conference on Multimedia Communications, Services and Security. AGH University of Science and Technology, Krakow, Poland, p. 5 (May 2010)
7. Ford, C.G., McFarland, M.A., Stange, I.W.: Subjective video quality assessment methods for recognition tasks. In: Rogowitz, B.E., Pappas, T.N. (eds.) SPIE Proceedings of Human Vision and Electronic Imaging., vol. 7240, p. 72400. SPIE, San Jose (2009)
8. Honess, T., Charman, E.: Closed circuit television in public places: Its acceptability and perceived effectiveness. Tech. rep. Home Office Police Department, London
9. ITU-T: Subjective Video Quality Assessment Methods for Multimedia Applications. ITU-T (1999)
10. Janowski, L., Romaniak, P.: Qoe as a function of frame rate and resolution changes. In: Zeadally et al. [14], pp. 34–45
11. Romaniak, P., Janowski, L.: How to build an objective model for packet loss effect on high definition content based on ssim and subjective experiments. In: Zeadally et al. [14], pp. 46–56
12. VQiPS: Video Quality in Public Safety Working Group, http://www.safecomprogram.gov/SAFECOM/currentprojects/videoquality/
13. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement. Signal Processing: Image Communication 19(2), 121–132 (2004),http://dx.doi.org/10.1016/S0923-5965(03)00076-6%20
14. Gomes, R., Junior, W., Cerqueira, E., Abelem, A.: A qoE fuzzy routing protocol for wireless mesh networks. In: Zeadally, S., Cerqueira, E., Curado, M., Leszczuk, M. (eds.) FMN 2010. LNCS, vol. 6157, pp. 1–12. Springer, Heidelberg (2010)

Integrating Applications Developed for Heterogeneous Platforms: Building an Environment for Criminal Analysts*

Arkadiusz Świerczek, Roman Dębski,
Piotr Włodek, and Bartłomiej Śnieżyński

Department of Computer Science
AGH University of Science and Technology
30 Mickiewicza Av., 30-059 Krakow, Poland
arek.swierczek@gmail.com, roman.j.debski@googlemail.com,
pwlodek@agh.edu.pl, bartlomiej.sniezynski@agh.edu.pl

Abstract. In this paper we present original approach for integrating systems on an example of LINK and Mammoth – criminal analysis applications. Firstly, a problem of integration is described with short description of integrated applications. Secondly, some theoretical information about integration is depicted. Paper continues with explanation of approach chosen for LINK and Mammoth integration and presents evaluation of achieved result. Eventually some final thoughts are stated.

Keywords: application integration, software engineering, criminal analysis.

1 Introduction

Application integration is a classical project type in the software engineering. It is well recognized and described. However, in practice, it is still a challenge.

Problem of integration appears when two or more applications are used in the same organization and cooperation between them helps to achieve organization's goals. Cooperation can be done manually, outside of the systems. Although, it is not a good solution. Integration of applications with use of files transfer, common database, remote procedure invocation or messaging, allows to provide automatic or at least semi-automatic cooperation which is much more convenient for the user.

Problem of integration appears in the criminal analysis domain. Most of typical criminal analyst's work scenarios involves executing of many applications and tools to import data, process it, and visualize results. Some of the tasks are executed in specialized applications like Analyst Notebook (see [6]). Others are done in standard applications like MS Excel or MS Access.

* The research leading to the results described in the paper has received funding from the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreement n° 218086.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 19–27, 2011.

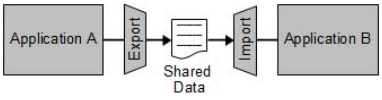


Fig. 1. File transfer

After consultations with criminal analysts working at the Voivodeship Headquarters of Police in Krakow, we decided to provide software assisting in their common work. As a result, the following two applications have been developed: LINK [4] and Mammoth [8].

LINK is a tool for importing, preprocessing and visualizing of data coming from various sources. It also provides validation tools, statistic data reports, basic search techniques, and simple analysis such as retrieving the most frequent callers.

Mammoth is a system for searching frequent patterns in the data such as phone billings or border passing records. Goal of the analysis is a detection of the cases in the data, in which some people interact with each other using the same communication pattern. Presence of such a pattern suggests that people, who appear in it, are members of cooperating criminal group.

After the development it appeared, that in many scenarios both applications may be used. As a result, they have been integrated using shared database solution, becoming an initial version of the environment for criminal analysts.

In the following section we describe application integration in general from the point of view of software engineering. After such theoretical introduction, we present details of the LINK and Mammoth integration. Next, we show how these applications can be used by criminal analysts after the integration.

2 Main Integration Styles

Application integration is nowadays a mature engineering discipline – good practices are cataloged in the form of patterns [5]. All the patterns can be generalized and divided into four main categories: *file transfer, shared database, remote procedure invocation* and *messaging*. The following sections describe briefly each of them.

2.1 File Transfer

This approach to integration utilizes files – universal storage mechanism available in all operating systems. One application (producer) creates a file that contains the information needed by the other applications. Next, the other applications (consumers) can read the content of the file (Fig. 1). Choosing this approach has the following consequences:

- it is *data sharing oriented* (not *functionality sharing oriented*),
- files are (effectively) the public interface of each application,

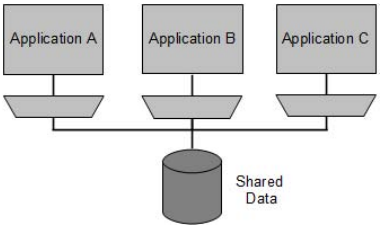


Fig. 2. Shared database

- choosing the right file (data) format is very important (nowadays it is often XML),
- applications are decoupled from each other,
- applications are responsible for managing the files (creation, deletion, following file-naming conventions etc.),
- updates usually occur infrequently and, as a consequence, the communicating applications can get out of synchronization,
- integrators need no knowledge of the internals of applications.

2.2 Shared Database

In this pattern the integrated applications store their data in a single (shared) database. The stored data can be immediately used by the other applications (Fig. 2). Choosing this approach has the following consequences:

- it is *data sharing oriented* (not *functionality sharing oriented*),
- data in the database are always consistent,
- defining a unified database schema that can meet the needs of many applications can be a really difficult task,
- any change of the shared database schema may have impact on all integrated applications (applications are strongly coupled),
- since every application uses the same database, there is no problem with *semantic dissonance* [5],
- shared database can become a performance bottleneck and can cause deadlocks.

2.3 Remote Procedure Invocation

In this approach each part of the integrated system (a set of cooperating applications) can be seen as a large-scale object (or component) with encapsulated data. Shared functionality of each application is accessible via its public interface[1] (Fig. 3). Choosing this approach has the following consequences:

[1] A number of technologies such as CORBA, COM, .NET Remoting, Java RMI and Web Services implement Remote Procedure Invocation (also referred to as Remote Procedure Call or RPC).

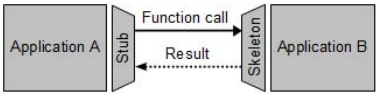


Fig. 3. Remote Procedure Invocation

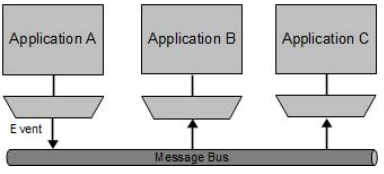


Fig. 4. Messaging

- it is *functionality sharing oriented* (not *data sharing oriented*),
- applications can provide multiple interfaces to the same data,
- applications are still fairly tightly coupled together (often each application of the integrated system perform a single step in many-step algorithms: in such a case one application's failure may bring down all of the other applications),
- communication is (usually) synchronous,
- developers often forget that there is a big difference in performance and reliability between remote and local procedure calls – it can lead to slow and unreliable systems.

2.4 Messaging

This approach combines all the benefits of the previous three and is often considered [5] as the best one. Messages transfer packets of data frequently, immediately, reliably and asynchronously using customizable formats (Fig. 4). Choosing this approach has the following consequences:

- sending small messages frequently allows applications to collaborate behaviorally as well as share data,
- applications are decoupled (it has many consequences e.g. integration decisions can be separated form the development of the applications),
- the integrated applications (usually) depend on a messaging middleware.

3 Integrating LINK and Mammoth

LINK and Mammoth integration was a time-consuming task. It involved adapting several important components, such as data models and data access layer. It is worth to mention that the integration was not planned during early phases of applications' development and the concept evolved as a side-effect of creating

flexible and efficient data access layer. Requirement of the integration emerged when applications were already functional, but missing some features offered by the other.

3.1 Choosing Integration Style

When deciding on the most appropriate integration style for LINK and Mammoth, several factors had to be taken under concern:

- Support – programming platforms of both applications should provide a mature and well-designed software components which support chosen integration method.
- Performance – a technology behind integration style should provide good performance for both reading and writing data.
- Autonomy – the integration method must not require both applications to be on-line when accessing data, thus *Remote Method Invocation* is not appropriate in this situation.

Considering above criteria led to an idea of using serverless database hosted in-process as an integration media, which constitutes a variation of *Shared Database* approach.

There were two candidating database solutions: *HSQL* and *SQLite*. The former has an advantage of native support from Java platform used by LINK, but, unfortunately, support from .NET Framework used by Mammoth is missing. The latter, on the other hand, is written in C/C++ and has support from both Java and .NET community. Moreover, because of its native implementation for most of the popular operating systems, it reaches better performance than HSQL.

Despite its significant limitations comparing to regular database servers, functionality provided by SQLite was broad enough to satisfy integration needs:

- The subset of supported *SQL-92* standard commands is sufficient for data querying and modification.
- Simple flat transactions, though limited, let keep data file in consistent state.
- A file locking-based concurrency is not a problem since only one instance of application should be able to access the file at a time.
- Data files generated by applications should not exceed hundreds of megabytes in size, thus problem of bottlenecks produced by memory overheads is not applicable here.

3.2 Implementing Integration Components

Instead of creating dedicated modules in applications destined to integration purposes only, it was decided to use previously described database as a main storage mechanism in both systems and treat it also as an integration media. This required synchronizing the structure of chosen applications' domain models

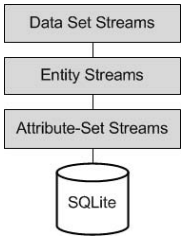


Fig. 5. Data access layer architecture

and creating similar data access layers (this requirement was solved by porting code from LINK to Mammoth). The approach has several advantages:

- Shared data models are designed once only and implemented similarly in both applications.
- Ported data access layer requires less testing than created from scratch.
- Data access layers' design and architecture are the same, which makes them easier to document and maintain.
- Beside compatible and portable data sources applications gains well-designed and tested data access layers.

However, during development some drawbacks were also discovered:

- Porting software components from one platform to another is an arduous and time-consuming task.
- Keeping implementations in sync leads sometimes to problems, as some global concepts in applications differ significantly.

Data Access Layer Architecture. Data access layer (Fig. 5) utilizes concept of streams drawn from popular file in-out handling approach. It consists of three main sublayers (in bottom-up order):

- Attribute-set streams – manage data grouped in sets of attributes (of primitive types), which are translated to database relations.
- Entity streams – transform sets of attributes to domain-specific objects and inversely.
- Data set streams – create references between objects and group them in data sets. They also split entities from data sets into sequences of elements for appropriate entity streams.

Porting Data Access Layer. The concepts were first implemented in LINK in Java language and ported to C# in Mammoth. In spite of Java and C# languages significant similarity, this process had some pitfalls:

- Lack of indexers in Java – it is common in C# collections to use indexers, i.e. [ ] symbols. Java uses indexers only in arrays and in collections usually uses *get(int index)* methods instead.

- Different generics implementations – because of more strict approach to generics in C#, some modifications were needed during porting process.
- Distinct handling of non-existing items in *map* collections – in Java standard library *HashSet<TKey, TValue>* class returns *null* when queried with non-existing key. On the opposite, .NET *Dictionary<TKey, TValue>* class throws an exception in this situation.
- *JDBC* and *ADO.NET* disparity – libraries differ in many areas, e.g. in *Statement/Command* parameter handling, *null* value handling and no support for commands' batching in ADO.NET.

Shared Data Models. As domain models constitute critical area of LINK and Mammoth, the schema of the data source had to conform to the structure of the models. It was achieved by providing an additional metadata to models, which allowed for dynamic generation of data source schema.

Each application uses programming language-provided mechanism to store metadata, i.e. LINK utilizes annotations and Mammoth – attributes. In future some other ways of providing metadata may be considered, e.g. ontology or plain XML file.

As far as shared data models are concerned, both applications use corresponding metadata information and have the same structure of these models. Models, which are not destined to be shared have no restrictions and can be treated as private for each application.

4 Integration Evaluation

Both LINK and Mammoth applications have strong and weak sides when it comes to data analysis and visualization. To name a few examples, both applications come with template-based data importing mechanism [2] which enables data loading from various file formats. However, data importer introduced in LINK is more advanced as it supports many additional options, and it allows to design import template in a visual manner. Both applications come with visualization components which enable to view imported and analysed data using graphical representation. Graphical component available in Mammoth provides rich but read-only data presentations as opposed to LINK where visual data can easily be modified by the analyst. Mammoth also incorporates some *data mining* techniques [8] including frequent itemset mining using both *Apriori* [2] and *FP-Growth* [3] algorithms as well as mining surprising periodic patterns [7] whereas LINK provides some data preprocessing routines.

To provide criminal analysts with a better approach to data analysis and visualization, we have combined strengths of both systems by integrating them on a database level. This allows both systems to be developed and delivered separately while leaving the possibility of using both tools to support criminal

[2] Each template defines a file structure which is the subject of import.

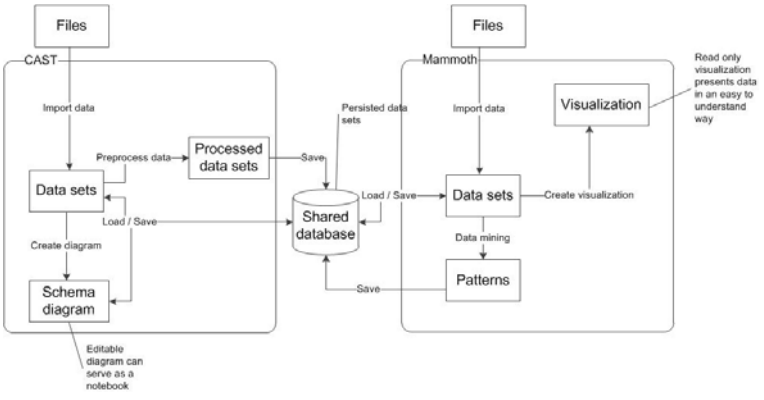


Fig. 6. Data flow between LINK and Mammoth including single shared database

investigation without having to prepare the same data twice in order to be used by both applications. Also, it allows to do data processing (like common prefix removal) in one application and to analyse resulting data in the other.

Typical data flow when using both applications is outlined in the Fig. 6. The shared database plays the vital role in the flow. It persists data sets in a single file allowing both applications to exchange data easily.

Analysis typically begins by importing data from external systems or resources. Current version of both applications support text files as a source of data. In addition, LINK also supports data stored in *Microsoft Excel* spreadsheets. It is usually a better option to import data using LINK because its data loading mechanism is more mature.

After importing relevant data in LINK the analyst can preprocess data, which includes common prefix removal or merging duplicated or similar phone numbers. The analyst can also create schema diagram which shows objects and their relations. Because the diagrams are fully editable, the analyst can add more information, like new people, phones, etc., when new facts in the investigation are discovered. The diagram can be then saved and restored later.

To perform automatic data analysis using some data mining techniques, the analyst can open the same database in Mammoth application. He can perform data mining tasks like discovering frequent sets and sequences or finding user defined patterns. After discovering potentially interesting patterns, the analyst can display them in a graphical form to better understand them. The analyst can use various filters to narrow the set of discovered patterns and to pick patterns which tend to be the most interesting. Once interesting patterns are identified, they can be presented in isolation using Mammoth timing visualization so that the analyst can see the exact occurrence time of the events which compose the pattern. Finally, patterns can be saved back to the shared database in order to be used in the LINK program.

5 Conclusions

This paper summarizes our experience in integration of two applications created for criminal analysts. We have presented theory related to integration and showed how it was applied in our case. Unfortunately, not all the issues are mentioned here because of the lack of space (e.g. security and ethical aspects).

We are planning to add new functionalities to both systems, like adding a kind of version control which would allow to test various hypothesis during analysis or searching of patterns with specific constraints. Based on this experience, we are also planning further integration of both applications with other systems developed in the security domain, like social network analysis, searching the web for personal profiles, to build a full environment for criminal analysts. However, this may demand of using SOA and more elaborated integration techniques based on ontologies.

Currently, after integration, LINK and Mammoth are tested by the Police and users' comments will have the biggest impact on our decisions about the future of the project.

References

1. Agrawal, R., Imieliński, T., Swami, A.: Mining association rules between sets of items in large databases. In: Proceedings of the 1993 ACM SIGMOD International Conference on Management of Data, pp. 207–216. ACM, New York (1993)
2. Bodon, F.: A fast apriori implementation. In: Proceedings of the IEEE ICDM Workshop on Frequent Itemset Mining Implementations (2003)
3. Borgelt, C.: An implementation of the fp-growth algorithm (2005)
4. Debski, R., Kisiel-Dorohinicki, M., Milos, T., Pietak, K.: Link – a decision-support system for criminal analysis. In: Danda, J., Jan, D., Glowacz, A. (eds.) IEEE International Conference on Multimedia Communications, Services and Security, MCSS 2010, pp. 110–116 (2010)
5. Hohpe, G., Woolf, B.: Enterprise Integration Patterns: Designing, Building, and Deploying Messaging Solutions. Addison-Wesley Professional, Reading (2003)
6. i2 Ltd. i2 Analyst Notebook (2011), http://www.i2group.com/
7. Jiyang, J.Y., Wang, W., Yu, P.S.: Infominer: Mining surprising periodic patterns. In: Proc. SIGKDD, pp. 395–400. ACM Press, New York (2001)
8. Wlodek, P., Swierczek, A., Sniezynski, B.: Pattern searching and visualization supporting criminal analysis. In: Danda, J., Jan, D., Glowacz, A. (eds.) IEEE International Conference on Multimedia Communications, Services and Security, MCSS 2010, pp. 212–218 (2010)

INACT — INDECT Advanced Image Cataloguing Tool

M. Grega, D. Bryk, M. Napora, and M. Gusta

AGH University of Science and Technology, Krakow, Poland
General Headquarters of Police, Warsaw, Poland

Abstract. Possession of images of child sexual abuse is forbidden in most countries. The ban generally applies not only to regular citizens but to police officers and units as well. This significantly complicates gathering and presenting evidence in police investigations. This presents a requirement for computer software to overcome this problem. A potential solution to this problem is presented in the paper. The outcome of the research and development work is INACT (INDECT Advanced Image Cataloguing Tool) software.

1 Introduction and Motivation

Possession of Child Pornography (CP) images is considered a crime in most countries. The law applies not only to regular citizens, but also to police units. This significantly complicates gathering and presenting evidence in police investigations. This presents a requirement for computer software to overcome this problem. As a response to this issue, a potential solution is presented below.

The outcome of the research and development work is INACT (INDECT[1] Advanced Image Catalogue Tool) software. In the proposed solution, if police officers have access to a hash database (extracted from the images) and the suspect's file system, they can run the INACT application and perform a full image search. This enables the police to expand their investigation into child sexual abuse pictures, the possession of which is forbidden.

Gathering and presenting evidence in police cases is one of the parts of investigation. Police units responsible for fighting child pornography catalogue evidence gathered during their investigations. Each time a police unit is in possession of such evidence, a police officer provides additional metadata (such as case number, case name, officer name, etc.) and runs the indexing module. The indexing module automatically extracts low-level metadata from the images and stores it in a database. This process can be time consuming if there is a high number of images; however, it is fully automated. After indexing is completed, the police unit can dispose of the evidence.

Metadata gathered in this way is fed into the main database. This allows the police to create a database useful in upcoming investigations, making it

[1] INDECT is an EC Grant that funded this work. We acknowledge the Grant in the last Section.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 28–36, 2011.

possible to identify pictures with child abuse material at the scene of a crime or while searching any device containing data (such as suspect's file systems on seized computers). It is in line with the Polish penal code, which states that possessing or storing child pornography is prohibited [3,4]. Additionally, this approach addresses provisions of the European Council Framework Decision 2004/68/JHA of 22 December 2003 on combating the sexual exploitation of children and child pornography.

INACT software is classified into a set of two applications utilising the Query by Example (QbE) approach. Several such applications have already been demonstrated, including a well-known, generic MIRROR content-based image retrieval system [11], as well as more specific systems such as GAMA, designed for accessing libraries of media art [7].

The approach presented in this paper relates to child pornography searches. Other tools, such as the Paraben Porn Detection Stick or the NuDetective [2], are specific in the analysis and detection of nudity, according to their descriptions. INACT is more generic as it deals with images similar to those to be retrieved.

The concept of using hash sets to identify suspicious files in digital forensics has been in use for a number of years, and is built into numerous forensic tools such as EnCase [5] and FTK (Forensic Tool Kit) [1]. While the above mentioned tools only allow the retrieval of images that are identical (they utilise MD5 sums), INACT also allows to retrieve similar images, since it utilises MPEG–7 [6] descriptor values. Although both forensic hash tools and the MPEG–7 standard are well-known techniques, their combination is novel.

The remainder of the paper is structured as follows. Section 2 presents a general concept of the application. Section 3 describes the MPEG–7 Standard. In Section 4, we presented the INACT system requirements. Sections 5 and 6 introduce the INACT INDEXER and INACT INSEARCHER, respectively. Section 7 outlines plans for further work. The paper is concluded in Section 8.

2 General Concept of Application

In order to solve the problem outlined in the introduction, a set of two complimentary applications was developed. The first application, the INACT INDEXER, is designed to be used at police headquarters. The police have at their disposal sets of images containing child pornography from various sources, including ongoing investigations and international channels of cooperation. Such sets are input into the INACT INDEXER. The INACT INDEXER processes the images and creates a catalogue containing information about the images (such as a description of the case in which the images were acquired) and a set of descriptors for each of the catalogued images. The descriptor set consists of MD5 hashes and MPEG-7 descriptors. The process of calculating hashes and descriptors is a one-way process, which means that the images cannot be recreated either from the hashes or from the descriptors. This allows the police to dispose of the images (as required by law) whilst retaining the information about the images themselves. The result of the INACT INDEXER analysis is a database which can be utilised by the INACT INSEARCHER application..

INACT INSEARCHER is an application designed to be used during a search of a suspect's file system. It utilises the database of hashes and descriptors created by the INACT INDEXER. The INACT INSEARCHER is very straightforward to use. It is sufficient to select the root of a directory tree to be searched. All images that are identical (utilising MD5 sums) or similar (utilising MPEG-7 descriptor values) will be retrieved and presented alongside the database information to the officer performing the search. This allows the police to draw conclusions regarding possible sources of the images and their distribution paths. The concept of two INACT applications is presented in Figure 1.

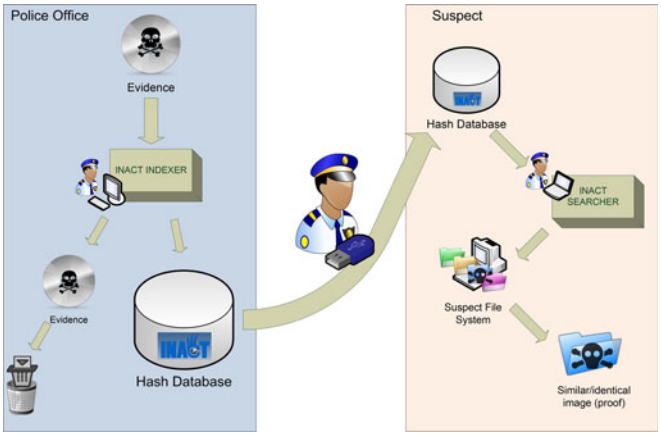


Fig. 1. The workflow of the INACT system

3 MPEG–7 Standard

The MPEG–7 standard was proposed by the Moving Pictures Experts Group (MPEG) as an ISO/IEC standard in 2001 [6]. The formal name of the standard is "Multimedia Content Description Interface". The main goal of the MPEG–7 standard can be described as following: "to standardise a core set of quantitative measures of audio–visual features [...]" [8].

One of the parts of the MPEG-7 standard defines algorithms that can be used for an advanced multimedia search method called Query by Example (QbE). These algorithms are called "descriptor" within the standard. In the QbE search method the query is a multimedia object. The results of the query are those multimedia objects from the dataset which are similar to the query. The similarity is defined by the utilised QbE method.

The presented application makes use of one of the MPEG-7 descriptors (the Colour Structure descriptor) for the similarity search. This allows the INACT INSEARCHER application to find images similar to the ones that are stored in the image catalogue.

4 The INACT System Requirements

The specifications of the police operation, the availability of personal computers, and the good practice of programming impose the following general requirements on the INACT System:

1. The indexing process is performed on the personal computers available to the police.
2. The search process is performed on the computer belonging to the suspect, utilising his/hers hardware and operating system.
3. Both INDEXER and INSEARCHER are intended to be operated by police officers who have received basic training in their use. Minimum knowledge of computer use is required from the operator.
4. Both applications must be responsive all the time when in use.

These requirements result in restrictions on the chosen hardware platform. In order to meet the first and second requirement, the application must be compatible with personal computers based on i486/i686 microprocessor core architecture. This architecture is common among users of personal computers equipped with Intel or AMD microchips [10], thus it is very likely that the suspect will also be using it.

In order to meet the software configuration aspect of the second requirement, the INSEARCHER must be able to run in different operating systems. Currently the most popular operating systems are Windows (91.09% of the market), Mac OS (5%) and Linux (0.86%)[9], therefore the application must be compatible with all of them. To do so, the Qt library is used for drawing the GUI, the OpenCV library and the MPEG-7 library are used for manipulating images, and Berkeley DB is used to create the local database. There are specific reasons for choosing Berkeley DB, alongside its search performance; they are explained in the fifth section of this article in context of the INDEXER application. The MPEG-7 library, written in ANSI-C, is an implementation of the tools specified in [6]. The libraries were selected because they have versions appropriate for those operating systems. This is required to deploy a stand-alone version of the INSEARCHER. These stand-alone versions are carried by police officers to the search location on a USB flash drive or any other mobile data storage device with a USB interface. INDEXER uses the same libraries so as to be compatible with INSEARCHER.

The third requirement is that the GUI of both application is as straightforward to use as possible. To start the search using INSEARCHER or the indexing using INDEXER, the police officer simply needs to select the root directory and press the start button. The third requirement dictates that indexing and the search process must be automated. Minimum interaction between the user and the application is required during these processes, and files are recognised by their true format rather than by their extension. This allows to find images that are hidden in an ordinary browsing process.

5 INACT INDEXER

According to the current legislation regarding police investigations in Poland, a police unit cannot store evidence containing prohibited content. Any data carrier with illegal images or videos must be destroyed after it is analysed. One of the INACT project goals is to create a tool which can automatically analyse illegal files and store metadata extracted during this process in a database. The database should include information describing the content of the images (evidence). Information about the content is acquired by utilising MPEG-7 descriptors and MD5 hashes. As mentioned in the previous section, calculating the descriptors is a time-consuming process. On the other hand, software used during investigations should work automatically. These considerations were the key arguments for the creation of INACT INDEXER.

Alongside the general requirements for the INACT system, the application must fulfil certain specific requirements. First, the software needs to use a portable database, as it is to be used in investigations outside of police stations. Second, an investigation application must allow users to view, edit and commit the database.

The INACT INDEXER allows a police officer to create a database which contains MPEG–7 descriptors of the images, MD5 sums of analysed files, important information about them and about the investigation they came from. The process of calculating the descriptors and storing them in database is called indexing. INACT INDEXER allows police officers to create a database containing MPEG-7 descriptors of images, MD5 sums of analysed files, and important information about the files and the investigation they originate from. The process of calculating the descriptors and storing them in a database is called indexing. The INDEXER software automatically calculates descriptor values and MD5 hashes of all images on hard disk drives, CDs/DVDs, pendrives or other data storage devices. Police officers simply need to select the root directory for the analysis to begin, and fill a metadata form. The form is needed to acquire information regarding the investigation, data storage device and date of the analysis. The fields are stored within the database and used when generating search reports.

6 INACT INSEARCHER

The purpose of the INACT system is to make police work more effective. INACT INSEARCHER (interface presented in Figure 2) is designed and implemented to improve quality of police force operations. Let's consider a situation in which the police have a strong suspicion that a serious crime has been committed; in this instance child abuse in the form of collection or distribution of photos depicting juveniles in a sexual context. Today, in an era of fast Internet and cheap and widely available digital cameras, such content is stored and shared in digitalised form. This suggests that the only possible way to gather evidence is to browse through all the files on a suspect's computer. To do so, the police officers arrive at a site where they can access the suspect's computer, or if they previously seized the suspicious file system, they can examine it in the laboratory.

These two approaches are called online and offline forensics respectively. In case of offline forensics, the data storage device is connected to forensic equipment through a hardware blocker, which ensures that the contents of the data storage device remains unchanged. In online forensics, when police officers have access to a running operating system, it is impossible to use a blocker as the device should not be powered off. Powering off a seized running device may result in encryption of previously unencrypted files. Therefore in case of online forensics all the required actions need to be carried out on a 'living' suspect's file system.

In the classical approach the search is performed by a police officer who manually browses through the file system. The success of this operation is highly dependent on the police officer's experience, accuracy and knowledge, where such content is usually stored, and finally luck. INSEARCHER is proposed as a solution, since it changes the entire approach to the search process.

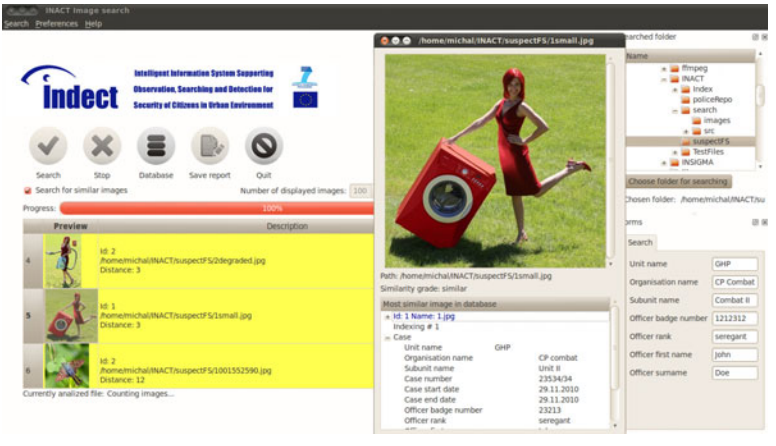


Fig. 2. Interface of the INSEARCHER application (shown with mock data)

The exact reason and form of the search process is regulated by legal systems in individual countries. When police officers are equipped with the INACT system, the procedure is carried out differently yet still according to the regulations in place. Firstly, the INDEXER database is updated on a mobile data storage device using INSEARCHER. Secondly, the officer arrives at the site where the suspect's computer is located. The officer plugs in the mobile storage device, runs INSEARCHER, selects a folder and initiates the search process. When the search is completed, the results are presented in an easily understandable form. They are displayed as a list, with exact matches listed at the top and highlighted in red. Similar results are listed below and organised by similarity to images in the database. The operator can access information stored in the database regarding the images, and preview any of the listed images in a separate window to see

the connection between them and those stored in the database. These functions allow INSEARCHER to meet the second and third general requirements for the INACT system.

The above description provides additional specific requirements for the INSEARCHER application. First, the operator must see the results of the search in real time. Just a single positive result is required for the search to be successful; the search can then be aborted. Charges can be pressed on the basis of the single match. Second, according to the current legislation regarding investigations, interventions must be as quick as possible. The search must be as non-disturbing to the suspect as possible.

There are no known publications describing a similar approach concerning the creation of an application performing a search process which allows for similarity searches. An implementation of the proposed application is available to the Polish police force, and is currently being developed in cooperation with them.

A preliminary performance analysis of the search procedure has been performed. The duration of the search is crucial, because conclusive results must be achieved before police intervention ends. The duration of the search process must not have a significant impact on the intervention time in order to meet the second specific requirement for INSEARCHER. Figure 3 shows the time consumption for the Scalable Colour Image Descriptor and MD5 hash calculation. A test was carried out using 202 images with a total physical size of 96MB and a cumulative image size of 249.5 megapixels. Images smaller than 50x50 pixels and smaller than 10 kB were filtered out. In the test the database contained two indexed images to minimise search time. Reading the test set from the disk took INSEARCHER 23 seconds. Reading the images and calculating the descriptor took 105 seconds. Calculating MD5 hashes took an additional 4 seconds. The computational overhead (82 seconds) of the descriptor calculation is dominant. Since the time for the descriptor calculation is directly proportional to actual

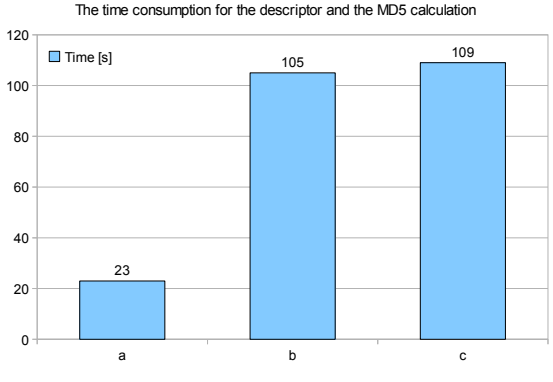


Fig. 3. Time consumption of the descriptor and the MD5 hash calculation: a – reading from disk; b – reading from disk, descriptor calculation; c – reading from disk, descriptor calculation and MD5 calculation

image size in megapixels (as observed in the performed experiment), the overhead is expected to grow linearly with actual image sizes. In comparison, the MD5 calculation is very quick, adding only approx. 4 seconds to the overall time. Any heavy computation processes could affect the GUI responsiveness. This cannot be permitted in order to meet the first specific requirement for INSEARCHER.

A 100K test was also performed, in which a subset of images was hidden among 100,000 other images with a total physical size of 3 GB. Test images are natural photographs downloaded from the www.flickr.com website. The goal of the test was to find this subset. Building of the image tree took approx. 25 minutes, while the actual search took an additional 2 hours 20 minutes. All tests were performed on desktop computers equipped with Intel Core 2 Duo T5600 1.83 GHz (dual core) processors and hard disks with ext3 partitions controlled by the Ubuntu 10.04 LTS operating system.

7 Further Work

The next stage is to progress to the INDEXER 2.0 version. Currently the main goal of INACT development is to create a networked version of the software. All the INDEXER copies distributed among police headquarters will be able to commit new indexing results into the central database, and all of the INSEARCHER copies will be able to verify externally-updated databases

As the duration of the search is crucial, more research work is planned in relation to optimal browsing algorithms aiming to reduce the duration of the search process. “Animal algorithms” will be implemented for this purpose.

8 Conclusions

This paper introduces a novel, advanced image cataloguing tool. INACT allows the creation of a database of images whose storage is prohibited, and an automated search for those images in a suspicious file system, as well as the creation of links between previous and ongoing cases. INACT should provide better protection for the victims, greater effectiveness in capturing the offenders, and, finally, less tedious work for police officers.

Implementation of the INDECT system based on INDEXER and INSEARCHER modules makes it possible for police units to set up databases holding data on child pornography. The concept of storing information on child pornography images is not new. The basis for identifying images with child abuse material is a hash value calculated using a dedicated file. It is the simplest yet not very progressive method of investigating such images. The main advantage of the INDECT system is enabling police forces to search for files resembling each other in their content, based on MPEG-7 descriptors. This functionality should allow the police to expand their investigation and unveil more evidence.

Acknowledgements

This work was supported by the European Commission under the Grant INDECT No. FP7–218086.

References

1. AccessData: Forensic Toolkit (FTK) Computer Forensics Software — AccessData, http://accessdata.com/products/forensic-investigation/ftk
2. de Castro Polastro, M., da Silva Eleuterio, P.M.: Nudetective: A forensic tool to help combat child pornography through automatic nudity detection. In: International Workshop on Database and Expert Systems Applications, vol. 0, pp. 349–353 (2010)
3. Collective work: (in polish) code of criminal procedure, (journal of laws 1997 no. 89, item 555)
4. Collective work: (in polish) penal code (journal of laws 1997 no. 88, item 553)
5. Guidance Software: Computer Forensics Solutions, Digital Investigations, e-discovery — Guidance Software, http://www.guidancesoftware.com/
6. ISO/IEC: Information technology – multimedia content description interface. ISO/IEC 15938 (2002)
7. Ludtke, A., Gottfried, B., Herzog, O., Ioannidis, G., Leszczuk, M., Simko, V.: Accessing libraries of media art through metadata. In: International Workshop on Database and Expert Systems Applications, vol. 0, pp. 269–273 (2009)
8. Nack, F., Lindsay, A.T.: Everything you wanted to know about mpeg-7, part 1. IEEE Multimedia 6(3), 65–77 (1999)
9. Net Applications: Operating system market share (November 2010), http://marketshare.hitslink.com/operating-system-market-share.aspx?qprid=8
10. Shilov, A.: Microprocessor market grows moderately in the third quarter (November 2010), http://www.xbitlabs.com/news/cpu/display/20101111195314_Microprocessor_Market_Grows_Moderately_in_the_Third_Quarter_Analysts.html
11. Wong, K.M., Cheung, K.W., Po, L.M.: Mirror: an interactive content based image retrieval system. In: ISCAS (2), pp. 1541–1544. IEEE, Los Alamitos (2005)

Distributed Framework for Visual Event Detection in Parking Lot Area

Piotr Dalka, Grzegorz Szwoch, and Andrzej Ciarkowski

Multimedia Systems Department, Gdansk University of Technology,
Narutowicza 11/12,80-233 Gdansk, Poland
{dalken,greg,rabban}@sound.eti.pg.gda.pl

Abstract. The paper presents the framework for automatic detection of various events occurring in a parking lot basing on multiple camera video analysis. The framework is massively distributed, both in the logical and physical sense. It consists of several entities called node stations that use XMPP protocol for internal communication and SRTP protocol with Jingle extension for video streaming. Recognized events include detecting parking vehicles supplemented with parking place and vehicle identification. Event detection is based on low-level image processing consisting of moving object detection, tracking and classification. Front-end of the framework is formed by the operator console that presents results of the framework accompanied with video streams from selected cameras in the real-time. Additionally, the operator console may be used to automatically aim any PTZ camera in the framework at the selected spot in video frames from fixed cameras and track moving objects as long as they move within the monitored area. The paper is concluded with the framework performance evaluation during real condition experiments.

Keywords: image processing, automatic event detection, object tracking, object classification, distributed system, video streaming.

1 Introduction

With increasing number of video cameras in every city, there is a need for automatic analysis of video streams in order to detect various events in the real time. There are various applications of such systems and depending on the requirements they may be applied to general trespassing discovery [1] or human activity analysis and suspicious behavior detection [2][3][4].

In case of parking lots, especially the larger ones with high traffic, an automatic video surveillance system is necessary for successful parking lot management. This paper presents a distributed framework for detecting various events occurring in a parking lot with particular emphasis on parking vehicle detection. Section 2 presents a description of the framework, including its architecture, communication protocols and the Operator Console. The next section describes software modules of the framework that make event detection possible. Framework performance is evaluated in Section 4. The last section concludes the paper.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 37–45, 2011.

2 Framework Description

The framework presented in the paper is designed to detect various visual events in video streams from multiple cameras placed in a parking lot. Detection of parking vehicles is the most important one. This includes identification of the parking time, parking place and identification of the parking vehicle.

Additionally, the system detects a few supplementary events that are useful to a surveillance operator:

- vehicle stopping outside of any parking place (i.e. on the road),
- detecting and counting of persons entering or leaving a building,
- detection and counting of vehicles entering or leaving a parking lot.

The framework design is modular and open which means that new event detection modules may be added in the future.

The framework consists of logical units called node stations. Each station forms an entity that can communicate with other nodes of the framework. A node station may be related with a single hardware component (i.e. computer) but there can be also many nodes running within one server.

There are two types of communication within the framework. The first one is related to short messages (e.g. notifications on events detected) sent between various nodes of the framework. The second type of the communication is dedicated to video streaming. Both types of communication have completely different properties, therefore their implementations are different and suited to the requirements.

The framework utilizes two types of cameras: fixed and PTZ ones. Video streams from fixed cameras are analyzed automatically within the framework in order to detect various events. PTZ cameras are controlled by the framework and may be automatically aimed at any object or region covered by fixed cameras.

2.1 Node Stations

There are three types of node stations in the framework. Video processing units are responsible for analyzing video streams from IP cameras connected to them. The central station integrates results of video processing, stores them in a database and generates messages for the third type of a node station – an operator console. The console forms the front-end of the framework and presents detailed results of the framework as well as notifications on events detected. It also allows sending commands to the system (e.g. related to PTZ camera control).

2.2 Operator Console

A system operator accesses the video analysis results by means of a dedicated module called the operator console (Fig. 1). This module constitutes the client side of the distributed analysis framework. Using the console, the operator requests access to data streams from selected cameras. This request is processed by the central server and the console receives on-demand streams consisting of encoded camera frames and metadata containing analysis results. The received streams are presented on the

monitor screen, with metadata overlaying the camera images. The type of metadata presented on the screen is selected by the operator, it may contain both high-level information (text messages indicating the detected event, frame encompassing the object that caused the event) and the low-level analysis results (object detection masks, tracker frames, detection areas, classification results, etc.). Additionally, the console provides the lists of formerly detected events, currently tracked objects and statistics of the analysis results. The operator may also use the console for sending control commands to the system server. Therefore, the console is both the tool for assessment of video analysis performance and the system front-end for the operator of the working monitoring system.

Fig. 1. Operator console application

2.3 Communications

The model based on node stations implies that the system is massively distributed, both in the logical sense – in the context of distribution of services and functionality onto multiple machines, as well as in the physical sense, since it is desirable that the node station was in close proximity to the operated cameras.

The consequence of such an approach may be a significant distance between the node station and monitoring center, preventing the use of a local area network (LAN) for communication between system elements. Thus it is reasonable to assume that the communication in the system may be conducted over a public network, including the Internet, through both wired and wireless channels of access.

Even cursory analysis of the system indicates that the communication scheme is twofold: on the one hand, real-time multimedia communication, related to the transmission of large amounts of binary data forming the video streams, on the other hand complex, interactive control communication (signaling) based on the exchange of short, structured messages. This characteristic brings the presented system towards the solutions typical for Internet telephony and Instant Messaging applications.

In order to fulfill these requirements, the XMPP (Extensible Messaging and Presence Protocol) has been implemented as a foundation of the system. [5] The XMPP transport layer uses a TLS 1.0 (Transport Layer Security), so it is possible to apply encryption and integrity control of the transmission easily. Data transmission is performed via so-called XML streams, which means that every message is a fragment (so-called stanza) of XML document. This allows, in principle, unlimited extensibility of the protocol, since the interpretation of a message depends only on the XML

namespace, by which its root element is qualified. This feature has been practically used in the presented framework for implementation of e.g. notifications on the detection of a specific event, defining the rules governing the work of analysis algorithms or remote management of node stations. It should be noted that in the context of the framework, the XMPP protocol provides a transport layer, and constitutes a platform for running of the services realizing proper functionality related to monitoring.

2.4 Multimedia Communications (Streaming)

An important extension of XMPP protocol, which is applied in the framework, is the Jingle subsystem [6]. XMPP, being a protocol based on TCP/TLS and exchange of XML-formatted messages is a suboptimal solution when it comes to the real-time transmission of multimedia data, since its use introduces a significant overhead associated with encoding of the binary data. It also causes prolonged delays during the transmission resulting from the use of TCP congestion control mechanisms, which is unacceptable from the real-time communication point of view. Therefore, the Jingle extension was developed. It allows to establish real-time communications session external to the XMPP session (out-of-band).

Jingle extension defines the control protocol (signaling), used for initiating and supervising of the proper communication session, which uses the RTP protocol for the transmission of multimedia content. RTP protocol is a standard tool for delivering real-time transmission. In fact, the Secure RTP (SRTP) extension is employed, enabling encryption and integrity control of the transmission. The actual transmission of video content is performed with the use of lossy compression, however the exact type and quality parameters of the compression algorithm depend on the capabilities of the target terminal and underlying network connection. They are negotiated during transmission establishment stage.

An important aspect of the transmission of multimedia material is the issue of transmission of metadata generated by the algorithms for automatic analysis of video streams. This is particularly important from the standpoint of verification and reproduction of decision-making process that resulted in detection of a specific event. This also allows for visualization of the algorithm. These metadata take the binary form and include intermediate and final results of the analysis for subsequent image frames, so that the operator console application may overlay on the original image additional elements, such as bounding rectangles around detected objects. Because of the temporal dependencies between packets of metadata and successive frames of the source material (and sometimes their substantial size) it is justified to treat them as an additional real-time stream transmitted within the same session, and synchronized with the stream they describe. In this case the fact is exploited that a single Jingle communication session may consist of multiple simultaneous media "lines". This is typically designed for video calls, when the picture and sound are sent in parallel "lines" of RTP, allowing for easy fallback to purely-voice call when not supported or not desired by one of the calling parties. During the Jingle session negotiation it is determined the same way whether metadata transmission is possible and necessary.

3 Framework Software Modules

The majority of software modules within the framework is dedicated to detection of events based on real-time analysis of video streams. The analysis is divided into two levels. During low-level image analysis, image processing is performed for each fixed camera independently and includes moving object detection, tracking and classification. Based on the results, high-level image analysis tracks moving object activity in order to detect various events happening in a parking lot. Software modules are complemented with PTZ camera control.

3.1 Low-Level Video Image Processing

Low-level video image processing consists of operations dedicated to moving object detection, tracking and classification. Low-level video image processing for the purpose of event detection includes operations performed independently in video frames acquired from each fixed camera. They are related to moving object detection, tracking and classification. Moving object detection is based on the background modeling method utilizing Gaussian Mixtures while object tracking is performed with Kalman filters. In case of tracking conflicts, visual object features are analyzed in order to provide correct association [7].

Each object moving in a camera field of view is classified into one of two categories: vehicles or other objects (i.e. persons). For the purpose of classification process, a set of SVM classifiers (Support Vector Machines) is used. Classifiers, for a specified camera observation angle, are trained using previously prepared synthetic 3D models. These models represent various vehicle types and people in different stages of movement. Such models are then projected to a 2D plane and parameterized. The description vector contains values corresponding to the normalized number of pixels falling into certain radial and distance bins defined relatively to the center of gravity (CoG) of an analyzed mask [8].

3.2 Event Detection

Detecting parking vehicles is the most important event recognized by the framework. This event includes identification of the parking event time, place and vehicle. Detection of the parking moment (i.e. when a vehicle has parked) is implemented as a set of rules. First of all, only moving objects classified as vehicles are analyzed. A vehicle is considered as parked if it entered any parking place, stopped while still being in the parking place and stayed in the vicinity of its current position for the defined amount of time. Parking places are defined as polygons "drawn" in the image frame (Fig. 2). Detailed description of the parking event detection is provided in [9].

Whenever a parking vehicle is detected, its parking place is identified. This task is complicated in case of small angles between the camera viewing axis and the ground because of the negative impact of the vehicle's height on its perspective projection on the ground level (Fig. 2). Furthermore, in case of wide angle lenses, significant differences are also visible within the same video frame.

Fig. 2. Parking places (light polygons) selected in video frames from two cameras with different viewing angles

In order to make parking place identification possible for small camera viewing angles, all cameras in the system are calibrated with Tsai's algorithm in order to allow bi-directional conversion of 3D coordinates in the real world into 2D coordinates in the video frame [10]. Each parking place for the purpose of its identification is modeled with a right prism (usually a cuboid) in 3D real world coordinate system. Its lower base is formed by a polygon labeled in a video frame that is converted to 3D coordinates, with the assumption that it is placed on the ground level. Upper base of the right prism is formed by its lower base lifted 1.5 meters above the ground. The height of the prism corresponds to the average height of cars. For each parking place, centers of the upper of and the lower bases of the prisms are calculated and projected back to the image plane. In order to determine where a vehicle has parked, the minimum distance from the upper and lower bases of all parking places is found.

Vehicles identification involves tracking the vehicle moving in the parking lot from the gate to its final stop. For this purpose, moving objects are identified correctly in different cameras' fields of view. This task is complex because of various camera orientations, white balance and lighting conditions. The developed algorithm for object identification and tracking in different video cameras utilizes object appearance features, such as statistical features of co-occurrence matrices, 2D colour histogram based on an rg chromaticity space and Speed Up Robust Features (SURF) local image descriptor. Additionally, spatial and temporal constraints are introduced into the matching algorithm that are related to cameras' placement in a parking area and its topography [9].

Besides detecting parking vehicles, a few supplementary events occurring in a parking lot are detected. First of all, vehicles stopping for a longer time on a road (i.e. outside any parking place) are detected. The algorithm is similar to parking vehicle detection. This makes possible to detect invalid parking that may lead to a road block. Furthermore, all vehicles that enter or leave a parking lot are detected in order to count them and to alert a system operator on a new incoming vehicle. This functionality is implemented by finding vehicles entering or leaving the field of view of the entrance camera in the correct hot spot.

Detection of people entering and leaving the office building is implemented by finding an object classified as a person who walks through the two hot areas marked in a video frame (the first one is placed at the door area and the second one - on the walking path leading to the door) in the correct direction.

3.3 PTZ Camera Control

The framework utilizes PTZ cameras. They may be aimed automatically at a desired spot in the monitored area by selecting any point in video frames from fixed cameras, presented in the operator console. Furthermore, if a moving object is selected, it is automatically tracked by all PTZ cameras regardless of possible transitions between fixed cameras' fields of view. Aiming PTZ camera at any point found in a fixed camera field of view requires a calibrated environment. Additionally, in case of moving object tracking, the delay caused by video image processing is compensated by aiming PTZ camera at the predicted object position [10].

4 Experiments and Results

A prototype video surveillance installation covering a part of the parking lot around an office building has been established. The installation contains 8 fixed cameras (and one additional camera pointed at the parking entrance only) with different orientations in relation to the ground, covering a total number of 54 different parking places. Polygons denoting parking place locations were labeled manually. The installation is supplemented with 2 PTZ cameras. Video streams from fixed cameras are analyzed by 4 servers.

Approximately 16 hours of video streams from all cameras, covering two weekdays have been analyzed by the system in the real time in order to detect all events. The results were verified manually.

Detailed results of event detection divided by event types are shown in Table 1. False negative ratio is calculated as the number of false negatives divided by the number of events. False positive ratio is defined as the share of false positives in the total amount of events generated by the system (false positives plus number of events). Results of event detection are characterized by small false negative rate and larger false positive rate. Approx. 97% of all events have been detected correctly. Missed events were caused by object detection and tracking errors in a crowded scene, when a vehicle bounding rectangle was not allowed to remain still for the required period of time.

False positives account for approx. 10% of events reported by the system. They are caused mainly by errors in object tracking (multiple trackers associated with the same real object) and by mistakes in object classification (e.g. a person stopping in a parking place). The highest number of false positives is related to parking and stopping on the road events. This causes significant difficulties for the system operators. Therefore, the ratio should be decreased with additional processing stages (e.g. suppressing reporting a parking event for the same place within a second).

Table 1. Detailed results of event detection

Event type	Number of events	Correct detections		False negatives		False positives	
Parking	152	143	94.1%	9	5.9%	37	20.6%
Stopping on the road	89	87	97.8%	2	2.3%	32	26.9%
Entering parking	271	265	97.8%	6	2.2%	8	2.9%
Leaving parking	231	217	93.9%	14	6.1%	31	12.5%
Entering building	208	205	98.6%	3	1.4%	17	7.7%
Leaving building	146	146	100.0%	0	0.00%	0	0.00%
All events	1097	1063	96.9%	34	3.1%	125	10.5%

Total accuracy of event detection calculated for each camera independently varies from 92% to 99% of correctly detected events which proves the algorithms can handle a large variety of camera orientations with respect to the ground.

Results of parking vehicle detection presented in the first row of Table 1 are related to identification of parking time only. Parking places have been identified correctly for all parking events, regardless of various camera viewing angles. For events requiring vehicle identification ("parking" and "stopping on the road"), it succeeded in over 85% of cases. It means that in these cases a vehicle has been tracked successively during its movement around the parking lot since its entering into the field of view of any camera to the moment the event has been detected. Depending on the vehicle route, vehicle identification requires correct vehicle tracking in and between the fields of view of up to 8 video cameras. The majority of errors in vehicle identification is caused by low-level object tracking errors, especially related to conflict resolving, and by unpredicted vehicle behavior (e.g. a long stop between cameras fields of view, stopping and going back, etc.). The latter case needs to be analyzed in order to tune the algorithms.

5 Conclusions

The paper presented design and implementation of the distributed framework for automatic detection of various events occurring in a parking lot. The system is based on visual analysis of video streams in the multi-camera environment. The results gathered during real condition experiments prove that the software is able to detect events with a high accuracy. However, the framework is not yet suitable for practical use because of relatively high amount of false positives generated. Therefore, future work will be focused on reducing these types of errors by introducing additional reasoning level to the system in order to filter false events. Another task will be devoted to increasing the accuracy of low-level image processing algorithms which will improve global effectiveness of the framework.

Acknowledgments

Research is subsidized by the European Commission within FP7 project "INDECT" (Grant Agreement No. 218086). Authors wish to thank the Gdansk Science and Technology Park and PPBW sp. z o.o. company for their help in establishing the test bed of the framework described in the paper.

References

1. Black, J., Velastin, S.A., Boghossian, B.: A real time surveillance system for metropolitan railways. In: Proc. IEEE Conf. Adv. Video Signal Based Surveillance, pp. 189–194 (2005)
2. Bird, N.D., Masoud, O., Papanikolopoulos, N.P., Isaacs, A.: Detection of loitering individuals in public transportation areas. IEEE Trans. Intell. Transp. Syst. 6(2), 167–177 (2005)
3. Ghazal, M., Vazquez, C., Amer, A.: Real-time automatic detection of vandalism behavior in video sequences. In: Proc. IEEE Int. Conf. Syst., Man, Cybern., pp. 1056–1060 (2007)
4. Datta, A., Shah, M., Lobo, N.D.V.: Person-on-person violence detection in video data. In: Proc. of the 16th Int. Conf. on Pattern Recognition, vol. 1, pp. 433–438 (2001)
5. Saint-Andre, P.: Extensible Messaging and Presence Protocol (XMPP): Core. RFC 3920 (October 2004)
6. Ludwig, S., Beda, J., Saint-Andre, P., McQueen, R., Egan, S., Hildebrand., J.: Jingle. XEP-0166 (Draft) (December 2009)
7. Czyzewski, A., Dalka, P.: Moving object detection and tracking for the purpose of multimodal surveillance system in urban areas. In: New Directions in Intelligent Interactive Multimedia, pp. 75–84. Springer, Heidelberg (2008)
8. Ellwart, D., Czyzewski, A.: Camera angle invariant shape recognition in surveillance systems. In: The 3rd International Symposium on Intelligent and Interactive Multimedia: Systems and Services, pp. 34–41 (2010)
9. Dalka, P., Ellwart, D., Szwoch, G.: Camera Orientation-Independent Parking Events Detection. In: Proc. of 12th Int. Workshop on Image Analysis for Mult. Interact. Services (WIAMIS), Delft, The Netherlands (2011) (accepted)
10. Szwoch, G., Dalka, P.: Automatic detection of abandoned luggage employing a dual camera system. In: Proc. of 3rd IEEE Int. Conf. on Multimedia Communications, Services & Security, Krakow, Poland, pp. 56–61 (May 2010)

INCR — INDECT Multimedia Crawler

Michał Grega, Andrzej Głowacz, Wojciech Anzel,
Seweryn Lach, and Jan Musiał

AGH University of Science and Technology

Abstract. The INDECT Project aims to develop tools for enhancing security of citizens and protecting confidentiality of recorded and stored information. One of the INDECT modules, Multimedia INDECT Crawler (INCR), has been developed to support crawling of public Internet resources with the aim of combating serious online crimes, including child pornography. In this paper we present three crawler plugins, which have responded to research challenges. The plugins described here are the shape detection plugin, the Optical Character Recognition (OCR) plugin, and the face detection plugin. We present the content-based image analysis methods chosen for each plugin to accomplish its task, as well as some results in terms of detection accuracy.

Keywords: Multimedia, crawler, security, Internet.

1 Introduction

The INDECT Project aims to develop tools for enhancing security of citizens and protecting confidentiality of recorded and stored information. In INDECT the threats are considered both in virtual and real environments. Furthermore, INDECT elaborates some horizontal technologies. As far as threats in a virtual environment are concerned, INDECT targets serious crimes such as Internet child pornography, trafficking in human organs, and the spread of botnets, viruses, malware.

One of the INDECT modules, Multimedia INDECT Crawler (INCR), has been developed to support crawling of public Internet resources, with the aim of combating serious online crimes such as child pornography. The application scenario assumes that (i) the user provides an address (set of addresses) to the crawler, (ii) images are automatically retrieved from the website for a given depth, (iii) images are analysed in search for a given visual pattern, and finally (iv) the user receives the list of the suspicious pages.

In this paper we present three crawler plugins which have responded to research challenges. The plugins described here are the shape detection plugin, the Optical Character Recognition (OCR) plugin, and the face detection plugin. We present the content-based image analysis method chosen for each plugin to accomplish its task, as well as some results in terms of detection accuracy.

Tools currently available to police forensics offer text-based web crawling only. One example of such a tool is the Picus crawling library [5]. To the best of the

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 46–54, 2011.

authors' knowledge, there are no forensic multimedia crawlers currently available. When needed, the police usually utilise general purpose crawlers and search services such as Google Images [2] and Google Videos [3]. However, these solutions do not offer sophisticated content-based analysis of crawled multimedia, except very simple analysis of image type (face, photo, clip art, line drawing) and colour (full colour, black and white).

The rest of the paper is structured as follows... Section 2 presents the general concept of the crawling application. The three crawler plug-ins are described in the three following sections: Section 3 — shape detection plug-in, Section 4 — OCR plug-in and Section 5 — face detection plug-in. Further work plans are presented in Section 6, while Section 7 concludes the paper.

2 General Concept of Application

This section describes the general concept of the application and its architecture (presented in Figure 1. The goal of INCR is to automate the tedious task of browsing the resources of the Internet in search of law-breaking content. As an example, the police forces are informed that an open Internet forum displays and disseminates images related to a serious criminal activity. Police officers need to manually browse through the forum (which may come to thousands or more pages) in order to gather as much evidence as possible. Such approach is very time-consuming and ineffective. However, if police officers are equipped with INCR software, they can spend much less time on mechanical tasks such as browsing and are able to focus on other important aspects of the investigation.

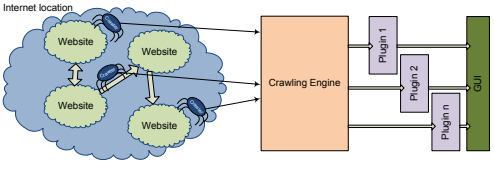


Fig. 1. Architecture of the INCR application

The application consists of several modules. At the heart of the application is the website crawler, which can be further split into the crawling engine and the crawlers. The task of the crawlers is to analyse and parse website code. The crawlers focus on detecting links to other web pages and detecting multimedia. If a link is detected, the crawler has an option to follow that link deeper into the website structure. If a media item (a movie or an image) is detected, it is passed on to the engine for further analysis.

The crawling engine is responsible for directing the crawlers within the Internet. Depending on the task the crawler may be prevented from or forced to follow specific links. For example, crawlers may be forced to remain within one domain, or be prevented from crawling into popular and trusted websites such as Wikipedia.

Once a media item is detected it is downloaded by the crawling engine and passed on to the media analysis plugins. These plugins are responsible for performing specific tasks defined by the system operator. For example, in the example presented in the beginning of the section, the police officers may use the visual pattern detection plugin.

Currently the following plugins are available or at the research stage or are being researched:

- *Detection of website elements.* A WWW website includes images forming part of its graphical design, as well as images and photographs related to the content. The former need to be filtered out before the latter are analysed by other plugins in order to increase the overall speed of the analysis. This plugin is currently available in INCR [6].
- *Face detection.* This plugin detects whether a face is present in a photograph. This plugin is currently available in the INCR.
- *Symbol detection.* This plugin is designed to detect whether a given symbol (such as a swastika) is present in an image downloaded from a website. Depiction of totalitarian symbols is forbidden in certain countries (e.g. Poland). This plugin is currently available in the INCR.
- *Universal shape detection.* This plugin will be capable of detecting any type of symbol present in a website based on examples provided by the police. This plugin is currently under development.
- *Child pornography detection.* This plugin will be able to classify whether a photograph is an example of child pornography, mainly based on nudity detection and age classification. This plugin is currently under development.
- *Automatic Speech Recognition(ASR).* This plugin will be capable of analysing audio tracks of videos embedded in websites in search for spoken phrases. This plugin is currently under development.
- *Optical Character Recognition(OCR).* This plugin will be capable of analysing images and videos in search for text phrases. This plugin is currently under development.
- *Face recognition.* This plugin will be capable of identifying a face based on provided examples. This plugin is currently under development and evaluation.

The last component of the INCR architecture is the Graphical User Interface designed for ease of use. It is accessible through any web browser.

It should be noted that while we believe in lawful INCR application, it has the potential for being misused. Instead of searching for criminals, it is possible to imagine the software being flooded with images originating from central or government databases such as passport databases or border control systems. The face recognition plugin especially could be misused if it is used for a purpose other than searching for criminals, such as finding individuals for private purposes and getting more of their personal data. Therefore access to the crawling tool should be limited and restricted in order to make it accessible only to persons in charge of investigations.

Access attempts and crawling requests should be logged in order to detect unauthorised access and usage attempts. All communication with the crawler tool should be secured in order to protect the confidentiality of the investigation. The system should also provide high levels of security for confidential personal data.

A good practice for the developers of Internet crawlers is to obey rules outlined in the robots.txt file that may be located in the website root directory. This file defines which publicly available parts of the website are to be off-bounds to crawlers. However, when designing a forensic tool to be used by the police during investigations, this file needs to be ignored. Using a real life example, no-one can avoid the policeman's eye by putting a "Don't look at me" sticker on their car's bumper.

3 Shape Detection

This section describes the processes of development and testing of a shape detector. The shape detector is already available in INCR.

3.1 Detector Training

The detector was developed using the Haar cascade technology. It is a very efficient and up-to-date object recognition method, based on machine learning. The concept is based on gathering an appropriate group of objects, followed by supervised training resulting in the creation of a vector of features. This vector contains descriptions of common trends detected in training groups. It uses a rejection cascade of nodes, where each node is a multi-tree classifier designed to have a high (say, 99.9%) detection rate (low false negatives or missed objects) and a low (50%) rejection rate (high false positives wrongly classified). For each node, a "not in class" result at any stage of the cascade terminates the computation, and the algorithm exits with no object existing at that location. Thus final detection is declared only if the calculation makes it through the entire cascade. At each node, the cascade compares Haar-like features in the region selected as a potential sign. The program takes features from the image, compares them with features stored in a special vector, and passes the results on to next node. Open source tools for training such detectors are included in the OpenCV library. The training was mainly performed by gathering a sufficiently large collection of images containing symbols, and iterative runs of a proper tool. The detection results were improved by adjusting training parameters and enlarging the size of the set of images containing the symbol. The outcome of the training was an XML file, containing Haar-cascade information.

3.2 Detection Results

The detector was tested in a number ways. At the development stage, it was repeatedly checked with an OpenCV statistics tool and small sets of images.

At the later stages, it was checked manually on larger image groups. At the final stage, an image segmentation algorithm was used before detecting.

Results Without Segmentation. The OpenCV statistical method compares the special description file (containing coordinates of the symbols in the images) with the detector's output listing. It only gives statistical information about the detector.

Manual checking means running the detector on a proper test-set (56 images containing the symbols and 36 non-containing) and looking at the result in every image. This gives good information on how the detector behaves for particular images. This shows domains where the detector behaves satisfactorily. Specificity and sensitivity measures were estimated (Table 1).

Table 1. Specificity and sensitivity measures

*S*pecificity [%]	*S*ensitivity [%]
80	50

Results with Segmentation. Segmentation subdivides an image into its constituent objects. The regions are homogenous with regard to selected features, such as:

- grey level
- colour
- texture recognition

The pre-segmentation process was used to extract symbols of interest from the input image. Segmented objects are passed to input of the recognition plugin. This operation may have helped the detector to find the appropriate mark. Methods of segmentation are described in detail in [7].

In order to conduct software tests, sets of images both containing and not containing symbols of interest were employed (Table 2). The segmentation algorithm was able to extract 37 correct symbols from images containing symbols.

Table 2. Test set

All images	Images with symbol
92	56

The assumption of the segmentation process was that the images were segmented before recognition was commenced. As shown in Table 3, this pre-segmentation action gives more selectivity in detection. On the other hand the standalone version was able to classify several cases more frequently than the previous method.

Table 3. Test results

	Result with pre-segmentation	Standalone detector
True positives	25	32
False positives	3	7

The best case was computed by three indicators as shown in Table 4.

$$F_1 = \frac{2TP}{(2TP + FP + FN)} \tag{1}$$

$$Accuracy = \frac{TP + TN}{(TP + FP + TN + FN)} \tag{2}$$

$$Balanced\ Accuracy = 0.5(\frac{TP}{(TP + FN)} + \frac{TN}{(FP + TN)}), \tag{3}$$

where
TP — true positive
FP — false positive
TN — true negative
FN — false negative

As shown in all three cases, the standalone version of the detector accomplishes better result than the version with a pre-segmentation process extension.

Table 4. Test results in space measures

	Result with pre-segmentation	Standalone detector
F_1	0.595	0.674
Accuracy	0.630	0.663
Balanced Accuracy	0.682	0.689

The experiment shows that the pre-segmentation process does not make a meaningful difference when compared to a standalone algorithm. A better solution is to try to focus on the standalone version of the detector and optimise it.

4 Optical Character Recognition

In this module Optical Character Recognition (OCR) methods are applied to visual analysis of multimedia content. The aim of this component is the automatic processing of videos and images in order to create a searchable representation of text that appeared in the content. Such representation (in the form of annotations or keywords) can be analysed further to detect dangerous objects. This is presented in an example below.

4.1 OCR in Video Content

Video sequences can contain potentially dangerous content. In this scenario, six video sequences showing explosions of home-made bombs were analysed. They included visible text instructions (in Polish) appearing on the screen, corresponding to the production and detonation of explosives being demonstrated. In order to analyse the content, the video file is acquired and split into frames. Frames are processed using the open-source Tesseract OCR engine [4]. The software is configured to accept Polish special characters. Contrary to common OCR processes performed in documents, raw output of video OCR is messy (many incorrectly recognised words), mainly because of video compression, low resolution and low quality. Therefore results require further filtering. After the whole process is completed, the output keywords can be analysed for known indicators of threats.

4.2 Filtration Process

Several algorithms are used in the process of filtering algorithms. The basic idea behind this is to develop universal word rejection mechanisms rather than using specific dictionary-based filtering. The selected approach allows for efficient searching, preserves proper nouns and is language-independent. First, raw text is pre-processed (using conversion of characters, removing non-printable characters, punctuation marks, etc.) for preliminary word classification. In the second step word candidates are selected based on the ratio of positive and false characters. Post-filtering employs analysis of consecutive characters to remove further false words (errors generated by some visual patterns similar to characters). A list of stopwords is then used to remove insignificant words such as “a”, “the”, “are”, “when”, etc. The final process produces a list of keywords and their location within the video sequence. Multiple keyword occurrences within a single video file can be compressed as well.

4.3 Recognition and Filtration Results

Each video sequence contained a few minutes of recording. The raw OCR output for analysed files was a total of 5667 lines of text. As expected, the majority of the results were messy. In fact most of the sequences showed explosions or preparations, and the exact text was a minor component. This low ratio of text to non-text frames posed an additional challenge to the filtering algorithm. However, the use of developed methods allowed for extraction of 93 searchable keywords. They were then analysed for known threat indicators. In result, all sequences could be correctly classified as potentially dangerous, based on words including “bomb”, “bombs”, “detonate”, “smoke”, “explosion”, “fireworks”, “GRG”, “charge”, “dangerous”, “nitrate”, and “shock”. The video OCR tool can automatically scan large volumes of multimedia data and is useful in terms of preliminary selection of dangerous content. However, textual context still requires human verification.

5 Face Detection

The Face Detection module is used to select candidate image regions to be indexed for face recognition in the next step. The input of this component is an image (or video frame), whereas the output is facial coordinates (upper-left and lower-right corners) and detection weight. Should there be more faces present in the image, the output contains multiple records. Face detection methods use two characteristic approaches. One relies on skin detection with characteristic colour and shape information. Another uses location of characteristic points (e.g. eyes) as they determine whole face position, even when rotated. Image-based face identification methods can be subdivided into appearance-based and model-based approaches. In this study we present results obtained with the open-source OpenCV library [1]. OpenCV provides various face detectors based on Haar cascades, which are evaluated on example below.

5.1 Detection Results

Analysis of threats, which is a top goal of the system, requires robust detection. It is assumed that it is better to have many less accurate results than few more accurate ones. Therefore OpenCV face cascades (*frontalface_default*, *frontalface_alt*, *frontalface_alt_tree*, *frontalface_alt2*, *profileface*) were compared and combined in a practical scenario. The test set consisted of 516 still frames of 22 different faces of hooligans. Photos from a stadium were taken in moderate conditions (light rain) using a 360x288 pixel camera. Due to the weather conditions some people were wearing hoods, caps, scarves etc., which made detection more difficult for the software. Frontal face detectors (*frontalface_default*, *frontalface_alt*, *frontalface_alt_tree*) achieved overall accuracies of: 72.73%, 77.27%, 54.55% and 68.18% respectively. The profile face detector achieved an accuracy of 72.73%.

5.2 Combined Face Detection Results

It is worth pointing out, that some faces (7 of 22) was detected either by single frontal detector or single profile detector. In order to fulfil criterion to maximize number of located faces, detectors are combined together. Only *frontalface_alt_tree* yielded results below average and no further advantage. Four other classifiers together provided the cumulative detection rate of 100% in this scenario, and each face was located on average by 2.91 of 4 selected detectors (72.75%). For comparison: a simple detector based on eye-location achieved 40.90% and another commercial software available to the authors achieved only 36.36% accuracy.

6 Further Work

Future work on the INCR software will focus mainly on integration of the new detection and recognition plugins with the crawler. Extensive end-user tests with the participation of police officers are planned.

One of the plugins currently under development is a universal shape detector based on shape descriptors of the MPEG-7 standard. Such a plugin will allow the detection of shapes based on a single example rather than a pre-trained Haar cascade.

Preparations are also underway to extend the functionality of the INCR software by the ability to download and analyse video files downloadable or embedded in websites.

7 Conclusions

In this paper, we presented three crawler plug-ins. For each of them, we presented the content-based image analysis method chosen to accomplish the task as well as some results in terms of detection accuracy. The shape detector plug-in was tested in a number of ways. Specificity (80%) and sensitivity (50%) measures were estimated. The face detector plug-in achieved an accuracy of 72.75%.

Acknowledgements

This work was supported by the European Commission under the Grant INDECT No. FP7–218086. Development of tools (swastika detector, OCR filtering and face detector) was supported by European Regional Development Fund within INSIGMA project no. POIG.01.01.02–00–062/09.

References

1. Collective work: Open Computer Vision Library (2010), http://opencv.willowgarage.com
2. Google: Google Images, http://images.google.com/
3. Google: Google Videos, http://video.google.com/
4. HP Labs, Google: Tesseract OCR (2010), http://code.google.com/p/tesseract-ocr/
5. Lubaszewski, W., Dorosz, K., Korzycki, M.: INDECT Deliverable D4.4. System for Enhanced Search: A Tool for Pattern Based Information Retrieval. Tech. rep. AGH University of Science and Technology (December 2009), http://www.indect-project.eu/files/deliverables/public/deliverable-4.4
6. Grega, M.: Automated classification of images into photographs and graphics. In: IEEE International Conference on Multimedia Communications, Services and Security (2009)
7. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Prentice-Hall, Englewood Cliffs (2002)

Detection and Localization of Selected Acoustic Events in 3D Acoustic Field for Smart Surveillance Applications

Józef Kotus, Kuba Łopatka, and Andrzej Czyżewski

Multimedia Systems Dept., Gdansk University of Technology,
Narutowicza 11-12, 80-233 Gdansk, Poland
{joseph,klopatka,andcz}@sound.eti.pg.gda.pl

Abstract. A method for automatic determination of position of chosen sound events such as speech signals and impulse sounds in 3-dimensional space is presented. The events are localized in the presence of sound reflections employing acoustic vector sensors. Human voice and impulsive sounds are detected using adaptive detectors based on modified peak-valley difference (PVD) parameter and sound pressure level. Localization based on signals from the multichannel acoustic vector probe is performed upon the detection. The described algorithms can be useful in a surveillance systems to monitor the behavior of participants of public events. The results can be used to detect the position of sound source in real time or to calculate the spatial distribution of sounds in the environment. Moreover, spatial filtration can be performed to separate the sounds coming from the chosen direction.

Keywords: sound detection, sound localization, audio surveillance.

1 Introduction

This paper introduces a method for detection and localization of acoustic events in 3-dimensional space. This solution is meant to be employed in a security surveillance systems to provide the functionality of recognizing and localizing unwanted events automatically. A possible scenario of using this kind of algorithms is the monitoring of audience during mass events, e.g. sport events. The aim of the described work is to develop a demonstration system employing above mentioned technology. The experimental system is installed in an auditory hall.. It comprises an acoustic vector sensors and cameras. The behavior of the audience is analyzed multimodally using video analytics and sound analysis [1, 2, 3]. This paper focuses on the acoustic part of this system. The concept diagram of the proposed demonstration system is presented in Fig. 1. The employed audio processing algorithms operate on signals from a 3D acoustic vector sensor (AVS). The detected events are related to audience activity, such as asking questions or disturbing the lecture. Therefore, it is necessary do detect the voice activity. If the voice activity is detected, the speaking person can be localized in the audience using sound localization algorithms. The output of the 3D AVS is composed of 4 signals: acoustic pressure p and particle velocity components for 3 directions (v_x, v_y, v_z) [5]. The modules of detection of impulse sounds and speakers use the acoustic pressure signal only. The localization algorithm processes all 4 signals to

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 55–63, 2011.

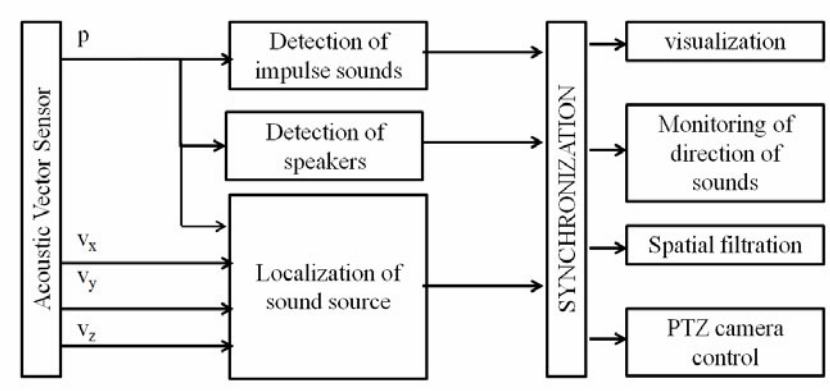


Fig. 1. Concept diagram of the system

determine the dominant direction of arriving sound in 3 dimensions. Knowing the shape of the room, the exact 3D position of the sound source can be determined.

Synchronization of results of event detection and sound source localization is needed. The results are used in different ways. The direction of incoming sound can be visualized in the user's interface. Information about localization of sound sources is also useful for monitoring of direction of sounds and creation of the map of acoustic activity. Spatial filtration of the sound field can also be performed to separate the sounds from the chosen direction. Finally, the angle of arriving sound is used to control the Pan-Tilt-Zoom (PTZ) camera to automatically point it towards the direction of the sound source [3].

2 Detection of Acoustic Events

The goal of the proposed system is to localize chosen acoustic events, i.e. speech signals and impulsive sounds. Therefore, an algorithm for detection of such events is needed. The system employs separate algorithms for detection of the two types of signals. The speech sounds are detected using an adaptive threshold algorithm based on the modified peak-valley difference (PVD) parameter. The detection of impulse sounds also utilizes adaptive threshold, however the parameter used for detection is simple, namely the equivalent level of sound pressure is utilized.

2.1 Detection of Speech Signals

The peak-valley difference parameter (PVD) is used in voice activity detectors (VAD), which are part of speech processing or recognition systems [4]. The parameter is based on the difference between spectral peaks and valleys of vowels, which are always present in speech. In the proposed method the parameter was modified due to following reasons:

- in speech processing the sound is usually sampled at 22050 samples per second. In the present application it is sampled at 48000 samples per second

- the localization frame covers 4096 samples. In most speech processing applications, shorter frames are used
- the 4096 Discrete Fourier Transform (DFT) representation of the signal is used to find the spectral peaks of sound
- the distribution of peaks and valleys in the spectrum of speech signals is dependent on the fundamental frequency of speech
- in classic PVD detection the model of peak distribution in vowels needs to be established before calculating the parameter.

To calculate the modified PVD first, the magnitude spectrum of the signal is estimated, using 4096 point DFT. Next, it is assumed that the fundamental frequency of speech (f_0) is located in the range of 80-300Hz (for speakers of both gender). The fundamental frequency is expressed in the domain of DFT indices and denoted as k_0. Consequently, the expected difference between spectral peaks equals k_0. Thus the distribution of peaks for the assumed fundamental frequency is known without the need for establishing a model of vowel spectral peak distribution. The PVD parameter is then calculated according to the formula (1):

$$PVD = \frac{\sum_{k=1}^{N/2} X(k) \cdot P(k)}{\sum_{k=1}^{N/2} P(k)} - \frac{\sum_{k=1}^{N/2} X(k) \cdot (1 - P(k))}{\sum_{k=1}^{N/2} (1 - P(k))} \tag{1}$$

where: $X(k)$ is the magnitude spectrum, N =4096 is the length of the Fourier Transform and $P(k)$ is the function, which equals 1 if k is the position of spectral peak, 0 otherwise.

The PVD parameter is extracted iteratively for every value of k_0 from the range corresponding to the assumed range of fundamental frequencies. The maximum value is achieved when the current k_0 matches the actual fundamental frequency of the present speech signal. This maximum value is assigned to the parameter for further processing. For non-periodic signals the PVD is bound to achieve smaller values than for periodic signals, due to smaller difference between the two components of the formula (1). Results from 16 frames are buffered and the mean value of $\overline{PVD}$ is calculated in order to automatically determine the detection threshold,. The instantaneous threshold T_i equals $m \cdot \overline{PVD}$ where m is the threshold multiplication factor. For example, $m=3$ means that the PVD should exceed 3 times the average value from 16 last frames to trigger speech detection. The parameter m can be adjusted to change the sensitivity of the detector. Finally, the adaptation of the threshold T in the frame number i is calculated using exponential averaging with the constant α=0.001 according to the formula (2):

$$T(i) = (\alpha - 1) \cdot T(i-1) + \alpha \cdot T_i \tag{2}$$

2.2 Detection of Impulsive Sounds

The impulsive sounds are detected based on the energy of the signal. The level L of the current frame is calculated according to the formula (3):

$$L = 20 \cdot \log\left(\sqrt{\frac{1}{N}\sum_{n=1}^{N}\left(x(n)\cdot L_{norm}\right)^2}\right) \tag{3}$$

where N = 4096 is the number of samples in the frame, L_{norm} is the normalization level corresponding to maximum sample value. The signal level is expressed in dB. It is assumed that full scale of the signal corresponds to 120 dB. The described system operates in an environment, where the level of acoustic background is low. Hence, frames with the signal level exceeding the 75 dB threshold value are assumed as containing impulse sounds.

3 Localization of the Sound Source

The algorithm of localization of the sound source comprises two operations. The first operation is the calculation of the *x y z* components of intensity vector of the acoustic field given the signals from the multichannel acoustic vector sensor. Then, knowing the angle, the position of the AVS inside the room and the shape of the room, localization of the sound source can be determined as defined in formula (4) [2, 5]:

$$\vec{I} = I_x\vec{e}_x + I_y\vec{e}_y + I_z\vec{e}_z \tag{4}$$

The employed method for detecting the sound source inside the room using the acoustic vector sensor is presented in Fig. 2.

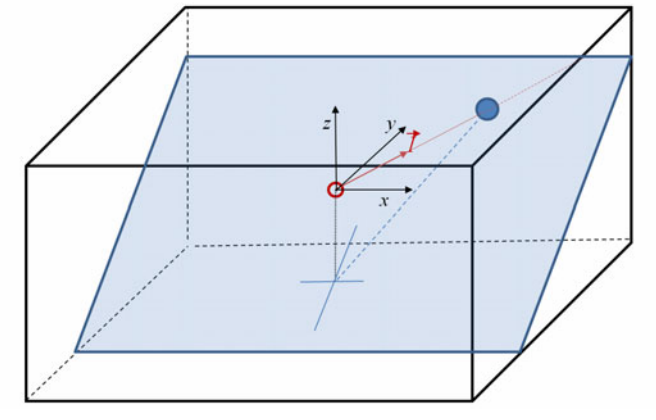


Fig. 2. Illustration of the employed method for detection of sound source inside a room using acoustic vector sensor

The large cuboid represents the shape of the interior. The rhomboid models the floor plane. The AVS is placed in the room above the floor in the place marked by the red empty dot. The dotted line corresponds to the height of the AVS placement. The two intersecting blue lines indicate the point of the perpendicular of the location of the AVS to the plane of the floor. The full dot marks the position of the sound source

inside the interior. The vector *I* of the intensity of the acoustic field calculated in the *xyz* space, has the direction of the arrow. The coordinate system, starting in the location of the AVS, is drawn. The intersection of the direction of the intensity vector with the floor plane indicates the position of the sound source.

4 Measurement System

A measurement system was set up in a lecture hall in Gdansk University of Technology, in the Faculty of Electronics building. It is composed of a fixed camera covering the audience, an acoustic vector sensor, the AVS conditioning module and a computer used for data acquisition. With the use of this demonstration system signals used for evaluation of algorithms were recorded. In Fig. 3 the setup of the measurement system is presented.

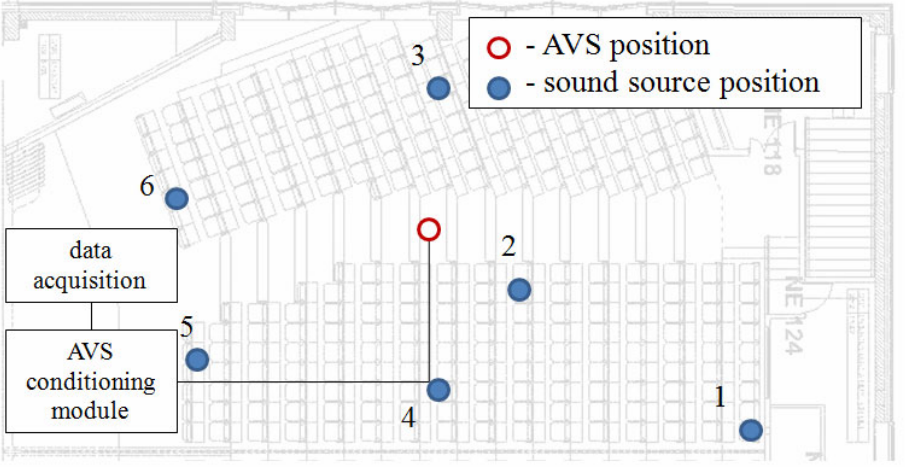


Fig. 3. Setup of measurement system

The placement of the acoustic vector sensor and the positions of sound sources (named 1-6) are presented on top of the layout of the lecture hall. In the photographs the practical case of the system at work is presented, in which the listeners occupy different places in the audience and act as sound sources.

5 Experimental Results

Signals registered during the experiment with the described measurement system were analyzed using aforementioned audio signal processing algorithms. Some example results of detection and localization of impulse sound sources are presented in this section. The output of the speech detection algorithm is shown. Basing on the results of the detection of speech sounds, an effort is made to localize the position of the speaker.

5.1 Detection Results

To assess the ability of the algorithm for detecting speech signals, a fragment of the measured signal of acoustic pressure was chosen. It contains words spoken by two speakers located at opposite areas in the auditory room (sources 2 and 3 in Fig. 3). The results of speech detection are presented in Fig. 4.

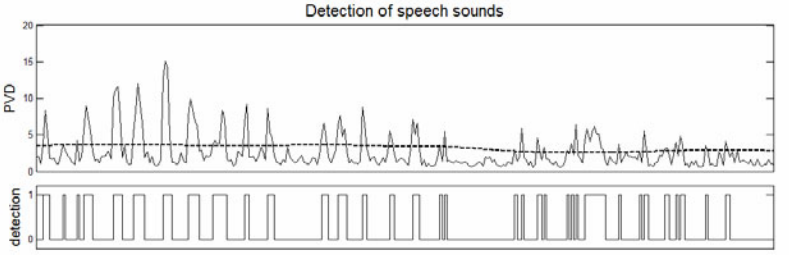


Fig. 4. Detection of speech sounds

The solid line on the top chart represents the plot of the PVD parameter. The dashed line is the adaptive threshold of detection. The bottom plot illustrates the decision of the detector. The detection of impulse sounds is presented using a fragment of the test signal containing clicks of the noise gun (shooting without bullets). The results are presented in Fig. 5.

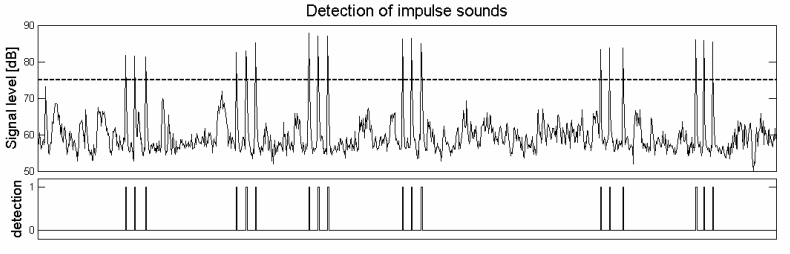


Fig. 5. Results of detecting impulsive sounds

A number of 18 shots was emitted - 3 shots from the position of every sound source 1-6. The dashed line o the top plot indicates the 75 dB threshold of detection of impulsive sounds.

5.2 Localization Results

A proper sound source localization can take place when the AVS captures the front wave of the acoustic event. Subsequent fragment of sound can include reverberant components produced by reflections from the walls and another objects present inside the room. For that reason, the sound event detection algorithm was modified to determine the attack phase properly. The impulsive sounds were analyzed employing frames of 1024 samples with 48 kHz sampling frequency. For speech sound events

4096 sample frame length was used. To improve the localization efficiency the band-pass filtration from 300 Hz to 3kHz was used. This frequency range was ideal not only to speech signals but also for impulsive ones. In this way many reflections, especially for higher frequencies were eliminated. In Fig. 6 the computed results of the sound source localization in two dimensions were presented. The left figure presents results for broadband impulsive sounds, the right plot was created for sounds processed with the pass-band filtration. In Fig. 7 computed results of the speech sound source localization in two dimensions were presented.

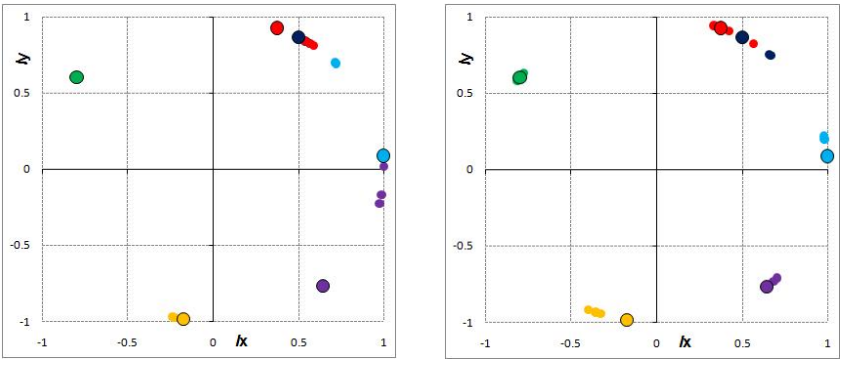


Fig. 6. 2D localization results for impulsive sound events

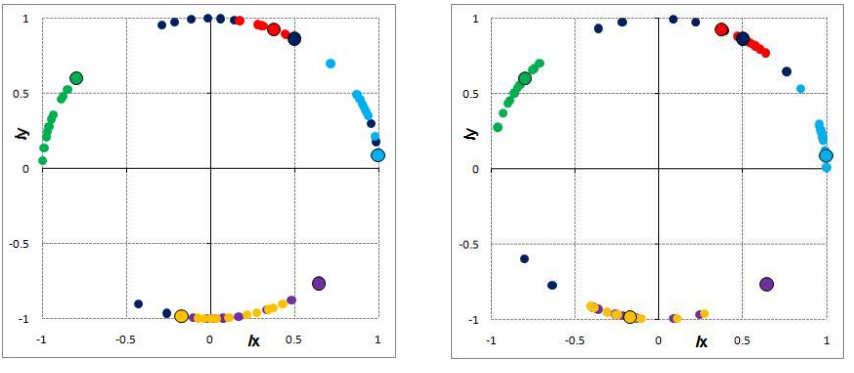


Fig. 7. Localization results for speech sounds events in 2D

The left figure presents results for broadband sounds, the right plot was obtained using pass-band filtration. The greater colored circles determine the position of the sound source. In Figs 8 and 9 the visualization of 3D localization results were shown. The lines represent the direction of the computed sound intensity vector. For impulsive sound we did not observe the crosscut between the intensity vector and the plane of the floor. It is because the impulsive sound events were produced using the signal pistol (only the spire sound was generated) above the volunteers head. It is well visible in Fig. 8 (right).

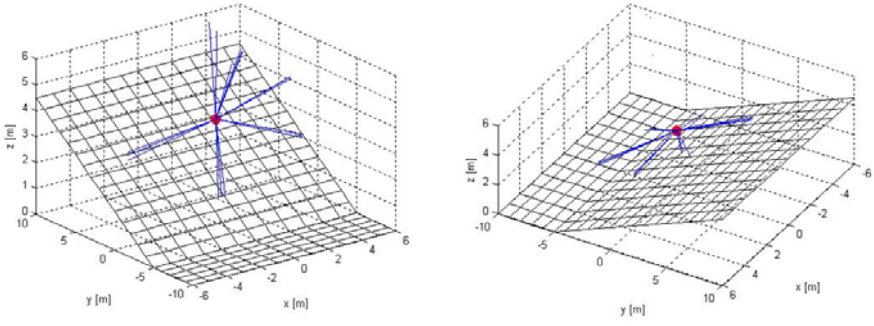


Fig. 8. Impulsive sound source localization results for all positions (in different orientations)

In Fig. 9 the obtained results for speaker localization were shown. The left part of Fig. 9 presents the proper sound source localization (the intensity vector crosscuts the plane of the floor). In the right part of Fig. 9 the intensity components for imaginary sources can by observed. Such a kind of vectors was produced because the speaker was directed into the front wall (the AVS was behind the head). For that reason the speaker's head produced the acoustic shadow. The sound, which was localized by the USP sensor was in fact the reflection from the wall.

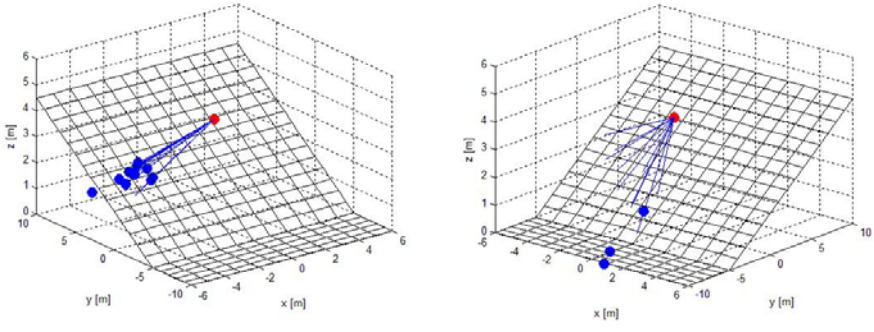


Fig. 9. Speaker localization, example results for speaker 3 and 6 (confront Fig. 3 for details)

The root mean squared error (RMSE) indicator was used for evaluation of the presented algorithm [7]. The computed values of RMSE for impulsive sounds were equal to 8.0 with filtration and 24.4 without filtration. For speech signals these values were equal to 39.1 and 42.5 respectively. Only dominant sound source was localized in the same time for typical acoustic background. More information about sensitivity and accuracy can be found in our previous work [1, 2].

6 Conclusions

A method for 3-dimensional analysis of acoustic field in interiors, allowing detection and localization of acoustic events was presented. The proper sound source localization

in the presence of reverberations in real time was possible, not only for impulsive sound but also for speech sound events. The sound source position was determined using a single 3D acoustic vector sensor. Since the attack slope of the sound includes the information of the sound source position the most important feature in this context is a proper detection of sound events.. The additional pass-band filtration significantly improved the localization of the sound source.

The experimental results are promising, as far as the functionality of monitoring the activity of people is concerned. The described algorithms can be useful in a surveillance systems to monitor the behavior of participants of public events. To completely assess the system's ability to detect events, a further detection efficiency evaluation is needed and some additional measurements of the precision of the localization should be carried out.

Acknowledgements

Research is subsidized by the European Commission within FP7 project "INDECT" (Grant Agreement No. 218086).

References

1. Kotus, J.: Application of passive acoustic radar to automatic localization, tracking and classification of sound sources. Information Technologies 18, 111–116 (2010)
2. Czyżewski, A., Kotus, J.: Automatic localization and continuous tracking of mobile sound source using passive acoustic radar, Military University of Technology, pp. 441–453 (2010)
3. Kotus, J., Łopatka, K., Kopaczewski, K., Czyżewski, A.: Automatic Audio-Visual Threat Detection. In: MCSS 2010: IEEE International Conference on Multimedia Communications, Services and Security, Kraków, Polska, May 6-7, pp. 140–144 (2010)
4. Yoo, I., Yook, D.: Robust voice activity detection using the spectral peaks of vowel sounds. Journal of the Electronics and Telecommunication Research Institute 31, 4 (2009)
5. Basten, T., de Bree, H.-E., Tijs, E.: Localization and tracking of aircraft with ground based 3D sound probes, ERF33, Kazan, Russia (2007)
6. Smith, S.W.: The Scientist and Engineer's Guide to Digital Signal Processing. California Technical Publishing (1997)
7. Lehmann, E.L., Casella, G.: Theory of Point Estimation, 2nd edn. Springer, New York (1998)

Brightness Correction and Stereovision Impression Based Methods of Perceived Quality Improvement of CCTV Video Sequences*

Julian Balcerek, Adam Konieczka, Adam Dąbrowski,
Mateusz Stankiewicz, and Agnieszka Krzykowska

Poznań University of Technology, Institute of Control and System Engineering,
Division of Signal Processing and Electronic Systems
Piotrowo 3a Street, 60-965 Poznań, Poland
{julian.balcerek,adam.konieczka,adam.dabrowski}@put.poznan.pl

Abstract. In this paper brightness correction and stereovision impression based methods of the perceived quality improvement of CCTV video sequences are presented. These methods are helpful for the monitoring operator in order to evoke attention during a long time observation. Clearness of picture is increased by using local brightness correction method. Another available option is a 3D visualization that can be used, e.g., to observe important image regions, which require an increased attention. Stereovision impression experiments and a real-time 2D to 3D video conversion tool are described.

Keywords: CCTV, brightness correction, stereovision, anaglyph, 2D to 3D video conversion.

1 Introduction

CCTV (*closed-circuit television*) is a system that is commonly used for surveillance in urban areas that may need monitoring. The resulting job of a monitoring system officer (operator), who must observe the monitor screen for many hours, is quite monotonous and exhausting. In this paper we propose some new image processing tools that can serve as optional aids for the monitoring operator, helping him / her to stay actively conscious for a long time. This is because we offer view diversity and improvement of the perceived quality of images. Two mechanisms are proposed:

- increase or decrease of local brightness in order to increase the image clearness and thus the information perception
- real-time 2D to 3D conversion in order to increase the subjective informative content of the observed scenes.

It has to be stressed that the proposed image transformations are optional, thus they will never burden the monitoring operator. They are helpful in special situations only, e.g., if there is an interesting detail, which requires an increased attention. An experimental software has been prepared for verification of the proposed concepts.

* The paper was prepared within the INDECT project.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 64–72, 2011.

Monitored events taking place in evenings and nights, filmed with artificial illumination, may result in low contrast images with poor informative content. The proposed local increase of brightness transforms, as a matter of fact, the view from a real to somehow artificial world but may essentially improve the human perception.

Stereovision impressions in a typical 2D monitoring system are possible, e.g., with the use of special (red / cyan glasses). The scenes are then perceived as if they contained collections of objects at different distances to the viewer. This phenomenon may be used for visualization of important events that have to be noticed and correctly interpreted.

Experiments on human 3D perception have been performed and the proper ways of image processing have been found [9]. Due to 3D impressions, the monitoring operator job is easier as his / her sight may easily be focused on the most important image regions and / or objects such as: people, moving vehicles, etc. This method can also help the monitoring operator to notice and remember important details.

The paper is structured as follows. After an introduction in this section, local brightness correction method is presented in Section 2. Section 3 is devoted to description of the 3D visualization tool together with the stereovision impression experiments and the real-time 2D to 3D video conversion. Conclusions are formulated in the last section.

2 Local Brightness Correction Method

Basic approaches to sharpness improvement, exposure correction, brightness, contrast, and color adjustment, and digital noise removing are presented in [1], [2], [3], and [4]. In this paper a new method for clearness and visibility improvement of CCTV frames captured under difficult circumstances (especially in dark places or at night), is proposed.

In first step of our method for each CCTV frame an RGB to HSV color space conversion is proceeded. H and S components remain not modified. V component image is divided into square blocks starting from the top left corner of the picture. Sizes of all blocks are the same. Number of blocks in the shorter side of the picture is defined by the user as a control parameter. In the next step, for each component block the minimal and the maximal V values are selected. Then all V values are linearly scaled to the whole available range using formula

$$V_{\mathrm{n}} = \left(\frac{V - V_{\min}}{V_{\max} - V_{\min}} \right), \tag{1}$$

where:

V_{n}– new V component value of a pixel,

V – V component value of this pixel in the original picture,

$V_{\max}$ – maximal V component value in the block,

$V_{\min}$ – minimal V component value in the block.

In order to avoid creation pseudo-edges in the final image, this operation must be repeated. Number of iterations is the same as the one dimensional pixel size of the block. Computations in each iteration (excluding the first one) start one pixel below

and one pixel to the right in relation to the previous iteration. In the next stage, new V component values must be averaged for each pixel. Near the picture edges original pixel values are left.

The proposed algorithm facilitates the viewer to perceive differences in brightness of dark objects. It helps in person identification by inspection, in reading inscriptions (i.e. license plates), or in distinguishing colors of very dark picture elements in CCTV recordings. Unfortunately, the proposed local brightness enhancement is a cause of digital noise enhancement in the picture. However, despite this inconvenience, recognition of noisy objects is much easier. This is an important advantage of the proposed method in comparison to classic manual or automatic brightness / contrast adjustment. Furthermore, our method is fully automatic.

As already mentioned, the only parameter, which user can modify, is the number of picture blocks n. If n is small, the visual effectiveness is low and the computational complexity is high (many iterations), however if n is too large, the final picture is not clear enough. It was experimentally proved, that the optimal value for n is about 10. The results of the proposed method for various values of n are presented in Fig. 1.

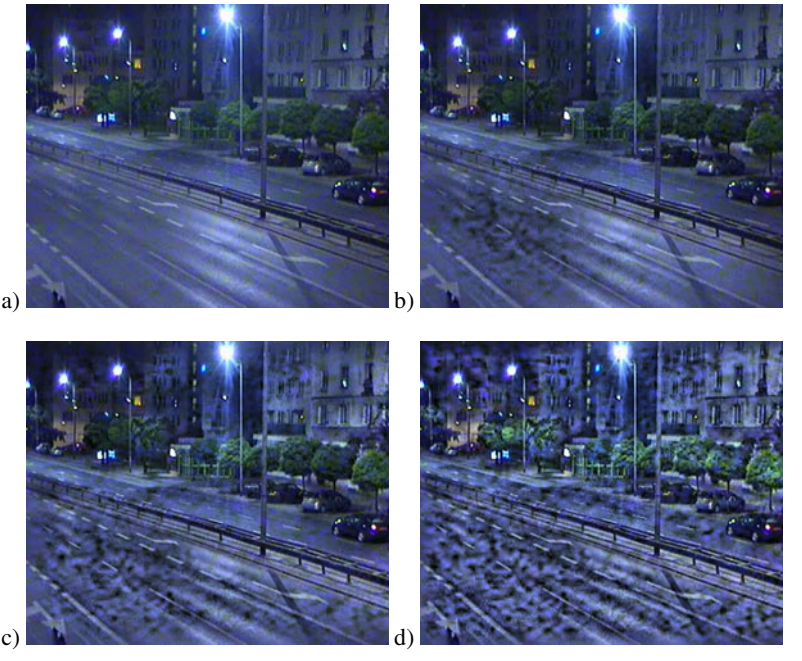


Fig. 1. Local brightness correction: a) original picture, b) picture processed with n=5, c) picture processed with n=10, d) picture processed with n=20

Compression loss of the original picture does not have any considerable influence on the quality of the final picture. In extreme cases, i.e., if the compression factor is greater than 100:1, compression noise artifacts may be visible.

3 3D Visualization Tool

3.1 Stereovision Impression Experiments

Differences in viewpoints between two two-dimensional (2D) images observed separately by the left and the right eye simultaneously are interpreted by human brain as a three-dimensional (3D) scene [5]. A strict 3D reproduction is influenced by the choice of the interaxial distance, focal length of the reference camera, and the convergence distance [6]. An interesting feature of human 3D perception is that in spite of some incomplete or even inconsistent 3D information, evoking the three-dimensional illusion is still possible.

Almost entire visible color space can be modeled with three color components as RGB. Even if one eye (e.g., the left one) reaches only a single (e.g., the red) component of the left eye image, and the right eye sees the two remaining components (i.e., green and blue) of the right eye image, our brain perceives not only a 3D scene but sees it in quasi-true colors [7], [8]. This 3D effect can be realized using classic 2D equipment by observing specially prepared flat images referred to as the *anaglyph images* with red and cyan filter glasses. The red glass removes blue and green components from the left eye view while the cyan glass filters out the red component from the right eye view.

We have experimentally observed that even if merely the red color component is simply horizontally shifted, a 3D effect appears. This is the way to transform 2D images into their suggestive, perhaps not exact but plausible 3D anaglyph versions. We have conducted a series of experiments. The first and the second experiments were only preliminary experiments in small group of viewers and the third experiment was the main experiment. PUT workers and students were examined.

In our first experiment we observed that during increasing the red component shift value, the 3D effect was stronger and stronger but the image quality was decreasing, because some annoying artifacts near to the image contours appeared to be stronger and stronger visible. Lastly, starting from a specific value of the shift, two separate images without any 3D effect were perceived. In 83% of cases the 3D effect occurred. Among viewers, who observed the 3D effect when the red component was shifted to the left, in 92% of cases viewers noticed that the distance to the overall image surface was increased. When, however, the red component was shifted to the right, in 67% of the examined cases, the viewers noticed that the distance to the overall image surface was decreased. The others viewers perceived the exactly opposite effect [9]. However, the subjective informative content of the image seemed in both cases to be increased.

In our second experiment among viewers who observed a 3D effect when the red component was shifted to the left, in 71% of the cases the viewers noticed the increasing distance to the overall image surface. When the red component was shifted to the right, in 66% of the cases they also noticed that the distance was increasing [9].

In our third experiment test images were observed by 30 viewers. Resolutions of original images were 1280×1024 pixels. In 4 test pictures the red color component was shifted from 0 to 20% (0–255 pixels) of the original image width, both to the left and to the right side. Shifting step was 20 pixels because in case of a smaller step

differences between pictures were too weak and experiment could be too tiring for the viewers. For each original picture there were 26 red component shifted pictures, 13 to the left and 13 to the right.

Application implemented in *C++* programming language and *QT* library was used (Fig. 2). Four different images were displayed on a 15.4 inches diagonal wide-screen with about 0.5 m distance from the viewer to the screen. Images were rescaled to the 960×768 pixels resolution in order to fit the viewing area for tests.

Viewers rated images by answering to given questions. The following questions and scoring criteria have been used. The first question was: "Can you see any 3D effect in the picture?" and there were two possible answers "yes" or "no" according to "1" or "0" value. The second question was: "Are there any visible artifacts?" and there were also two answers "yes" or "no" according to "1" and "0" value. The third question was "How does the distance between the viewer and the displayed objects change?" and two possible answers "increases" and "decreases" according to "1" and "0" value. If the viewer did not see any 3D effect, the button for this question was deactivated. The last question was "What is the perceptible quality of the picture?" and there were three possible answers "good", "medium" and "bad" according to the values "3", "2" and "1".

Fig. 2. Application window of for image testing

Results of our experiments explain how the human brain interprets and perceives quality of the image, occurrence of the 3D effect, and the level of artifacts after simple red component shifts (Fig. 3). They are the basis for development of simple methods for real-time 2D to 3D image or video conversion in CCTV systems.

3.2 Real-Time 2D to 3D Conversion

An example of a 3-dimensional CCTV application is a 3D motion detection and assessment system described in [10]. Among possible advantages of using 3D instead of a 2D classic system is reduction of false alarms in case of difficult environment conditions like shadows, reflections, rain, or snow. Another advantage of using the

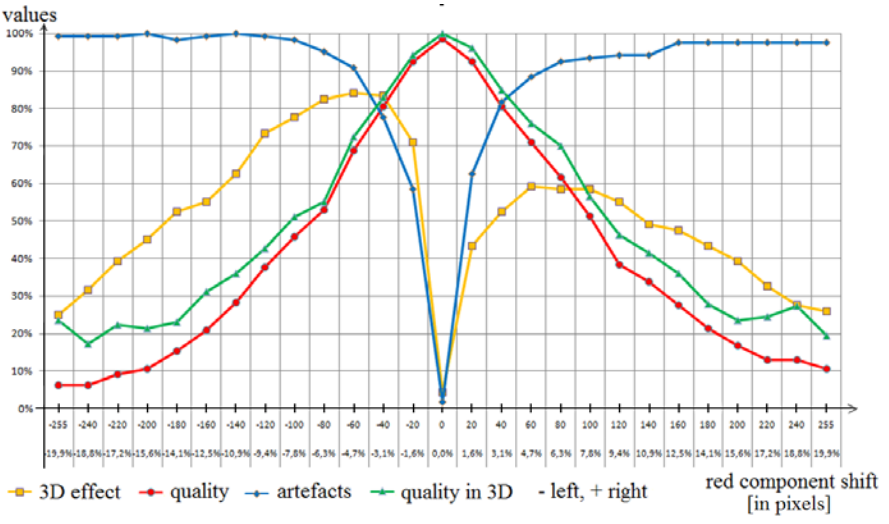


Fig. 3. Results of experiments with red component shifts

3D approach, i.e., a system with more than one camera, is reliability. Indeed, in case of a camera failure the system can still operate with the remaining camera(s).

Both classic and stereovision systems can be used for obstacle and vehicle detection in traffic [11], [12]. In case of a single-camera system the vehicle localization and tracking should be based on a 2D image sequence [11]. However, the precise vehicle distance can only be computed using a stereovision system based on the offset, measured between the left and the right images. Pairs of the corresponding points in the left and the right images have to be found and mapped into a calibrated 3D world model using the stereo geometry. Then moving objects can detected and tracked using a quite involved algorithm [12].

We suggest another approach, which, instead of the described automatic detection, can help people to extract information from the classic 2D monitoring system. Thus in our case a 3D effect is realized with 2D to 3D conversion methods, which must be simple enough to operate in real-time on a video sequence. We assume that the anaglyph images are generated and are observed with red and cyan filter glasses. The proposed real-time 2D to 3D conversion schema is illustrated in Fig. 4.

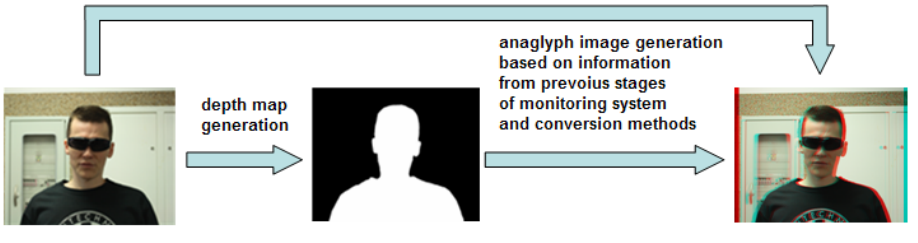


Fig. 4. A real-time 2D to 3D conversion schema for monitoring system

An *OpenCV* computer vision library is used. Examples of windows of a real-time 2D to 3D conversion application are presented in Fig. 5.

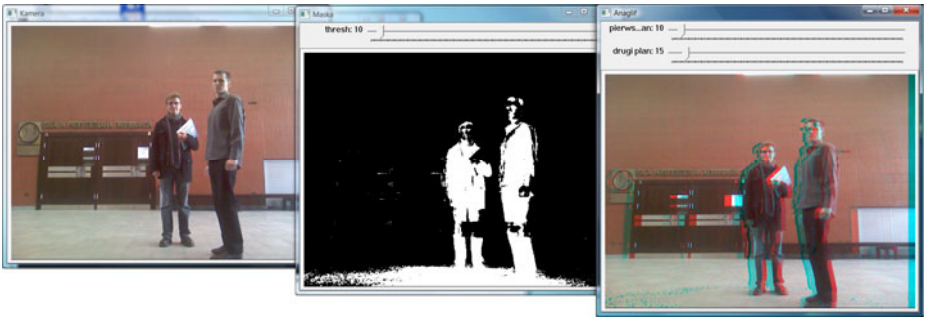


Fig. 5. Examples of windows of a real-time 2D to 3D conversion application using a direct shift method

We proposed simple techniques for creating effective 3D impressions (stereovision illusion of depth) from 2D original images. They are based on the simplest, i.e., binary depth map [9]. The following algorithms were suggested: direct shift, direct shift with interpolation, segment scaling, and segment shifting (Figs. 6 and 7).

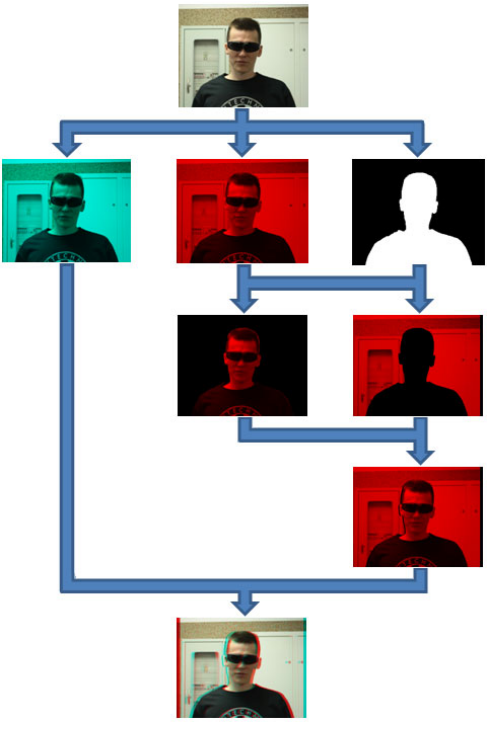


Fig. 6. Anaglyph generation for direct shift method [9]

Fig. 7. Red color component and anaglyph generation using: a) direct shift with interpolation, b) segment scaling, c) segment shifting [9]

The anaglyphs in Figs. 6 and 7 are of astonishingly good quality due to strong and tolerant human abilities of perception the 3D effects despite various kinds of distortions. Our two simple methods (segment scaling and segment shifting) based on a binary depth map and specific image deformations allow to obtain quite natural 3D effects with almost imperceptible artifacts [9].

4 Conclusions

The described approach for image clarity improvement and 2D to 3D transformations can be used in CCTV systems for processing both video and for still images.

Current state of experiments with the proposed operator aids proved their effectiveness and encourages for further work on this subject. In future more involved depth maps and more advanced image enhancement methods might be used.

References

1. Yibin, Y., Boroczky, L.: A new enhancement method for digital video applications. IEEE Transactions on Consumer Electronics 48(3), 435–443 (2002)
2. Cuizhu, S., Keman, Y., Jiang, L.: Shipeng Li Automatic image quality improvement for videoconferencing. In: Proceedings of 2004 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), vol. 3, pp. 701–704 (2004)
3. Zicheng, L., Cha, Z., Zhengyou, Z.: Learning-based perceptual image quality improvement for video conferencing. In: 2007 IEEE International Conference on Multimedia and Expo, pp. 1035–1038 (2007)
4. Hough, K., Marshall, S.: Soft Morphological Filters Applied to the Removal of Noise From CCTV Footage. In: The IEE International Symposium on Imaging for Crime Detection and Prevention, ICDP, pp. 61–66 (2005)
5. Onural, L.: An Overview of Research In 3DTV. Systems, Signals and Image Processing. In: 14th International Workshop on 2007 and 6th EURASIP Conference Focused on Speech and Image Processing, Multimedia Communications and Services, June 27-30, p. 3 (2007)

6. Fehn, C.: Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV. In: Proc. of the SPIE Stereoscopic Displays and Virtual Reality Systems XI San Jose, CA, USA, pp. 93–104 (2004)
7. Dumbreck, A.A., Smith, C.W.: 3-D TV displays for industrial applications. In: IEE Colloquium on Stereoscopic Television, October 15, pp. 7/1–7/4 (1992)
8. Dubois, E.: A projection method to generate anaglyph stereo images. In: Proc. (ICASSP 2001) 2001 IEEE International Conference on Acoustics, Speech, and Signal Processing, May 7-11, vol. 3, pp. 1661–1664 (2001)
9. Balcerek, J., Dąbrowski, A., Konieczka, A.: Simple efficient techniques for creating effective 3D impressions from 2D original images. In: Proc. of Signal Processing NTAV/SPA 2008, Poland Section, Chapter Circuits and Systems IEEE, Poznań, Poland, September 25-27, pp. 219–224 (2008)
10. Nelson, C.L.: 3-dimensional video motion detection and assessment. In: Proc. of IEEE 38th Annual 2004 International Carnahan Conference on Security Technology, October 11-14 (2004)
11. Bertozzi, M., Broggi, M., Fascioli, A., Nichele, S.: Stereo Vision-based Vehicle Detection. In: Proc. of the IEEE Intelligent Vehicles Symposium 2000, Dearbon (MI), USA, October 3-5 (2000)
12. Nedevschi, S., Danescu, R., Frentiu, D., Marita, T., Oniga, F., Pocol, C., Schmidt, R., Graf, T.: High Accuracy Stereo Vision System for Far Distance Obstacle Detection. In: 2004 IEEE Intelligent Vehicles Symposium, University of Parma, Parma, Italy, June 14-17, pp. 292–297 (2004)

Semantic Structure Matching Recommendation Algorithm

Andrzej Szwabe, Arkadiusz Jachnik, Andrzej Figaj, and Michał Blinkiewicz

Institute of Control and Information Engineering, Poznan University of Technology, ul. M. Curie-Sklodowskiej 5, Poznan, Poland
{andrzej.szwabe,andrzej.figaj,arkadiusz.jachnik}@put.poznan.pl, michal.blinkiewicz@gmail.com

Abstract. The paper presents a new hybrid schema matching algorithm: *Semantic Structure Matching Recommendation Algorithm* (SSRMA). SSRMA is able to discover lexical correspondences without breaking the structural ones — it is capable of rejecting trivial lexical similarities, if the structural context suggests that a given matching is inadequate. The algorithm enables achieving results that are comparable to those obtained by means of state-of-the-art schema matching solutions. The presented method involves an adaptable pre-processing and flexible internal data representation, which allows to use a variety of auxiliary data (e.g., textual corpora) and to increase the accuracy of semantic matches accommodated in a given domain. In order to increase the mapping quality, the method allows to extend the input data by auxiliary information that may have the form of ontologies or textual corpora.

Keywords: metadata, XML, schema matching.

1 Introduction

Although the information, especially in a multimedia domain, is currently often tagged or semantically described, the usage of such metadata is limited because of the diversity of data models spread in various systems. The data schema matching is an important problem that appears in different domains, like database schema integration, web services integration, bioinformatics, etc. [1]. Over the years, a variety of matching algorithms and systems have been proposed. Their heterogeneity is a result of the diversity of application areas and characteristic of the input data as well as non-triviality of achieving matching scalability, while taking into account the number of attributes, cardinality of alignments, and other features and factors.

This paper describes a schema matching algorithm, called SSMRA (*Semantic Structure Matching Recommendation Algorithm*). In our method, the input data has a form of graphs (e.g. acyclic directed graphs – trees) representing different types of relations between nodes (e.g. *is-a*, *has-a* relations). The graphs are

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 73–81, 2011.

then transformed into an adjacency matrix. The method will be examined in EU INDECT Project[1] [2].

1.1 Problem Statement

The approach presented in this paper is in line with the trends in the current research on schema matching and ontology alignment. In particular, the usage of external data sources increases the *Precision* and *Recall* [3], [4].

According to [1], in most cases the best matching systems take into account both structural (constraints-based) and lexical information. Designers of matchers try to increase the *Precision* and *Recall* of their systems by including auxiliary information in the analysis. They use lexical knowledge (e.g., analysing word similarities between concepts), domain knowledge (e.g., ontologies, taxonomies etc.) and structural knowledge (e.g., derived from the structure of the input).

There are two basic approaches to processing data about a given data structure. Matching systems like *COMA++* [5], *LSD* [6] use different matchers for each knowledge source. In consequence, lexical similarities are calculated independently of the similarities that are derived from structures of input schemas. Such an approach forces end-users to decide how outputs from different matchers should be used together (parallel, sequential, etc.) and how the final results are created. Another approach is to integrate different knowledge sources together as is done in the case of hybrid matchers, like *SEMINT* [7], *Cupid* [8]. The drawback of the hybrid approach is typically the inflexibility of matching methods.

In the two abovementioned cases, a fundamental problem arises: how to achieve a matching quality that is good enough to help users not only in trivial (e.g., lexical) but also in more complex (e.g., structure-dependant, contextual) matching tasks. In this paper, a *Semantic Structure Matching Recommendation Algorithm* is presented, which, thanks to the flexibility of the proposed data representation, effectively uses data that may be inaccessible to other methods (e.g., having the form of a vector-space), what, in turn, improves the matching accuracy.

2 State of the Art

The current matching solutions have a lot in common. The usage of thesauri and dictionaries as an auxiliary knowledge base is very common [9]. An exception is *Similarity Flooding* [10], which uses only string similarities between element names. Graph representation (acyclic graphs, trees, labelled graphs) is the most popular model of internal representation (e.g. *SEMINT*, *LSD*, *COMA++*, *SF*, *Cupid*). It stands in contrast to the method described in this paper, which uses

[1] The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement n°218086 — Project INDECT.

unified vector space representation. Some of the approaches demand a learning phase in order to provide alignments (*LSD*, *GLUE* [11], *SEMINT*), which is not necessary in the presented solution. Only few of the existing methods (e.g CtxMatch [12]) are capable of giving information about the types of relations revealed between structures [13]. The presented method is able to suggest the direction of the relationships between nodes and the type of the relations. However, this aspect is still under study. Most of the matching approaches mentioned above use *Latent Semantic Analysis* (LSA) [14] for computing semantic similarity of documents.

3 Semantic Structure Matching Recommendation Algorithm

The input data for the algorithm are as follows: parsed graphs (source graph G^I and target graph G^{II}; e.g. XMLs), singular vectors of terms (collections V^{G^I} and $V^{G^{II}}$) and weights (a set of weights of sub-matrices S^I, S^{II}, C^I, C^{II}, the weight of diagonals d_m connecting sub-matrices of relations, the weight of term vectors $v_{t^I_{i,j}}$ and $v_{t^{II}_{k,l}}$ and the weight of diagonal d_v 'combining' vectors of terms). The output data for the algorithm is a list of matches retrieved from the reconstructed matrix R'^s.

3.1 Step 1: Building Sub-matrices of Relations

Figure 1 presents an example of the structure where N_x represents a node, $t_{x,y}$ represents a token of the x node (each node may be described by a set of terms); lines stand for the relations between nodes.

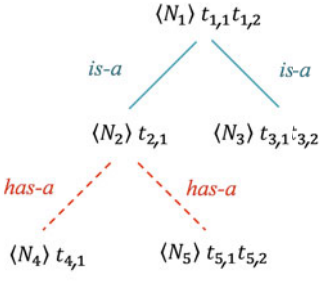


Fig. 1. Example of a mapped structure

Each type of relation (e.g. "is-a", "has-a", etc.) is processed separately. Each sub-graph of each relation is represented by a separate adjacency matrix. An adjacency matrix is used as a means of representation, in which vertices of a graph are adjacent to other vertices (nodes). The adjacency matrix allows us to

model reflective/unreflective and directed as well as undirected relations. Entries equal to 1 denote adjacencies between nodes. When a relation connects nodes that consist of more than one term the value is divided between each cell. Table 1 presents the relation matrices for the presented example.

Table 1. Relations matrices for the given example

has-a		N_1		N_2	N_3		N_4	N_5	
		$t_{1,1}$	$t_{1,2}$	$t_{2,1}$	$t_{3,1}$	$t_{3,2}$	$t_{4,1}$	$t_{5,1}$	$t_{5,2}$
N_1	$t_{1,1}$	1	0	0	0	0	0	0	0
	$t_{1,2}$	0	1	0	0	0	0	0	0
N_2	$t_{2,1}$	0	0	1	0	0	0	0	0
N_3	$t_{3,1}$	0	0	0	1	0	0	0	0
	$t_{3,2}$	0	0	0	0	1	0	0	0
N_4	$t_{4,1}$	0	0	**1**	0	0	1	0	0
N_5	$t_{5,1}$	0	0	**0,5**	0	0	0	1	0
	$t_{5,2}$	0	0	**0,5**	0	0	0	0	1

is-a		N_1		N_2	N_3		N_4	N_5	
		$t_{1,1}$	$t_{1,2}$	$t_{2,1}$	$t_{3,1}$	$t_{3,2}$	$t_{4,1}$	$t_{5,1}$	$t_{5,2}$
N_1	$t_{1,1}$	1	0	0	0	0	0	0	0
	$t_{1,2}$	0	1	0	0	0	0	0	0
N_2	$t_{2,1}$	**0,5**	**0,5**	1	0	0	0	0	0
N_3	$t_{3,1}$	**0,3**	**0,3**	0	1	0	0	0	0
	$t_{3,2}$	**0,3**	**0,3**	0	0	1	0	0	0
N_4	$t_{4,1}$	0	0	0	0	0	1	0	0
N_5	$t_{5,1}$	0	0	0	0	0	0	1	0
	$t_{5,2}$	0	0	0	0	0	0	0	1

In order to cope with multiple terms in one node, an artificial *coString* relation has been introduced. It is a term-term matrix where each row and its corresponding column describes the relation of co-existence of string tokens representing the node's multi-term name. The *coString* relation groups the terms from the single node. For example, as in Table 2, terms $t_{1,1}$ and $t_{1,2}$ are in the *coString* relation, because they co-create the string of N_1 node name.

Table 2. Example of *coString* relation

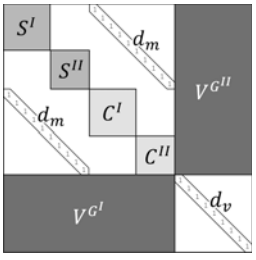
coString		N_1		N_2	N_3		N_4	N_5	
		$t_{1,1}$	$t_{1,2}$	$t_{2,1}$	$t_{3,1}$	$t_{3,2}$	$t_{1,1}$	$t_{1,2}$	$t_{2,1}$
N_1	$t_{1,1}$	**1**	**1**	0	0	0	0	0	0
	$t_{1,2}$	**1**	**1**	0	0	0	0	0	0
N_2	$t_{2,1}$	0	0	**1**	0	0	0	0	0
N_3	$t_{3,1}$	0	0	0	**1**	**1**	0	0	0
	$t_{3,2}$	0	0	0	**1**	**1**	0	0	0
N_4	$t_{4,1}$	0	0	0	0	0	**1**	0	0
N_5	$t_{5,1}$	0	0	0	0	0	0	**1**	**1**
	$t_{5,2}$	0	0	0	0	0	0	**1**	**1**

3.2 Step 2: Preparation of Vectors for all Pre-processed Terms

Vector space obtained from auxiliary data as a result of *LSA* is stored in the database. Thanks to applying LSA (SVD — *Singular Value Decomposition* or RSVD — *Randomized Singular Value Decomposition* [15] for large corpora) the dimensionality of such vector space is limited (by applying a k-cut).

3.3 Step 3: Creation of the Input Matrix *A*

The final matrix of input for the algorithm is the merger of all the relation matrices and textual vector-space. Schema of the full input matrix A and notations are presented in the Figure 2.



G^I	first (source) structure graph
G^{II}	second (target) structure graph
S^I	sub-matrix of *matches-to* relations of the graph G^I
S^{II}	sub-matrix of *matches-to* relations of the graph G^{II}
C^I	sub-matrix of *coString* relations of the graph G^I
C^{II}	sub-matrix of *coString* relations of the graph G^{II}
V^{G^I}	set of term vectors of graph G^I
$V^{G^{II}}$	set of term vectors of graph G^{II}
d_m	diagonal matrices 'connecting' sub-matrices of relations
d_v	diagonal matrix 'connecting' vector-space

Fig. 2. Scheme of an initial matrix

The input matrix, denoted by A is a square matrix. Figure 2 shows how matrix A combines two sub-matrices of relations (i.e. structural information) and vector-space (i.e. lexical similarity). The rows and corresponding columns describe the same terms of the graph's node. Each sub-matrix representing relations of this graph is described by such rows and columns. For this reason, these sub-matrices (more precisely — terms) must be 'connected' by diagonal sub-matrices (d_m), for which the values on the diagonal can be further weighted (they are equal to 1 by default). Moreover, columns of terms' vectors must be combined with corresponding rows by a diagonal sub-matrix located on their intersection (d_v). The node terms are then extended by vectors describing those terms constructed in step 1. All the other elements of the matrix, which are not defined, are set to zero.

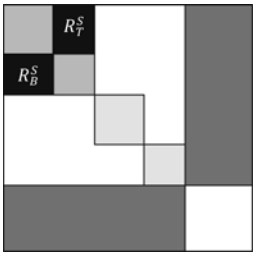
The elements of initial matrix A may be weighted. The operator can customize the 'strength' of each relation (sub-matrix) or the diagonal values. Different mapped structures have different numbers of levels, and various levels of name complexity. For this reason, the ability to customize the weight of the structural and lexical type of information is well founded.

3.4 Step 4: *Singular Value Decomposition* and Dimensionality Reduction of the Input Matrix

This step is similar to Step 2, but instead of auxiliary data the A matrix is processed. The number of terms in large-size structures may exceed several thousand, which may pose difficulties with their efficient processing. In order to solve this problem, the *Singular Value Decomposition* technique is used. Input matrix A is decomposed into three matrices $A = U\Sigma V^T$ and then the k-cut procedure is applied. Next, the opposite process is applied in order to reconstruct the initial dimensionality and matrix indexes.

3.5 Step 5: Summation of the Resulting Sub-matrix Cells and Returning the Output of the Reconstructed Matrix

As a consequence of LSA applied to the input matrix A the reconstructed matrix A_k contains the estimated values representing "matches-to" relations, which are



R_B^S results sub-matrix represents relations in the direction of 'the node of G^I matches-to the node of G^{II},

R_T^S results sub-matrix represents relations in the direction of 'the node of G^{II} matches-to the node of G^I,

Fig. 3. Scheme of a full output matrix

the result of the presented matching system procedure. Therefore, due to the mapping direction, the result is contained in the corresponding sub-matrices (of these relations) below the main diagonal.

A graph node is represented by multiple rows and columns, corresponding to terms in the name of the node. As it was described, in step 1, the value used to represent the relation between two nodes was divided by the number of terms of these nodes. Therefore, in order to prepare the final results, the values in 'matches-to' that describe the nodes composed of several terms need to be added up. Since the calculated value is between 0 and 1, it may be treated as a confidence that corresponding nodes are connected with "matches-to" relations.

4 Experiments

In this chapter, an experimental evaluation of the proposed matching algorithm (SSMRA) is presented. The publically available web version of *COMA++* — a mapping system for flexible combination of match algorithms — was used as the reference in the presented comparison. The methodology for the evaluation of matching accuracy assumes selecting n most similar matches between nodes, while not restricting the number of correspondences that may be matched to a single node. This strategy is used because the proposed mapping method is intended to be used in a 'real-world' semi-automatic schema alignment scenario.

4.1 Experimental Data

One of the main problems that appear in the process of schema matching is dealing with linguistic similarities that do not correspond to the structural relations. The experiment concentrates on a small schema matching problem analysed in [16]. The data consists of fragments of Google and Yahoo web directories, concerning the subject of 'art' that accurately addresses the mentioned issues.

The textual corpora was used as an auxiliary knowledge. It contained abstracts of the most popular articles from Wikipedia, which served as input documents in the process of LSA-based vector space generation. A corpus of $84,158$ articles with $102,750$ unique words was analysed by means of LSA in order to

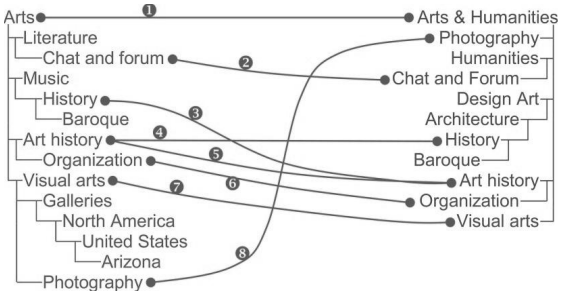


Fig. 4. Experimental data and expert matches

model lexical similarities of words appearing in the node names. Each lemma was represented as a vector of dimensionality reduced to 500 'concepts' (the most principal components). The SSMRA combines data about nodes' relations with the terms' vectors generated in LSA process. All rows and columns of the matrices describing *matches-to* and *coString* relations were extended by a vectors represention of given terms. Matrix representations of parsed and pre-processed graph structures as well as vector-space node representations were entered into the matrix A.

4.2 Experimental Evaluation of SSMRA

The *Semantic Structure Matching Recommendation Algorithm* not only provides correct matching of entities that are lexically similar (such as "Arts history" and "History"), but also (which is even more important) the quality of matching does not depend on the existence of the exact correspondences at the lexical level. For example, when analysing the matches of nodes "History" and "Baroque", it may be seen that those nodes were not matched to lexically identical nodes in the second structure. The reason why those two entities were not matched with identical entities in the second structure is the fact that, despite their 100% lexical similarity, their structural 'origins' are not corresponding to each other. (In the first structure "History" and "Baroque" have a common ancestor "Music", and in the second structure "History" and "Baroque" are "connected" with "Architecture"). This case shows that the presented matching algorithm takes structural information into consideration when generating the results. What is characteristic of SSMRA is the fact that the node "History", located below the node "Music", is properly matched to "Art history". This indicates that for the same node labelled "History", both structural and lexical information has been taken into account.

SSMRA achieved high values of *Precision*, *Recall* and F_1 score equal to 0.875. In the analysed experimental mapping task, results of SSMRA were better than the best results generated by *COMA++*, which is illustrated in the Figure 5.

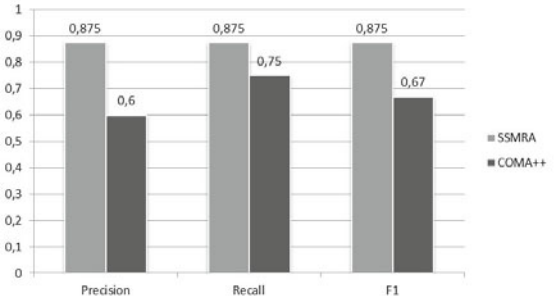


Fig. 5. Experiment results of SSMRA and COMA++ — both operating in default configurations

5 Conclusions

The presented matching method (SSMRA) identifies correspondences between structures based on the similarity of node names (lexical information) and their structural relations. Such an operation is possible because of the following reasons:

- Semantic concept space (obtained using the RSVD [15] method applied to the term-document matrix built from auxiliary documents) is used instead of using simple literal similarities between node names.
- Input data representation reflects structural relations between nodes. SSMRA unifies all data sources storing them in the form of sparse matrices. Thanks to using such an integrated data model, SSRMA is able to discover non-trivial lexical correspondences which are coherent with structural correspondences. Even for literally identical terms, the algorithm is able to exclude such correspondences from the results, e.g. identical terms exist in unrelated branches.

The experiments show that the algorithm reaches a satisfactory value of F_1. The outcomes are based on a vector space created from the Wikipedia corpus consisting of over $102,000$ unique words. Certainly, the use of larger and more detailed corpora may possibly result in the improvement of the matching quality.

In the presented algorithm, the input data is processed in two phases. The first one, although resource-consuming, has to be performed only once. In the second phase, match results are calculated based on semantic vector-space and structural data, which makes the method relatively fast. As a result, SSMRA may be used in user-interactive tools.

The presented algorithm may support the process of solving a general problem of data integration and semantic interoperability in *Service Oriented Architecture* (SOA). The solution may be used to create mappings that translate data to a new *Enterprise Data Model* (an ambitious approach) as well as to translate messages between individual entities, which is frequently necessary in peer-to-peer lightweight service interoperability systems, like SIX P2P [17].

References

1. Rahm, E., Bernstein, P.: A survey of approaches to automatic schema matching. The VLDB Journal — The International Journal on Very Large Data Bases (2001)
2. INDECT Homepage, http://www.indect-project.eu
3. van Rijsbergen, C.: Information Retrieval. Butterworths, London (1979)
4. Manning, C., Raghavan, P., Schütze, H.: An Introduction to Information Retrieval. Cambridge University Press, Cambridge (2009)
5. Abteilung Datenbanken Leipzig: Abteilung Datenbanken Leipzig am Institut für Informatik. In: COMA++ Web Edition, http://139.18.13.36:8080/coma/WebEdition
6. Doan, A., Domingos, P., Halevy, A.: Learning to Match the Schemas of Data Sources: A Multistrategy Approach. Machine Learning 50(3), 279–301 (2003)
7. Li, W., Clifton, C.: SEMINT: a tool for identifying attribute correspondences in heterogeneous databases (2000)
8. Madhavan, J., Bernstein, P.A., Rahm, E.: Generic schema matching with Cupid. In: Proc 27th Int Conf On Very Large Data Bases, pp. 49–58 (2001)
9. Do, H., Melnik, S., Rahm, E.: Comparison of schema matching evaluations. In: NODe 2002 Web and Database-Related Workshops on Web, Web-Services, and Database Systems (2003)
10. Melnik, S., Garcia-Molina, H., Rahm, E.: Similarity flooding — a versatile graph matching algorithm. In: Proc 18th Int Conf Data Eng. (2002)
11. Doan, A., Madhavan, J., Domingos, P., Halevy, A.: Learning to Map between Ontologies on the Semantic Web. VLDB Journal, Special Issue on the Semantic Web (2003)
12. Bouquet, P., Serafini, L., Zanobini, S.: Semantic Coordination: A New Approach and an Application. In: Fensel, D., Sycara, K., Mylopoulos, J. (eds.) ISWC 2003. LNCS, vol. 2870, pp. 130–145. Springer, Heidelberg (2003)
13. Euzenat, J., Mochol, M., Shvaiko, P., Stuckenschmidt, H., Šváb, O., Svátek, V., van Hag, W., Yatskevich, M.: Results of the Ontology Alignment Evaluation Initiative 2010. In: ISWC workshop on Ontology Matching OM–2006, pp. 73–95 (2006)
14. Landauer, T., Foltz, P., Laham, D.: Introduction to Latent Semantic Analysis, 259–284 (1998)
15. Ciesielczyk, M., Szwabe, A., Prus-Zajączkowski, B.: Interactive Collaborative Filtering with RI-based Approximation of SVD. In: Proceedings of PACIIA 2010 (2010)
16. Bouquet, P., Serafini, L., Zanobini, S.: Semantic coordination: a new approach and an application, Trento, Italy (2003)
17. Pankowski, T.: XML data integration in SixP2P — a theoretical framework. DaMaP 2008 Proceedings of the 2008 international workshop on Data management in peer-to-peer systems (2008)

Hierarchical Estimation of Human Upper Body Based on 2D Observation Utilizing Evolutionary Programming and "Genetic Memory"

Piotr Szczuko

Multimedia Systems Department, Gdansk University of Technology,
Narutowicza 11/12, 80-233 Gdansk, Poland
szczuko@sound.eti.pg.gda.pl

Abstract. New method of the human body pose estimation based on single camera 2D observation is presented. It employs 3D model of the human body, and genetic algorithm combined with annealed particle filter for searching the global optimum of model state, best matching the object's 2D observation. Additionally, motion cost metric is employed, considering current pose and history of the body movement, favouring the estimates with the lowest changes of motion speed comparing to previous poses. The "genetic memory" concept is introduced for the genetic processing of both current and past states of 3D model. State-of-the art in the field of human body tracking is presented and discussed. Details of implemented method are described. Results of experimental evaluation of developed algorithm are included and discussed.

Keywords: pose estimation, evolutionary optimization.

1 Introduction

Human action recognition lies in the scope of computer vision research for years [12]. It can be utilized in human-computer-interaction methods (HCI), for gesture navigated user interfaces [10], for markerless motion capture systems, and threats recognition in surveillance systems [3]. This process often comprises of following stages: background modelling, detection of foreground objects, classification and tracking of objects, and finally analysis of the performed action for event recognition. In single camera systems (monocular) the estimation of 3D features of the object based on 2D projection is ambiguous. Therefore, multi-camera approaches are introduced, dealing with ambiguity by fusion of data from multiple 2D projections obtained from various observation points. Those techniques perform very well, and already have found a commercial application in markerless motion capture [2].

In the reported work a monocular vision is considered, as a basis for general purpose computer vision system, that can be useful in any conditions without strict requirements on the number and positions of cameras. First, state-of-the-art in the pose estimation algorithms is described, then a proposal of the new single camera method based on evolutionary programming, motion dynamics and hierarchical pose estimation is presented. Finally the approach is tested and results are discussed.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 82–90, 2011.

2 State-of-the-Art in Monocular Human Body Tracking

Pose estimation methods utilize multidimensional parameterization of the body, where its state is described by degrees of freedom (DOFs) related to bones rotations and body position in space. Possible approaches are not consistent, as models can have 25- 34 DOFs, while 3D animated models employ over 40 DOFs.

A 3D model of the body is considered, characterized by body proportions, number of DOFs, and angle restrictions. The optimal state of such model is being sought, the one which best matches the shape of current 2D observation of the real body and fulfils body biomechanical constraints.

2.1 Pose Estimation

The model **X** to object **Z** matching degree $w(\mathbf{X},\mathbf{Z})$ can be expressed as a coverage of their silhouettes (1) or their edges (2) [6][8][9]. Cumulated metric is proposed [4](3):

$$w^r(X,Z) = \exp\left(-\frac{1}{N}\sum_{i=1}^{N}(1-p_i^r(X,Z))^2\right) \tag{1}$$

$$w^e(X,Z) = \exp\left(-\frac{1}{N}\sum_{i=1}^{N}(1-p_i^e(X,Z))^2\right) \tag{2}$$

$$w(X,Z) = w^r(X,Z)\cdot w^e(X,Z) \tag{3}$$

where: **X** - model state,
Z – observation (object shape or edges extracted from the video frame),
N – number of comparison points in the model image and object image
p_i^r, p_i^e – pixel-wise AND operator between regions and edges respectively

Optimization of model state based on matching metric $w(\mathbf{X},\mathbf{Z})$ is performed with following methods.

Particle Filter

Probabilistic modelling of possible multidimensional states of the tracked body is performed with particle filtering approach (PF), known also as CONditional DENSity propagATION (CONDENSATION). It was first introduced for tracking outlines of hands [7], and later extended for whole body [4][13][15]. PF method is useful for analysis of multiple hypotheses, here called “particles”. Even if some model states are less probable, considering previous states, they are taken into account, resulting in more robust tracking. The new particles are located densely around previous match, but also are randomly scattered in the whole range. The one with highest match (likely global optimum of $w(\mathbf{X},\mathbf{Z})$) is taken as a result. This method considers many modalities with varying probability, therefore it tests various action courses at the same time. Unfortunately, finding global optimum requires employing large amounts of particles, and long computation times are reported (e.g. 1 video frame in 18 minutes) [4].

Annealed Particle Filter
An interesting modification of PF method is Annealed Particle Filter (**AFP**), where the optimum search is performed in M stages, so called layers. The layer m is characterized with annealing speed β_m, where $1 \geq \beta_0 > \beta_1 > \ldots > \beta_M$ and analyzed metric is $w_m(\mathbf{X},\mathbf{Z}) = w(\mathbf{X},\mathbf{Z})^{\beta m}$. The higher the β_m the more coarse $w_m(\mathbf{X},\mathbf{Z})$ is, and the search is less susceptible to local optima. In consecutive layers the particles are located in the most probable areas, where high values of $w_m(\mathbf{X},\mathbf{Z})$ occur, with some randomization introduced. This method performs significantly better than FP [4].

2.2 Predefined Action Recognition

For distinct motions such as walking, instead estimating every consecutive pose, a common approach is to compare the estimated pose to some presets, e.g. phases of walking [5], but applications are limited to detection of predefined actions [3], therefore preset matching lies out of scope of this work.

3 Hierarchical Upper Body Pose Estimation and Motion Constrains Utilizing "Genetic Memory"

Proposed new upper body pose estimation algorithm is based on genetic evolution of 3D models population tested against current 2D silhouettes from single camera. The models generation and fitness testing are performed with respect to evolutionary programming paradigm [1][11], utilizing genetic algorithm extended with new concept of "genetic memory", combined with APF for additional optimization. The version tested here assumes only upper body changes, but the algorithm can be modified to account whole body pose estimation.
In the proposed model-to-object matching method following ideas are employed:

- the matching procedure should be performed in stages:
    - first the local optima of higher hierarchy of the body are sought (torso, head, i.e. the parts influencing location of arms),
    - next, for each found optimum a second search run is performed, considering also lower hierarchy of the body parts (forearms, arms, and hands),
- the observed motion (model and object state changes in consecutive frames) is continuous and fluent, abrupt speed vector changes are not plausible (yet still are possible),
- therefore, for selection of best estimates during genetic optimum search, the history of motion and current estimated motion should be considered,
- the history of motion is inscribed into "genetic memory" of the object
- utilizing genetic operators of cross-over and mutation a new, possibly better, estimates can be obtained based on previous estimates.

The concepts introduced above are summarized in the following sections.

3.1 3D Model of Human Body

Implemented body model consists of 17 elements, modelled as balls and cuboids, some with limited Degrees of Freedom (DOF), 40 DOFs total. For presented here

initial experiments the pose it was limited to upper body alterations. The structure is presented in Fig. 1. and Tab. 1. Model state is described with g=40 values (genes) of current state, subject to modification by genetic algorithm and optimum search, and H*40 values of H-long history of previous states ("genetic history"), which is neither crossed-over nor mutated. The model state can be described as:

$$\mathbf{X}=\{x_{1,0}, x_{2,0}, \ldots x_{g,0}; x_{1,1}, x_{2,1}, \ldots x_{g,1}; \ldots; x_{1,H}, x_{2,H}, \ldots x_{g,H}\} \quad (4)$$

where: the subscript 0 depicts current moment in the history,
genes $x_{1,j} \div x_{16,j}$ higher hierarchy,
genes $x_{17,j} \div x_{30,j}$ lower body (legs - not accounted in current research)
genes $x_{31,j} \div x_{40,j}$ lower hierarchy.

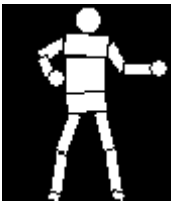


Fig. 1. 3D model of human body (segment edges only for visualization)

Table 1. Parts of the modelled body and angle limits for joints (degrees). 40 values are used to describe full state of the model

								Hierarchy
Hip location is space:		*x*		*y*		*z*		Higher
Rotations:		α		β		γ		
Body part	**Model**	α_{min}	α_{max}	β_{min}	β_{max}	γ_{min}	γ_{max}	
Hip	Cuboid	-180	180	-180	180	-180	180	Higher
Abdomen	Cuboid	-15	15	-110	15	-15	15	Higher
Torso	Cuboid	-5	5	fixed		-5	5	Higher
Neck	Cuboid	-15	15	-30	30	-50	50	Higher
Head	Ball	-45	45	-30	15	fixed		Higher
Thigh x2	Cuboid	-15	90	70	125	75	75	n/a
Calf x2	Cuboid	fixed		145	0	fixed		n/a
Foot x2	Cuboid	-15	15	-45	30	-15	15	n/a
Forearm x2	Cuboid	fixed		fixed		-150	0	Lower
Arm x2	Cuboid	0	360	0	360	-155	35	Lower
Hand x2	Ball	-20	20	fixed		fixed		Lower

3.2 Genetic Fitness Function

Used matching metric is an extension of shape and edge coverage metrics (3). Additionally it considers "motion cost" $v(\mathbf{X})$, related to motion dynamics and movement speed changes (5):

$$w(X,Z) = w^r(X,Z) \cdot w^e(X,Z) \cdot v(X) \quad (5)$$

where:

$$v(X) = \exp(-\frac{1}{N}\sum_{i=1}^{N}\sum_{h=1}^{H}(v_{0,i} - v_{h,i})^2) \quad (6)$$

where: N – number of bones analysed in current hierarchy level

$v_{0,i}$ – current angular motion speed for i-th value of state model calculated as $x_{i,0}$–$x_{i,1}$ (the time span is 1 frame, therefore no denominator is written)

$v_{h,i}$ – historical angular motion speed for i-th value of state model calculated as $x_{i,h}$–$x_{i,h-1}$.

Presented metric (5) is used as a fitness function for evolutionary processing. $w(\mathbf{X},\mathbf{Z}) \in (0,1>$, and the perfect match is obtained when $w(\mathbf{X},\mathbf{Z})=1$.

3.3 Crossing-over and Mutation of the Model State

Two model states **X** and **Y** are crossed-over by exchanging one, randomly selected i-th ($i \in <1; g=40>$) value from current state: $x_{i,0}$ with $y_{i,0}$. The mutation is performed on randomly selected i-th value from current state, changing it by random value $\Delta \in <-5;5>$, with the requirement that the result stays in allowed angle range.

3.4 Hierarchical Pose Matching

Hierarchical matching of 3D model to 2D shape of the body is performed by successive consideration of model parts, starting high in the hierarchy, proceeding to lower levels. In each run the model is simplified to represent only the parts that are on the current hierarchy level (Fig. 2). The algorithm performs following steps (also show on block diagram in Fig. 3):

1. N random objects are generated. Initially the history contains static pose, i.e. for every i: $x_{i,0} = x_{i,1} = x_{i,2} = \ldots = x_{i,H}$.
2. Each estimate is evaluated utilizing genetic fitness function (5) (shape coverage and motion cost) for higher hierarchy of the body.
3. M=2 stage Particle Filtering is performed for local optima search over the fitness function. N particles in 16-dimensional space are used, initialized with current state of higher hierarchy model, and adjusting only current state (not the history). Head and shoulders shape is very distinct and optimum search converges easily (Fig. 2b).
4. $N' < N$ best estimates (particles) of last PF stage are selected.
5. Utilizing each of N' estimates all (N–N') worse objects are readjusted, by substituting $x_{1,0} \div x_{16,0}$ genes in the object with genes of randomly selected higher hierarchy estimate. Probability of selection is proportional to the value of estimate $w(\mathbf{X},\mathbf{Z})$ calculated in PF in step 3. The results are N readjusted objects with well fitting higher hierarchy of the body.
6. Optimum search is performed with M=2 stages PF with N particles in 10-dimenstional space of lower hierarchy bones are considered. Local optima of those bones rotations are found (Fig. 2c).
7. Each estimate is evaluated utilizing genetic fitness function (shape and motion cost).

8. Best $L \geq 1$ estimates are presented on the screen and compared with reference data for objective estimate rating and algorithm benchmarking.
9. All N estimates from step 7. are aged (in the history the last state of age H is removed, all states are shifted right by g cells). New state of the model is created by crossing-over (with probability 0.5) all worse (N–L) estimates with randomly selected one of L best ones. Finally the mutation of the current state is performed (with probability 0.1). It is then taken as a starting point for matching the model to next video frame.
10. The process repeats starting at step 2.

The algorithm is implemented in C++ utilizing own pose generation library (it provides binary image of 3D model silhouette based on its state description and camera position) and OpenCV image processing library [14] for image processing (normalization, matching measure calculation, result visualization).

4 Algorithm Evaluation

For the experiment and objective assessment of the results a set of 10 poses of upper body were created utilizing 3D model posed by hand, each accompanied with H=3 long motion history, comprised of 3 poses before target pose (Fig. 4).

Current poses and their histories were saved as bitmap files, and supplemented with the XML-formatted reference data in a form of bones angles values for particular poses. Then the pose estimation algorithm was initialized with T-shape pose (all angles equal to zero) and performed runs for higher and lower hierarchy (Sec. 3.4). After the pose estimation the model state was compared to respective reference data and cumulative Sum of Squared Differences (SSD) of angles for particular bones was calculated to rate the pose estimate.

In the experiment the following values were used: number of objects N=160, best N'=16 objects were selected for reproduction, motion history length H=3, M=2 stages AFP was performed, and best L=1 object was used for comparison to reference values. The relation between those parameters and computation time is defined, therefore more precise or more coarse analysis can be performed in particular time requirements.

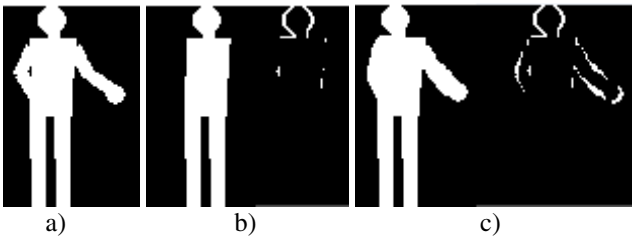


Fig. 2. Hierarchical matching of 3D model to 2D observation of human body: a) sample pose, b) 3D model posed by first run (high hierarchy) and its matching metric, c) 3D model posed by second run (lower hierarchy) and its matching metric

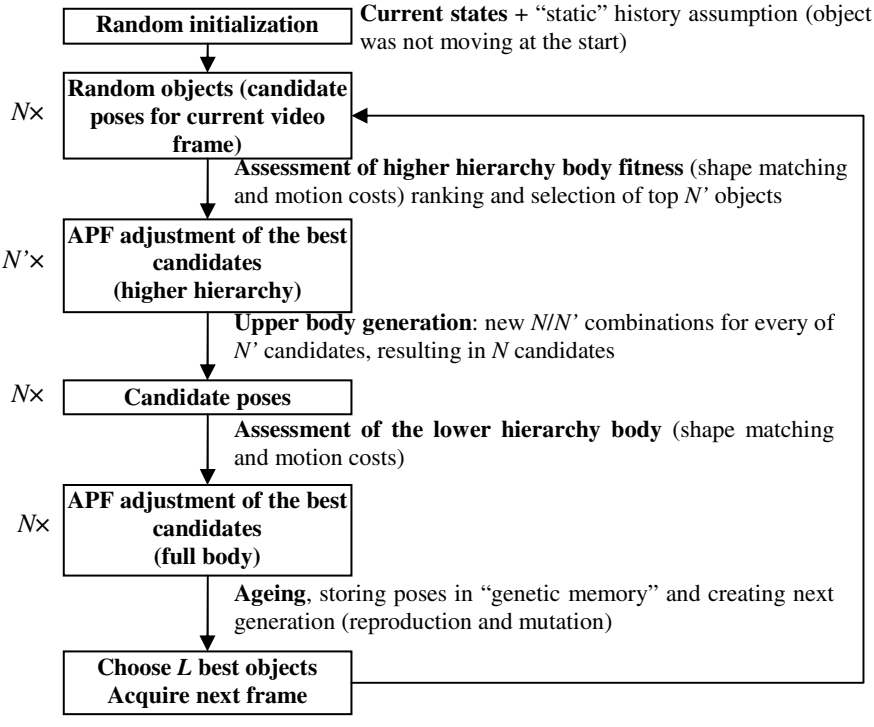


Fig. 3. Block diagram of the evolutionary algorithm (description in the text)

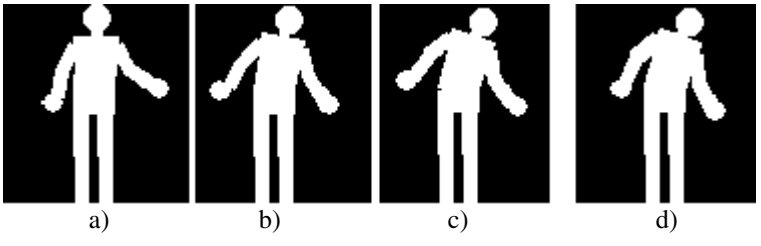


Fig. 4. One of the analysed poses with H=3 long history: a)-c) 3 previous poses stored in the history, d) currently estimated pose

Obtained results of estimation of 10 poses with 3 step history and hierarchical matching are presented in Tab. 2.

If in the pose shape arms connect to the torso (poses 7÷10 in Tab. 2), then the matching obtained in the first stage of higher hierarchy estimation is low, because of the attempt of matching "handless" model to full body shape. Then, the matching value increases in the second stage, when full model is used and the hands positioned correctly provide correct matching of shape and edges.

Table 2. Results of hierarchical matching of body model for various poses. SSD is a Sum of Squared Differences of bones angles compared to the reference values.

Pose	Higher hierarchy matching $w(\mathbf{X},\mathbf{Z})$	Lower hierarchy matching $w(\mathbf{X},\mathbf{Z})$	Bones angles errors: SSD
1	0.900	0.852	11.72
2	0.915	0.874	11.08
3	0.944	0.903	13.14
4	0.943	0.885	16.78
5	0.919	0.832	24.80
6	0.954	0.930	8.95
7	0.883	0.961	3.48
8	0.927	0.977	2.31
9	0.947	0.979	2.30
10	0.880	0.965	2.49

Contrary, if the arms are stretched away from the torso (poses 1÷6 in Tab. 2), then the higher hierarchy matching result is high, due to precise localization of the torso and head shape. Then, for the whole body, errors of shape and edges matching higher and lower hierarchy (for torso and arms) sums, therefore decreasing total matching result.

Poses 1 and 2 were created with abrupt motion change comparing to the historical poses, therefore the matching result is lowered despite average SSD values. More thorough experiments are advised for determination of the correct influence of motion cont $v(\mathbf{X})$ on total $w(\mathbf{X},\mathbf{Z})$. Currently shape, edges, and motion cost metrics are considered as equally important, while for longer sequences and histories this approach may lead to motion continuity preference over shape estimation precision.

The least effective matching process was observed for pose 5, where the arms are embracing the torso, and large degree on ambiguity is present, as very low information is contained in the shape. This type of the pose (self occlusion, limbs very close to the torso) stays currently as the main challenge in pose estimation research.

5 Summary

Hybrid genetic and Annealed Particle Filtering method was proposed and tested. New metric for model-to-object matching was proposed. The concept of "genetic memory" was introduced, facilitating processing of the motion history and accounting the history in estimate fitness measurement. The genetic crossing-over operator forces APF algorithm to assess other modes of the model matching metric, and genetic mutation introduces randomness, important for avoiding local optima. The proposed algorithm can be used for various body hierarchies with more than two hierarchy levels, and various genetic history lengths H.

The future work will focus on extending the procedure to whole body, not only upper body, and implementing some ways of limbs occlusions handling. Moreover, the matching metric will be further extended by introducing pixel-level motions (e.g. based on Optical Flow or Motion History Imaging).

Acknowledgements

Research funded within the project No. POIG.02.03.03-00-008/08, entitled "MAYDAY EURO 2012 - the supercomputer platform of context-depended analysis of multimedia data streams for identifying specified objects or safety threads" subsidized by the European regional development fund and by the Polish State budget.

References

[1] Bäck, T.: Evolutionary Algorithms in Theory and Practice: Evolution Strategies, Evolutionary Programming, Genetic Algorithms. Oxford University Press, Oxford (1996)
[2] CONTOUR: Markerless Motion Capture System, http://www.mova.com
[3] Czyżewski, A., Ellwart, D.: Camera angle invariant shape recognition in surveillance systems. In: Proc. KES IIMSS 2010, Baltimore, USA (2010)
[4] Deutscher, J., Blake, A., Reid, I.D.: Articulated body motion capture by annealed particle filtering. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, pp. 126–133 (2000)
[5] Efros, A.A., Berg, A.C., Mori, G., Malik, J.: Recognizing action at a distance. In: 9th Inter. Conf. Computer Vision (ICCV 2003), Nice, France, pp. 726–733 (2003)
[6] Gavrila, D.M., Davis, L.S.: 3D model-based tracking of humans in action: A multi-view approach. In: Proc. Computer Vision and Pattern Recognition (CVPR 1996), pp. 73–80 (1996)
[7] Isard, M., Blake, A.: CONDENSATION—conditional density propagation for visual tracking. Int. Journal of Computer Vision 29(1), 5–28 (1998)
[8] Kakadiaris, I., Metaxas, D.: Model-based estimation of 3D human motion. IEEE Tran. Pattern Analysis and Machine Intelligence 22(12), 1453–1459 (2000)
[9] Kehl, R., Van Gool, L.: Markerless tracking of complex human motions from multiple views. Computer Vision and Image Understanding 104(2-3), 190–209 (2006)
[10] Lech, M., Kostek, B.: Fuzzy Rule-based Dynamic Gesture Recognition Employing Camera & Multimedia Projector. In: Proc. Int. Conf. Multimedia Network Information Systems (2010)
[11] Michalewicz, Z.: Genetic Algorithms+Data Structures=Evolution Programs. Springer, Heidelberg (1998)
[12] Moeslund, T., Hilton, A., Kruger, V.: A survey of advances in vision-based human motion capture and analysis. Computer Vision and Image Understanding 104(2-3), 90–126 (2006)
[13] Ong, E.-J., Micilotta, A.S., Bowden, R., Hilton, A.: Viewpoint invariant exemplar-based 3D human tracking. Computer Vision and Image Understanding 104(2-3), 178–189 (2006)
[14] OpenCV Image Processing and Compute Vision Library, http://opencv.willowgarage.com
[15] Sidenbladh, H., Black, M.J., Fleet, D.J.: Stochastic tracking of 3D human figures using 2D image motion. In: 6th European Conference on Computer Vision, pp. 702–718 (2000)

Assessing Task-Based Video Quality — A Journey from Subjective Psycho-Physical Experiments to Objective Quality Models

Mikołaj Leszczuk

AGH University of Science and Technology, al. Mickiewicza 30,
PL-30059 Krakow, Poland

Abstract. This paper reviews the development of techniques for assessing video quality. Examples have been provided on the quality of video applications ranging from popular entertainment to new trends such as applications in broad public systems, not just those used by police forces but also for medical purposes. In particular, the author introduces two typical usages of task-based video: surveillance video for accurate licence plate recognition, and medical video for credible diagnosis prior to bronchoscopic surgery. The author also presents the field of task-based video quality assessment from subjective psycho-physical experiments to objective quality models. Example test results and models are provided alongside the descriptions.

Keywords: video, quality, experiments, models.

1 Introduction

The storage and transmission of video is used for many applications outside of the entertainment sector; generally, this class of video is used to perform a specific task (task-based video). Examples of these applications include public safety including surveillance, medical services, remote command and control, and sign language.

Development efforts reveal a significant potential behind platforms allowing access to digital recordings of surveillance or medical video sequences. Video compression and transmission are the most widespread problems in those platforms. Surveillance and medical applications add a new dimension, as lossy compression techniques need to be both resource-effective and credible from the point of view of practitioners in the field of public safety and medical services.

Anyone who has experienced artefacts or freezing play while watching a film or live sporting event on TV is familiar with the frustration accompanying sudden quality degradation at a key moment. Video services with blurred images may have far more severe consequences for practitioners.

This paper introduces two typical usages of task-based video (Section 2): surveillance video used for accurate licence plate recognition, and medical video used for credible diagnosis prior to bronchoscopic surgery. The remainder of the

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 91–99, 2011.

article introduces the field of task-based video quality assessment from subjective psycho-physical experiments (Section 3) to objective quality models (Section 4). Example test results and models are provided alongside the descriptions. Section 5 outlines the conclusions and plans for further work, including standardization.

2 Use Cases

This section introduces two typical usages of task-based video: surveillance video used for accurate licence plate recognition, and medical video used for credible diagnosis task prior to bronchoscopic surgery.

Accurate Licence Plate Recognition Task. Recognizing the growing importance of video in delivering a range of public safety services, let us consider a licence plate recognition task based on video streaming in constrained networking conditions. Video technology should allow users to perform the required function successfully. The paper presents people's ability to recognize car registration numbers in video material recorded using a CCTV camera and compressed with the H.264/AVC codec. An example frame from the licence plate recognition task is shown in Figure 1. The use case is sufficiently presented in [10].

Fig. 1. Example frame from the licence plate recognition task

Credible Bronchoscopic Diagnosis Task. The presented task targets video bronchoscopy, a type of recording made during surgery. In such videos the image can remain almost motionless for prolonged periods of time. This predisposes such recordings to compression. However, it is possible to use just those motion images in which lossy compression has not caused any distortion visible to the

Fig. 2. Example frame from the bronchoscopic diagnosis task (source: [8])

physicians [1,9,11,12]. The paper describes the degree of compression which introduces quality impairment visible to the physicians. An example frame from the bronchoscopic diagnosis task is shown in Figure 2. The use case is sufficiently presented in [9,2].

Video streaming services still face the problem of limited bandwidth access. While for wired connections bandwidth is generally available in the order of megabits, higher bit rates are not particularly common for wireless links. This poses problems for mobile users who cannot expect a stable high bandwidth.

Therefore a solution for streaming video services across such access connections is transcoding of video streams. The result is transcoding bit-rate (and quality) scaling to personalize the stream sent according to the current parameters of the access link. Scaling video sequence quality may be provided in compression, space and time. Scaling of compression usually involves operating the codec quantization coefficient. Scaling of space means reducing the effective image resolution resulting in increased granularity when one tries to restore the

original content on the screen. Scaling of time amounts to rejection of frames, i.e. reducing the number of frames per second (FPS) sent. However, frame rates are commonly kept intact as their deterioration does not necessarily result in bit-rate savings due to inter-frame coding [7].

The abovementioned scaling methods inevitably lead to lower perceived quality of end-user services (Quality of Experience, QoE). Therefore the scaling process should be monitored for QoE levels. This gives the opportunity to not only control but also maximize QoE levels depending on the prevailing transmission conditions. In case of failure to achieve a satisfactory QoE level, an operator may intentionally interrupt the service, which may help preserve network resources for other users.

3 Procedures for Subjective Psycho-Physical Experiments

To develop accurate objective measurements (models) for video quality, subjective experiments must be performed. The ITU-T[1] P.910 Recommendation "Subjective video quality assessment methods for multimedia applications" (1999) [5] addresses the methodology for performing subjective tests in a rigorous manner.

However, such methods are currently only targeted at entertainment video. In task-based applications, video is used to recognize objects, people or events. Therefore the existing methods, developed to assess a person's perception of quality, are not appropriate for task-based video.

The QoE concept for video content used for entertainment differs considerably from the quality of video used for recognition tasks. This is because in the latter case subjective user satisfaction is improved by achieving a given functionality (event detection, object recognition). Additionally, the quality of video used by a human observer for recognitions tasks is considerably different from objective video quality used in computer processing (Computer Vision).

Task-based videos require a special framework appropriate to the video's function — i.e. its use for recognition tasks rather than entertainment. Once the framework is in place, methods should be developed to measure the usefulness of the reduced quality video rather than its entertainment value.

Issues of quality measurements for task-based video are partially addressed in the ITU-T P.912 Recommendation "Subjective video quality assessment methods for recognition tasks" (2008) [6]. This Recommendation introduces basic definitions, methods of testing and ways of conducting psycho-physical experiments (e.g. Multiple Choice Method, Single Answer Method, and Timed Task Method), as well as the distinction between Real-Time- and Viewer-Controlled Viewing scenarios. While these concepts have been introduced specifically for task-based video applications in ITU-T P.912, more research is necessary to validate the methods and refine the data analysis methods.

[1] International Telecommunication Union — Telecommunication Standardization Sector.

3.1 Psycho-Physical Experiment for the Accurate Licence Plate Recognition Task

A subjective experiment was carried out in order to perform the analysis. A psycho-physical evaluation of the video sequences scaled (in the compression or spatial domain) at various bit-rates was performed. The aim of the subjective experiment was to gather the results of human recognition capabilities. Thirty non-expert testers rated video sequences influenced by different compression parameters. ITU's Absolute Category Rating (ACR), described in ITU-T P.800 [4], was selected as the subjective test methodology.

Video sequences used in the test were recorded in a car park using a CCTV camera. The H.264 (x264) codec was selected as the reference as it is a modern, open, and widely used solution. Video compression parameters were adjusted in order to cover the recognition ability threshold. The compression was done with the bit-rate ranging from 40 kbit/s to 440 kbit/s.

The testers who participated in this study provided a total of 960 answers. Each answer could be interpreted as the number of per-character errors, i.e. 0 errors meaning correct recognition. The average probability of a licence plate being identified correctly was **54.8%** (**526/960**), and **64.1%** recognitions had no more than one error. **72%** of all characters were recognized.

For further analysis it was assumed that the threshold detection parameter to be analyzed is the probability of plate recognition with no more than one error. For detailed results, please refer to Figure 3.

3.2 Psycho-Physical Experiment for the Credible Bronchoscopic Diagnosis Task

Well established methods of subjective assessment are based on Receiver Operating Characteristic (ROC) curves. Anyway, another, different but tested [1,9,11,12] subjective method for qualifying lossy compressed still images to a visually undistorted subset is based on sorting compressed images by their quality. The same approach was adopted in the investigation of video sequences. Expert testers (clinicians) were randomly presented with several video sequences: the original, and seven copies compressed with various bit-rate values. As a result of ordering, it was generally possible for an experiment supervisor to clearly distinguish two subsets of video sequences [1,9,11,12]. The first consisted of the highest quality video sequences in a random order. The other video sequences appeared in the second lowest quality subset. Only the video sequences belonging to the first subset were considered to be of a quality suitable for diagnostic purposes [2].

A subjective evaluation of video sequences compressed at various bit-rates was performed. The MPEG-4 codec was selected as the reference as it is still the most widely used solution for telemedicine. The codec has also been successfully applied to surgery video compression [1,9,11,12]. A test based on the abovementioned quality-based ordering method was carried out in order to obtain the bit-rate for which a clinician cannot distinguish between the original and the compressed video sequences. Each of the three original (uncompressed)

video sequences was supplemented with a few compressed video sequences with the same content, thus constituting three investigated video sequences [2]. The compression was done with the bit-rate approximately ranging from 80 kbit/s to 1280 kbit/s.

Eight clinicians were asked independently to arrange the video sequence in each of the sets following the well-known bubble sort algorithm. The only information gathered was the order of the sequences. The clinicians used purpose-developed software in the sorting. The software was run on an ordinary personal computer located at the clinic. No time restrictions were applied to the evaluations [2].

For further analysis it was assumed that the threshold detection parameter to be analyzed is the likelihood that a clinician cannot distinguish between the original and compressed video sequences. For detailed results, please refer to Figure 4.

4 Modelling Perceptual Video Quality

In the area of entertainment video, a great deal of research has been carried out on the parameters of the contents that are the most effective for perceptual quality. These parameters form a framework in which predictors can be created such that objective measurements can be developed through the use of subjective testing [13].

Assessment principles for task-based video quality are a relatively new field. Solutions developed so far have been limited mainly to optimizing network Quality of Service (QoS) parameters. Alternatively, classical quality models such as the PSNR [3] or SSIM [15] have been applied, although they are not well suited to the task. The paper presents an innovative, alternative approach, based on modelling detection threshold probabilities.

4.1 Quality Modelling of the Accurate License Plate Recognition Task

It was possible to fit a logarithmic function in order to model quality (expressed as detection threshold probability) of the licence plate recognition task. This is an innovative approach. The achieved R^2 is 0.81 (see Figure 3). According to the model, one may except 100% correct recognition for bit-rates of around 350 kbit/s and higher. Obviously accuracy of recognition depends on many external conditions and also size of image details. Therefore 100% can be expected only if other conditions are ideal.

4.2 Quality Modelling of the Credible Bronchoscopic Diagnosis Task

Before presenting the results for the second quality modelling case, it should be noted that a common method of presenting results has been used. This is

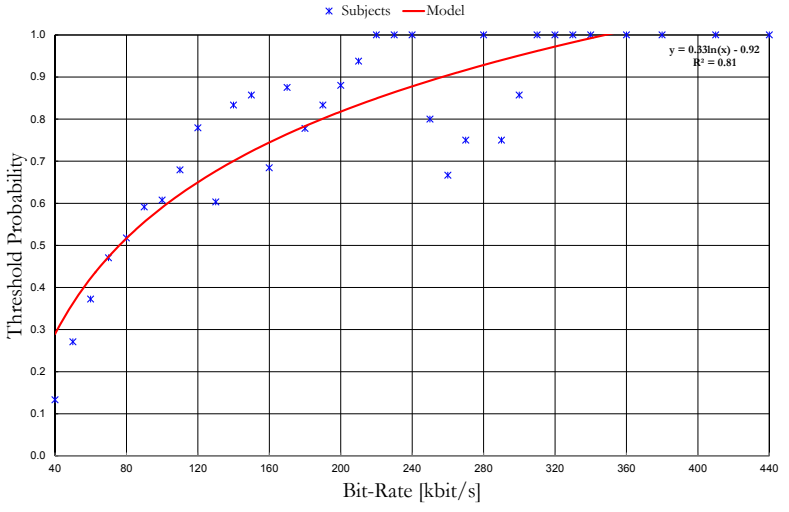


Fig. 3. Example of the obtained detection probability and model of the licence plate recognition task

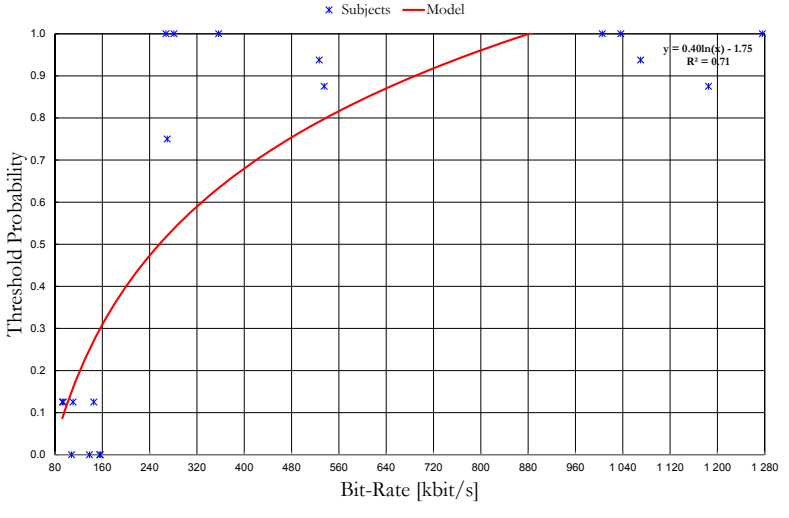


Fig. 4. Example of the obtained detection probability and model of the bronchoscopic diagnosis task

possible through the application of appropriate transformations, allowing the fitting of diverse recognition tasks into a single quality framework.

Again, it was possible to fit a logarithmic function in order to model quality (expressed as detection threshold probability) of the bronchoscopic diagnosis

task. This is again an innovative approach. The achieved R^2 is 0.71 (see Figure 4). According to the model, one may except 100% correct recognition for bit-rates of around 900 kbit/s and higher.

Unfortunately, due to relatively high diversity of subjective answers, no better fitting was achievable in either case. However, a slight improvement is likely to be possible by using other curves.

5 Conclusions and Future Work

The methodologies outlined in the paper are just a single contribution to the overall framework of quality standards for task-based video. It is necessary to define requirements starting from the camera, through the broadcast, and until after the presentation. These requirements will depend on scenario recognition.

So far, the practical value of contribution is limited. It refers to limited scenarios. The presented approach is just a beginning of more advanced interesting framework of objective quality assessment, described below.

Extensive work has been carried out in recent years in the area of consumer video quality, mainly driven by the Video Quality Experts Group (VQEG) [14]. A new project, Quality Assessment for Recognition Tasks (QART), was created for task-based video quality research at a recent meeting of VQEG. QART will address the problems of a lack of quality standards for video monitoring. The initiative is co-chaired by NTIA (National Telecommunications and Information Administration, an agency of the United States Department of Commerce) and AGH University of Science and Technology in Krakow, Poland. The aims of QART are the same as the other VQEG projects — to advance the field of quality assessment for task-based video through collaboration in the development of test methods (including possible enhancements of the ITU-T P.912 Recommendation [6]), performance specifications and standards for task-based video, and predictive models based on network and other relevant parameters.

Acknowledgments

This work was supported by the European Commission under the Grant INDECT No. FP7-218086.

References

1. Bartkowiak, M., Domanski, M.: Applications of chrominance vector quantization to intraframe and interframe compression of colour video sequences (2001)
2. Duplaga, M., Leszczuk, M., Papir, Z., Przelaskowski, A.: Evaluation of quality retaining diagnostic credibility for surgery video recordings. In: Sebillo, M., Vitiello, G., Schaefer, G. (eds.) VISUAL 2008. LNCS, vol. 5188, pp. 227–230. Springer, Heidelberg (2008)
3. Eskicioglu, A.M., Fisher, P.S.: Image quality measures and their performance. IEEE Transactions on Communications 43(12), 2959–2965 (1995), http://dx.doi.org/10.1109/26.477498

4. ITU-T: Methods for subjective determination of transmission quality. ITU-T, Geneva, Switzerland (1996)
5. ITU-T: Subjective Video Quality Assessment Methods for Multimedia Applications. ITU-T (1999)
6. ITU-T: Recommendation 912: Subjective video quality assessment methods for recognition tasks. ITU-T Rec. P.912 (2008)
7. Janowski, L., Romaniak, P.: QoE as a function of frame rate and resolution changes. In: Zeadally, S., Cerqueira, E., Curado, M., Leszczuk, M. (eds.) FMN 2010. LNCS, vol. 6157, pp. 34–45. Springer, Heidelberg (2010)
8. Leszczuk, M., Grega, M.: Prototype software for video summary of bronchoscopy procedures with the use of mechanisms designed to identify, index and search. In: Piętka, E., Kawa, J. (eds.) Information Technologies in Biomedicine. AISC, vol. 69, pp. 587–598. Springer, Heidelberg (2010)
9. Leszczuk, M.: Analiza mozliwosci budowy internetowych aplikacji dostepu do cyfrowych bibliotek wideo. Ph.D. thesis, AGH University of Science and Technology, Krakow (April 2006)
10. Leszczuk, M., Janowski, L., Romaniak, P., Glowacz, A., Mirek, R.: Quality assessment for a licence plate recognition task based on a video streamed in limited networking conditions. In: Fourth Multimedia Communications, Services and Security (MCSS 2011), Krakow, Poland, pp. 10–18 (June 2011)
11. Przelaskowski, A.: Falkowe metody kompresji danych obrazowych. Ph.D. thesis, Oficyna Wydawnicza Politechniki Warszawskiej, Warsaw (2002)
12. Skarbek, W.: Multimedia — algorytmy i standardy. PLJ, Warsaw (1998)
13. Takahashi, A., Schmidmer, C., Lee, C., Speranza, F., Okamoto, J., Brunnström, K., Janowski, L., Barkowsky, M., Pinson, M., Staelens, Nicolas Huynh Thu, Q., Green, R., Bitto, R., Renaud, R., Borer, S., Kawano, T., Baroncini, V., Dhondt, Y.: Report on the validation of video quality models for high definition video content. Tech. rep., Video Quality Experts Group (June 2010)
14. VQEG: The Video Quality Experts Group, http://www.vqeg.org/
15. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement. Signal Processing: Image Communication 19(2), 121–132 (2004), http://dx.doi.org/10.1016/S0923-5965(03)00076-6%20

Graph-Based Relation Mining

Ioannis P. Klapaftis, Suraj Pandey, and Suresh Manandhar

Department of Computer Science
University of York
United Kingdom
{giannis,suraj,suresh}@cs.york.ac.uk

Abstract. Relationship mining or Relation Extraction (RE) is the task of identifying the different relations that might exist between two or more named entities. Relation extraction can be exploited in order to enhance the usability of a variety of applications, including web search, information retrieval, question answering and others. This paper presents a novel unsupervised method for relation extraction which casts the problem of RE into a graph-based framework. In this framework, entities are represented as vertices in a graph, while edges between vertices are drawn according to the distributional similarity of the corresponding entities. The RE problem is then formulated in a bootstrapping manner as an edge prediction problem, where in each iteration the target is to identify pairs of disconnected vertices (entities) most likely to share a relation.

Keywords: Natural Language Processing, Relation Extraction.

1 Introduction

Relation Extraction (RE) seeks to identify a relation of interest that might exist between two or more named entities. For instance, consider the following three sentences:

- *George Smith* works for *KKC*.
- *Karen Smith*, daughter of *George Smith* works for the same company.
- *Karen Smith* married *George Brown* last year.

A RE system should be able to identify the relation between the PERSON entity, *George Smith*, and the ORGANIZATION entity, *KKC*. A similar relation exists between *Karen Smith* and *KKC*, as well as, between *Karen Smith* and *George Brown*. Relation extraction can be exploited in order to enhance the usability of a variety of applications. The most profound application of RE would be in Information Retrieval (IR), where RE could assist current search engines in handling queries, whose correct answers highly depend on the semantic interpretation of the data. Similarly, RE can be beneficial for Question/Answering (QA) systems by allowing them to answer questions related to specific attributes/relations of a particular named entity [11], such as: *Who is the father of X ?* or *When was X born* etc.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 100–112, 2011.

Additionally RE is a key technology for security-related applications, since it would allow the detection of relationships between people (e.g. known criminals) or between people and organizations from unrestricted corpora. The extracted relationships, as well as the patterns leading to these relationships will be beneficial for an automated system aimed at the automatic detection of threats or illegal behavior. Furthermore, RE could also benefit the development of methods for offender profiling, since a possible set of behavioral characteristics could be reflected on the relations that a particular entity shares with another one.

2 Background

Relation extraction systems can be broadly divided into three categories, i.e supervised, semi-supervised and unsupervised ones. Supervised methods exploit annotated data to train a classifier that is then applied to unseen corpora to predict relations between entities. These methods [2,3,14] model the RE problem as a classification task, in which the target is to assign a candidate relation between two or more entities to one relation type or none given the features associated with this candidate relation. The range of features used varies from simple word-co-occurrence and collocations to shallow parse trees and WordNet-based features. Supervised models applied to RE include naïve Bayes and voted perceptron [14], Support Vector Machines (SVM) [3,14], Conditional Random Fields (CRF) [2] and logistic regression [2].

The work on RE has shown that discriminative models such as SVM and CRF perform better than generative models such as naïve Bayes or voted perceptron due to the fact that they directly model the conditional probability $P(Y|X)$, where Y refers to relation labels and X to observations (features). These models differ from generative models which specify a joint probability distribution over observation and relations labels, $P(Y, X)$, and then form a conditional distribution, $P(Y|X)$, using Bayes rule. To keep the problem tractable and specify the joint distribution generative models often make independence assumptions that hurt the performance, in effect making them less suitable for classification tasks. Additionally, both types of supervised methods require hand-tagged training data, which in effect makes their portability to different domains or applications problematic.

Semi-supervised approaches [5,10,11] attempt to overcome the main limitation of supervised systems, i.e. their portability to domains where annotated data do not exist. Typically, these approaches require a few manually-created relation examples as well as the types of the relations the initial examples describe. For example, the pattern ***PERSON*** *works for* ***ORGANIZATION*** is a relation example of an employment relation. New relations are then extracted by applying such patterns on large corpora and then weighting the patterns and extracted relations using statistical association measures (mutual information, log-likelihood and others). Although these methods have a smaller dependency on hand-tagged data, they do require some initial relation examples as well as

the relation types that these examples represent. As such, these pattern-based approaches are biased towards finding specific-type relations and often achieve high levels of recall at low levels of precision and vice versa.

Unsupervised methods do not make use of any hand-tagged data in effect being portable to different domain or applications. Typically, these approaches [6,12] associate with each pair of entities a vector features that serve as the dimensions of the vector space. These features are the intervening words between the two entities of the pair in a given corpus. Different statistical measures (log-likelihood, tf/idf and others) are used to weight each feature, and finally the resulting feature vectors are clustered. Each cluster groups pairs of entities and represents one relation type. Clustering methods applied so far do not deal with the problem of identifying the number of clusters, i.e. the number of different relations. Additionally, they only exploit local information (pairwise similarities between entities).

In a graph-based framework, however, in which entities are represented as vertices in a graph and edges denote contextual similarity, one could exploit both local and global information. This framework could possibly result in improved performance. In our work, we present a novel graph-based method that exploits both word co-occurrence and syntactic features, whose combination has shown to be perform better than each one individually. This setting allows the formulation of a framework, where RE is cast as an edge prediction problem. This framework exploits both text-mined features, as well as relational features (relations between entities) that might have been extracted in previous iterations or be known a priori.

3 Proposed Method

An undirected graph $G = (V, E)$ is a representation of a set of objects $V = \{v_0, \ldots, v_n\}$, where some pairs of the objects are connected by undirected links. These links are represented by a set of edges $E = \{e_0, \ldots, e_m\}$, where each edge (relation) between two objects (vertices) v_i and v_j is denoted by $e_k = \{v_i, v_j\} = \{v_j, v_i\}$. An undirected graph can be used to illustrate the relationships between entities. Given the three sentence example in Section 1, we can construct the graph shown in Figure 1, in which entities are represented as vertices, while edges between vertices are drawn if their corresponding entities co-occur in a sentence.

While it is straightforward to capture the relation between two entities appearing in the same context, this is not the case for capturing the implied relationships between entities. For instance, there is a direct relationship between *Karen Smith* and *KKC*. However, this is not captured, since *KKC* does not co-occur with *Karen Smith*. The same applies for *George Smith* and *George Brown*. The task of relation extraction can be cast into a graph-based framework in which the aim is to identify the missing edges (missing relations) in a graph based on the relations that already exist in the graph. For instance in Figure 1, *George Smith* and *George Brown* share a relation with *Karen Smith*. Hence, we

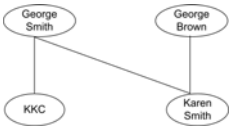


Fig. 1. Named entities & graph example

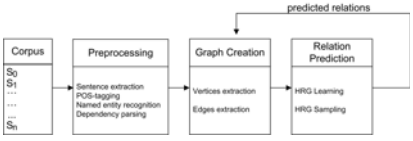


Fig. 2. Stages of our method

could suggest with a certain confidence level that a relation also exists between *George Brown* and *George Smith*. Discovered relations can be added to the initial graph, so that new predictions can also be made by exploiting relational features discovered in previous iterations. Figure 2 shows the conceptual architecture of the proposed method. Each of these stages is described in the following sections.

3.1 Pre-processing

The input to this stage is raw text. Initially, we apply Part-Of-Speech (POS) tagging, lemmatization and Named Entity Recognition (NER). The output for the first sentence of our example would be the following:

George_Smith/PERSON work/VERB for/IN *KKC*/ORGANIZATION

Each processed sentence is syntactically analyzed using the Stanford dependency parser [7]. The produced parse trees are mapped onto a directed graph representation, in which words in the sentence are vertices and grammatical relations are edge labels. Figure 3 shows the dependency graph for the first sentence of our example.

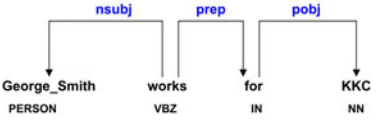


Fig. 3. Dependency graph for the second sentence of our example

Extracting Most Significant Words. The aim of this stage is to identify for a given target entity type, the words most strongly associated to that type. In our work, we focus on 3 target entity types, i.e. *Person*, *Organization* and *Location*. Initially, each named entity instance of the processed corpus is replaced with its corresponding named entity type. Hence, the first sentence of our example would be transformed to the following representation:

PERSON work/VERB for/IN *ORGANIZATION*

Given a target entity type E (e.g. *Person*), we perform a log-likelihood test [4] for each word that co-occurs with the particular entity type. Let w be a word that co-occurs with E in one or more sentences. The first step in the log-likelihood

calculation is to create a table (Table 1 (Left)) that contains the frequencies observed in our corpus. In Table 1 (Left), the word w and the entity type E co-occur in 150 sentences (cell (2,2)), while w occurs in 2430 sentences in which E is absent (cell (2,3)).

In the next step, one formulates the hypothesis that w and E co-occur at chance, i.e. $P(w, E) = P(w)P(E)$. Using the model of independence a second table can be created (Table 1 (Right)) that provides the expected values for the co-occurrence of E and w under that model. The counts of the tables can then compared using the log-likelihood statistic (Equation 1), where o_{ij} refers to the observed value of cell i, j and e_{ij} refers to the expected value of cell i, j. High log-likelihood values indicate strong relationship between E and w. For each target entity type, we produce a list of words ranked by their log-likelihood weights. We select the top n words (parameter p_1) from each list to proceed to the next stage.

$$G^2 = 2 \sum_{i,j} o_{ij} \log \frac{o_{ij}}{e_{ij}} \tag{1}$$

Extracting Dependency Paths. Given the set $l(E)$ of related words for a target entity type E, the aim of this stage is to extract the dependency paths connecting E and each $w \in l(E)$. To perform this task, we use the dependency graphs generated by the Stanford parser [7] as described in subsection 3.1. Specifically, we extract the shortest path connecting E and each $w \in l(E)$ in each of the sentences that both E and $w \in l(E)$ appear together. For example, given the dependency path in Figure 3, the target entity type *Person* and the word *work*, we would extract the path: *work.verb* ***-nsubj-*** *Person*, where the strings in bold represent the edges. Additionally, we extract all the shortest dependency paths connecting a target entity type with any other target entity type. Table 2 shows the two most frequent paths associated with the *Person* entity type in our corpus.

3.2 Creating the Graph of Entities

The aim of this stage is to create the initial graph of entities. Therefore, the identified named entities are represented as vertices in an undirected graph G. To calculate the relatedness between two named entities e_i and e_j and draw the corresponding edge, we assume that entities occurring in similar syntactic contexts tend to be engaged into one or more relations.

Table 1. Observed frequencies (Left) - Expected frequencies (Right)

	E	$-E$
w	150	2430
$-w$	23270	25000

	E	$-E$
w	1188.27	1391.73
$-w$	22231.73	26038.27

Table 2. Most frequent paths for *Person* entity type

Path	Frequency
-nsubj- tell.VERB -dobj- reporter.NOUN	248
-nsubj- tell.VERB -dobj- ORGANIZATION	247

Let e_i be a named entity instance, $T(e_i)$ the name entity type of e_i, $D(t)$ the set of dependency paths associated with named entity type t and $m(D(T(e_i)), e_i)$ the set of dependency paths in $D(T(e_i))$ that entity e_i instantiates. Note that e_i instantiates a dependency path, when it appears either as the first or as the last token of the path. For instance, in Figure 3, *George Smith* instantiates the path *work.verb* ***-nsubj-*** *Person*. Given two entities e_i and e_j, their similarity is defined as the Jaccard coefficient of their corresponding instantiated paths, i.e $S(e_i, e_j) = \frac{|m(D(T(e_i)),e_i) \cap m(D(T(e_j)),e_j)|}{|m(D(T(e_i)),e_i) \cup m(D(T(e_j)),e_j)|}$. If $S(e_i, e_j)$ is higher than a pre-specified threshold (parameter p_2), then an edge is included in the graph.

3.3 Predicting Relationships Using Hierarchical Random Graphs

This section describes the process of predicting relations between entities using Hierarchical Random Graphs (HRG) [1]. HRG is an unsupervised method for inferring the hierarchical structures (e.g. a binary trees) of an undirected graph. These structures divide vertices (entities) into groups (internal nodes) that are further subdivided into groups of groups, and so on. Each inferred hierarchical structure provides additional information as opposed to flat clustering by explicitly including organization at all scales of a graph [1]. The inferred structures are combined in order to associate with each pair of disconnected vertices the probability that an edge (relationship) exists between them.

Inferring Hierarchical Structures. Given an undirected graph G, each of its n vertices is a leaf in a dendrogram. The internal nodes of that dendrogram indicate the hierarchical relationships among the leaves. Such an organization is denoted by $D = \{D_1, D_2, \ldots D_{n-1}\}$, where D_k is an internal node. Hierarchical random graphs [1] associate with each pair of nodes (v_i, v_j) a probability θ_k that an edge $\{v_i, v_j\}$ exists in G. This probability is in turn associated with D_k, the lowest common ancestor of v_i and v_j in D. In this manner, the topological structure D and the vector of probabilities θ define an ensemble of HRGs given by $H(D, \theta)$ [1]. Figure 4 shows two dendrograms, with 4 leaves and 3 parents, for the graph in Figure 1. Our aim is to compute the parameters of D and θ, so that the chosen HRG is statistically similar to the observed graph of entities G. Let D_k be an internal node of dendrogram D and $f(D_k)$ be the number of edges between the vertices of the subtrees of the subtree rooted at D_k that actually exist in G. For example, in Figure 4(A), $f(D_2) = 1$, because there is one edge in G connecting vertices *George Smith* and *Karen Smith*. Let $l(D_k)$ be the number of leaves in the left subtree of D_k, and $r(D_k)$ be the number of leaves in the right

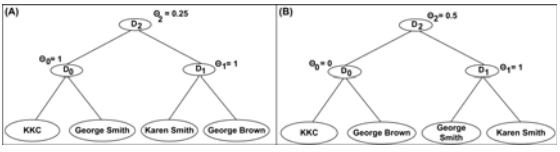


Fig. 4. Two dendrograms for the graph in Figure 1

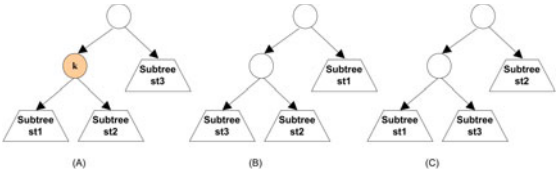


Fig. 5. (A) current configuration for internal node D_k and its associated subtrees (B) first alternative configuration, (C) second alternative configuration

subtree. For example in Figure 4(A), $l(D_2) = 2$ and $r(D_2) = 2$. The likelihood of the hierarchical random graph $L(D, \theta)$ that we wish to maximize is defined in Equation 2, where $A(D_k) = l(D_k)r(D_k) - f(D_k)$.

$$L(D, \theta) = \prod_{D_k \in D} \theta_k^{f(D_k)} (1 - \theta_k)^{A(D_k)} \tag{2}$$

The probabilities θ_k maximizing the goodness of fit of a dendrogram D can be estimated using the method of MLE i.e $\overline{\theta}_k = \frac{f(D_k)}{l(D_k)r(D_k)}$. The likelihood function is maximized when θ_k approaches 0 or 1, i.e. high-likelihood dendrograms partition vertices into subtrees, in which the connections among their vertices in the observed graph are either very rare or very dense. Regarding the dendrograms in Figures 4(A) and 4(B), the first one ($L(D_1) = 0.105$) is more likely than the second ($L(D_2) = 0.062$) providing a better division of the graph entities.

Finding the values of θ_k using the MLE method is straightforward. In the next step, we need to maximize the likelihood function over the space of all possible dendrograms. To perform this task, a Markov Chain Monte Carlo (MCMC) method is applied. The MCMC method samples dendrograms from the space of dendrogram models with probability proportional to their likelihood. To apply MCMC we need to pick a set of transitions between dendrograms. HRGs define a transition as a re-arrangement of the subtrees of a dendrogram. Specifically, given a current dendrogram D_{curr}, each internal node D_k of D_{curr} is associated with three subtrees of D_{curr} derived from its sibling and its two children (Figure 5). Given a current dendrogram, D_{curr}, the algorithm proceeds by choosing an internal node $D_k \in D_{curr}$ uniformly at random, and then choosing uniformly at random one of its two possible new configurations (Figure 5) to generate a new dendrogram D_{next}. The transition is accepted according to the Metropolis-Hastings rule [9], i.e. a transition is accepted if $\log L(D_{next}) \geq \log L(D_{curr})$; otherwise it is accepted with probability $\frac{L(D_{next})}{L(D_{curr})}$.

Following the work of Clauset et al. [1], the convergence heuristic consists of: (1) calculating the difference between the average log-likelihood over X steps and the average log-likelihood over the next X steps, and (2) checking whether this difference is below a fixed threshold (1.0). We use the same X parameter as in the experiments of Clauset et al. [1], where $X = 65536$.

Identifying Novel Relations & Bootstrapping. Each inferred tree associates with each pair of vertices i, j a probability θ_k, which is in turn associated with D_k, the lowest common ancestor of i, j. This probability reflects the probability that an edge exists in our initial graph G. Upon convergence of the MCMC method, we can use the highest likelihood dendrogram, in order to find out the probability that an edge exists between two vertices.

By following such a strategy, however, we would be ignoring the fact that many dendrograms have competing likelihoods, which suggests that rather than focusing on the most likely one, it would be more appropriate to sample the distribution of dendrograms and then predict new edges or relations based on the average produced by the sampled trees. Hence in our setting, once the MCMC method has converged, we sample a total of m dendrograms at regular intervals. Given a pair of vertices i, j for which an edge does not exist in our graph, we can calculate the average probability that an edge should exist using Equation 3. In Equation 3, $LCA(i, j, D_a)$ provides the probability θ_k associated with the lowest common ancestor of i, j in D_a and $L(D_a)$ is the likelihood of dendrogram D_a used to weight $LCA(i, j, D_a)$.

$$P(i,j) = \frac{\sum_{a=1}^{m} LCA(i,j,D_a)L(D_a)}{\sum_{a=1}^{m} L(D_a)} \tag{3}$$

Following the work of Clauset et al. [1], in our experiments: (1) the number of sampled dendrograms is set to 10000, and (2) the top 1% of the predicted edges are extracted as new relations between entities. The extracted relations can be further exploited, in order to create a modified graph G', which now contains the new edges discovered. The modified graph, G', is the new input to HRG, which can now predict new relations. In each iteration of the method, relation extraction is based on two type of features: (1) attribute features extracted during the pre-processing stage, and (2) relational features discovered in previous iterations of the method. The bootstrapping process continues for a pre-defined number of iterations.

4 Evaluation

We evaluate our method on the TAC-2009 corpus [8] that consists of 1.2 million news articles. To create our evaluation corpus, we sampled 40000 sentences containing the words *terrorism*, *terror* and *violence*. We followed this strategy, in order to create a dataset reflecting an area of research in INDECT[1] project. The

[1] http://www.indect-project.eu/

500 most frequent entity mentions were used to identify the relations between them. To assess the performance, in each iteration we sampled uniformly 100 relations from the ones that were predicted by HRGs. Three human annotators manually judged whether these were correct or wrong.

In each iteration, the precision of our method is the ratio of correct relations to the total number of relations considered (100). For all iterations, precision is the ratio of the sum of correct relations in all iterations to the total number of relations considered in all iterations (a maximum of 1000 since we perform 10 iterations). In our case, it is not feasible to measure recall, since the total number of relations in the dataset is not known. Despite that, we evaluate the precision of our method relative to the number of correctly predicted relations, in order to show the trend between precision and recall.

We compare the proposed approach with a baseline, in which the initial graph of entities is created by following a state-of-the-art pattern-based method [13]. The particular pattern-based approach participated in the TAC-2009 slot filling challenge [8] achieving the best accuracy among unsupervised and semi-supervised systems. The initial graph is the input to the relation prediction stage (subsection 3.3). The comparison with such a baseline allows us to observe, whether state-of-the-art pattern-based features [13] are able to perform better than the dependency features used in our approach for graph creation.

4.1 Aggregate Results

Table 3 shows the results of our method under all parameter combinations and all ten iterations. The first parameter of our method selects the top p_1 words from each target entity type list (subsection 3.1). The second parameter of our method (p_2), is the edge similarity threshold between two vertices (subsection 3.2). As can be observed the highest number of correct relations (682) is extracted at a precision of 68.2% ($p_1 = 1500$, $p_2 = 0.02$). Similar performance in terms of the number of correct relations (648) is achieved at 64.8% precision ($p_1 = 1000$, $p_2 = 0.03$), as well as for $p_1 = 500$ and $p_2 = 0.02$. The highest precision (75.71%) is achieved when extracting 70 relations ($p_1 = 1500$, $p_2 = 0.08$).

Figure 6(A) shows the number of correct relations for $p_1 = 500$ and different values of p_2. As can be observed, increasing the edge similarity threshold leads to sparser graphs that include a few relations between entities. The sparsity in turn affects the relation extraction stage predicting a small amount of correct relations. The same picture is also observed for $p_1 = 1000$ and $p_1 = 1500$ respectively. In all cases the maximum amount of correct relations is achieved when p_2 is close to its lowest point.

A similar case exists when p_2 is fixed and parameter p_1 varies, although the impact of parameter p_1 is small, since in most cases the difference between the number of correct relations in different parameter settings is small. Despite that, the general tendency that we observe (especially for $p_2 \geq 0.05$) is that lower p_1 values lead to a higher number of correct relations. Specifically, lower p_1 values

Table 3. Overall results for all parameter combinations

p_2	$p_1 = 500$		$p_1 = 1000$		$p_1 = 1500$	
	Correct relations (#)	Precision	Correct relations (#)	Precision	Correct relations (#)	Precision
0.02	**646**	**64.6**	631	63.1	**682**	**68.2**
0.03	646	64.6	**648**	**64.8**	629	62.9
0.04	645	64.5	609	60.9	603	60.3
0.05	584	64.6	620	62	480	62.34
0.06	522	57.49	284	60.43	181	56.21
0.07	263	51.57	140	56.91	92	60.53
0.08	118	50.86	80	64.52	**53**	**75.71**
0.09	74	59.68	24	60	12	60
0.1	18	60	5	50	0	0

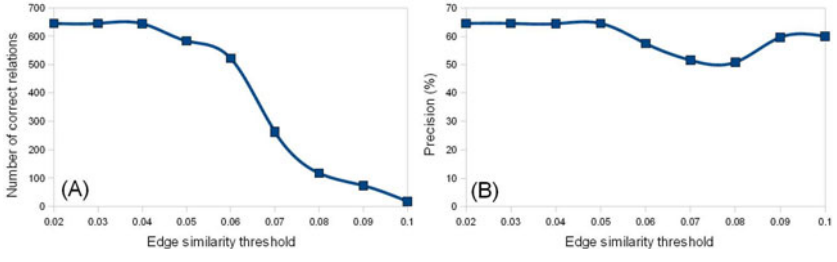


Fig. 6. (A) Number of correct relations, (B) Precision, for $p_1 = 500$, different p_2 values

result to entities whose vectors have fewer dimensions (instantiated dependency paths) than when setting p_1 to higher values. This results in more dense graphs, which in turn assists HRG to predict a higher number of correct relations.

In the previous discussion, we observed that high levels of recall can be achieved by setting p_1 and p_2 close to their lowest points. One would expect that a biased system would achieve its lowest (or a very low) precision at these points. However, this is not the case for our method. Figure 6(B) shows the precision of our method for $p_1 = 500$ and different values of p_2. We observe that for $p_2 = [0.02, 0.05]$ our approach achieves competing levels of precision, while for $p_2 \geq 0.06$ precision drops. This picture is similar for $p_1 = 1000$ and $p_1 = 1500$.

Overall, our results in this section have shown that the highest levels of recall can be achieved by setting our two parameters at their lowest points. At the same time, these lowest points achieve competing levels of precision compared to the rest of parameter combinations. Hence, the conclusion is that the proposed method is able to achieve high coverage of relations without compromising accuracy. From a practical point of view, our findings suggest that a potential user could set the parameters of our methods to low values (as the ones used in this experiment) and expect to retrieve a large number of correct relations.

4.2 Results Per Iteration

Figure 7 shows the precision of the proposed method in each of the ten iterations for two high performing parameter settings (Table 3). As can be observed, the

first iteration achieves higher precision than most of the iterations. Despite that, we also observe that the differences between the highest performing iterations are very small. One would expect that in each iteration the precision would drop, if we were including so many noisy incorrect relations in the modified graph of each iteration as to negatively affect the HRG clustering and the prediction of new edges. However in Figure 7, we observe that the precision in each iteration does not follow such a trend. For example in Figure 7(A), precision drops after the second and third iteration, and then increases in the fourth and fifth. This picture is similar in Figure 7(B). This result in combination with the high precision in the

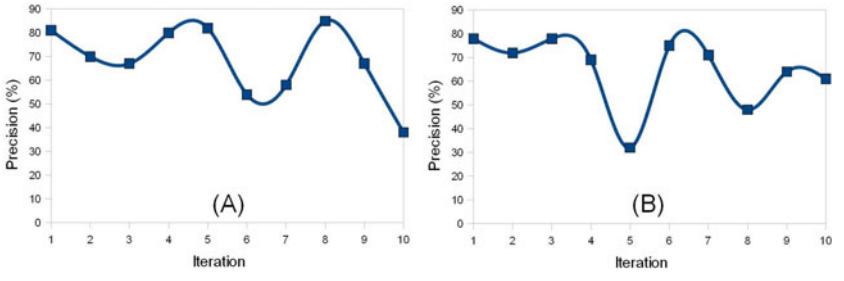


Fig. 7. Precision per iteration, (A) $p_1 = 1500$ & $p_2 = 0.02$, (B) $p_1 = 1000$ & $p_2 = 0.03$

first iteration, suggests that the amount of added relations in the first iteration does not have a direct positive impact in the HRG clustering. This is possibly due to the fact that in each iteration we only select a small number of relations to add to the graph (1% of the candidate edge ranked list). In contrast, HRGs predict some incorrect relations (edges) that were also predicted in the previous iteration, but were not included in the selected top 1%. Despite that, in Figure 7(A) we also observe that after the third iteration, precision increases steadily until the fifth iteration. This indicates that the correctly added relations assist HRGs, in order to predict new correct relations in next iterations to the point, in which the precision reaches and sometimes outperforms the precision in the first iteration. Overall, it seems that the use of a dynamic parameter for selecting the new edges in each iteration is more appropriate than a fixed threshold. This parameter could be set to a high value in the first iterations, reflecting the belief that initial iterations provide accurate relations, and then gradually decrease.

4.3 Comparison with the Baseline

Figure 8(A) shows the precision of the proposed method for $p_1 = 500$, $p_1 = 1000$, $p_1 = 1500$ and different values of p_2 against the precision of the pattern-based baseline for the same p_2 values. Figure 8(B) presents the number of correct relations of the proposed method against the baseline for the same parameter setting as in Figure 8(A). We observe that the proposed method performs always better than the baseline in terms of precision, apart from one extreme parameter combination ($p_1 = 1500$, $p_2 = 0.10$) in which the proposed method suggested

only 10 relations that were incorrect. Additionally, the proposed method extracts a higher number of correct relations than the baseline at its highest precision points. Specifically, the proposed method is able extract its highest number of correct relations (682) at a 68.2% precision ($p_1 = 1500$, $p_2 = 0.02$). In contrast, the baseline extracts its highest number of correct relations (503) at 50.30% precision ($p_2 = 0.08$). Interestingly, the precision of the baseline increases as

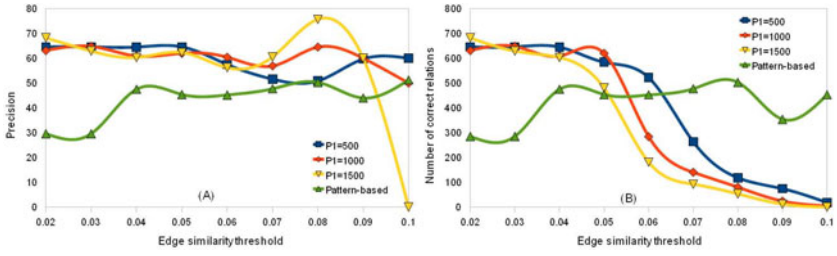


Fig. 8. Comparison of the proposed method against the pattern-based baseline

the edge similarity threshold increases. This means that that the pattern-based features, used for creating the graph, provide noisy edges between vertices in the graph. These edges are filtered out when increasing the edge similarity threshold, in effect providing a higher precision. We also observe that the highest number of correct relations for the baseline are captured when the edge similarity threshold is in the range from 0.04 to 0.08.

The described behavior of the baseline contrasts with the behavior of the proposed method, whose highest levels of recall and precision are captured when the parameters' values are set close to their lowest points. Overall, the conclusion is that our features (dependency paths extracted from words contextually related to a target entity) are: (1) more discriminative than the pattern-based features of the baseline providing a higher precision, and (2) more general than the pattern-based features of the baseline providing a higher number of correct relations.

5 Conclusion

We presented a novel unsupervised and graph-based method for relationship mining. Our evaluation on a large sample of news articles has shown that the highest levels of recall are achieved at high levels of precision, as opposed to semi-supervised RE methods, where precision and recall moved towards opposite directions. The iterative process allows us to infer more relations between entities, although the precision in each iteration varies, which in effect suggests that there is a need to have a dynamic parameter on the selection of predicted relations. This parameter would reflect our belief on the amount of correct relations that each iteration may provide, where initial iterations would be more accurate. The features of our method (dependency paths extracted from words

contextually related to a target entity) are more discriminative as well as more general than the pattern-based features employed by a state-of-the-art slot filling system [13].

References

1. Clauset, A., Moore, C., Newman, M.E.J.: Hierarchical Structure and the Prediction of Missing Links in Networks. Nature 453(7191), 98–101 (2008)
2. Culotta, A., McCallum, A., Betz, J.: Integrating probabilistic extraction models and data mining to discover relations and patterns in text. In: Proceedings of HLT-NAACL 2006, pp. 296–303. ACL, USA (2006)
3. Culotta, A., Sorensen, J.: Dependency tree kernels for relation extraction. In: Proceedings of ACL 2004. ACL, USA (2004)
4. Dunning, T.: Accurate Methods for the Statistics of Surprise and Coincidence. Computational Linguistics 19(1), 61–74 (1993)
5. Etzioni, O., Cafarella, M., Downey, D., Popescu, A.M., Shaked, T., Soderland, S., Weld, D.S., Yates, A.: Unsupervised named-entity extraction from the web: an experimental study. Artif. Intell. 165, 91–134 (2005)
6. Hasegawa, T., Sekine, S., Grishman, R.: Discovering relations among named entities from large corpora. In: Proceedings of ACL 2004. ACL, USA (2004)
7. Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: Proceedings of ACL 2003, pp. 423–430. ACL, USA (2003)
8. McNamee, P., Dang, H.T.: Overview of the tac 2009 knowledge base population track. In: Proceedings of TAC 2009. NIST, USA (2009)
9. Newman, M., Barkema, G.: Monte Carlo Methods in Statistical Physics. Clarendon Press, Oxford (1999)
10. Pantel, P., Pennacchiotti, M.: Espresso: leveraging generic patterns for automatically harvesting semantic relations. In: Proceedings of ACL 2006, pp. 113–120. ACL, USA (2006)
11. Ravichandran, D., Hovy, E.: Learning surface text patterns for a question answering system. In: Proceedings of ACL 2002, pp. 41–47. ACL, USA (2002)
12. Shinyama, Y., Sekine, S.: Preemptive information extraction using unrestricted relation discovery. In: Proceedings of HLT-NAACL 2006, pp. 304–311. ACL, USA (2006)
13. Varma, V., Bharat, V., Kovelamudi, S., Bysani, P., GSK, S., Kumar, K.N., Reddy, K., Kumar, K., Maganti, N.: IIIT hyderabad at tac 2009. In: Proceedings of TAC 2009. NIST, USA (2009)
14. Zelenko, D., Aone, C., Richardella, A.: Kernel methods for relation extraction. J. Mach. Learn. Res. 3, 1083–1106 (2003)

On Occlusion-Handling for People Detection Fusion in Multi-camera Networks

Anselm Haselhoff[1], Lars Hoehmann[1], Christian Nunn[2], Mirko Meuter[2], and Anton Kummert[1]

[1] Communication Theory, University of Wuppertal, D-42119 Wuppertal, Germany
{haselhoff,hoehmann,kummert}@uni-wuppertal.de
[2] Delphi Electronics & Safety, D-42119 Wuppertal, Germany
{christian.nunn,mirko.meuter}@delphi.com

Abstract. In this paper a system for people detection by means of Track-To-Track fusion of multiple cameras is presented. The main contribution of this paper is the evaluation of the fusion algorithm based on real image data. Before the fusion of the tracks an occlusion handling resolves implausible assignments.

For the evaluation two test vehicles are equipped with LANCOM® wireless access points, cameras, inertial measurement units (IMU) and IMU enhanced GPS receivers. The people are detected by using an AdaBoost algorithm and the results are tracked with a Kalman filter. The tracked results are transmitted to the opponent vehicle in appropriate coordinates. The final stage of the system consists of the fusion and occlusion handling.

Keywords: Object Detection, Sensor Data Fusion, Occlusion Handling, Camera Networks.

1 Introduction

Surveillance camera systems as well as driver assistance systems become more and more important. The used technologies are image processing, machine learning, sensor fusion, and communication. Thus the boundary between surveillance systems and driver assistance systems disappears. By means of Car2Infrastructure communciation both research fields can be connected. While in the automotive field systems like forward collision warning and communication units are already available, the next logical step is to connect these individual systems and generate an enriched view of the environment. Therefore a sensor data fusion can be applied. When dealing with object detection from different views, like in a multi-vehicle or camera network scenario, it is important to incorporate occlusion information. If occlusion handling is neglected, incorrect track-assignment can lead to strong errors of the object position.

In this work a system for fusing object detections in multi-camera networks is presented [2]. In addition the topic of occlusion handling is treated. It is not important for the system whether the cameras included in the network are static

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 113–119, 2011.

or mounted in a vehicle. The remainder of the paper proceeds as follows. It is started with the description of the overall system and the test vehicles in section 2. Afterwards the components of the fusion system are described separately, including the track assignment, occlusion handling and the fusion of people detections. Finally, the evaluation and the conclusions are presented in sections 4.

2 System Overview

For demonstration the system setup consists of two test vehicles equipped with LANCOM® wireless access points, cameras, inertial measurement unit, GPS, and a regular PC. The monochrome camera is mounted at the position of the rear-view mirror and is connected to the PC. The vehicle bus enables the access to the inertial measurement unit and GPS. The GPS is used to generate timestamps for the data that is subject to transmission. The GPS unit (AsteRxi system) delivers a position accuracy of a around $2cm$ and a heading accuracy of $1°$. The position information is obtained relative to one vehicle that is defined to be the dedicated master. Finally, the LANCOM® unit is responsible for the data transmission.

Fig. 1 illustrates the system that is used for multi-camera people detection and fusion. First, the image is scanned via an AdaBoost detection algorithm. The detection results are then tracked using a Kalman filter that is working in image coordinates. The tracked detections including the uncertainties are then transformed to appropriate coordinates that can be used for the fusion. The transformation of the uncertainties is implemented using the scaled unscented transformation (SUT) [4]. The calibration component delivers the needed camera position by means of the GPS unit. The transformed detections, including their uncertainties, are then transmitted to the other cameras as well as the camera position. The data is then synchronized using the GPS timestamps and passed to the track-assignment module. Afterwards it is checked for occlusion. Corresponding tracks are finally fused by means of a Track-To-Track fusion algorithm.

3 Fusion of People Detections

Each node in the multi-camera network sends the tracked detection results to the other nodes and receives the tracked detection results of the opponents. The detections are described by their position and their uncertainties in the world coordinate system (WCOS). The state estimates of each track are denoted by $\boldsymbol{\mu}_1$ and $\boldsymbol{\mu}_2$, whereas the fused state is $\boldsymbol{\mu}$. The according covariance matrices are denoted by $\mathbf{C}_1$, $\mathbf{C}_2$, and $\mathbf{C}$. The fusion is subdivided into two components, namely the track assignment and the Track-To-Track fusion.

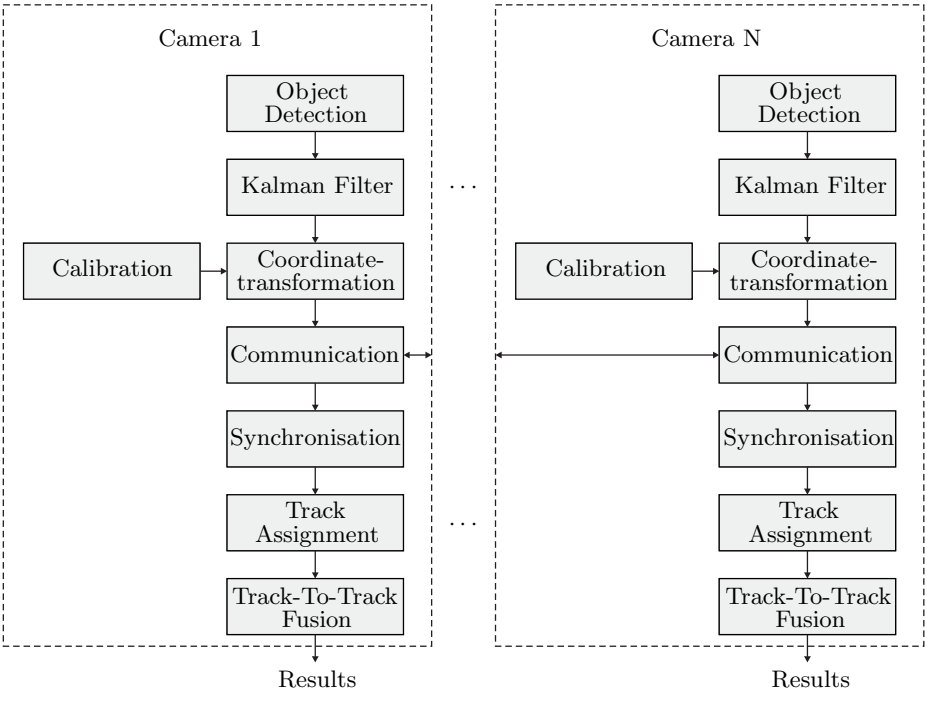


Fig. 1. System for Object Detection Fusion

3.1 Track Assignment and Occlusion Handling

The track assignment is based on a measurement Δ that defines the distance between each track

$$\tilde{\boldsymbol{\mu}} = \boldsymbol{\mu}_1 - \boldsymbol{\mu}_2 \tag{1}$$

$$\Delta = \tilde{\boldsymbol{\mu}}^T \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \tilde{\boldsymbol{\mu}}. \tag{2}$$

Tracks are only fused if the track distance falls below a certain threshold and the tracks must fall into the field of view of both cameras. If these conditions are met the Hungarian method [5] is applied to perform the actual track assignment. After the initial track assignment it is checked for occlusion, to prevent the fusion of implausible assignments.

In Fig. 2 a typical occlusion scenario is shown. There are three ground truth objects, where from V_2 object 2 is occluded by object 1 and from V_1 object 3 is occluded by object 2 (see Fig. 2a). Fig. 2b illustrates a wrong track assignment, where object 3 and object 2 are labeled to belong to the same ground truth object. To resolve this incorrect assignment the projections of the bounding boxes are used (Fig. 2c). The idea is, that if the base point of an object projects into the polygon of any bounding box projection, occlusion has occured.

It is appropriate to fuse two objects that occlude each other, but if in addition one of the objects is occluded by another object, the fusion should be avoided. A simple algorithmic describtion is given in Algorithm 1.

Algorithm 1. Occlusion handling

O_i^1, O_j^2 objects from both vehicles ($V1$ and $V2$) that are subject to fusion
L_i^1, L_j^2 list occluding objects, determined by bounding box projections
if ($L_i^1 = \emptyset$ OR $L_i^1 = \{O_j^2\}$) AND ($L_j^2 = \emptyset$ OR $L_j^2 = \{O_i^1\}$) **then**
  ALLOW FOR FUSION
else
  AVOID FUSION
end if

3.2 People Detection Fusion

Once the track assignment is completed, the tracks are subject to fusion. The choice of the fusion method depends on the data at hand and to what extent the data or sensors are correlated. A comparison of fusion methods on simulated data is given in [3], a survey is given in [6] and a general treatment on data fusion can be found in [1]. Three typical methods for the fusion are *Covariance Fusion* (CF), *Covariance Intersection* (CI), and *Covariance Union* (CU). For our application the CF outperforms the other algorithms [2].

The *Covariance Fusion* (CF) can be defined by the following equations

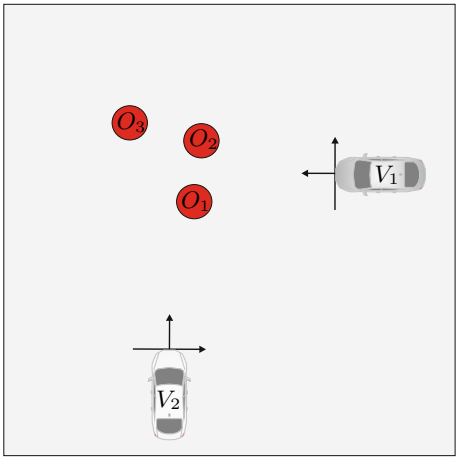
$$\mathbf{C} = \mathbf{C}_1 \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \mathbf{C}_2 \tag{3}$$

$$\boldsymbol{\mu} = \mathbf{C}_2 \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \boldsymbol{\mu}_1 + \mathbf{C}_1 \left(\mathbf{C}_1 + \mathbf{C}_2\right)^{-1} \boldsymbol{\mu}_2. \tag{4}$$

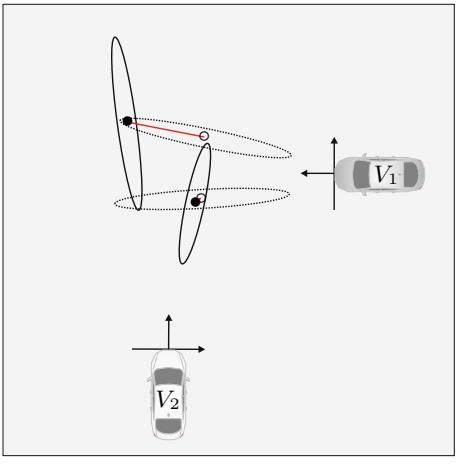
4 Evaluation

The evalution of the fusion method is performed on real data (see Fig. 3) with ground truth information in world coordinates. The system is tested on various video-sequences and different scenarios. The used scenarios are based on typical cross-way situations. In the first scenario both vehicles are approaching the object with an angle difference of 90° and in the second scenario the vehicles are placed at an angle difference of 180°. The distance of the vehicles to the objects at the starting position goes up to around $50m$.

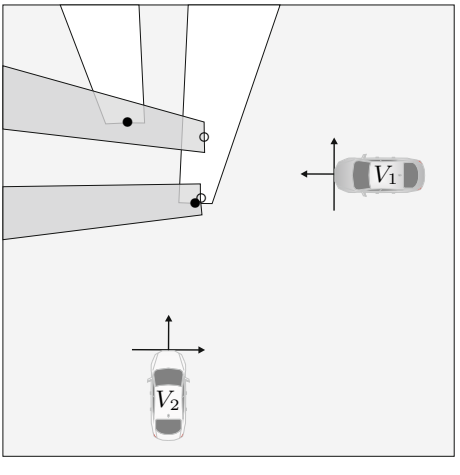
An example of the detection results with and without the occlusion handling is presented in Fig. 3. It becomes clear that the fused detections in Fig. 3a and 3b do not correspond to the ground truth. In contrast if the fusion is avoided for inconsistent tracks the results presented in Fig. 3c and 3d are obtained. These results reflect the ground truth data. To demonstrate the advantage of the fusion, first the detection results of the individual vehicles are evaluated, where a root mean square error (RMSE) of around 4 meters is obtained. The RMSE is calculated using the distance of the ground truth objects to the detections



(a) Ground truth data of three objects. Each camera can perceive two objects. Object 2 is occluded by object 1 and object 3 by object 2.

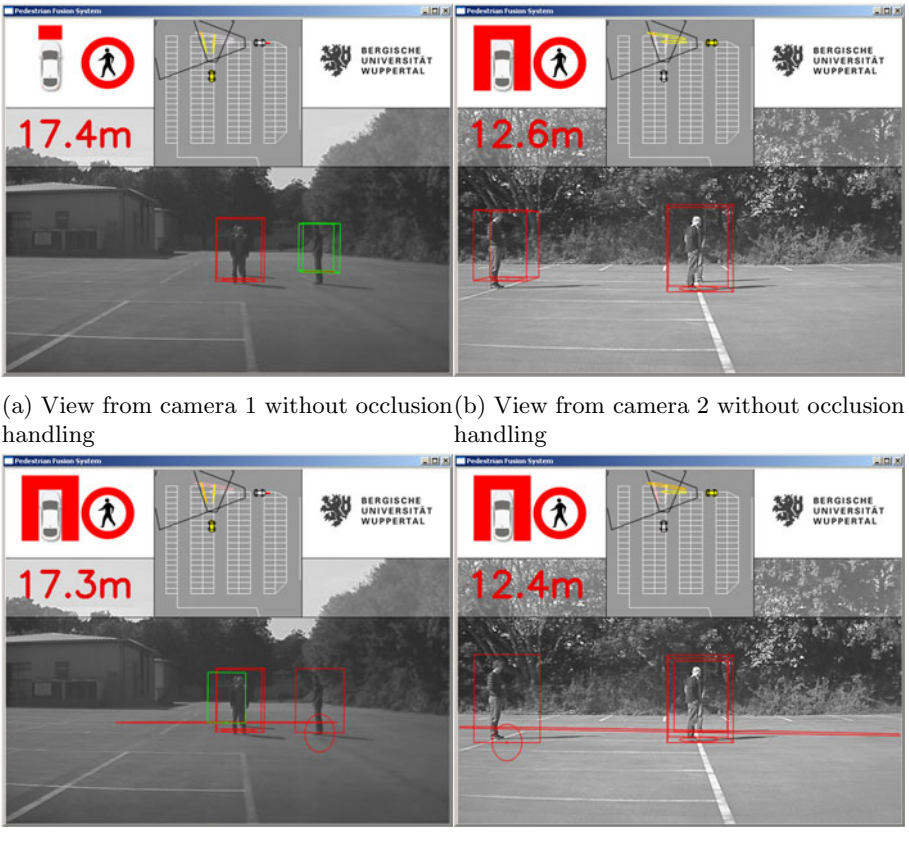


(b) Detected objects including the uncertainties. The track assignment is visualized by the red line. In this case a wrong track assigment has occured.



(c) Bounding box projections onto the ground plane can be used to resolve the wrong track assignment

Fig. 2. Example where the standard track assignment fails



(a) View from camera 1 without occlusion handling

(b) View from camera 2 without occlusion handling

(c) View from camera 1 using occlusion handling

(d) View from camera 2 using occlusion handling

Fig. 3. Fusion results of two cameras mounted in vehicles. Fig. 3a and 3b belong to the same timestamp, where a fusion leads to wrong results. The ego vehicle is gray and the remote vehicle is yellow. The trajectory of the vehicles is denoted by red dots. Fig. 3c and 3d belong to the same timestamp, where the occlusion handling is applied.

in world coordinates. As expected the lateral position (with reference to the vehicle coordinate system) of the objects can be measured precisely, whereas the depth information is inaccurate. The results in Table 1 reveal that the fusion dramatically improves the overall precision. The RMSE of d_w (distance of ground truth object and prediction) gets improved.

Table 1. RMSE: Fusion of two cameras

method	x_w [m]	y_w [m]	d_w [m]
scenario 1			
CF	0.35	0.55	0.69
scenario 2			
CF	0.15	1.00	1.02

5 Conclusions

A promising system for fusing detection results from multiple cameras was presented. The performance results underline the practicability of the approach. The CF method is appropriate for our application since the correlation of the data is very small, e.g. correlation coefficient of around 0.3. The occlusion handling avoids the fusion of objects that would result in erroneous ground truth objects.

References

1. Bar-Shalom, Y., Blair, W.D.: Multitarget-Multisensor Tracking: Applications and Advances. Artech House Inc., Norwood (2000)
2. Haselhoff, A., Hoehmann, L., Kummert, A., Nunn, C., Meuter, M., Mueller-Schneiders, S.: Multi-camera pedestrian detection by means of track-to-track fusion and car2car communication. In: VISAPP (to be published) (2011)
3. Matzka, S., Altendorfer, R.: A comparison of track-to-track fusion algorithms for automotive sensor fusion. In: Proc. of International Conference on Multisensor and Integration for Intelligent Systems, pp. 189–194. IEEE, Los Alamitos (2008)
4. Merwe, R.V.D., Wan, E.: Sigma-point kalman filters for probabilistic inference in dynamic state-space models. In: Proceedings of the Workshop on Advances in Machine Learning (2003)
5. Munkres, J.: Algorithms for the assignment and transportation problems. Journal of the Society for Industrial and Applied Mathematics 5(1), 32–38 (1957)
6. Smith, D., Singh, S.: Approaches to multisensor data fusion in target tracking: A survey. IEEE Transactions on Knowledge and Data Engineering 18(12), 1696–1710 (2006)

An Informatics-Based Approach to Object Tracking for Distributed Live Video Computing

Alexander J. Aved, Kien A. Hua, and Varalakshmi Gurappa

Department of Electrical Engineering and Computer Science
University of Central Florida, Orlando, FL USA
aaved@mail.ucf.edu,kienhua@cs.ucf.edu,varu2u@gmail.com

Abstract. Omnipresent camera networks have been a popular research topic in recent years. Example applications include surveillance and monitoring of inaccessible areas such as train tunnels and bridges. Though a large body of existing work focuses on image and video processing techniques, very few address the usability of such systems or the implications of real-time video dissemination. In this paper, we present our work on extending the LVDBMS prototype with a multifaceted object model to better characterize objects in live video streams. This forms the basis for a cross camera tracking framework based on the informatics-based approach which permits objects to be tracked from one video stream to another. Queries may be defined that monitor the streams in real time for complex events. Such a new database management environment provides a general-purpose platform for distributed live video computing.

Keywords: Distributed Computing, Network, Camera, Query Evaluation, Tracking.

1 Introduction

Camera networks have been the subjects of intensive research in recent years, and can range from a single camera to a network of tens of thousands of cameras. Thus, usability is an important factor for its effectiveness. As an example, the camera network in London costs £200 million over a 10-year period. However, police are no more likely to catch offenders in areas with hundreds of cameras than in those with hardly any [6]. This phenomenon is typical for large-scale applications and can be attributed to the fact that the multitude of videos is generally not interesting and constantly monitoring these cameras for occasional critical events can quickly become fatiguing. Due to this limitation in real time surveillance capability, cities mainly use video networks to archive data for post-crime investigation.

To address this issue, automatic video processing techniques have been developed for real time event detection under various specific scenarios [5]. For instance, a camera network that is deployed for traffic monitoring uses customized software to process the video streams for detecting car accidents in real time. This customized approach, however, is not applicable to a general-purpose camera network with many user groups who use the same camera network for different purposes. As an example,

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 120–128, 2011.

university police can use a university camera network to monitor security around the campus while a professor may use a camera on the same camera network to be informed when his research assistant arrives at the lab. It is also desirable to provide the capability to enable rapid development of various customized applications for domain-specific users, such as counting cars along a section of highway, or informing an employee when a nearby conference room becomes available.

To achieve the aforementioned properties, we propose a live video database model for live video computing. In this framework, the cameras in the network are treated as a special class of storage with the live video streams viewed as database content. With this abstraction, a general-purpose query language can be designed to support ad hoc queries over the live video streams. The availability of the general-purpose query language also enables rapid development of camera-network applications, much like many of today's applications are developed atop a relational database management system. Based on this live video database model, we have designed and implemented a Live Video Database Management System (LVDBMS) [7].

In this paper, we present our work on extending the LVDBMS prototype with (1) a multifaceted object model to better characterize objects in live video streams, and (2) a cross camera tracking framework based on an informatics-based approach which permits queries to be defined that span multiple video streams.

The remainder of this paper is organized as follows. In Section 2 we present an overview of the LVDBMS, we introduce the proposed multifaceted object model and the cross-camera tracking technique in Section 3. Object detection and tracking is presented in Section 4, and related work in Section 5. We present our performance evaluation in Section 6. Finally, we provide our conclusions in Section 7.

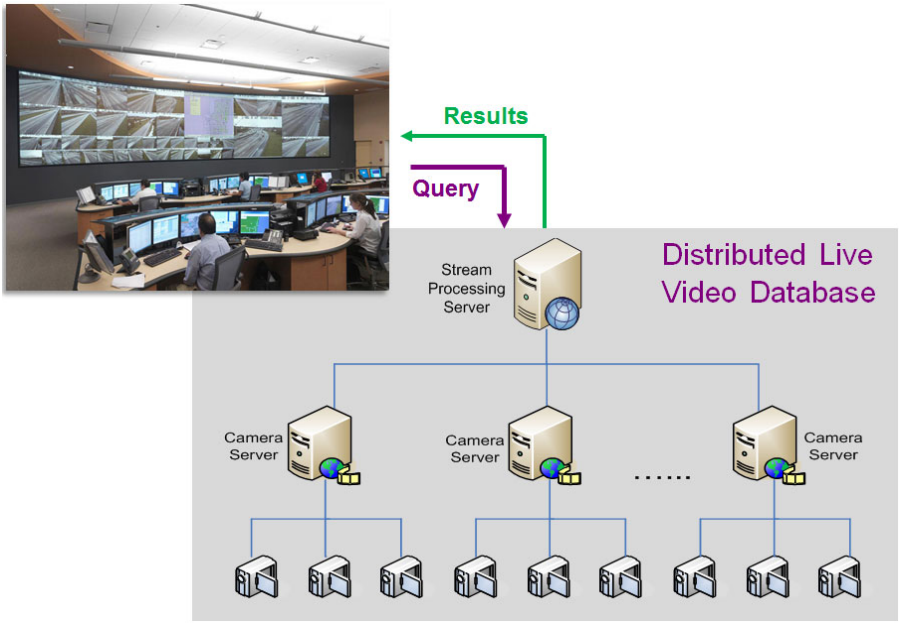


Fig. 1. The LVDBMS hardware environment. Camera Servers are spatial processing layer hosts.

2 LVDBMS System Overview

In this section we briefly introduce LVDBMS, a video stream processing Live Video Database (LVD) environment, and refer the reader to [7] for details. A query is a spatiotemporal specification of an event, posed over video streams and expressed in Live Video SQL (LVSQL), a declarative query language. By implementing a general query language and web services interface the LVDBMS can facilitate rapid application development in video stream processing.

The LVDBMS is based upon a distributed architecture which is logically grouped into four tiers which communicate through web services interfaces. The *camera layer* consists of physical devices that capture images, paired with a camera adapter which runs on a host computer (illustrated on the lowest level of **Fig. 1**). The camera adapter allows any type of camera to be used with the LVDBMS. The camera adapter first performs scene analysis on the raw image data, making processed image descriptor information available to the higher layers in the architecture. *Spatial processing layer* hosts evaluate spatial and temporal operators over the streams of image descriptors and send results to the stream processing layer. The *stream processing layer* accepts queries submitted by clients and partial query evaluation results from the spatial processing layer. As queries are decomposed and pushed down to relevant camera servers in the spatial processing layer for partial evaluation, a query may require processing on multiple camera servers depending on which video streams it is posed over. As sub queries are evaluated, results are streamed from camera servers up the hierarchy and the final query result is computed in a stream processing layer host. Users connect to the LVDBMS and submit queries using a graphical user interface in the *client layer*.

3 Multifaceted Object Model

The proposed object tracking technique is based on a multifaceted object model. In this environment, objects are tracked from frame to frame using a traditional tracking technique (e.g. [11]). In this work we refer to this as a frame-to-frame tracker, since it tracks objects within a single video stream. When an object appears in a consecutive sequence of frames, the frame-to-frame tracker assigns a unique identifier to the object as a part of the tracking process. A feature vector based upon the object's appearance is also calculated. More formally, an object is represented as a bag of multiple instances [12-13], where each instance is a feature vector based upon an object's visual appearance at a point in the video stream. Thus, an object can be viewed as a set of points in the multidimensional feature space, referred to as a point set (**Fig. 2**). In our model, the size of a bag or point set is fixed. The k instances in a bag are derived from k samplings of the object (which may not necessarily be taken from consecutive frames).

We implemented a first in first out (FIFO) database to hold the multiple-instance bags of objects recently detected by the different cameras in the system. As new observation becomes available, the bag itself is updated by adding the new instance and removing the oldest instance. As surveillance queries generally concern real-time events that have occurred recently, the FIFO database is typically very small, and in

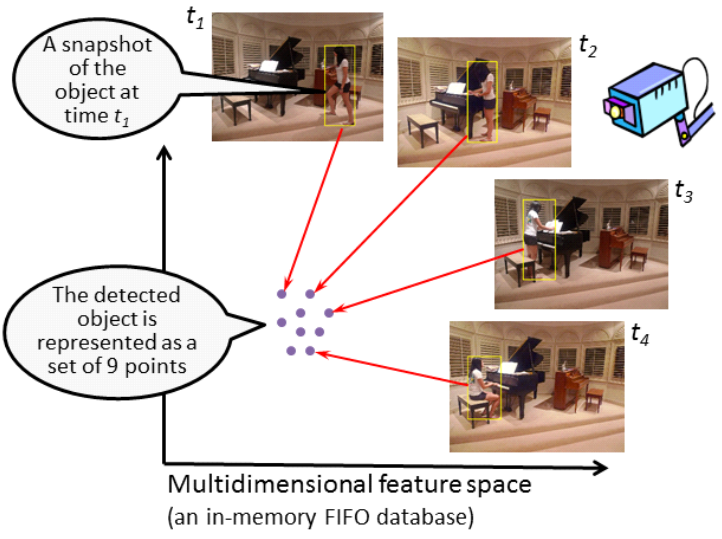


Fig. 2. Objects are identified in video streams and tracked from frame to frame. As their appearance changes over time the collective appearance is represented in a multidimensional feature space as a point set (bag of instances).

our prototype we implemented it as a distributed in-memory database system (distributed among spatial processing layer hosts). The number of bags retained is a function of currently active queries and bounded by a system threshold.

4 Cross-Camera Object Tracking

Cross-camera object tracking is performed as follows. When an object is detected by a camera, its multiple-instance bag is extracted from the video stream and used as an example to retrieve a sufficiently similar bag in the distributed object-tracking database. If the retrieval returns a bag (i.e., , the newly detected object is considered as the same as the object returned by the database. On the other hand, if the system does not find a sufficiently similar bag in the database, the occurrence of this newly detected object is considered as its first appearance in the system recently. By logging the sequence in which the different cameras detect the same object, we can track the trajectory of this object as it moves from camera to camera.

To support the retrieval operations, the distributed in-memory object-tracking database needs to compute similarity between multiple-instance bags. Given two bags of multiple instances:

$$X = \{\vec{x}_1, \vec{x}_2, \ldots, \vec{x}_k\} \text{ and } X' = \{\vec{x}'_1, \vec{x}'_2, \ldots, \vec{x}'_k\},$$

where k is the cardinality of the bags, we compute their similarity as follows:

$$d^m(X, X') = min_{\tau_i, \tau'_i} \sum_{i=1}^{m} \left\| \vec{x}_{\tau_i} - \vec{x}'_{\tau_{i'}} \right\|^2,$$

where $m \leq k$ is a tuning factor and $\left\| \vec{x}_{\tau_i} - \vec{x}'_{\tau_{i'}} \right\|^2$ is the squared distance between the two vectors. This distance function computes the smallest sum of pairwise distance

between the two bags. Although we can set $m=k$, a smaller k value is more suitable for real-time computation of the distance function. For instance, if $m=1$, two objects are considered the same if their appearances look similar according to some single observation. We set $m=5$ in our study. Traditionally, each object is represented as a feature vector, i.e., a single point, instead of a point set, in the multidimensional feature space. This representation is less effective for object recognition. For clarity's sake, let us consider a simple case in which two different persons currently appear in the surveillance system. One person wears a 2-color t-shirt with white in the front and blue in the back. Another person wears a totally white t-shirt. If the feature vectors extracted from these two persons are based on their front view, the two would be incorrectly recognized as the same object. In contrast, the proposed multifaceted model also takes into account the back of the t-shirt and will be able to tell them apart. The bag model is more indicative of the objects. The parameter *m,* discussed above, controls the number of facets considered by the system.

For computation efficiency, The FIFO database consists of a series of distributed queues and indices residing in spatial processing layer hosts. Each host maintains indices of the bags of objects observed in video streams from corresponding camera adapters. Indices associate objects with specific video frames in which they were observed, video frames with video streams, objects to bags, objects with queries over the objects' corresponding video streams, etc. Objects appearing in two separate video streams will have two separate bags in the index and two separate identifiers (camera adapter identifier, local object tracking number and host identifiers concatenated into a string). If the two objects are determined to be the same object, their bags are merged and the index updated such that both object identifiers point to the same (merged) bag of observations.

5 Evaluation

To test the effectiveness of the object model and tracking technique we used videos from the CAVIAR project [14]. Evaluations are conducted with pre-recorded videos to permit repeated measurements with different parameters for system tuning.

Object recognition is essential to object tracking. Our system attempts to identify objects using the proposed object tracking database. In our experiments, if the database returns a result and it is the correct (same) object, true positive (TP) is incremented; else false positive (FP) is incremented. Likewise, if no result is returned, TP is also incremented if this object has not appeared before; else we increment false negative (FN). We present the performance results in terms of the *quality, precision* and *recall* metrics:

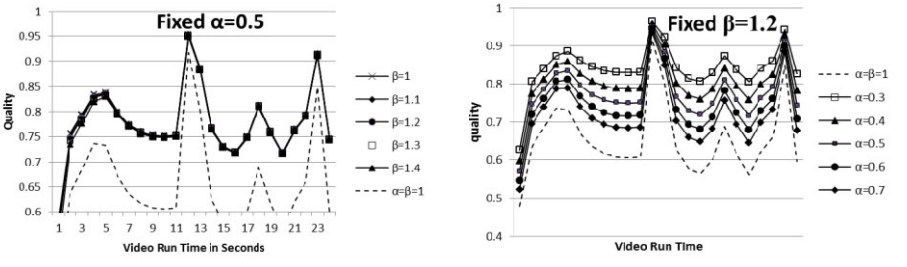
$$\text{Quality} = \frac{\text{TP}}{\alpha \cdot \text{FP} + \beta \cdot \text{FN} + \text{TP}}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

where $0 \le \alpha \le 1 \le \beta$. In the case $\alpha=\beta=1$, quality is equated to the *Accuracy* metric commonly used in evaluating content-based image retrieval systems. Video surveillance

applications have slightly different concerns. Although FP is an inconvenience for causing an unnecessary alert or recording to take place, it is not as serious as a FN which is more damaging because a specified event may occur and fail to trigger a notification action. The proposed quality metric is, therefore, suited for performance evaluation in applications such as surveillance. It discriminates between FP and FN by adjusting their weighting factors α and β, respectively.

We present the sensitivity analysis with respect to these two weighing factors in **Fig. 3**. In **Fig. 3**(a), α is fixed and the *Quality* measurements for various values of β are plotted. In this video, "OneShopOneWait2cor," there were initially some FN's; then as more observations were added to the database, no more FN's occurred resulting in good *Quality* at about 0.80. For comparison we include *accuracy* (α=β=1, the dashed line).



(a) (b)

Fig. 3. Quality plotted for videos (a) "OneShopOneWait2cor", varying β, and (b) "Walk3", varying α (FP)

In **Fig. 3**(b), we show the *Quality* where β is fixed and we vary α. The inflection points in the graph correspond to objects in the video leaving or entering a scene. **Table 1** provides running precision and Recall tracking results for approximately the first third of the Walk3 video, giving additional details that cannot be determined by the plot in **Fig. 3** (b).

Table 1. Cumulative precision and recall results from the Walk3 video

TP	FP	FN	Precision	Recall
223	**38**	**0**	**0.85**	**1**
285	**73**	**0**	**0.79**	**1**
391	**108**	**0**	**0.78**	**1**
492	**146**	**0**	**0.77**	**1**
588	**191**	**0**	**0.75**	**1**
706	**230**	**0**	**0.75**	**1**
759	**265**	**0**	**0.74**	**1**
844	**314**	**0**	**0.72**	**1**

In terms of lessons learned, the quality of video is important as well as the performance of the underlying operators which model the scene background and extract foreground objects. The CAVIAR dataset was chosen as it provides numerous scenes from two views: front and side. However, as the video resolution is relatively small (384 by 288 pixels), and pixels representing the people depicted generally comprise a fraction of the frames containing them. The result is that objects are represented by relatively few pixels, meaning relatively few pixels are used in the feature vector calculations, which represent the objects in the database. Similarly, a high resolution video in which the objects of interest are far away from the camera would illustrate a similar effect. Regarding the low level computer vision operators which segment foreground objects, the ability to recognize when a disconnected set of pixels belongs to the same object can affect performance of the tracker. If a single object is disconnected by the underlying operators it will appear as two distinct objects to the higher-level operators which track and calculate feature vectors, they will treat the objects as separate. However, other cameras or later in the same scene the object may eventually be recognized as a single object. This could happen, for instance, when a person enters the camera's view such that they initially appear large, and eventually smaller as they walk farther away from the camera's location. In this scenario, their head and their body might initially be segmented as separate objects. Though this problem has been essentially solved in computer vision research, it still is a challenge to implement as a component in a real-time system.

6 Related Work

Our LVDBMS approach manages live video streams coming from hundreds of video cameras, much like a traditional database management system managing captured data sets stored on a large number of disk drives. However, in the context of live video databases, the video data is generated online, and very little prior processing of the live video data is possible. This is unlike conventional video retrieval systems [15-18]; LVDBMS has performance concerns associated with segmenting and efficiently tracking objects in real time. For example, the VDBMS presented in [15] allows the user to search video based on a set of visual features and spatiotemporal relationships. The automatic feature extraction in this system is done offline so it is not suitable for the real-time needs of our LVDBMS. To the best of our knowledge, there is no other work targeting a general purpose live video database management system that provides real-time spatiotemporal query support over a camera network. These systems require a prior processing and indexing of the entire video collection.

Existing multi-camera tracking environments [1-4, 8-10] expect information about the spatial relationships between the various cameras. Furthermore, they assume overlapping fields of view of the cameras, or non-random object movement patterns. In the latter scenario, when an object moves from the field of view of one camera into another, it can be recognized in the second camera by taking into consideration its speed and trajectory as it exits the first camera's field of view [8-9]. This strategy is only applicable to non-random movement patterns, for example when objects are constrained by roads or walls. Considering cameras deployed in the rooms of a

building: a person leaving a room can enter any other room, or reenter the original room. Since our technique does not rely on assuming a trajectory, such a scenario would not affect our result.

7 Conclusions

Camera networks have many important applications in diverse areas including urban surveillance, environment monitoring, healthcare, and battlefield visualization. In this paper, we present a distributed object recognition and cross-camera tracking technique based upon a multifaceted object model. Unlike existing multi-camera object tracking techniques that rely on various restrictions and assumptions, our general-purpose in-memory FIFO database approach allows general spatiotemporal queries to be posed over objects appearing in live video streams. We evaluated the tracking technique in the LVDBMS system and the results indicate the proposed technique is effective to support queries over multiple video streams.

References

1. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. ACM Comput. Surv. 38(4) (2006)
2. Du, W., Piater, J.: Multi-camera People Tracking by Collaborative Particle Filters and Principal Axis-Based Integration. In: Asian Conf. on Computer Vision (2007)
3. Song, B., Roy-Chowdhury, A.: Stochastic Adaptive Tracking in a Camera Network. In: IEEE Intl. Conf. on Computer Vision (2007)
4. Tieu, K., Dalley, G., Grimson, W.: Inference of Non-Overlapping Camera Network Topology by Measuring Statistical Dependence. In: IEEE Intl. Conf. on Computer Vision (2005)
5. Velipasalar, S., Brown, L.M., Hampapur, A.: Detection of user-defined, semantically high-level, composite events, and retrieval of event queries. Multimedia Tools Appl. 50(1), 249–278 (2010)
6. The London Evening Standard. Tens of thousands of CCTV cameras, yet 80% of crime unsolved (2007), http://www.thisislondon.co.uk/news/article-23412867-tens-of-thousands-of-cctv-cameras-yet-80-of-crime-unsolved.do
7. Peng, R., Aved, A.J., Hua, K.A.: Real-Time Query Processing on Live Videos in Networks of Distributed Cameras. International Journal of Interdisciplinary Telecommunications and Networking 2(1), 27–48 (2010)
8. Hu, M.K.: Visual Pattern Recognition by Moment Invariants, IRE Trans. Info. Theory, vol. IT-8, pp.179-187 (1962)
9. Javed, O., Rasheed, Z., Shah, M.: Tracking Across Multiple Cameras with Disjoint Views. In: The Ninth IEEE International Conference on Computer Vision (ICCV), Nice, France (2003)
10. Javed, O., Rasheed, Z., Shah, M.: Modeling Inter-Camera Space-Time and Appearance Relationships for Tracking across Non-Overlapping Views. Computer Vision and Image Understanding Journal 109(2) (February 2008)

11. Hampapur, A., Brown, L., Connell, J., Ekin, A., Haas, N., Lu, M., et al.: Smart Video Surveillance, Exploring the concept of multi-scale spatiotemporal tracking. IEEE Signal Processing Magazine (March 2005)
12. Cheng, H., Hua, K.A., Yu, N.: An Automatic Feature Generation Approach to Multiple Instance Learning and its Applications to Image Databases. In: The International (2009)
13. Chen, X., Zhang, C., Chen, S., Chen, M.: A latent semantic indexing based method for solving multiple instance learning problem in region-based image retrieval. In: Seventh IEEE International Symposium on Multimedia, pp. 8, 12–14 (2005)
14. CAVIAR: Context Aware Vision using Image-based Active Recognition (2011) http://homepages.inf.ed.ac.uk/rbf/CAVIAR/
15. Adali, S., Candan, K.S., Chen, S., Erol, K., Subrahmanian, V.S.: Advanced video information systems: data structures and query processing. ACM Multimedia Systems 4, 172–186 (1996)
16. Donderler, M.E., Saykol, E., Ulusoy, O., Gudukbay, U.: BilVideo: A video database management system. IEEE Multimedia 1(10), 66–70 (2003)
17. Flickner, M., Sawhney, H., Niblack, W., Ashley, J., Huang, Q., Dom, B., et al.: Query by image and video content: The QBIC system. IEEE Computer 28, 23–32 (1995)
18. Jiang, H., Montesi, D., Elmagarmid, A.K.: VideoText database systems. In: Proceedings of IEEE Multimedia Computing and Systems, pp. 344–351 (1997)

M-JPEG Robust Video Watermarking Based on DPCM and Transform Coding

Jakob Wassermann

Dept. of Electronic Engineering
University of Applied Sciences Technikum Wien
Hoechstaedtplatz 5, 1200 Vienna
jakob.wassermann@technikum-wien.at

Abstract. Robust Video watermarking for M-JPEG data stream based on DCT-Transform and DPCM Encoder is introduced. For this purpose the DPCM encoder is modified and feeds with DCT spectral coefficients of the incoming frames. The difference of the spectra of two sequential frames are used for the embedding the watermarks. To enhance the embedding performance and the security of watermarks some rearrangement of the pixels of the video frames and of the watermarks are introduced. The first permutation was applied to generate a random distribution of the pixels inside a block. The second was applied on the watermark and the third is responsible for the rearrangement of the frame order inside the GOP. The permutation inside the frame and the frame order increases the encryption ability and reduces visible degradation due to watermarking embedding procedure. To realize robust transmission the modified spectral coefficients are encoded into an M-JPEG data stream.

Keywords: Video Watermarking, DPCM; DCT, JPEG, Steganography.

1 Introduction

A watermarking of video is a new and very quickly developing area. It can be used to guarantee authenticity, to identify the source, creator and owner or authorized consumer. Also it could be used to transport hidden information and can be used as a tool for data protection. There are a lot of water-marking technologies that specially fit to the demand of video requirements. Most of them are frame based, it means they using the I-Frame only of the GOP's. (Group of Picture's) [1] and the DCT (Discrete Cosine Transformation) approach of still images [2].

In this paper a new method for robust video watermarking optimized for M-JPEG data stream is introduced. It's based on classical DPCM (Differential Pulse Code Modulation) and DCT Transform. Instead of ingest the watermarks into the spectrum of every frame (frame based approach) the hidden information is embedded into the differences between the spectra of two sequential frames, which is generated by DPCM. The DPCM approach for Watermarking in time domain was already introduced in [3]. The results show that it's enables the height capacity watermarking, however with low robustness. By introducing the same principles in spectral domain enhance the robustness and can be adapted to M-JPEG data stream for error free transmission.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 129–136, 2011.

2 Classical DPCM

The classical DPCM is a data compression method that removes the existing redundancy between the frames. In Figure 1 closed loop DPCM encoder is depicted.

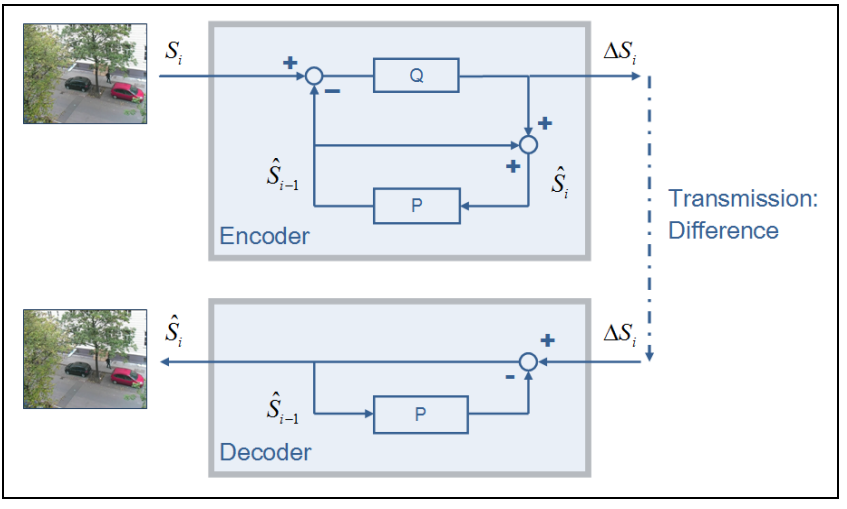


Fig. 1. Classical DPCM

From the incoming video frame S_i the previous frame $\hat{S}_{i-1}$, which is stored in the predictor memory P, is subtracted. We obtain the difference between these two frames that undergoes the quantization process by quantizer Q
(Eq. (1)).

$$\Delta S_i = S_i - \hat{S}_{i-1} + \delta_i \tag{1}$$

δ_i stands for the added quantization noise. Simultaneously the quantized difference image is used to reconstruct the original frame on the encoder side, by adding the previous reconstructed frame to it.

$$\hat{S}_i = \Delta S_i + \hat{S}_{i-1} \tag{2}$$

From the transmitted difference image the DPCM decoder is able to reconstruct the frames. The reconstruction procedure is shown in Eq. (2). The closed loop is used to avoid the accumulation of the quantization errors.

3 W-DPCM in Spatial Domain

To use the DPCM techniques for watermarking the encoder and decoder were modified and well adapted for the needs of watermarking technology. Instead of hiding watermarks in every frame separately, the differences between the frames are watermarked. To realize this, a modification of closed loop DPCM was done. In the following Figure 2 new developed DPCM based watermarking algorithm in time domains is depicted.

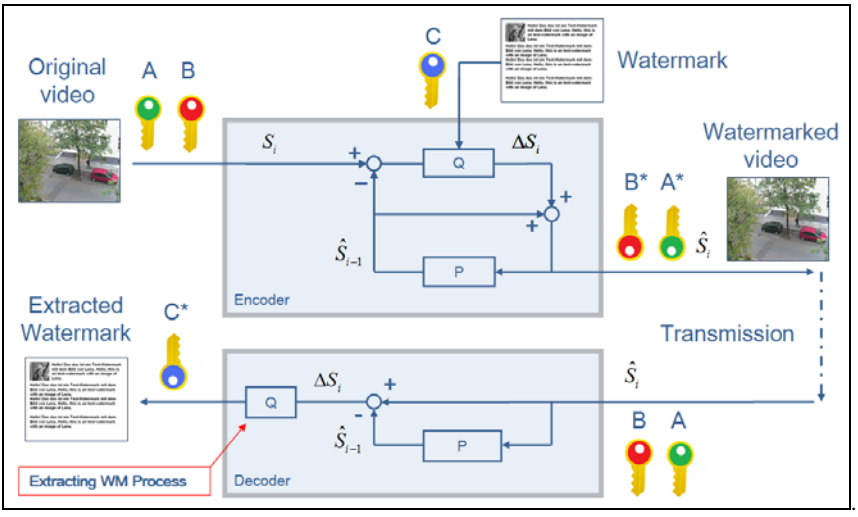


Fig. 2. DPCM based Watermarking Algorithm

From the actual frame the predicted frame from P is subtracted and the so called differential image is undergoing the quantization process Q (Eq. (1)). Through steering the quantizer functionality by watermark content, it is possible to embed watermarks into these differential images. It is done by modifying the LSBs (Least Significant Bits) of the differential values in dependency of the watermark content

The differential image is created by subtracting the incoming video frames from the predicted ones. The following equation (3) describes the decoding procedure.

$$\Delta S_i = \hat{S}_i - \hat{S}_{i-1} \tag{3}$$

Very important is to underline the fact, that in this scheme the predictor on the encoder side works identically as the predictor on the decoder side (property of closed loop DPCM). The obtained differential image is undergoing the extraction procedure by analysing their LSB values according to the codebook of the watermark.

The extraction procedure is very simple and is reverse to the embedding process. If $|\Delta S_i|$ is even, watermark pixel is black otherwise it is white.

The Permutation can improve the performance dramatically. Three permutation keys are introduced: Key A is permutating the order of the frames, key B permutating the pixel of the frames and key C the watermark itself.

4 W-DPCM in Frequency Domain

The big disadvantage of W-DPCM in spatial domain is the lack of robustness. Therefore the W-DPCM codec was modified for spectral domain operations, which enable to realize a transmitting of watermarked video width reasonable data rate in M-JPEG data format. In Figure 3 the modified W-DPCM codec is depicted.

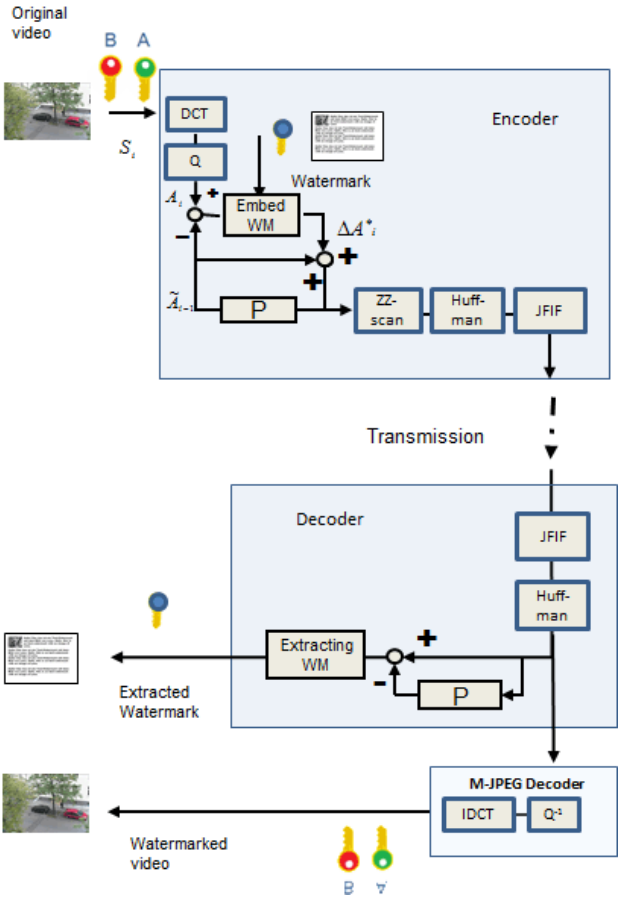


Fig. 3. W-DPCM Codec in Frequency Domain

It's works similar like W-DPCM in spatial domain. The main distinction is, that the difference between the actual and predicted frame is transformed to spectral domain by the DCT unit. After quantization of these DCT coefficients some selected frequencies are modified by the DPCM quantizer to embed the watermark information. In this manner gained new set of coefficients are undergoing additional JPEG compression procedures like Zig-Zag Scan, RLE and Huffman Codings and a data stream with corresponding JFIF header is created for transmission. It contains the modified JPEG coefficients which carries the watermark information.

On the decoder side the access to these DCT coefficients are done by decrypting the incoming data stream through Huffman decoder. It is important to emphasize that in this stage the decrypted coefficients on the encoder and decoder side are absolutely identical. Now the watermarks can be extracted by inverse dpcm operation (dpcm decoder).Parallel to watermark extraction the decoded DCT Coefficients undergoes the inverse DCT to visualize the received video.

The DCT Transform is block based. It uses the size of 8*8 pixels per block. The number of significant coefficient can be increases by introducing GOP and Block based Permutation.

The watermarks are placed into significant coefficients of a DCT Block by modifying their LSB. The significant coefficients are defined by a threshold. Most of them are located nearby the DC. Actually the number of such coefficients is small, which has negative impact on the capacity.

5 Using Permutation to Improve the Capacity and Security of the Watermarked Sequence Domain

To enhance the embedding performance and the security of watermarking procedure some rearrangement of the pixels of the video frames and of the watermark pixels are introduced. For this reason a key B was applied to generate a random permutation of the pixels inside the DCT block. The block himself has the size of 8x8 pixels. The task of key C is a rearrangement of the watermark pixels (Figure 4). The key A is responsible for the rearrangement of the frame order inside the GOP. It improves significantly the quality and the security of the video. Of course on the decoder side all this permutations should be reversed with the same keys.

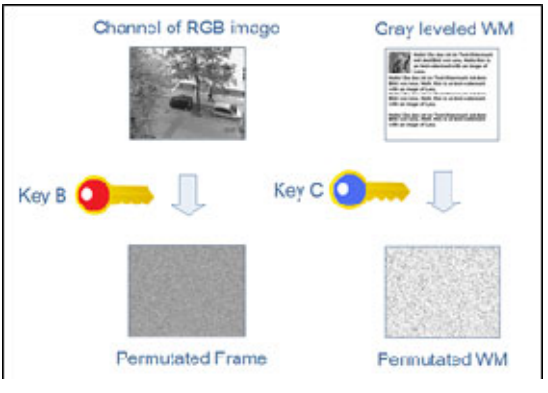


Fig. 4. Functionality of Key A and Key C

6 Results

The investigation was done with a video sequence in DV-Format, which has a resolution of 720x576 pixels and 25 fps. Only the green channel of RGB was used for embedding procedure. The embedded watermarks have the resolution of 90x72 pixels with different depths (grey levels). It was investigated how many grey levels the embedded watermarks can have without causing visible degradations. The

degradation of the watermarked output video was measured with SSIM (Structural Similarity) index. SSIM is based on the human eye perception and so the expressiveness about distortion is better than in the traditional methods like PSNR (Peak Signal to Noise Ratio) or MSE (Mean Square Error) [7]. The output video has a data rate of 4 Mbit/s.

In the results of the table 1 the measured SSIM Index, represent the similarity between the original and watermarked and with M-JPEG compressed sequences.

Table 1. Results of Watermark Capacity Embedded in M-JPEG Sequence

WM Depth in Bits	Embedded Data (Bit)	SSIM	Data rate of embedded Video
1	6480	0,9961	4 Mbit/s
2	12960	0,9828	4 MBit/s
3	19440	0,8891	4 Mbit/s
4	25920	0,7302	4Mbit/s

In the Figures 5 and 6 the decoded M-JPEG Frames with different amount of hidden Information are depicted. Especially in figure 6 distortions are visible. It seems to be the capacity limit.

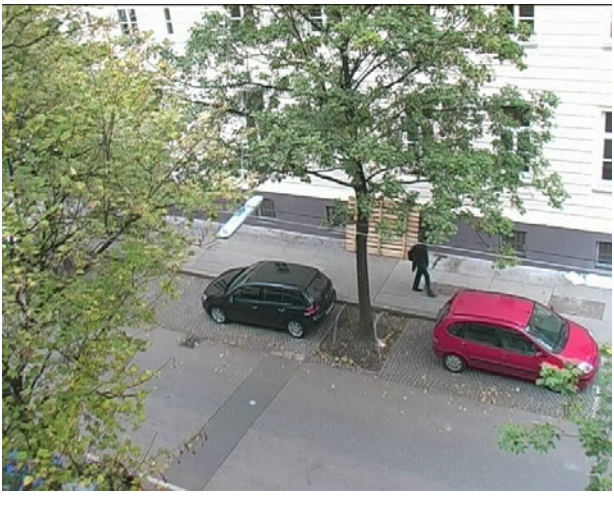


Fig. 5. Output of M-JPEG coded Frame with hidden 12960 bit of WM Information, SSIM=98,82%

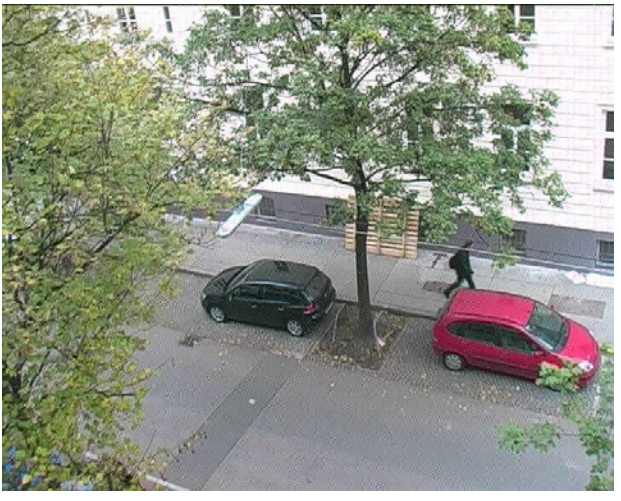


Fig. 6. Output of M-JPEG coded Frame with hidden 19440 bit of WM Information, SSIM=88,91%

7 Conclusions

A new method for robust embedding of watermarks into video sequences based on classical DPCM and spectral transform was introduced. The watermarks can be fully reconstructed without any degradation although the video was compressed by M-JPEG codec. The amount of embedded data is about 12.960 bits. It means a watermark of 12960 bit can be hidden into compressed video without visible degradations. The similarity index SSIM is 99,98%. The output video has a data rate of 4 Mbit/s and it could be transmitted with hidden information through a narrow channels. Any further increasing of embedded information cause significant degradation in the output sequences.

References

1. Benham, D., Memon, N., Yeo, B.-L., Yeung, M.M.: Fast Watermarking of DCT-based compressed images. In: Proceeding of International Conference and Imaging Science, Systems and Applications, pp. 243–252 (1997)
2. Hartung, F., Girod, B.: Digital watermarking of raw or compressed video. In: Proceedings of European EOS/SPIE Symposium on Advanced Imaging and Network Technologies, Digital Compression Technologies and Systems for Video Communication, pp. 205–213 (1996)
3. Wassermann, J., Moser, G.: New Approach in High Capacity Video Watermarking based on DPCM Coding and MPEG Structure. In: Proceeding of Multimedia Communications, Services and Security, MCSS 2010, May Krakow, pp. 229–233 (2010)
4. Chung, T.Y., Hong, M.S., Oh, Y.N., Shin, D.H., Park, S.H.: Digital watermarking for copyright protection of MPEG-2 compressed video. IEEE Transaction on Consumer Electronics 44, 895–901 (1998)

5. Jayant, N.S., Noll, P.: Digital coding of waveforms-principals and applications to speech and video. Prentice Hall, Englewood Cliffs (1984)
6. O'Neal Jr., J.B.: Differential pulse code modulation with entropy coding. IEEE Trans. Inform. Theory IT-21, 169–174 (1976)
7. Sun, M.-T., Reibmann, A.R.: Compressed videos over network, Signal Processing and Communication Series, Marcel Dekker, NY (2001)
8. Wang, Z., Bovik, A. C., Sheikh, H. R., Simoncelli, E. P.: Image Quality Assessment: From Error Visibilty to Structural Similarity. IEE Transaction on Imageprocessing 13(4) (April 2004)

Assessing Quality of Experience for High Definition Video Streaming under Diverse Packet Loss Patterns

Lucjan Janowski, Piotr Romaniak, and Zdzisław Papir

Department of Telecommunications,
AGH University of Science and Technology, Krakow, PL-30059 Poland
Tel.: +48 12 617 48 06
{janowski,romaniak,papir}kt.agh.edu.pl
http://qoe.kt.agh.edu.pl/

Abstract. An approach for derivation a QoE model of High Definition video streaming in existence of packet losses of different patterns is presented. The goal is achieved using the SSIM video quality metric, temporal pooling techniques and content characteristics. Subjective tests were performed in order to verify proposed models. An impact of several network loss patterns on diverse video content is analyzed. The paper deals also with encountered difficulties and presents intermediate steps to give a better understanding of the final result. The research aims at the evaluation of a perceived performance of IPTV and video surveillance systems. The model has been evaluated in the Quality of Experience (QoE) domain. The final model is generic and shows high correlation with the subjective results.

Keywords: Objective evaluation techniques, Subjective evaluation techniques.

1 Packet Losses in HDTV Video Streaming

IP transmission of video streams can be affected with packet losses even if some kind of a resource reservation algorithm is used. Therefore, an influence of packet loss on the perceived quality has to be carefully considered.

Recent premiere of High Definition IPTV brought new requirements on terms of bit-rate and quality of service assurance. Competition on the markets is fierce and service providers desperately seek video quality monitoring and assurance solutions in order to satisfy more and more quality aware customers. The impact of network losses on the perceived video quality is still challenging task because (among others) "not all packets are equal" as claimed in [2].

Evaluation of packet loss effect on video content has been extensively analyzed over recent years. However, hardly few results can be found on a High Definition content. One of the first substantial publications on this particular topic describes performance of the NTIA General Video Quality Metric (VQM) in the task of High Definition TV (HDTV) video quality assessment [12]. Another research published recently is dedicated exclusively to network losses [4], which

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 137–143, 2011.

is in contradiction with the recent results published by the VQEG group under the HDTV Final Report [6]. The report shows that even for full reference metrics it is very difficult to obtain an accurate objective model.

In the presented research an influence of different network loss patterns on the perceived video quality is investigated. We focus on Full HD videos being diverse in terms of content characteristics. The objective model of video quality is based on the SSIM metric so it represents an image analysis approach. Moreover, it has been shown that SSIM averaged over all video frames is not a sufficient quality indicator. More advanced strategies for temporal pooling aided with content characteristics calculation have been proposed. Derivation the final model is presented step by step with all the problems met along the way. The final results prove the correctness of the applied methodology and a high quality prediction accuracy of the model.

2 Experiment Description

Tests of HD video streaming have been carried out in an emulation environment. It was assumed that the emulation environment has to fulfill the following set of requirements:

- Emulation of a multi-node network,
- Unicast and multicast transmission,
- Ability of real-time 100 Mbps transmission of HD video,
- Simple and efficient control of an experiment run.

Experiments were carried out for two levels of aggressiveness of best effort traffic (pace of increasing throughput of the best effort traffic in case of lack of losses) in ten different scenarios under diverse packet loss patterns [1].

The idea behind the emulation system assumes that video streaming operations should be separated from operations of emulation of a multi-node network. As a result, the emulation system was deployed in three separate computer systems (Fig. 1). Video server and video client systems are elements of video streaming system, while yet another server emulates a multi-node network (see Fig. 2).

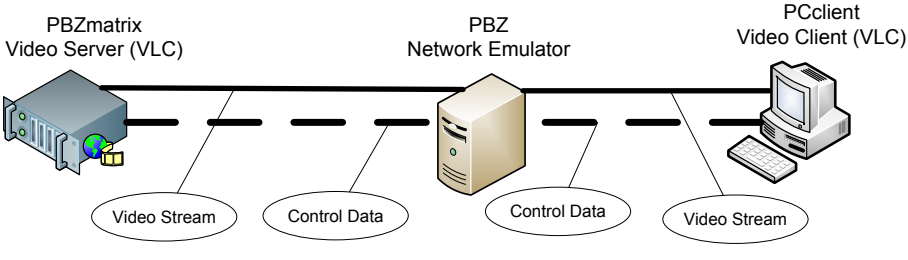


Fig. 1. Block diagram of the emulation system — logical connections

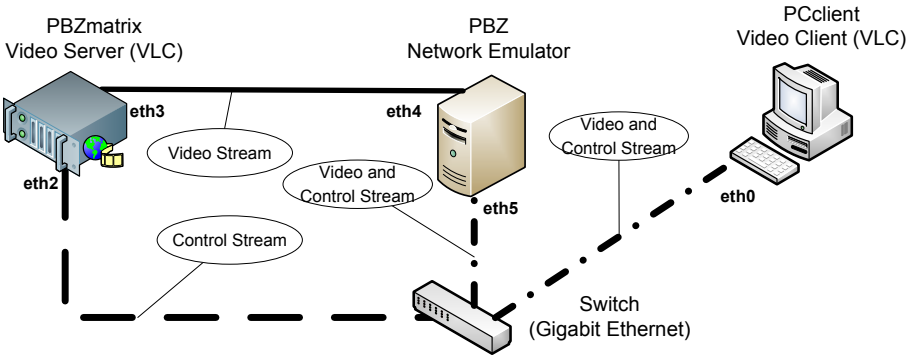


Fig. 2. Block diagram of the emulation system — physical connections

In the experiment eight VQEG test sequences were used [7,8]. These sequences target a variety of spatial and temporal activities within different applications (movies, sports, advertisement, animation, broadcasting news, home video and general TV material: e.g., documentary, sitcom, series of television shows).

The psychophysical experiment followed the VQEG HDTV Test Plan methodology [7], while the test room conformed to ITU-R Rec. BT.500-11 [3] (see Fig. 4).

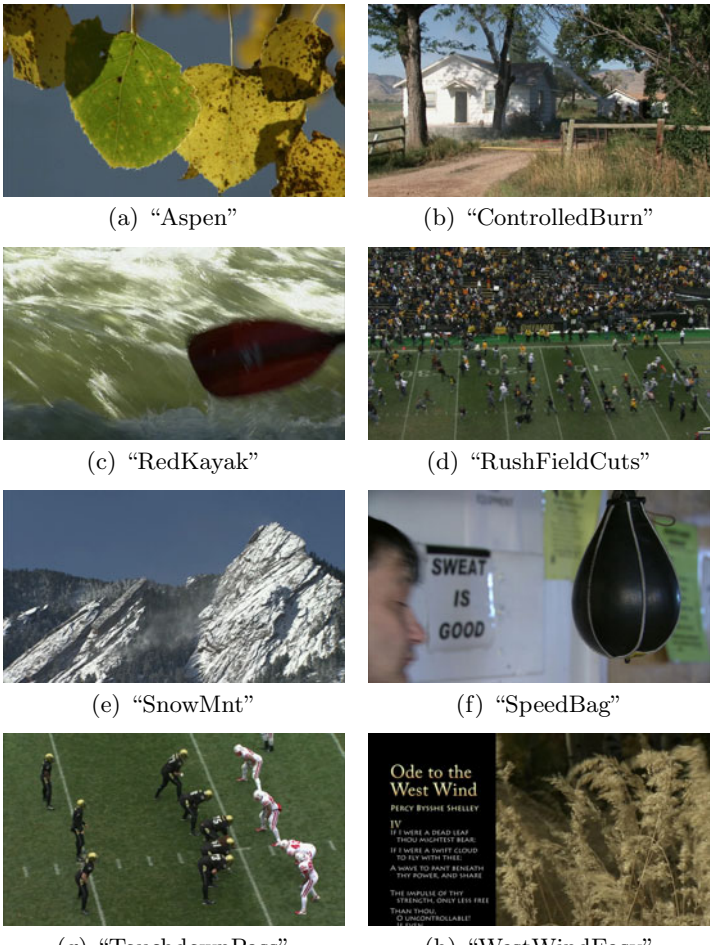
3 Video Quality Metric and Synchronization of Sequences

The Structural Similarity Index Metric (SSIM) operating in a full reference mode has been selected for evaluation of the experiment results as it is easily available, correlates with a human perception well so is widely accepted in the video QoE community [9,10,11]. The additional rationale follows Wang's results [10] that the human visual system (HVS) is sensitive to structural information (provided on an image in the viewing field) and packet losses do inject structural changes to video frames.

An important task to be fulfilled prior to SSIM calculation is synchronization of the reference and distorted sequences. The "reference" sequences are the ones streamed over the "perfect" scenario, while the "distorted" are those streamed over lossy scenarios. In result, SSIM = 1 for the corresponding frames not affected with packet loss should be obtained. In the investigated case synchronization stands for temporal alignment of video frames in order to assure that only the corresponding video frames from both sequences are analyzed. Two important problems to be solved were detection of missing frames in distorted sequences and synchronization recovery after long freezes caused by an extensive packet loss.

4 Derivation of a Model

In the first straightforward attempt the average SSIM value, calculated over all 300 frames for each sequence, was considered. The resulting correlation with



(a) "Aspen" (b) "ControlledBurn"

(c) "RedKayak" (d) "RushFieldCuts"

(e) "SnowMnt" (f) "SpeedBag"

(g) "TouchdownPass" (h) "WestWindEasy"

Fig. 3. Eigth VQEG Test Sequences

DMOS values obtained in the subjective experiments was not satisfactory ($R^2 = 0.55$). A better fitted model may be proposed only after a visual inspection of SSIM plots resulting in additional parameters addressing different loss patterns. An analysis of SSIM plots indicates that a number of separate losses, single loss of a long duration, and spatial extend of a loss — all these factors can possibly matter. Hence, in the second attempt these factors were addressed by introducing new parameters: 1) average SSIM — AvgSSIM, 2) SSIM averaged over the worst second (the lowest SSIM value) — Worst(1s), 3) number of separate losses — NLoss, and 4) number of frames with the SSIM value below 0.9 — NF(0.9). The resulting correlation of the model with DMOS was much higher ($R^2 = 0.84$).

Fig. 4. Psychophysical experiment environment

Two of the selected parameters show the highest statistical significance (i.e. p-value equal to 0), namely Worst(1s) and NLoss. Therefore we decided to simplify our model and make it more generic and eliminate possible over-fitting to the data. In order to account for diverse video content we used spatial and temporal characteristics. In statistical analysis temporal activity was removed as being insignificant (p-value higher than 0.05, according to [5]). Therefore, the final model includes spatial activity SA and two other significant parameters. The model is given by Eq. 1:

$$\begin{aligned} \mathrm{D} = &-5.10 * \mathrm{WorstSq(1s)} - 0.077 * \mathrm{NLoss} \\ &+0.0031 * \mathrm{SA(1s)} + 4.65 \end{aligned} \tag{1}$$

where D is DMOS value predicted by the model and WorstSq(1s) is given by Eq. 2.

$$\mathrm{WorstSq(1s)} = \sqrt{1 - \mathrm{Worst(1s)}} \tag{2}$$

The proposed model correlates with DMOS at a satisfactory level of $R^2 = 0.87$ while preserving a generic form.

5 Conclusions

Tests of HD video streaming application in an emulation environment revealed that the perceived quality of HD video streaming is prone not only to packet losses but then to their patterns. The relevant DMOS predicting model includes three parameters: spatial activity, number of separate losses, and SSIM averaged over the worst second. Other suspected parameters appeared statistically insignificant. A simple form of the model based on the available SSIM metric makes it recommendable for assessment of Quality of Experience (QoE) of IPTV and surveillance systems.

Acknowledgment

The work presented in this paper was supported by the Polish Ministry of Science and Higher Education under the Grant "Future Internet Engineering".

References

1. Chodorek, R., Leszczuk, M.: QoE Validation of a RSVP Protocol Extension Enabling Efficient Resource Reservation for Aggregated Traffic in Heterogeneous IP Networks. In: QoMEX 2010 Second International Workshop on Quality of Multimedia Experience. Centre for Quantifiable Quality of Service in Communication Systems at the Norwegian University of Science and Technology, Trondheim, Norway (June 2010)
2. Greengrass, J., Evans, J., Begen, A.C.: Not all packets are equal, part 2: The impact of network packet loss on video quality. IEEE Internet Computing 13, 74–82 (2009), http://dx.doi.org/10.1109/MIC.2009.40
3. ITU-R: Recommendation 500-10: Methodology for the subjective assessment of the quality of television pictures. ITU-R Rec. BT.500 (2000)
4. Li, W., Issa, O., Liu, H., Speranza, F., Renaud, R.: Quality assessment of video content for hd iptv applications. International Symposium on Multimedia 0, 517–522 (2009)
5. Natrella, M.: NIST/SEMATECH e-Handbook of Statistical Methods (July 2010), http://www.itl.nist.gov/div898/handbook/
6. VQEG: Report on the Validation of Video Quality Models for High Definition Video Content, http://www.vqeg.org/
7. VQEG: The Video Quality Experts Group, http://www.vqeg.org/
8. VQEG: VQEG HDTV TIA Source Test Sequences, ftp://vqeg.its.bldrdoc.gov/HDTV/NTIA_source/
9. Wang, Z.: The SSIM Index for Image Quality Assessment, http://www.cns.nyu.edu/~zwang/

10. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004), http://dx.doi.org/10.1109/TIP.2003.819861
11. Wang, Z., Lu, L., Bovik, A.C.: Video quality assessment based on structural distortion measurement. Signal Processing: Image Communication 19(2), 121–132 (2004), http://dx.doi.org/10.1016/S0923-5965(03)00076-6
12. Wolf, S., Pinson, M.: Application of the NTIA General Video Quality Metric (VQM) to HDTV Quality Monitoring. In: Proceedings of The Third International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM). Scottsdale, Arizona, USA (January 2007)

LDA for Face Profile Detection

Krzysztof Rusek, Tomasz Orzechowski, and Andrzej Dziech

AGH University of Science and Technology,
Department of Telecommunications Krakow, Poland
{krusek,tomeko}@agh.edu.pl,
dziech@kt.agh.edu.pl
http://kt.agh.edu.pl

Abstract. This paper presents a new approach to face profile classification. It was shown that the Linear Discriminant Analysis combined with the Principal Component Analysis can be used to construct an accurate face profile predictor. Proposed classifier achieves 85% accuracy in face profile prediction.

Keywords: LDA, PCA, Profile, Classification.

1 Introduction

Existing face recognition techniques, based on Principal Components Analysis (PCA), are not translation or rotation invariant [5]. Even a small face displacement results in a substantial decrease of accuracy. The same situation is with images taken from a different angles. However, those methods are content agnostic, i.e. they return the most similar object, regardless of the object type. This led us to the idea that different types of face pictures can be segregated into few groups. For each group then, we can have the separate search system.

In general the face image belongs to the one of three class: the left half-profile, the right half-profile and the frontal (en face). From the face recognition point of view the frontal face is the best one. However is some situations we have the only left or the right profile image. Supposing we want find who is on that picture. Using the database containing only the frontal images is pointless, because the recognition ratio is too low. On the other hand if there were a database of left and right faces, we would have a better recognition accuracy. In order to build such a system we need a robust face profile classifier.

The another use case where profile classification is useful is following. Supposed we have a few surveillance cameras, and from each we have a face picture. Using the face profile classifier we can choose the best image for our system.

In this paper we present the new profile classifier based on the Linear Discriminant Analysis. There were some related work see for example [2]. However, according to best authors knowledge, this is the first attempt at using Linear Discriminant Analysis for face profile detection.

This work is divided into six sections. In Section 2 presents some brief description the Principal Component Analysis. In Section 3 the face discriminant

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 144–148, 2011.

function is discussed. Obtained results are presented and assessed in Section 5. Finally, Section 6 concludes the paper.

2 PCA

Principal Component Analysis (PCA) is a statistical technique that transforms linearly given data set to the new one with substantially less number of dimensions without significant loss of the information contained in the original data [4,6].

Suppose we have m independent observation of p dimensional random vector. In this paper those observations are the reshaped images from the training set. These observation can be represented as a $m \times p$ matrix $\boldsymbol{x}$ whose rows are the observed vectors. Let us denote these rows by $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m$.

The PCA finds the transformation matrix $\boldsymbol{W}$ transforming $\boldsymbol{x}$ to the new feature space $\boldsymbol{y} = \boldsymbol{W}^T\boldsymbol{x}$. This transformation has the property, that sample variance of the first direction in $\boldsymbol{y}$ has the greatest variance, the second direction has the second greatest variance and so on. The k-th coordinate of $\boldsymbol{y}$ is the k-th principal component of $\boldsymbol{x}$. In order to construct the matrix $\boldsymbol{W}$, we need the sample covariance matrix of $\boldsymbol{x}$, which is for the unknown mean defined as [4]:

$$\boldsymbol{S} = \frac{1}{m-1}\boldsymbol{x}^T\boldsymbol{x}. \tag{1}$$

It is possible to construct up to p principal component, however, the most variation of $\boldsymbol{x}$ is accounted in the first n of them, and $n << p$ therefore, the matrix $\boldsymbol{W}$ has dimensions $m \times n$. Each column of $\boldsymbol{W}$ is the eigenvector of $\boldsymbol{S}$ corresponding to one of the m largest eigenvalues. The matrix $\boldsymbol{W}$ can be easily computed even for a very large p; see [5] for references. Therefore, the PCA is very efficient tool used in the dimensionality reduction problems.

3 The Discriminant Function

Fisher [3] was the first to propose using the linear combination of features as a class predictor. In this paper we follow his idea,to the linear discriminant function to discriminate among the face profiles.

The traditional discriminant function obtained for the maximum a posteriori has the form [6]

$$D_i^2(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{\mu}_i)^T\boldsymbol{S}^{-1}(\boldsymbol{x} - \boldsymbol{\mu}_i), \tag{2}$$

where μ_j is the sample mean of class j, and S is the covariance estimate. For estimate the covariance matrix we used the pooled covariance matrix defined as:

$$\boldsymbol{S} = \sum_{i=1}^{c} \frac{n_i\boldsymbol{S_i}}{n-c}, \tag{3}$$

where $\boldsymbol{S_i}$ is the sample covariance matrix for class i.

The new observation x is assigned to the class for which

$$L_i(\boldsymbol{x}) = -D_i^2(\boldsymbol{x})/2 + log(p_i) \tag{4}$$

is a maximum, where p_i is the prior probability for class i. Alternatively, the assigment may be based on the class probabilities given by:

$$P(\text{class } i|\boldsymbol{x}) = \frac{p_i e^{-D_i^2(\boldsymbol{x})/2}}{\sum_{j=1}^{c} p_j e^{-D_j^2(\boldsymbol{x})/2}}, \tag{5}$$

where c is the number of classes. It can be proved that if all populations have multivariate normal distribution with a common covariance matrix and prior probabilities, the classification based on maximum L_i is equivalent to classifying observation based on the Fisher's discriminant function [6]. In practice the prior probabilities are ignored and the new statistic

$$L_i^*(\boldsymbol{x}) = \boldsymbol{\mu}_i^T \boldsymbol{S}^{-1} \boldsymbol{x} - \boldsymbol{\mu}_i^T \boldsymbol{S}^{-1} \boldsymbol{\mu}_i/2 \tag{6}$$

is used for classification. This new statistic can be expressed as a linear transform of the shifted original data or equivalently as a linear transform of the data merged with the extra column of ones.

4 The Face Profile Classifier

Detection of the face profile is an example of the typical classification problem. We have three types of classes: the left, the right and the frontal. The problem is to assign the new face to the one of the classes. We propose to use the LDA to solve this problem. The LDA can not be applied to the entire image because the matrix $\boldsymbol{S}$ gets singular. If the pooled covariance matrix were not singular, the calculations would require too much processing power. Therefore, we reduce the dimensionality of the problem by using only a few principal components of the image.

At the beginning the algorithm has to be trained. Therefore, we need to prepare three $\boldsymbol{x}^{(i)}, i = 1, 2, 3$ sets containing only the single profile images. Because the database we used contains only a few right profile images, we will consider only two classes, namely the frontal and the left profile (later called a profile). However, it is straightforward to extend this method to case of the three classes.

The training is being done in two stages. In the first stage all training sets are merged into one training set $\boldsymbol{x}$. Then, the PCA is performed on $\boldsymbol{x}$ and the few principal components are piked. Using the PCA projection matrix, all training sets $\boldsymbol{x}^{(i)}, i = 1, 2$ are mapped to the new sets $\boldsymbol{y}^{(i)}, i = 1, 2$ with highly reduced dimensionality. In the second stage the new sets $\boldsymbol{y}^{(i)}, i = 1, 2$ are used to find the discriminant function.

Once the training is done, the algorithm can be used to find the profile of the new face. The classification scheme is presented in Fig. 1. The incoming picture, after some processing is projected onto the PCA subspace and then onto the

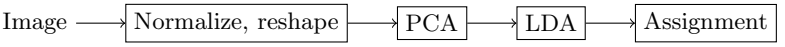


Fig. 1. The classification scheme for face profile classification

LDA space. At the end of the classification process, the image is assigned to the class for which the probability it belongs to is the highest. Computing the matrices for the PCA and the LDA is quite slow, but once it is finished, the matrices can be stored for the later use. This makes the classification process from Fig. 1 very fast.

5 Experimental Results

The results to be presented in this section were obtained using the missing persons database [1]. One hundred images were downloaded and split into two subsets, the training one and the testing one. The training sets, consisted of 15 images of each profile. The remaining 70 images were used as a testing set. Each image was normalized to standard size 100×100px.

Table 1. Classification rates

PCA dimensionality	15	10	5	3
Total rate	85.7 %	82.9 %	71.4 %	64.3 %
Frontal rate	85.7 %	82.9 %	88.6 %	82.9 %
Profile rate	85.7 %	82.9 %	54.3 %	45.7 %

Obtained classification rates are presented in Table 1. It is interesting that the classification rate for the frontal profile is almost invariant under the PCA subspace dimensionality reduction. This is because some profile images are quite similar to the frontal face (low rotation angle).

The classification ratios presented in Table 1 are not extremely high, however, they are slightly better the those presented in [2]. It is also important, that this classification intents to improve the face recognition system. So if profile face is classified as a frontal one there is a chance that it is similar to some frontal faces in the database. The future work is to plug the proposed classifier into the face recognition system and check how this procedure affects is accuracy.

6 Conclusion

In this paper we raised the concept face recognition system for different face profiles. For such a system we proposed and evaluated the face profile classifier. We showed that the LDA combined with the PCA can be used to construct an accurate face profile predictor. Proposed classifier achieves 85% rate in face profile prediction.

Acknowledgment

This work has been performed in the framework of the EU Project INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment*) – grant agreement number: 218086 and co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme, INSIGMA project no. POIG.01.01.02-00-062/09. Development of algorithm and implementation have been fund by EU Project INDECT. Development of system, tests and result analysis have been fund by INSIGMA project.

References

1. http://zaginieni.policja.pl/
2. Chang, C.Y., Li, J.S., Kuo, J.Y.: Face view recognition and facial feature extraction. In: 2010 First International Conference on Pervasive Computing Signal Processing and Applications (PCSPA), pp. 452–455 (2010)
3. Fisher, R.A.: The statistical utilization of multiple measurements. Annals of Eugenics 8, 376–386 (1938)
4. Jolliffe, I.: Principal component analysis. Springer series in statistics. Springer, Heidelberg (2002)
5. Martinez, A., Kak, A.: Pca versus lda. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 228–233 (2001)
6. Timm, N.: Applied multivariate analysis. Springer texts in statistics. Springer, Heidelberg (2002)

Performance Evaluation of the Parallel Codebook Algorithm for Background Subtraction in Video Stream

Grzegorz Szwoch

Gdansk University of Technology, Multimedia Systems Department
Narutowicza 11/12, 80-233 Gdansk, Poland
greg@sound.eti.pg.gda.pl

Abstract. A background subtraction algorithm based on the codebook approach was implemented on a multi-core processor in a parallel form, using the OpenMP system. The aim of the experiments was to evaluate performance of the multithreaded algorithm in processing video streams recorded from monitoring cameras, depending on a number of computer cores used, method of task scheduling, image resolution and degree of image content variability. The results of the tests are presented and discussed. The main purpose of the research is application of the tested algorithm in a real-time video content analysis system, e.g. for automatic detection of important security threats.

Keywords: background subtraction, image content analysis, multi-core parallel processing.

1 Introduction

Background subtraction is usually the first operation performed in automatic video content analysis systems [1]. This procedure operates in the image pixel space, comparing each pixel value with the background model and deciding whether the pixel belongs to the moving object or to the background. The background model is usually constructed using statistical methods, as in the Gaussian Mixture Model algorithm [2], or as a codebook, consisting of codewords representing possible pixel values [3]. The background subtraction is performed by testing each image pixel versus the related background model element, comparing the pixel value with a number of Gaussian models or codewords, depending on the method. If any background model element matching the examined pixel is found, the pixel is assigned to the background, otherwise the pixel belongs to the moving object. After the analysis, the background model is usually updated. For each image pixel, a number of operations related to matching the pixel versus the background model, has to be performed. As a result, background subtraction is usually the most computationally expensive and time consuming part of the whole image processing chain. Therefore, obtaining real-time background subtraction in modern high-resolution cameras is a practical problem.

Modern computer systems are equipped with multi-core processors. Quad core CPUs in desktop computers are becoming a standard and processors having 16 and more cores are expected in the near future. In complex multi-camera systems it is

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 149–157, 2011.

possible to use a 'supercomputer' – a cluster of multi-core computing nodes. The task of background subtraction in multiple camera images and further processing stages (object tracking, classification, event detection, etc.) may be distributed amongst a number of nodes. Another possibility is to off-load the task of background subtraction to the graphic processing unit, using a platform such as OpenCL or CUDA [4]. All the mentioned solutions are based on concurrent, parallel processing of data (image pixels in the discussed case) using multi-core processors and multithreaded system architecture. An important feature of the two background subtraction algorithms mentioned earlier is that each pixel is processed independently, so this task is well-suited for parallel processing. On the other hand, the processing time of a single pixel is not constant, it depends on the variability of pixel value, as discussed further in the paper. Therefore, proper assignment of data to computing threads (task scheduling) should be selected carefully in order to provide proper work balance and avoid idle threads. OpenMP is an open standard for parallel computing which allows for parallelization of existing algorithms with minimal changes on code and provides several task scheduling methods [5].

The background subtraction algorithm using the codebook approach, proposed by KIm et al., was selected for evaluation, because it is more robust to dynamic changes in lighting and more efficient in memory usage than the Gaussian method [3]. The algorithm was implemented in C++ in a parallel form using OpenMP and tested on an eight-core computing unit. Efficiency of the algorithm was tested depending on the number of used processor cores, task scheduling method, camera image resolution (number of processed pixels) and variability of the input image. The aim of these experiments was to assess whether computational power of a single multi-core node of a supercomputer is sufficient for real-time processing of video streams, especially in case of high resolution cameras and in high frame rate video streams.

2 Background Subtraction Using the Codebook Method

In the codebook modeling method proposed by Kim et al. [3], each pixel in the background model is represented using a number of vectors called the codewords. A single codeword contains nine parameters:

$$\mathbf{c}_i = \left\langle \overline{R}_i, \overline{G}_i, \overline{B}_i, \check{I}_i, \hat{I}_i, f_i, \lambda_i, p_i, q_i \right\rangle , \tag{1}$$

where $\left(\overline{R}_i, \overline{G}_i, \overline{B}_i\right)$ are background pixel values in RGB color space, $\left(\check{I}_i, \hat{I}_i\right)$ define a range of brightness values covered by the codeword, f_i is a number of times the codeword was matched, λ_i is a longest period which the codeword was not matched, p_i is the time the codeword was added to the model, q_i is the time of the last codeword update.

Detection of moving objects with the codebook algorithm is a two-stage process. During the training phase, a background model is built by adding new codewords as the pixel value changes. After processing a defined number of frames the training ends, inactive codewords are deleted from the model and thus the fixed background model is constructed. In the detection phase, image pixels are compared with the background model. New codewords are added to the short-term model (the cache) and inactive codewords are periodically removed from the cache [6].

In order to test whether an image pixel belongs to the background or not, its (R, G, B) value is matched versus all the related codewords in the background model. The match is found if, and only if, two conditions are met. First, color difference must be within range defined by an imposed constant ε:

$$\sqrt{\left(R^2+G^2+B^2\right)-\frac{\left(R\overline{R}+G\overline{G}+B\overline{B}\right)^2}{\overline{R}^2+\overline{G}^2+\overline{B}^2}} \le \varepsilon \,. \tag{2}$$

The second condition imposes that pixel brightness, defined as

$$I=\sqrt{R^2+G^2+B^2}\,, \tag{3}$$

has to be within limits defined by brightness range of codeword:

$$\alpha\hat{I} \le I \le \min\left\{\beta\hat{I}, \frac{\breve{I}}{\alpha}\right\}, \tag{4}$$

where α and β are constants that allow for removal of shadows and highlights.

If a matching codeword is found according to Eqs. 2–4, codeword (R, G, B) values are updated with pixel values, using a running average. The brightness range of the codeword is extended if needed, so that the pixel brightness is within this range. The remaining parameters are updated using the current image frame number. If no matching codeword was found, a new codeword is added to the model, initialized with current pixel RGB values and brightness, and the current frame number [3].

Fig. 1 shows an example of background subtraction result. The output from the procedure is usually encoded in a binary or grayscale image in which zero values mark background pixels (a match was found in the codebook), non-zero values mark moving object pixels (no match was found). It can be seen that the result is noisy, so usually it has to be post-processed using morphological operations, e.g. opening for pixel noise removal, followed by closing for filling small gaps [1].

The number of operations (multiplication/division, addition/subtraction and calculation of square root) needed for testing the color condition, brightness condition and for updating the codeword, is presented in Table 1. Total processing time for a single pixel is:

$$T = N_B T_B + N_C (T_B + T_C) + N_U (T_B + T_C + T_U), \tag{5}$$

where N_B is the number of codewords not fulfilling brightness condition, N_C is the number of codewords fulfilling the brightness condition but not the color condition, N_U is 1 if a matching codeword was found and 0 otherwise. Processing times T_B, T_C, T_U are given in Table 1.

The procedure for matching the image pixel to the background model is depicted in Fig. 2. It can be seen that number of performed processing steps (and, as a result, time of a single pixel processing), depends on how the pixel value changes. If pixel color and brightness are nearly constant, the first codeword is matched and the processing ends ($T = T_B + T_C$). If the pixel value changes constantly, none of the large number of codewords may match the pixel and the processing time is significantly longer, especially if codewords are rejected after checking both conditions. Therefore,

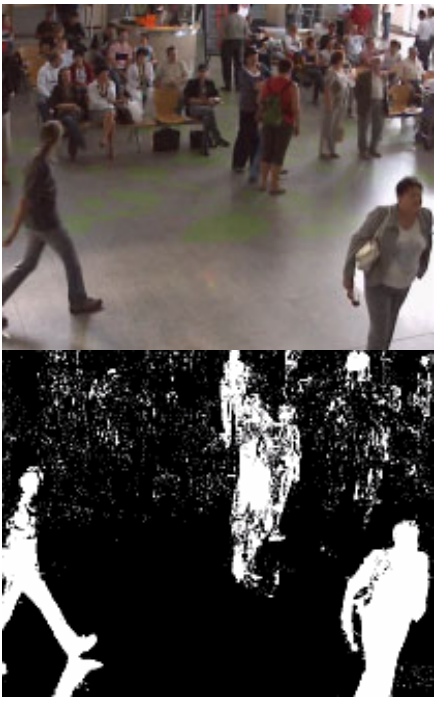


Fig. 1. Example of background subtraction using codebook method: input camera image (top), result of processing (bottom, white pixels mark moving objects)

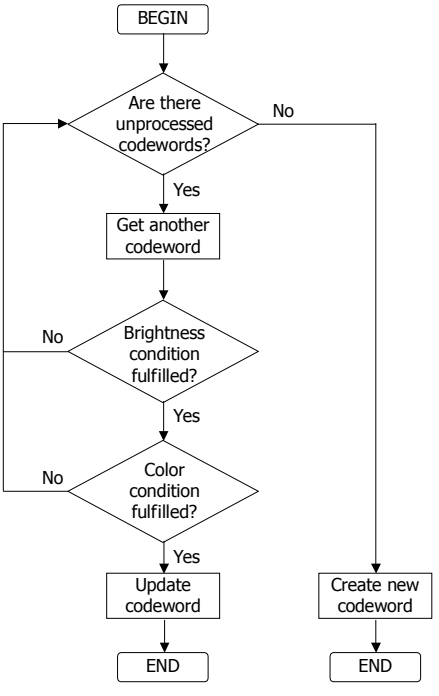


Fig. 2. Block diagram of procedure for searching the codeword that matches the pixel

processing time in image regions that exhibit large and frequent changes (e.g. walking persons) is expected to be significantly longer than in regions with stable background. Therefore, proper assignment of image regions to processing threads in order to obtain proper work balance is an important problem. This will be addressed further in the paper.

Table 1. Number of multiplication/division, addition/subtraction and square root operations in procedure searching for codeword matching the pixel

Operation	Time	×	+	√
Testing brightness condition	T_B	6	2	1
Testing color condition	T_C	9	5	1
Updating codeword	T_U	6	9	0

3 Experiments

The algorithm was implemented in C++ using OpenMP system (GOMP library with GCC compiler version 4.3.2). The algorithm was run on a single supercomputer node,

consisting of two motherboards, each equipped with Intel Xeon Quad Core 2.33 GHz processor and 16 GB of memory, controlled by Linux operating system. The task of decoding input video stream was handled by an external framework. The examined algorithm received a decoded video frame, performed background subtraction and returned the result to the framework. Three tests were performed using three video streams with different content and resolution (Table 2).

Table 2. Video files used in the performance evaluation tests

Test	Resolution	Fps	Number of frames	Description
Test 1	640×480	10	2440	Street, moderate traffic in middle part of the image
Test 2	720×576	10	1452	Airport hall, many people constantly moving
Test 3	1600×900	10	3000	Same as Test 2, but higher resolution

Three tasks scheduling methods available in OpenMP system [8] were tested. *Static* scheduling is the simplest one: the image is divided into several parts (by rows of pixels) and each thread receives its own image part to process. The assignment is made before the algorithm starts processing the image. This method is typically used by the programmer who performs task scheduling 'by hand', managing threads by himself, without OpenMP help. As it was already mentioned, processing time of a single pixel depends on variability of the pixel value. Therefore, the thread that processes stationary part of the image will finish its work before the thread analyzing another part of the image. containing moving objects, leading to non-optimal utilization of available resources. *Dynamic* scheduling assigns groups of items (pixel rows) to threads while the algorithm is running. When the thread finishes its work, it receives a new group of items. The drawback of this method is increased computing overhead, related to task scheduling during the program run [8]. *Guided* scheduling is supposed to balance advantages and drawbacks of static and dynamic scheduling.

Only the time of performing background subtraction in detection phase, for each video frame (in the main processing loop of the algorithm) was measured in order to exclude irrelevant factors from the analysis. The measured total processing time was converted to a number of frames processed per second (fps). The tested scheduling methods were: static, guided and dynamic with different number of image rows assigned at a time (1, 2, 4, 8, 16). For each case, the algorithm was run multiple times, using 1, 2, 4, 6 and 8 processor cores (each computing thread runs on its own core). Each test run was repeated 12 times, which yielded a total of 420 algorithm runs per file. Of the 12 measurements made for a single case, the smallest and the largest fps value was discarded and the remaining 10 results were averaged. The results of all the tests are presented in Table 3.

Table 3. Results of tests evaluating performance of the codebook algorithm for different task scheduling methods and number of used processor cores. Thr means a number of threads. Values in the table are mean fps ± standard deviation.

Test 1

Thr	static	guided	dynamic1	dynamic2	dynamic4	dynamic8	dynamic16
1	15.65±0.08	15.73±0.10	15.73±0.09	15.64±0.07	15.70±0.07	15.76±0.14	15.64±0.14
2	21.46±0.19	21.31±0.26	27.63±0.34	27.77±0.45	27.96±0.40	27.62±0.52	28.15±0.18
4	26.35±0.21	25.97±0.22	41.84±1.34	41.52±1.49	41.29±0.79	43.82±1.47	31.83±0.30
6	28.11±0.29	27.84±0.19	49.99±3.16	49.04±2.19	50.96±3.06	47.18±5.36	31.82±0.31
8	29.43±0.30	28.83±0.24	53.22±5.32	49.99±0.40	52.07±4.81	46.74±1.89	31.66±0.34

Test 2

Thr	static	guided	dynamic1	dynamic2	dynamic4	dynamic8	dynamic16
1	11.19±0.08	11.13±0.09	11.10±0.10	11.04±0.11	11.18±0.09	11.09±0.18	11.04±0.11
2	15.63±0.22	15.66±0.13	18.62±0.25	18.77±0.21	18.78±0.25	18.77±0.26	18.90±0.10
4	23.97±0.27	23.71±0.10	29.83±0.16	29.92±0.11	30.05±0.26	30.00±0.29	29.87±0.22
6	29.93±0.54	29.98±0.24	36.44±0.46	36.25±0.62	36.95±0.48	36.24±0.68	36.54±0.40
8	33.74±0.72	33.57±0.41	40.26±0.70	41.14±0.68	40.79±0.84	40.51±1.12	39.58±0.69

Test 3

Thr	static	guided	dynamic1	dynamic2	dynamic4	dynamic8	dynamic16
1	2.73±1.29	3.13±2.16	2.78±1.35	3.13±2.15	3.17±2.23	3.03±1.98	4.79±5.64
2	2.52±0.24	2.57±0.24	3.20±0.57	3.17±0.55	3.21±0.57	3.19±0.56	3.14±0.53
4	3.64±0.49	3.65±0.51	5.13±0.86	5.15±0.84	5.12±0.83	5.13±0.83	5.13±0.86
6	4.64±0.33	4.56±0.34	6.57±0.51	6.53±0.52	6.45±0.45	6.48±0.49	5.54±0.59
8	5.26±0.22	5.12±0.20	7.30±0.32	7.20±0.30	7.29±0.27	7.31±0.26	6.01±0.45

In Test 1, video containing a view of city street was examined. The middle rows of pixels presented the street with moving vehicles, while the top and bottom image parts were mainly stationary. Therefore, it was expected that static task scheduling will not be optimal. Test results confirmed this assumption: dynamic task scheduling was much more efficient (Fig. 3). Due to relatively small number of pixel rows, the *dynamic16* case was close to the static one. There was no significant difference between the other dynamic cases. The guided scheduling method did not work as expected, it performed similarly to the static method. Using two or four computing threads instead of one gave a significant increase in algorithm performance, while using a larger number of threads gave smaller advantage. In the case of video stream similar to the examined one – with relatively low resolution and variability limited to a part of the image, using four cores and *dynamic1* scheduling method seems to be optimal.

In Test 2, a video recorded inside the arrival hall of the Poznan-Lawica airport was examined. In this case, there was constant movement of persons, approximately uniform within the camera view. Therefore, a better performance was expected from the static method in this test. Although this was indeed the case, the dynamic method

performed better for each number of threads (Fig. 3). Moreover, the gain in performance with increase of threads number is almost linear. Because of larger image variability compared to Test 1, a much larger number of calculations is necessary, so increasing the number of threads to 8 still gives the performance gain. Again, no significant difference between different dynamic methods was observed and the guided method has performance comparable to the static one. In order to ensure that the video stream is processed in real time, using 8 computing cores with *dynamic1* scheduling method is recommended in case of video streams similar to the one used in Test 2.

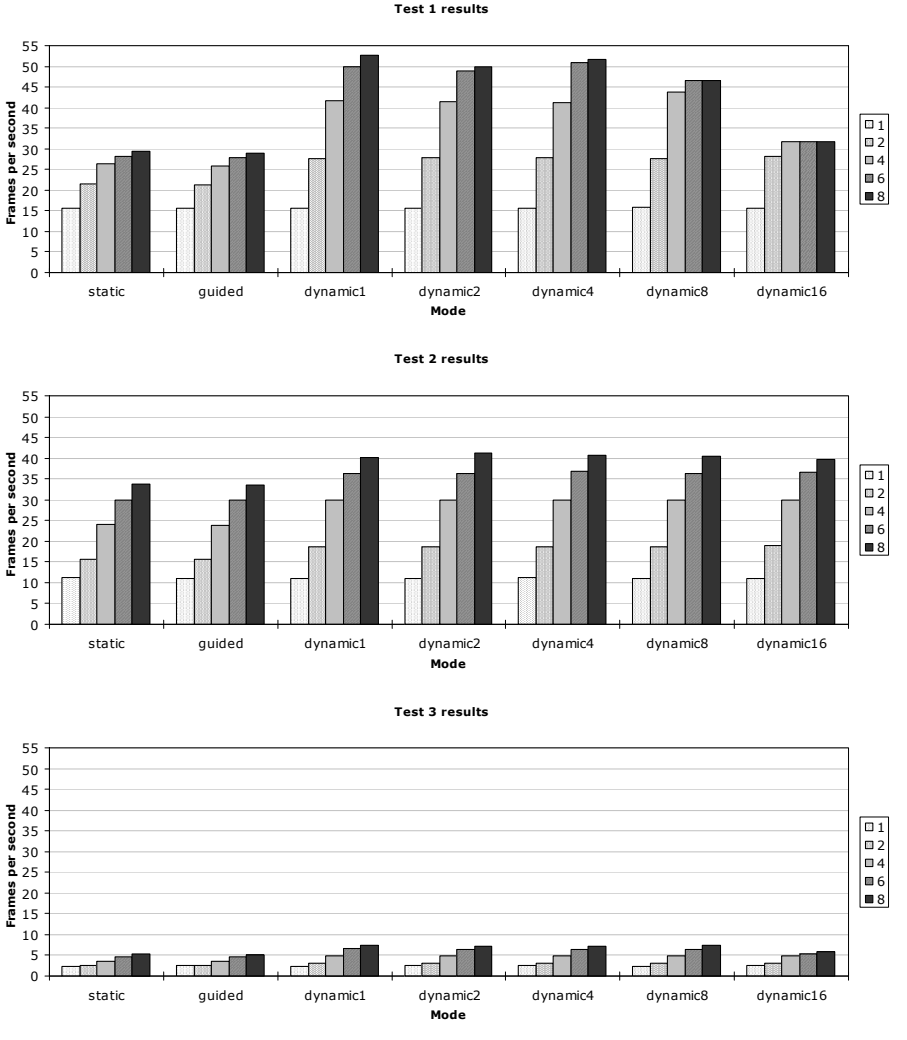


Fig. 3. Results of performance tests of the codebook algorithm for different task scheduling methods and varying number of computing threads (1, 2, 4, 6, 8)

Test 3 was performed in order to examine whether the hardware used in the experiments is able to process a high resolution video stream in real time. The test video presented similar scene to Test 2, but captured with a megapixel camera. It was found that the best-case performance (8 cores used and dynamic scheduling) is well below the input frame rate. Since the measurements do not include the payload introduced by the framework, practical frame rate is even lower. Therefore, it is not possible to perform background subtraction using the tested hardware in real time without downscaling the video frames. The results of Test 3 are consistent with Test 2 – increasing the number of threads results in larger fps number and the dynamic scheduling performs better than the static or the guided method.

4 Conclusions

The results of performance evaluation tests indicate that the background subtraction algorithm based on the codebook method is well suited for parallel, multithreaded implementation on a multi-core system. Increasing the number of concurrent computing threads that run on dedicated processor cores, results in improved performance, measured in terms of analyzed image frames per second. The main advantage of using the OpenMP system instead of implementing own task managing system, is the dynamic task scheduling mechanism which proved to be more efficient than the straightforward static scheduling. This is especially evident in processing of a video stream in which some rows of image pixels contain moving objects and other represent a static background. In this case, static scheduling is suboptimal and it results in idle threads. Dynamic scheduling provides better balance between tasks assigned to threads. Even if image variability is more uniform in the camera frame, dynamic scheduling performs better than the static one. The choice of the item size in dynamic scheduling is not important, provided that it does not exceed 8 image rows. Single rows of pixels may be assigned to threads in dynamic scheduling. It was observed that the guided task scheduling method does not yield any advantages over the static method. These conclusions are valid for the current implementation of the OpenMP system in the GCC 4.3.2 compiler, used on the Linux platform.

Processing time of a single image frame depends on the image resolution (number of pixels to process) and on variability of the image content. For real-time analysis of video streams having low to moderate resolution (up to 720×576 pixels), with dynamic task scheduling, four computing threads are sufficient for processing the stream in which only part of the image changes, and 6 to 8 threads are recommended if image content variability is high. Computing power of a eight-core system used in tests is sufficient for video processing in these cases. However, even utilization of all available cores is not enough for real-time processing of high resolution video streams.

In the experiments described in this paper, only a single supercomputer node was used. Resources provided by a single node are insufficient for real time processing of high resolution video streams. However, it is also possible to utilize multiple nodes, communicating with each other using MPI system. This approach requires solving practical problems, such as memory sharing, task synchronization, etc. One example solution may be dividing the image frame into vertical strips, assigning each strip to a

different node and performing parallel analysis within each node using method described in this paper. This approach, based on the master–slave task architecture, should provide a proper work balance between nodes (static scheduling) and node cores (dynamic scheduling). The proposed algorithm modification will be tested in the next stage of the experiments.

Acknowledgements

Research is subsidized by the European Commission within FP7 project INDECT (Grant Agreement No. 218086).

References

1. Czyżewski, A., Szwoch, G., Dalka, P., et al.: Multi-stage Video Analysis Framework. In: Lin, W. (ed.) Video Surveillance, pp. 147–172. Intech (2010)
2. Stauffer, C., Grimson, W.E.: Adaptive background mixture models for real-time tracking. In: Proc. IEEE Conf. Computer Vision and Pattern Recognition, Fort Collins, USA, pp. 246–252 (1999)
3. Kim, K., Chalidabhongse, T.H., Harwood, D., Davis, L.: Real-time foreground-background segmentation using codebook model. Real-time Imaging 11, 167–256 (2005)
4. Pham, V., Vo, P., Hung, V.T., Bac, L.H.: GPU Implementation of Extended Gaussian Mixture Model for Background Subtraction. In: IEEE RIVF International Conference on Computing and Communication Technologies, Research, Innovation, and Vision for the Future (RIVF), pp.1–4 (2010), doi: 10.1109/RIVF.2010.5634007 (2010)
5. OpenMP Specification, `http://openmp.org/wp/openmp-specifications/`
6. Kim, K., Harwood, D., Davis, L.: Background updating for visual surveillance. In: Bebis, G., Boyle, R., Koracin, D., Parvin, B. (eds.) ISVC 2005. LNCS, vol. 3804, pp. 337–346. Springer, Heidelberg (2005)
7. TASK, Galera supercomputer, `http://www.task.gda.pl/kdm/sprzet/Galera`
8. Chapman, B., Jost, G., van der Paas, R.: Using OpenMP. Portable Shared Memory Parallel Programming. MIT Press, Cambridge (2007)

Automated Optimization of Object Detection Classifier Using Genetic Algorithm

Andrzej Matiolański and Piotr Guzik

Department of Telecommunications,
AGH University of Science and Technology,
Cracow, Poland
{matiolanski,guzik}@kt.agh.edu.pl

Abstract. The problem of optimizing classifiers for object detection has already been discussed in several publications. In order to achieve better results, it was decided to use genetic algorithms to optimize the classifiers. By applying this approach optimization is automatic in respect to image (or group of images). For test issues the haar-like object detection features were used. Genetic model has been created over the field of solutions and evolved to provide better results. Proposed algorithm was tested and results are presented. The proposed solution can be applied to another type of classifier and adapted to optimize any detection parameter.

Keywords: Genetic algorithms, object detection, optimization, classifiers.

1 Introduction

There are plenty of methods used in optimization problems. One of most popular is genetic algorithm. It can be used in problems that can be described by a set of numbers – parameters to optimize.

In this paper a simple genetic algorithm is used for finding optimal parameters of face detection classifiers as an example of object detection classifiers. Optimization seems to be essential there as there are some parameters describing used classifiers that noticeably affect detection rates. There are also usually a few different classifiers that are used in detection of one particular object. What is important there is the fact that in each situation there are different classifiers and different parameters best for that particular problem. The aim here is to prepare a tool with as high positive detection ratio as possible and at the same time as low false-positive detection ratio as possible. The fact that at each particular situation the solution is different indicates the need of optimization.

2 State-of-the-Art

One of the popular methods for objects detection is use of haar-like features. Algorithm presented at [1] uses cascade (trained on positive and negative content) as a classifier. This classifier decides if searched object is detected. Detection rate is

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 158–164, 2011.

from approx. 78% (for false-positive ratio near zero) up to 93.7% (when number of false-positive detections is significant) [2]. To increase positive and decrease false-positive object detection ratio two classifiers or more can be merged and work together. Concept of boosting algorithm by combining many classifiers is called Adaboost [3]. Boosted algorithm for object detection using three classifiers (cascades) at the same time gives 96.4% positive detection ratio [4].

3 Novelty

In this paper a modification to standard object detection method is introduced. There are only two classifiers used at the same time. That allows to save some computational time but on the other hand there are still a set of classifiers that may be used as those two classifiers which are best are chosen from whole set. At the end of computation there are two optimal classifiers described by optimal parameters. To reduce false-positive detection rate, each detection by one classifier is confirmed by the other one, hence it gives low false-positive detection ratio. Additionally the system is automatic and it doesn't need human supervision. Optimization technique ensures high positive detection ratio.

4 Realization

Genetic algorithm is a computational model that is inspired by nature and is used to solve optimization problems. It belongs to a group of algorithms called evolutionary algorithms as it simulates process of natural evolution. Genetic algorithms were developed and introduced to science in 1970s by Holland's book [5].

Genetic algorithm firstly needs a genetic representation of solution domain. Single solution is represented as an array of bits of numbers. Such an array is often called an individual. Each individual is described by its fit function that defines the quality of this solution.

Searching for best solution using genetic algorithm is an iterative process that is expected to improve the solution in each iteration. At the beginning population of random individuals is generated. Then the iterative process starts. There are three steps at each iteration. At first, all of individuals are sorted in respect of their fitness function. Next part of them is selected to breed a new generation. Usually the probability of selection depends on the value of fitness function of an individual. Sometimes only best individuals are selected. After selection process genetic operators like crossover, mutation and others are performed on selected individuals to produce new generation of solutions. Those three steps are conducted repeatedly as long as needed. At the end, population of individuals is expected to be composed of only such individuals that represent solutions close to optimal.

In this work single solution consists of 3 parameters of classifiers represented by an array of 11 bits and 2 numbers describing which classifiers (out of 5) to use. The model of an individual is shown in Fig. 1.

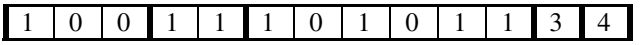


Fig. 1. Representation of simple solution

Each parameter has its own number of bits. Realization of crossover is shown in Fig. 2. Every two individuals called parents gives 2 other individuals called offspring. This provides constant population size. After crossover, mutation is performed. In this step there is a small (1% – 5%) probability for every bit (out of 11) to became its own reverse and at the same time there is the same probability that chosen classifier changes into another one.

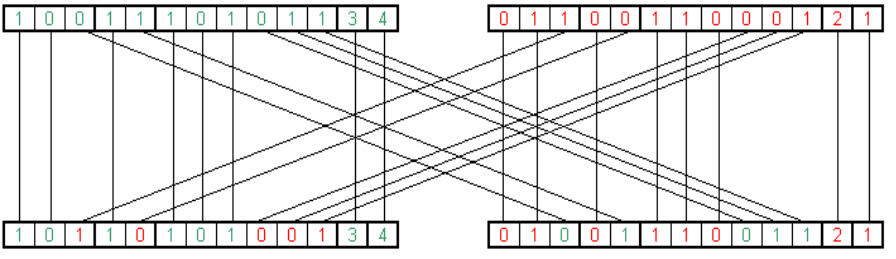


Fig. 2. Crossover model

Genetic algorithm optimizes population with respect to fitness function. In classic object detection optimization we are interested in:

- increasing positive detection rate
- decreasing false-positive detection rate
- decreasing detection time

Simple fitness function is presented below:

$$F = \frac{AD_P - BD_{FP}}{1 + CT}. \tag{1}$$

where: D_P - number of positive detections, D_{FP} – number of false-positive detections, T – detection time, A, B, C – weights of parameters (A, B, C ≥ 0).

The algorithm allows the possibility of modifying the fitness function, both: the parameters and weights. The fitness function that is used should allow to optimize the parameters for a specific application (by modifying values of weights).

For optimization issue we are interested in deciding whether the detected object is positive one or not in an automatic way. We use merge results of two different classifiers (from one individual). When object is detected by both classifiers we are sure that this detection is a positive one. The detection by two classifiers is regarded as one object when the center of smaller object lies inside the bigger one, as we can see at Fig. 3. We would like to achieve positive match ratio as high as possible. This assures us that the object is detected correctly.

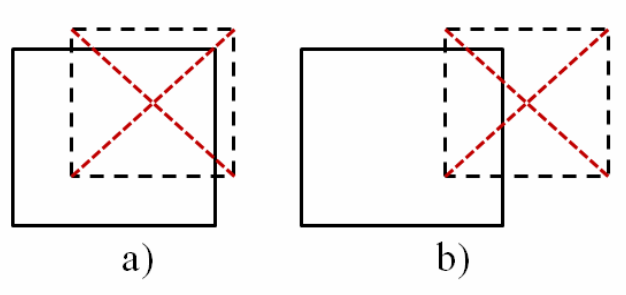


Fig. 3. a) – positive math, b) – non-positive math

The optimization process takes a certain time (limited by the number of generations). It stops after obtaining a suitable factor of positively detected objects or when the variability vanishes. The speed of the optimization process depends on the number of detected objects and the classifier speed.

Proposed algorithm is implemented using OpenCV (version 2.2) [6]. Library has its own haar-like object detection [1] implementation and is distributed with basic cascades. We decide to use five different face detection cascades as five different classifiers. Detection parameters that we are dealing with are:

- haar scale: interval [1.05 – 1.2],
- minimum number of neighbors: interval [1 – 8],
- minimum object size: interval [15 – 47].

We create population of eight individuals which ensures sufficient dispersion of parameters values to cover the solution space. Individuals are representing parameters and chosen classifiers. Number of generations is set at 20 which is sufficient to observe improvement in object detection.

5 Results

Testing scenario assumes that there are better and worse classifiers for specific content (or content type). Therefore, the optimization is based on pre-defined content. In our work we choose a collection of images containing multiple faces. Tests were conducted on real images which increases their credibility. In addition, tests carried out on images from the perspective that is more difficult for face detection gives a larger field for optimization and achieve visibly better results (Fig. 4).

Detection based on two classifiers allows merging results of each of them. Results can be treated in different ways depending on the emphasis on certainty of detection of as many objects as possible (combining the results with the conjunction OR) or confirmation that the detection is positive (combining the results with the conjunction AND).

We conducted tests on images containing 355 human faces. In each image there are at least 10 faces. Detection test results are presented in the Table 1. It shows the detection rates of both classifiers alone and the rate of detections common for both

Fig. 4. An exemplary image for algorithm testing [7]

Table 1. Face detection test results

Number of faces	Number of detected faces		
	OR (better classifier)	OR (worse classifier)	AND
355	352	346	343
100%	99.15 %	97.46 %	96.62 %

classifiers. It is worth noting that on 46% of the images all faces were detected (Fig. 5). Tests show the appearance of the false-positive ratio to be no more than 4.8%.

The obtained results are promising. Unfortunately, the apparent amount of false-positive detections is not eliminated. During the tests, it was proven that it is important to appropriately cover the solution space at the beginning. This allows for rapid (in approx. 18 generations) increase of the fitness function value. Tests show that for assumed parameters and classifiers the optimal population size is 6-10 and the number of individuals in the population is in the range of 18-25.

Fig. 5. An exemplary face detection results with 100% detection ratio. There is only one false detection – the big rectangle in the left part of an image (image from [8]).

6 Conclusion and Further Research

Demonstrated solution to the optimization of classifiers to detect objects helped to reduce the number of classifiers used to a minimum. The results that have been obtained are better than the results obtained for multiple classifiers using Adaboost algorithm. In addition, a simple way to automatically compare the results of the classifiers to facilitate machine processing of results is proposed. Proposed algorithm can be used for optimization of classifiers different than haar-like.

Further work will concern the extension of optimization using genetic algorithm. In particular, the creation of a separate set of parameters for each of the classifiers and the use of other methods of reproduction of the population (crossover, mutation) may give some new results. Emphasis will be put on a reduction of false-positive detection ratio.

Acknowledgment

The work was realized as a part of MAYDAY EURO 2012 project, Operational Program Innovative Economy 2007-2013, Priority 2 „Infrastructure area B+R". It was also partially supported by European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 218086.

References

1. Viola, P., Jones, M.: Rapid object detection using boosted cascade of simple features. Computer Vision and Pattern Recognition, 511–518 (2001)
2. Viola, P., Jones, M.: Robust Real-time Object Detection. In: Second International Workshop on Statistical and Computational Theories of Vision – Modeling, Learning, Computing and Sampling, Vancouver, Canada, July 13 (2001)

3. Freund, Y., Schapire, R.E.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi, P.M.B. (ed.) EuroCOLT 1995. LNCS, vol. 904, pp. 23–37. Springer, Heidelberg (1995)
4. Lienhart, R., Liang, L., Kuranov, A.: A detector tree of boosted classifiers for real-time object detection and tracking. In: International Conference Multimedia and Expo (ICME 2003) (2003)
5. Holland, J.H.: Adaptation in Nature and Artificial Systems. The University of Michigan Press (1975)
6. Open source computer vision library (2011.03.01), http://opencv.willowgarage.com/
7. Byron Katie Blog 2011.03.01, http://www.byronkatie.com/
8. Sackler Institute 2003 (2011.03.01), http://www.sacklerinstitute.org/cornell/summer_institute/2003/

Traffic Danger Ontology for Citizen Safety Web System*

Jarosław Waliszko, Weronika T. Adrian, and Antoni Ligęza

Institute of Automatics,
AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
jaroslaw.waliszko@gmail.com, {wta,ligeza}@agh.edu.pl

Abstract. In this paper a novel approach to development of a Web-based threat information system for citizens is presented. The method assumes a significant role of a domain ontology and Description Logics (DL) reasoning. In order to test this approach a prototype ontology-driven application has been developed. The main objective of the application is to provide real time information for users about threats represented with 2D GIS system. The system integrates a database and an ontology for storing and inferring desired information. While the abstract of traffic danger domain is described by the ontology, the location details of traffic conditions are stored in the database. During run-time the ontology is populated with instances stored in a database and used by a DL reasoner to infer required facts. The paper gives an overview of the system design and implementation.

Keywords: security, citizens, threat ontology, INDECT.

1 Introduction

Information systems have been widely used to facilitate communication and distribution of information in a rapid and efficient way. Whether through official news portals or social systems like Facebook or Twitter, people inform each other about the events, dangers or changes important to them. One of the area which still needs careful attention from the information systems is the local safety of citizens in the urban environment.

Within the INDECT Project important problems related to security and information intelligent systems are investigated. Task 4.6 of the project focuses on development of a *Web System for citizen provided information, automatic knowledge extraction, knowledge management and GIS integration* [3]. Within this task several initial system prototypes have been developed, one of which is described in this paper.

* The research presented in this paper is carried out within the EU FP7 INDECT Project: "Intelligent information system supporting observation, searching and detection for security of citizens in urban environment" (http://indect-project.eu).

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 165–173, 2011.

As it has been noted in [4], "In recent years the development of ontologies has been moving from the realm of Artificial-Intelligence laboratories to the desktops of domain experts.". Over the last decade there has been a significant increase in the number of organizations and disciplines that developed standardized ontologies that domain experts can exchange and reuse. An ontology, "explicit formal specifications of the terms in the domain and relations among them" [2], specifies a set of common vocabulary researchers can share within a domain of knowledge.

Following the trends of the increased significance of ontology-based solutions, a prototype threat information system for citizen has been developed [5]. The system main assumption is an integration of database and ontology approach for storing and inferring desired information about some domain in real time. In this paper such a composite approach is presented in which location details of traffic conditions are stored in a database, while the abstract of traffic danger concept is described by an ontology. Such an integration results in dynamic deduction possibility of desired knowledge, using the ontology based approach, instead of using static relations defined in the database only.

The paper is organized as follows: In Section 2 the proposed solution is described, with its functionality (Sect. 2.1), threat ontology (Sect. 2.2), database-ontology integration (Sect. 2.3) and the reasoning process (Sect. 2.4) outlined. Implementation and deployment information is given in Section 3. The paper is summarized in Section 4 and future work is outlined in Section 5.

2 Proposed Solution

2.1 Functionality

The main objective of the proposed system is to provide citizens with real-time data about dangers occurring in a chosen area. Part of this information is entered into the system by trusted users and another part of it is inferred based on the threat ontology and current conditions (facts). General idea with main functionalities outlined can be observed in Figure 1.

The application cooperates with traffic danger ontology and synchronizes that ontology with information stored in database. Upon synchronization a Description Logic [1] reasoner infers dangers which occur on selected locations. Locations of traffic conditions occurrences are defined by postal codes. In database, postal codes are connected with streets, which in turn are connected with districts. As a result, ontology gives the possibility of answering, which kind of traffic danger can occur on the desired location. The deduction is based on traffic conditions connected to specific postal codes.

System users can address the system with dynamic questions and get results of inferred traffic dangers. This functionality is provided by a simple web-based interface. Additionally, the system allows trusted users to make changes to locations of traffic conditions occurrence. For working with most recent data, provided by trusted users, the synchronization mechanism integrates core ontology, describing the abstract of traffic dangers, with specific real time data. The process is executed for the first time on application startup – the first request to the

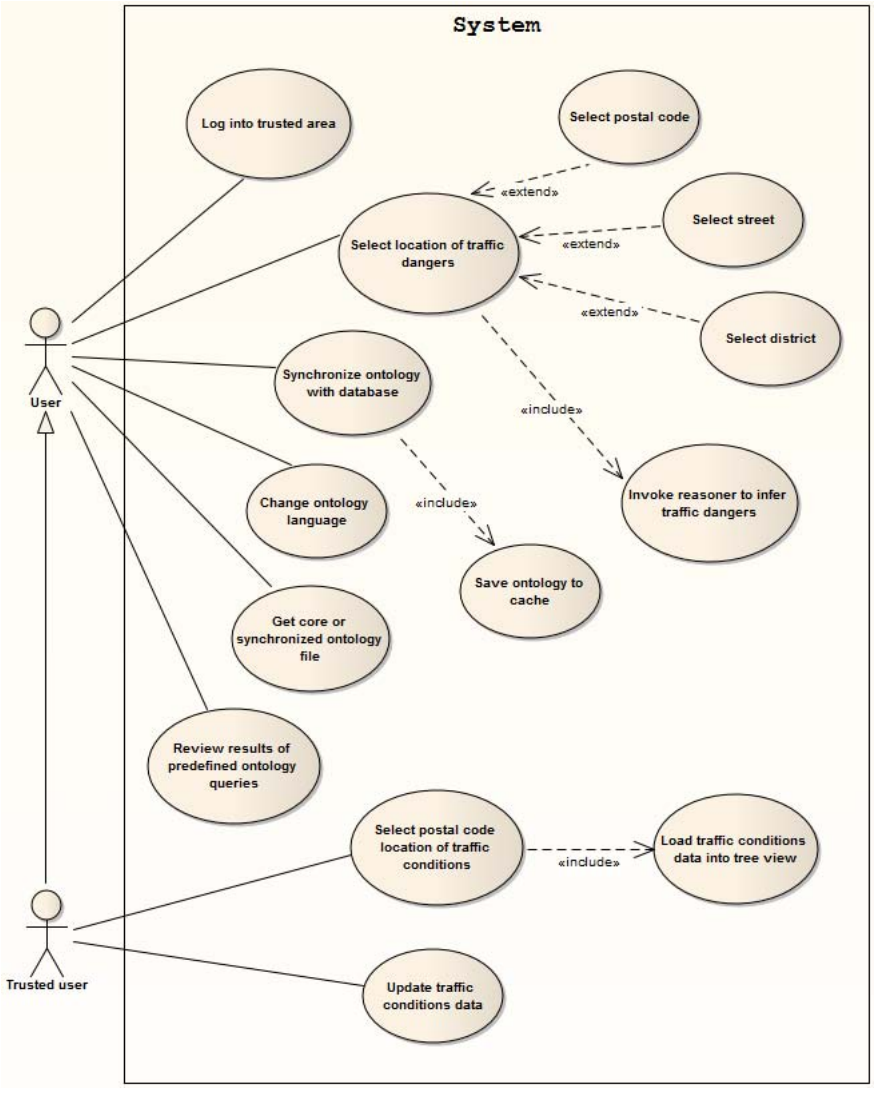


Fig. 1. Use cases of the system

server while accessing main page. This functionality is also available on demand. After synchronization, the ontology cached in memory is used for inferencing.

2.2 Traffic Danger Ontology

Sharing common understanding of knowledge domain is one of the most critical reason for developing ontologies. Explicit domain assumptions provide clear description of domain knowledge and simplify knowledge extensibility for domain.

While developing the traffic danger ontology, the following fundamental rules described in [4] have been taken into consideration:

- There is no one correct way to model a domain – there are always viable alternatives. The best solution almost always depends on the application in mind and the anticipated extensions.
- Ontology development is necessarily an iterative process.
- Concepts in the ontology should be close to objects (physical or logical) and the relationships in the domain of interest. These are most likely nouns (objects) or verbs (relationships) in sentences that describe the domain.

The ontology has been developed in a top-down process with the Protégé[1] editor integrated with HermiT DL Reasoner[2]. An excerpt of the ontology is shown in Figure 2. Sample questions the system is able to answer are:

Fig. 2. Asserted class hierarchy

[1] See http://protege.stanford.edu/
[2] See http://owlapi.sourceforge.net/

- What are the traffic dangers that can be met within specific area ?
- Are there any dangers connected with specific condition (e.g. low friction) in specific area ?
- What are the subareas of specific location ?
- What kind of dangers are connected with specific atmospheric conditions ?
- Is there any danger connected with specific postal code on specific district ?
- Are there any traffic conditions provided for specific location ?

In order to answer these questions either a semantic reasoner is used (which performs classification of concepts within the ontology) or appropriate DL queries are constructed.

2.3 Integration of the Ontology and the Database

One of the most important aspects of the system is the possibility of integration data from database and ontology. For the purpose of this paper the process is called synchronization. It allows to populate the traffic danger ontology with additional information from database.

This additional knowledge consists of locations structure and traffic conditions occurrence in that locations. The approach provides loose coupling between core ontology, describing the abstract of traffic dangers (see Figure 2), and synchronized ontology, which is filled with specific data connected with real time conditions on specific area.

Core ontology describes clear concept of traffic danger, while synchronized one is related to specified environment. Consequently, synchronized ontology can differ between the various environments where it is deployed. For example, traffic conditions information for Cracow are significantly different than those for Warsaw. There is a possibility of a single installation and synchronizing of the ontology at once with all global data, but it can result in system overloading and decreasing performance while inferring dependencies.

2.4 Reasoning in the System

Reasoning in the system is provided by invoking HermiT DL Reasoner through The OWL API Library[3], on synchronized ontology. The sequence diagram (see Figure 3) shows what steps are required, for the reasoning to occur.

Firstly, trusted users have to provide traffic conditions for some areas. This process is transparent for other, so-called "public" users. When such a user wants to check what threat he or she can expect in specific area, he or she has to download the up-to-date facts into locally stored ontology (synchronize the ontology), and then query the ontology by selecting desired locations from the web-based interface. After location selection, appropriate query is created using the OWL API Library and HermiT DL Reasoner is invoked to process such query on cached ontology. Inferred set of information is then provided to the user.

[3] See http://owlapi.sourceforge.net/

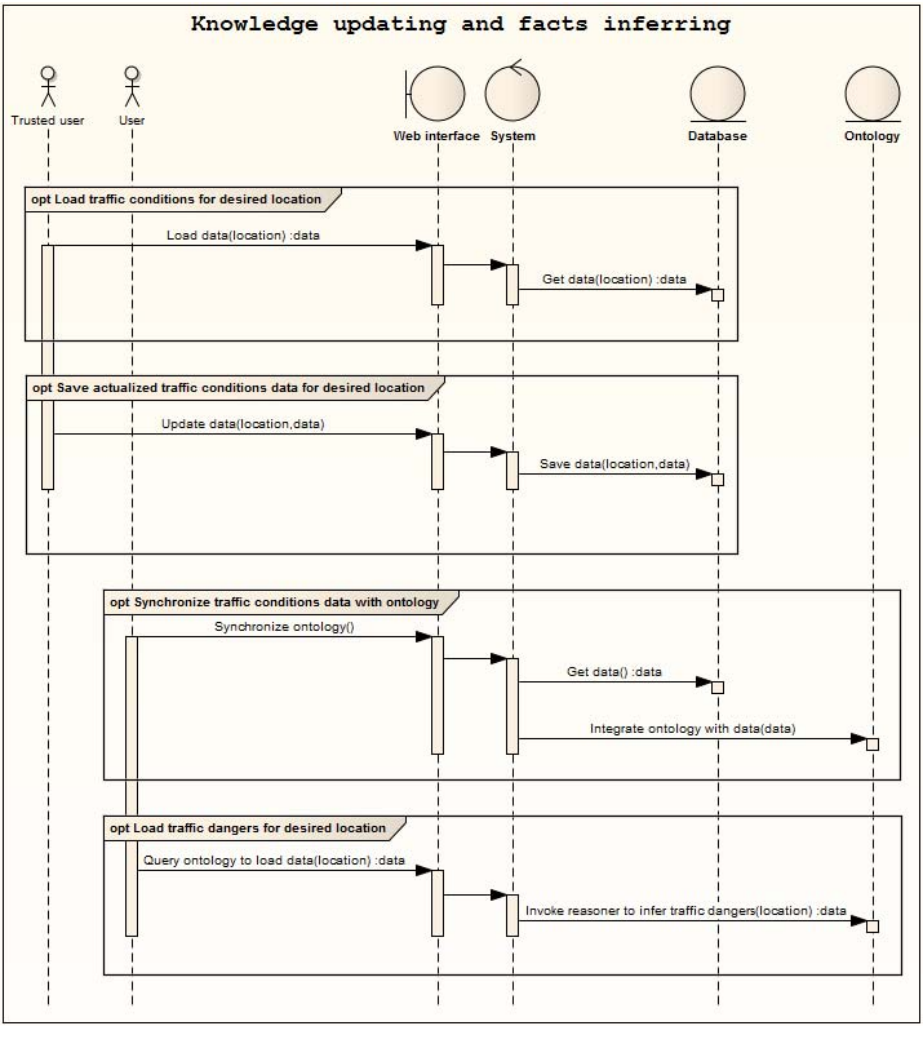


Fig. 3. Sequence diagram for updating and inferring data

3 Implementation and Deployment

The system has been implemented in a layered architecture that consists of: a web dashboard layer cooperating with users through browser clients, platform layer being the core of system, and storage layer where data is stored in both database and ontology (see Figure 4).

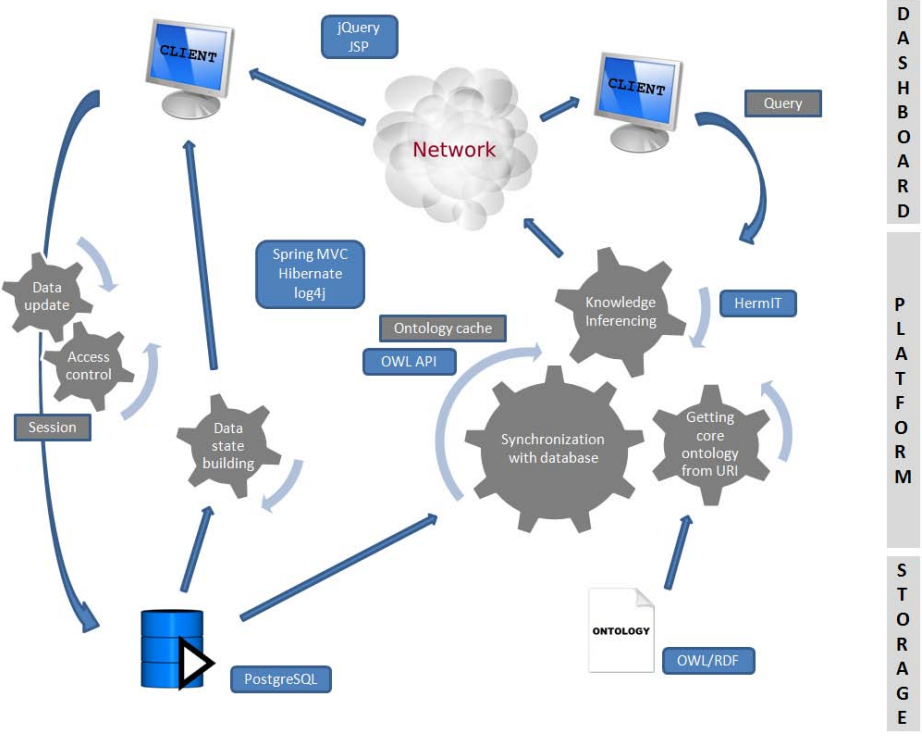


Fig. 4. Data flow in the system

The ontology can be stored on local or remote server and is accessed by URI. Cooperation with database is provided through Hibernate ORM[4] technology. User interface is built with JavaServer Pages (JSP)[5] and jQuery JavaScript Library[6], while requests from users and appropriate responses, are controlled by Spring MVC[7]. For logging the results of particular operations, log4j Java-based logging utility[8] is used. Ontology is provided in different formats like

[4] See http://www.hibernate.org/
[5] See http://www.oracle.com/technetwork/java/javaee/jsp/index.html
[6] See http://jquery.com/
[7] See http://static.springsource.org/spring/docs/3.0.x/
[8] See http://logging.apache.org/log4j/

OWL2 XML[9], RDF/XML[10] or OWL2 Manchester Syntax[11]. Synchronization is based on the OWL API Library, and provides up-to-date cached information for HermiT DL Reasoner to infer. PostgreSQL[12] is chosen as SQL database server. The project is written in Java language using the Eclipse Java IDE[13]. Dependencies management and versioning is the task of Apache Maven tool[14]. All of the mentioned technologies are free and open source.

The architecture of the system, with loose coupling of the database and ontology, enables using the same core ontology in various installations. Synchronized ontologies (populated with real time data) can differ between various environments. This decentralized way of cooperation, when each client have cached in memory its own synchronized-on-demand ontology instance, is chosen for performance optimization reasons. Comparisons to situation in which single instance of ontology is the centralized part, accessible for synchronization to all clients, provide obvious performance drawbacks. This is why it was considered as an antipattern for such system and depreciated by design.

4 Summary

Development of Semantic Web applications with use of recent powerful tools like Protégé is an inspiring but a challenging task. Ontology-driven development with the support of agile methodologies ensures an efficient implementation process. Ontologies can be developed directly by domain experts. Because of direct access to executable systems, feedbacks can be made frequently. Best practices from agile methodologies, effective at delivering particular outcomes, can be used in the development of high-quality domain models. Domain experts may work together with programmers, which can guarantee coherent testing and fast implementation processes.

5 Future Work

Future work should focus on providing map-based interface for interaction with users, such as OpenStreetMap[15] or Google Maps[16]. Another directions of project development could be focused on extensions for heterogeneous application-to-application communication. The RESTful Web Services[17] can be taken under consideration. These external systems could be perceived as software agents.

[9] See http://www.w3.org/TR/owl2-xml-serialization/
[10] See http://www.w3.org/TR/rdf-syntax-grammar/
[11] See http://www.w3.org/TR/owl2-manchester-syntax/
[12] See http://www.postgresql.org/
[13] See http://www.eclipse.org/
[14] See http://maven.apache.org/
[15] See http://wiki.openstreetmap.org/wiki/Main_Page
[16] See http://code.google.com/apis/maps/index.html
[17] See http://www.ics.uci.edu/~fielding/pubs/dissertation/rest_arch_style

Their tasks could be focused on periodic connections to the primary system, getting some information set, and creating statistics about traffic dangers. Statistics could visualize frequencies of desired dangers on specific area or classification of safety in desired district.

References

1. Baader, F., Calvanese, D., McGuinness, D.L., Nardi, D., Patel-Schneider, P.F. (eds.): The Description Logic Handbook: Theory, Implementation, and Applications. Cambridge University Press, Cambridge (2003)
2. Gruber, T.R.: A translation approach to portable ontology specifications. Knowledge Acquisition 5(2), 199–220 (1993)
3. Ligęza, A., Ernst, S., Nowaczyk, S., Nalepa, G.J., Furmańska, W.T., Czapko, M., Grzesiak, P., Kałuża, M., Krzych, M.: Towards enregistration of threats in urban environments: practical consideration for a GIS-enabled web knowledge acquisition system. In: Jacek Dańda, A.G., Derkacz, J. (eds.) MCSS 2010: Multimedia Communications, Services and Security: IEEE International Conference: Kraków, May 6-7, pp. 152–158 (2010)
4. Noy, N.F., McGuinness, D.L.: Ontology Development 101: A Guide to Creating Your First Ontology, Stanford University, Stanford, CA, 94305
5. Waliszko, J.: Knowledge representation and processing methods in Semantic Web. Master's thesis, AGH University of Science and Technology (2010)

A Multicriteria Model for Dynamic Route Planning*

Wojciech Chmiel, Piotr Kadłuczka, and Sebastian Ernst

AGH University of Science and Technology
{wch,pkad,ernst}@agh.edu.pl

Abstract. In this paper we introduce INSIGMA project and its part related to road transport. We propose methods and parameters for route optimization in dynamic urban environments. An overall motivation for development of proposed dynamic system is discussed with detailed efficiency analysis of placed algorithms. From the end users' perspective, several practical use cases are identified. The system takes advantages of so-called dynamic maps, which provide information about up-to-date traffic conditions and visualization functionality as well.

Keywords: route optimization, dynamic maps, multimedia services.

1 Introduction

The INSIGMA project [1] aims at development and implementation of an advanced information system for monitoring of moving objects and detection of threats. The goals can be divided into three groups. The first and foremost goal is analysis of traffic parameters on basis of dynamic maps. A dynamic map is defined as representation of the road transport infrastructure combined with up-to-date information about traffic intensity and historical traffic data. Such a combination includes information stored in a database and map visualisation, which can be presented to the end user via a web-based or mobile interface. Algorithms for dynamic route optimisation are applied to the system; they aid traffic control systems and are particularly useful in urban environments.

The second goal of the project is development of automatic methods for object observation and registration of their parameters. Particular application of these systems will be targeted at providing low-level data for generation of dynamic maps. Finally, the search system will enable interfaces to define route optimisation tasks, intelligent monitoring and other multimedia services.

In the time when more and more cities are facing the problem of traffic jams and worsening ecological conditions, INSIGMA will provide tools for efficient traffic control and threat detection. In case of existing solutions for map creation and traffic management, the main problem is related to the lack of efficient tools

* This work has been co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme, INSIGMA project no. POIG.01.01.02-00-062/09.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 174–182, 2011.

that enable conversion of object location data and audiovisual data onto dynamic maps. In practice, traffic control is often based on low-frequency components of traffic dynamics. Thus, in case of road accidents or other threats, these systems are incapable of fast response. In such scenarios, traffic problems can be first detected after 30 minutes. Another issue is the limited area of the managed region. Traffic detectors are mostly deployed in major highways but in limited scope in access roads.

On the other hand, vehicle users demand efficient route planning and optimisation. Existing navigation solutions consist in analysis of static parameters (e.g. total route length or road type), and even recent solutions rarely use statistical data (e.g. maximum driving speed in selected hours). The purely dynamic traffic component still remains unused and relevant traffic conditions need to be personally recognised by the user in advance or, worse, in the place of event. Thus, in case of a jammed metropolis, the efficiency of route optimisation is far from what is expected. This is particularly important in daily operations of emergency services, fire brigades, police, etc.

In this paper, a solution for route optimisation using heuristic algorithms is presented, along with detailed analysis.

The remainder of this paper is organised as follows. Section 2 refers to related work and presents motivation for the development of the system. Section 3 presents theoretical background for presented aspects of the system. Section 4 includes description of search for shortest paths in dynamic environments. We conclude in Section 5. References are included at the end of the paper.

2 State of the Art and Motivation

Computer-aided, commercially-available route planning systems date back to the 1980s. At that time, products such as AutoRoute had to cope with harsh limitations of available resources. The improvement of processing power even in portable personal computers helped solve that problem, but the search algorithms and optimality criteria remained mostly unchanged since that time.

Today, route planning algorithms are based on informed graph-search methods, such as various modifications of the famous A^* algorithm and its descendants (e.g. IDA^*) [11,8]. The optimality criteria used by such route planners are, in most cases, limited to computation of: (a) shortest routes (based solely on street network geometry); (b) time-optimal solutions (where the travel time is usually calculated using simple criteria based on the type of route, not taking the prevailing traffic conditions or incidents into account); (c) least expensive routes (by adding estimated fuel consumption and road fees/taxes to the time-optimal solution). These possibilities can be supplemented by requiring the inclusion or avoidance of via points, avoiding left-turns, bridges, as well as taking other user preferences and physical vehicle limitations (size, weight) into account.

Recently, commercial navigation systems have been supplemented with collaborative traffic data collection features, either in an implicit (by recording

GPS tracks) or an explicit (by user input) manner. However, such systems make rather limited use of the collected data as far as route profiling and the ability to react dynamically are concerned.

Route planning is the topic of a lot of research being currently done; however, little work pertains directly to the case of navigation in cities. The characteristics of urban environments are different from those of present in long-distance route planning, such as [6,5]:

- **Vehicle-unfriendly layout.** Cities built before cars have appeared were designed with pedestrian, horse and rail transport in mind; the streets are narrow, and buildings or other obstacles make it impossible to create an appropriate road network.
- **Non-traversable obstacles.** Rivers, railways and parks are cannot be traversed without providing special means, such as bridges.
- **Non-homogeneous structure.** Navigation is a straightforward task in cities with a regular structure, such as Manhattan in New York. However, historic European cities feature regions of various street density, often with limited means of transit from one region to another.
- **Uncertain traffic conditions.** Overcrowded streets form a network that is very prone to suffer from bottlenecks. As a result, traffic jams occur irregularly; it is difficult to predict their location and extent.

These characteristics require application of specialised route planning methods, as described in the following sections.

3 Theoretical Background

The proposed model of the transport network, called RPS (*Route Planning Subsystem*) is a directed multigraph with weights (labels) defined for both arcs and vertices.

Such approach makes it possible to take alternative (multivariant) routes, optimised with respect to the various types of users with different preferences and criteria of optimality, into account. The foundation for this approach is the requirement to provide individual problem solutions for dynamic map customers (as described in Section 1). Any two vertices can be connected with more than one edge having the same return and can be assigned different weights (labels).

For example, a multigraph is built by defining the set A as a generalisation of the set of arcs in the form of a multiset, or by introducing a vector of weights and labels on the edges.
Let

$$G_{A,W} = \{G, \mathbf{A}, \mathbf{W}\} \tag{1}$$

define a network, where $\{\mathbf{A}_{i,j}^{p,t}\}$ is a matrix family, containing the description of edges of multigraph G, and $\{\mathbf{W}_{i}^{p,t}\}$ is a matrix family decribing the properties of vertices, where:

- i, j are the indexes of multigraph vertices,

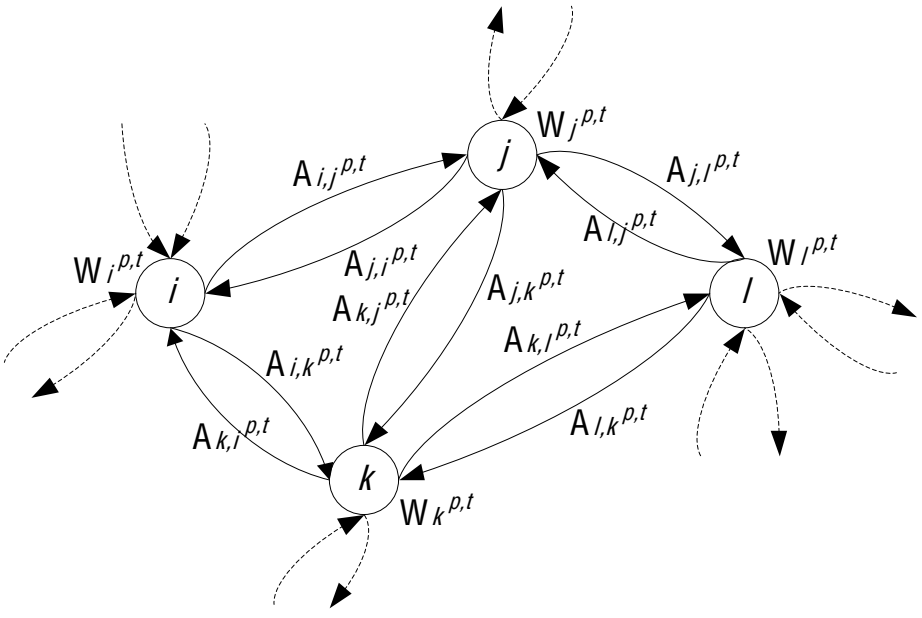


Fig. 1. Model of the transport network for the INSIGMA project

– p is an edge label,
– t is the time interval index.

In the considered model, matrix $\mathbf{A}_{i,j}^{p,t}$ is interpreted as parameters of the edge (i, j), specified for the label (arc-labelled multigraph) p in time interval t. In the simplest case, this value can be interpreted as the transition time of edge (i, j), for user of category p. This parameter is variable in time (discrete) and defined for a specified time horizon $t = t_0, \ldots, t_{MAX}$ repeated periodically (e.g. every week). It contains the historical current and predicted values of this parameter. Analogously, we can interpret $\mathbf{W}_i^{p,t}$ as a set of parameters describing times of transition between the multigraph arcs. These arrays are assigned to vertices i, category label p and time t.

In other words, for each vertex of a graph describing the transport network, we define a matrix which describes its properties from the perspective of the network.

Matrix $\mathbf{W}_i^{p,t}$ is of the form:

$$\mathbf{W}_i^{p,t} = \begin{bmatrix} w^{p,t}_{in_1,i,out_1} & w^{p,t}_{in_1,i,out_2} & \cdots & w^{p,t}_{in_1,i,out_m} \\ w^{p,t}_{in_2,i,out_1} & \ddots & \cdots & w^{p,t}_{in_2,i,out_m} \\ \vdots & \vdots & \ddots & \vdots \\ w^{p,t}_{in_n,i,out_1} & w^{p,t}_{in_n,i,out_2} & \cdots & w^{p,t}_{in_n,i,out_m} \end{bmatrix} \tag{2}$$

where:

- n is the number arcs connecting to vertex i and starting in the vertices $in_1, in_2, \ldots, in_n$,
- m is the number of *outgoing* arcs from vertex i, ending in vertices $out_1, out_2, \ldots, out_m$.

Matrix elements $w_{i,j,k}^{p,t} \in \mathbf{R}$ are, in general, a real continuous (or discrete) function of time. If interpreted as the time of transition, the three indexes i, j, k describe the transition time the vertex i through vertex j to vertex k.

The matrix $\mathbf{W}_i^{p,t}$ describes all possible transitions through the vertex (e.g. corresponding to various manoeuvres on a junction, such as going straight on, or turning left/right) from every possible direction. The parameters included in the vector $\mathbf{A}_{i,j}^{p,t}$, can describe road properties such as:

- passing time,
- density,
- flow (crowding),
- amount of occupancy,
- length queues,
- length path,
- likelihood of overcrowding of the road.

These parameters can be considered independently or together in one multicriteria optimisation model. These parameters can be aggregated in a generalised cost function or interpreted as a constraint for the defined model.

Any vertex of the multigraph is assigned a pair (x_i, y_i), defining Cartesian coordinates. Cartesian coordinates enable calculating the azimuth of directions defined by pairs of vertices and their absolute distances. These values, among others, are used in several types of greedy algorithms used in RPS. Depending on the conditions in which the vehicles are driving, the system enables using different models of transport networks:

- In the *static model*, the parameters of the model describing the roads are fixed and independent of the flow of time. It corresponds to the situation where road traffic is free and does not depend on interactions with other vehicles on the road. In this case, the vehicle can be affected with the technical features of roads, speed limits, its own parameters and atmospheric conditions.
- In the *statistical model*, parameters describing the roads and road traffic change periodically over time. In this case, road traffic is partially dependent [10].
- In the *dynamic model*, traffic takes place in a quickly-changing environment. Parameter values of the transport network model are based on prediction. It is possible thanks to information about the current state of the road, provided by a set of sensors.

Sensors provide data for the system, used to calculate the traffic parameters. This information includes not only the traffic flow, but also information about

incidents and phenomena (e.g. weather), which may have an impact in the near future on the state of the transport network.

In any case, the model features should be adapted to the type of roads and user preferences. For each path (vertices) based on observation, the best-fit model is selected. This approach allows for reduction of computational effort to determine the best routes and, consequently, for potential reduction of computational performance requirements.

The proposed multigraph model generalises the mathematical description of all the above types of models of transport networks.

4 Algorithm Framework for Dynamic Environments

The classical shortest path problem belongs to the complexity class *P*. However, additional requirements and restrictions in the considered problem model of transport network shift that problem to a higher complexity class [2,3].

Determining globally optimal solutions using exact methods, due to the required computation time, is possible only for problem instances of small size. To solve larger problems, dealt with by the INSIGMA project, efficient approximate methods should be applied.

The general scheme of a two-step multicriteria algorithm for determination of a suboptimal path in a transport network is shown in Figure 2.

In the first phase of computation, a collection of greedy algorithms is used. The obtained solutions form the starting set for the next optimisation phase. Different criteria used in construction algorithms ensure diversification of the initial solutions. The general principles of this construction algorithm are as follows:

1. Specify the start and end vertices of the path and the criteria vector (weights) associated with the user preferences.
2. Build a solution – a path in the graph. Start at the initial vertex and continue until the final vertex is reached.
3. In each iteration, determine a new element of the solution by adding one of the vertices reachable from the end of the current path. The new vertex is determined using one or more user-selected criteria, for example:
   - *Criterion 1:* the smallest deviation from the azimuth to target vertex.
   - *Criterion 2:* the distance from the target.
   - *Criterion 3:* appropriate width of the road (or any other restrictions specified for the edges).
   - ...
   - *Criterion n:* maximum capacity or class of the road.
4. Vertices meeting the criteria receive weights corresponding to values provided by the predefined weight vector. As a result of this procedure, a node with largest total weight is selected. Another possible solution is to select vertices based on probabilities proportional to obtained weights.

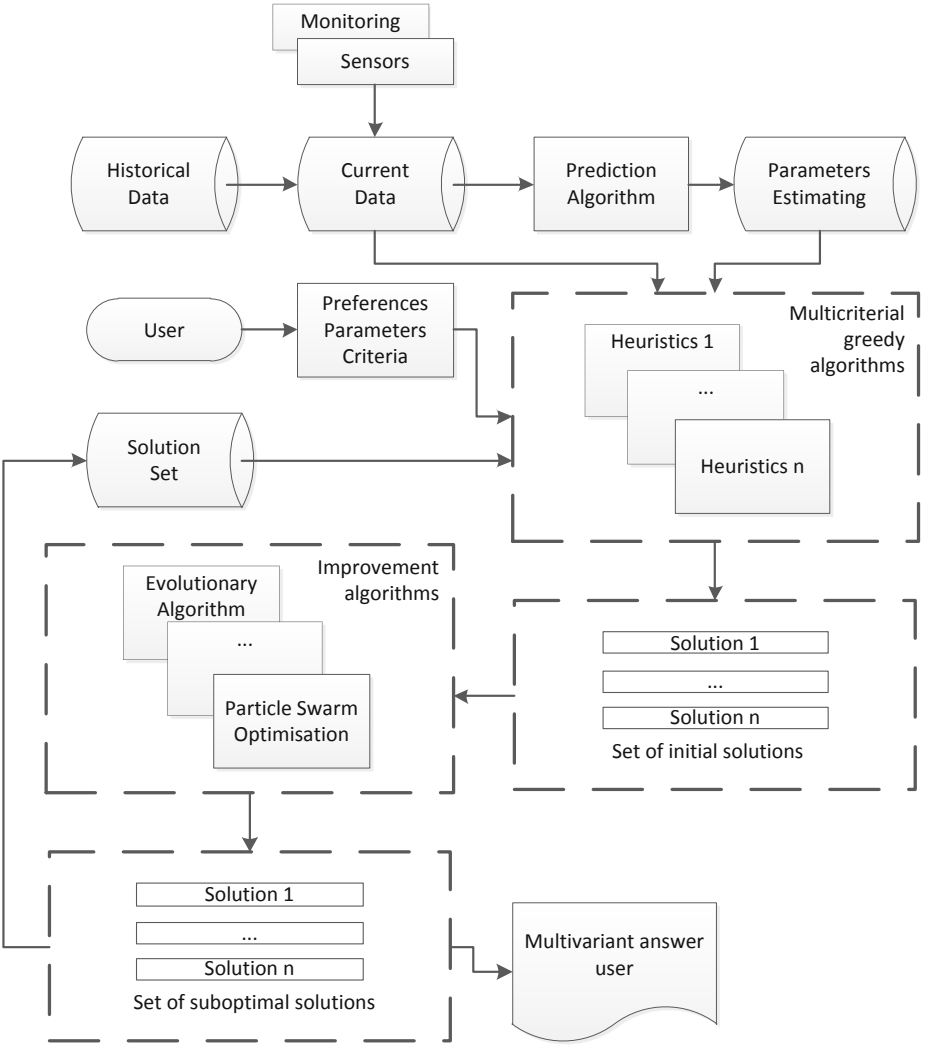


Fig. 2. Multivariant algorithm for determination of routes in the transport network

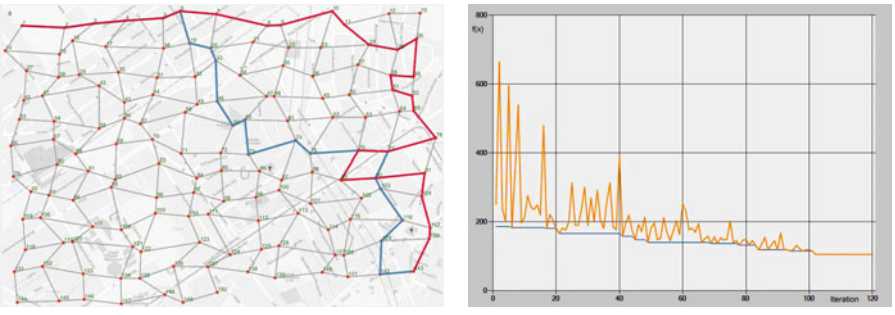


Fig. 3. Example of a solution to the shortest path problem (start solution and best solution) and the process of optimisation (best particle in a swarm)

5. If a vertex appears twice in the path (a cycle), the path is reduced by removing the last added set of vertices. The last iteration is repeated with the vertices removed from the neighbour set. This step is repeated as long as the cycle appears in the calculated path. This allows for detection of graph inconsistencies.

In the second phase of the algorithm, population-based methods are used to improve the initial solution.

Algorithms used to process a set of solutions preserve the possibility of finding alternative routes. For this purpose, a selection of population algorithms, such as the evolutionary algorithm and PSO (Particle Swarm Optimisation) is used. In the construction of pseudo-genetic operators like crossover and mutation, greedy and local optimisation methods are used among many others [4].

Particle Swarm Optimisation (PSO) is an optimisation procedure based on the social behaviour of a set of individuals, such as birds or fish. Individual solutions in a population are viewed as "particles", which evolve or change their positions in time. Each particle modifies its position in the search space, according to both its own and the neighbouring particle's experience by remembering the best position visited by itself and its neighbours, thus combining local and global search methods [9]. In the proposed solution, a the PSO algorithm was adapted for the special discrete problem like the shortest path problem [7].

The algorithm operates on current data, generated using historical data (due to the cyclic nature of the road traffic process). Information from sensors enables the system to reflect changes in dynamic parameters of the transport network model.

Taking emergency situations (accident, road works) and the trend of the number of vehicles on the road into account, new values of the model parameters are predicted. Data associated with the customer data profile allows determination of the most appropriate multigraph labelled edge subset. This data includes: vehicle types (motorcycle, car, truck, bus, vehicle carrying dangerous substances, the privileged or special vehicle), vehicle size, weight, etc., as well as search criteria (quickest, shortest, special, off-road, cross, walking).

Customer profiling enables determination of the criteria weight vector used by the greedy algorithms, constraints, and solution evaluation functions.

5 Conclusions

This paper describes the application of multicriteria algorithms to dynamic route optimisation. General motivations, assumptions and layered structure of the proposed dynamic system are presented and discussed.

A heuristic algorithm framework aimed at utilisation of dynamic data and provision of profiled solutions is presented. These algorithms provide the required flexibility while maintaining reasonable computation time.

In further work, we plan to extend processing and analysis of dynamic data and further optimise selected algorithms using real-life data and performance evaluation techniques.

References

1. INSIGMA Project (2009), http://insigma.kt.agh.edu.pl
2. Adamski, A., Chmiel, W.: Optimal service synchronization in public transport. Transportation System. In: Papageorgiou, M., Pouliezos, A. (eds.), vol. 3, pp. 1283–1287 (1994)
3. Ceder, A.: Public Transit Planning and Operation: Theory, Modeling and Practice. Elsevier, Butterworth-Heinemann, Oxford, UK (2007)
4. Chmiel, W., Kadłuczka, P.: Algorytm hybrydowy z długoterminową pamięcią czestotliwościową. Automatyka 3(1), 59–70 (1999) (in polish)
5. Ernst, S.: Artificial Intelligence Techniques in Real-Time Robust Route Planning. Ph.D. thesis, AGH University of Science and technology (2010)
6. Ernst, S., Ligeza, A.: Knowledge Representation for Intelligent and Error-Prone Execution of Robust Granular Plans. A Conceptual Study. In: FLAIRS Conference. Sanibel Island, Florida, USA (2009)
7. Filipowicz, B., Chmiel, W., Kadłuczka, P.: Guided search of the solution space in swarm algorithm. Automatyka 13(2), 247–255 (2009)
8. Ghallab, M., Nau, D., Traverso, P.: Automated Planning: Theory & Practice. Morgan Kaufmann Publishers Inc., San Francisco (2004), http://portal.acm.org/citation.cfm?id=975615
9. Kennedy, J., Eberhart, R.C.: Swarm Intelligence. Morgan Kaufmann Publishers, San Francisco (2001)
10. Komar, Z., Wołek, C.: Inżynieria ruchu drogowego: wybrane zagadnienia. Wydawnictwo Politechniki Wrocławskiej (1994) (in Polish)
11. Russell, S.J., Norvig, P.: Artificial Intelligence: A Modern Approach, 3rd edn. Pearson Education, London (2010)

Extensible Web Crawler – Towards Multimedia Material Analysis

Wojciech Turek, Andrzej Opalinski, and Marek Kisiel-Dorohinicki

AGH University of Science and Technology, Krakow, Poland
{wojciech.turek,doroh}@agh.edu.pl, opal@tempus.metal.agh.edu.pl

Abstract. Methods of Web pages content monitoring come increasingly in the interest of law enforcement services, searching for Web pages contain symptoms of criminal activities. The information can be hidden from indexing systems by embedding in multimedia materials. Finding such materials is a large challenge of contemporary criminal analysis. A concept of integrating a large scale Web crawling system with a multimedia materials analysis algorithms is described in this paper. The Web crawling system, which is processing a few hundred pages per second, provides a mechanism for plugin inclusion. A plugin can analyze processed resources and detect references to multimedia materials. The references are passed to a component, which implements an algorithm for image or video analysis. Several approaches to the integration are described and some exemplary implementation assumptions are presented.

Keywords: Web crawling, image analysis.

1 Introduction

The World Wide Web is probably the greatest public ally available base of electronic data. It contains huge amounts of information on almost every subject. Last two decades brought significant changes in the source of information – so called "Web 2.0" solutions allowed users of the Web to become authors of Web pages content. This phenomenon has many side effects, including partial anonymousness of information source. As a result, the content of Web pages comes increasingly in the interest of law enforcement services.

It is relatively easy to find or analyze text on a Web page. There are many popular and powerful tools for that present on the Internet. Publicly available tools, like Google Search, have some limitation and features, however an experienced user can make a good use of this. Because of that, people rarely place textual symptoms of criminal activities on publicly available pages.

It is much easier to hide information, which is embedded into a multimedia material, like an image or a movie. Analysis of such materials is much more complex and time consuming. Public search systems can only perform basic operations of this kind, leaving lots of information unrecognized.

In this paper a concept of integrating a large scale Web crawling system with a multimedia materials analysis algorithms is described. The concept is being

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 183–190, 2011.

tested on a crawling system, which has been developed at the Department of Computer Science, AGH UST [1].

The paper starts with a short motivation concerning the complexity of Web analysis tasks. In the next sections a general structure of the crawling system is presented followed by the crawlers' resource processing scheme. Then selected aspects of integration with multimedia analysis algorithms are discussed together with possible applications of the approach.

2 Motivation

In order to perform large-scale tests in the domain of Web pages content analysis, a Web crawling system is needed. It is relatively easy to create or adopt a solution, that would be capable of processing tens of thousands of Web pages. Some examples of such tools, which can be easily downloaded from the Internet, are:

- WebSPHINX [2], which can use user-defined processing algorithms and provides graphical interface for crawling process visualization,
- Nutch, which can be integrated with Lucine [3], a text indexing system,
- Heritrix [4], used by The Internet Archive digital library, which stores Web pages changes over recent years.

However, if a crawling system is required to process tens of millions of Web pages, a complex scale problem arises. Some of available solutions can handle this kind of a problem, however deploying such a system is not as easy as running an off-shelf application.

The most obvious scale problem is a number of pages in the Web. On 2008 Google engineers announced, that the Google Search engine discovered one trillion unique URLs [5]. Assuming, that one URL has over 20 characters, the amount of space required for storing URLs only is more than 18 terabytes.

Surprisingly, a list of unique words found on Web pages should not cause scale problems. A system, which is indexing over 10^8 pages found only $1.7 * 10^6$ words. This amount cannot be held in memory, but can be located on a hard drive of a single computer.

However searching Web pages by words requires the system to build an inverted index, witch stores a list of URLs for each word. Assuming, that each URL is identified by an 8-bytes integer, and an average Web page contains 1000 words, the index requires more than 7,2 petabytes.

Another issue related to the scale of the problem is page refresh time, which determines how often a crawler should visit a particular page. If each page is to be visited once a year (which is rather rare), the system must process more than 31000 pages per second. Assuming, that an average page is bigger than 1 KB, a network bandwidth at 31 MB per second is a minimum requirement.

It is obvious, that such a huge amount of data cannot be processed by a single computer. It is necessary to design a distributed system for parallel processing of well-defined subtasks.

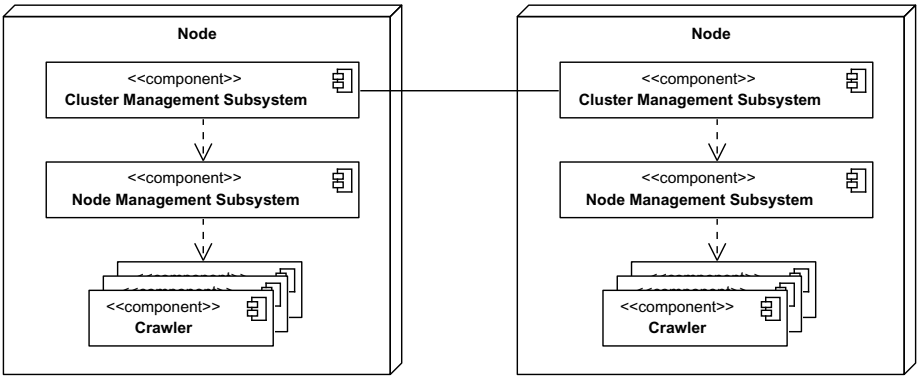


Fig. 1. The structure of the crawling system

3 Architecture of the Crawling System

An exemplary crawling system, which has been created at the Department of Computer Science, AGH UST [1], is designed to handle hundreds of millions of Web pages. It runs on a cluster of 10 computers and handles up to a few hundred pages per second. Performance of the system can be improved easily, by adding more computers to the cluster.

The system is implemented in Java, it runs on a Lunux OS in JBoss application server [6]. Communication between computers in the cluster is based on EJB components. Most of the storage uses MySQL DBMS, which has been carefully tuned for this application. Each of the servers has a single, four-core CPU, 8 GB of RAM and 3 TB of storage.

Use of relational database is very convenient, however crawling systems use different persistence solutions, due to performance reasons. Indexes created by typical DBMS are becoming too slow when datasets grow huge. However, if a dataset does not exceed 10^6 elements the performance is sufficient. The system described in this paper splits large datasets, like word vocabulary, into 256 subsets using a hashcode of elements. Performed tests prove, that use of a RDBMS is possible in this type of system, provided that datasets are split into proper fragments.

There are two main subsystems of the crawling system (see figure 1):

1. cluster management subsystem and
2. node management subsystem.

The cluster management subsystem is responsible for integration of all nodes into one uniform system. It provides an administrator Web interface and controls the process of Web pages crawling an all nodes. The process is controlled by assigning Internet domains to particular nodes. Each domain (and all its subdomains) can be assigned to one node only, therefore there is no risk of processing

the same URL on multiple nodes. The domain-based distribution of tasks is used also for load balancing – each node gets approximately the same number of domains to process.

The cluster management subsystem is also responsible for providing an API for search services. The services use the node management system's search services to perform search on all nodes, aggregate and return results. Two basic elements of the API allow users to search for:

1. URLs of Web pages containing specified words,
2. content of indexed Web page of a specified URL.

The node management subsystem manages the crawling process on a single node. It is controlled by assigned set of domains to process. Each domain has a large set of URLs to be downloaded and analyzed each time the domain is refreshed. During the processing, new URLS can be found. One of three cases can occur:

1. the URL belongs to the domain being processed – it is added to the current queue of URLs,
2. the URL belongs to a domain assigned to this node – it is stored in a database of URLs,
3. the URL belongs to a different domain – it is sent to the cluster management subsystem.

Each domain is being processed by a single crawler. The node management subsystem runs tens of crawlers simultaneously, starting new ones if sufficient system resources are present.

4 Crawler Resource Processing

Each crawler has a list of URLs to be processed. URLs are being processed sequentially – the crawler has a single processing thread. Sequential processing of a single domain can take a long time (a few days) if the domain is large or the connection is slow. However in this application long processing time is not a drawback, as particular time of indexing is not important. Moreover, this solution is less prone to simple DOS filters, detecting number of queries per second. It is also easier to implement due to reduction of thread synchronization required.

The sequence of actions performed by a crawler during single URL processing is shown in figure 2.

There are five main steps of resource processing:

1. Downloading the resource performed by a specialized component, which limits its maximum time and controls the size of the resource.
2. Lexical analysis performed on a source of the Resource, which creates a unified "resource model". The source is:
   (a) converted to Unicode,

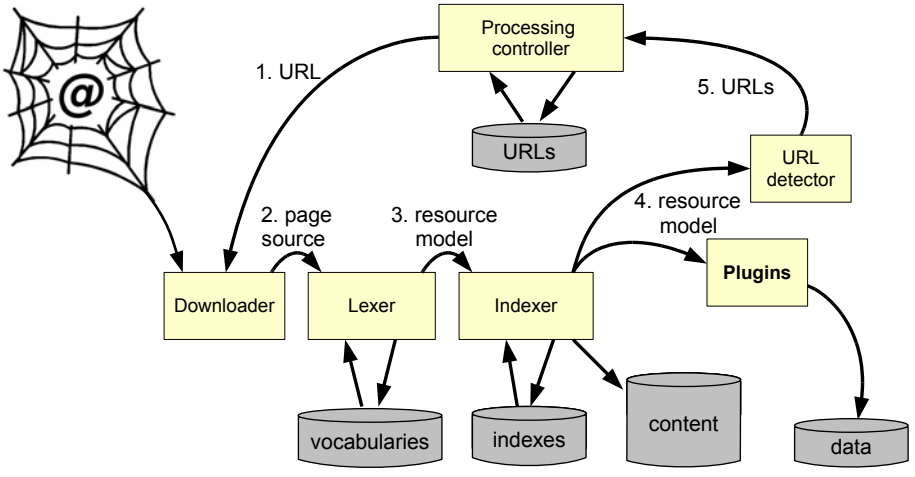


Fig. 2. Resource processing performed by a crawler

(b) split into segments during HTML structure analysis,
(c) divided into tokens and words sequences,
(d) converted to tokens and words identifiers, which are stored in a dynamically extended vocabularies.

3. Content indexing performed on a resource model by an Indexer component. An inverted index of words and stored content of the resource are being updated.
4. Content analysis performed by specialized plugins. Each plugin receives a resource model, performs particular operations and stores results in own database. One of the plugins, the URL detector, locates all URLs in the resource.
5. URLs management performed by the Processing controller. URLs are added to a local queue stored in a database or sent to the cluster management subsystem, depending on domain.

The resource model created by the Lexer is a data structure used for convenient processing of content of any Web page. The structure of a Web page source is used for creating a tree of content segments. During content parsing, HTML tags defining tables, paragraphs and lists items are interpreted as segment delimiters. Example of a very simple Web page content and a tree generated for it is presented in figure 3.

Resource model provides a convenient interface for Web page content analysis. An algorithm can analyze textual content only, or traverse through the structure of the tree. Additional information concerning formatting or embedded objects can be easily extracted. All words and special characters in the content are converted to identifiers, which significantly improves the performance of analysis.

```
<html>
  <body>
    <p> First p </p> <p> Second p </p>
    <br><br>
    <table border=1>
      <tr>
        <td> <b>Michael </td> <td> Ston </td>
      </tr><tr>
        <td>Lynx St. 23/1</td><td>London</td>
      </tr>
    </table>
  </body>
</html>
```

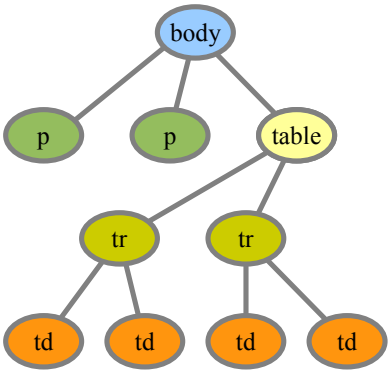


Fig. 3. Exemplary HTML source and a tree created by the Lexer

Currently the system is indexing pages in Polish language. Polish language detection is based on the count of Polish-specific characters detected in the Web page content.

5 Integration with Multimedia Analysis Algorithms

The crawling system, presented in the previous section, can be integrated with a multimedia analysis algorithm in several ways. In general, a plugin for the crawling system has to be developed. It has to detect multimedia materials references in the resource model. After a reference is found, several options are available:

- Download the material and perform the analysis by the plugin,
- Use a remote service to download and analyze the material; collect results,
- Feed a remote, independent service with references.

Each of these solutions has some advantages and drawbacks. The integrated solution is limited to the Java technology, therefore is not universal. Many processing algorithms are written in different languages and cannot be directly integrated with the crawler. Moreover, this type of algorithm can have very large requirements, which can cause uncontrolled JVM errors and shutdown.

A remote synchronous service, which downloads and processes a material autonomously, can be implemented in any technology and run on a remote computer. However, this type of solution can cause some synchronization problems. The crawling system has to wait for results, not knowing how much time is it going to take. The crawling system uses several computers and hundreds of crawlers working simultaneously. Adjustment of multimedia processing performance can be a hard task.

The last solution, which uses a remote asynchronous service, has the smallest impact on the crawling system. However, feeding the remote service with too many references to multimedia materials can easily result in service overload.

To reduce the number of queries to the multimedia analysis service, the plugin integrated into the crawling system cam perform basic filtering. An exemplary application of the approach described in this paper will be responsible for finding images, which contain faces of people useful for identification. Therefore the plugin will analyze a surrounding of located images references, verifying specified criteria. This should reduce the number of analyzed images by removing most of graphical elements of a Web page.

6 Possible Applications

There are many practical applications of the solution described in this paper. Some of those are already implemented in publicly available Web search engines, like Google Images (http://images.google.com/). A user can find images of a specified size or type or even images which contain faces.

However many different applications, which may be useful in criminal analysis, are not available. Therefore use of a dedicated system is required.

Three different groups of applications have been identified:

1. text detection and recognition,
2. face detection and recognition,
3. crime identification.

Plain text, located in a source of a Web page (HTML) can be found and analyzed easily. However, one could use images or movies to publish information on the Web, which would not be indexed by a text-based crawler. An image analysis algorithm, which can detect this kind of images and recognize the text, can be used to enrich index of information hidden in the image.

Face detection and face recognition can be used by identity analysis systems. If an individual, whose Internet activities are suspected, is to be identified, a system can help to locate and associate pages containing information about the individual. Moreover, face recognition can help to equate a person, active at different locations on the Web, by finding her/his face on several images.

People can be associated with each other by finding significant elements present in multimedia materials. Car plates are a good example of such elements.

In some cases, evidence of criminal activities are being published on the Web. The most typical example of such publications are movies submitted to video-sharing services. This type of material can be analyzed by an advanced video processing component, searching for defined patterns or text elements.

Another very important case of illegal multimedia material on the Web is child pornography. Despite the fact that publishing such material is a serious crime in most countries, such situations still occur. An algorithm for detecting child pornography in multimedia materials, integrated with a crawling system could significantly improve detectability of such crimes.

7 Conclusions

The concept described in this paper is based on an idea of integrating a Web crawling system with multimedia analysis algorithm. These two elements can create a system for detecting various contents present in the Web, which are impossible to find using text-based indexes. Such tools can be used by the law enforcement services in many different situations.

The crawling system presented in this paper is being integrated with an image analyzing component, developed at the Department of Telecommunications, AGH UST [7]. The component is able of finding the number of faces in a given image. In the near future another features of the image analyzing component, like fascist contents identification or child pornography detection, will be used. Many technical issues concerning the integration have already been solved. Further research and tests of the system are aimed at providing high quality tools useful in criminal analysis and crime detection.

Acknowledgments

The research leading to these results has received funding from the European Communitys Seventh Framework Program (FP7/2007-2013) under grant agreement nr 218086.

References

1. Opalinski, A., Turek, W.: Information retrieval and identity analysis. In: Metody Sztucznej Inteligencji w Dzialaniach na Rzecz Bezpieczenstwa Publicznego, pp. 173–194 (2009); ISBN: 978-83-7464-268-2
2. Miller, R.C., Bharat, K.: SPHINX: A Framework for Creating Personal, Site-Specific Web Crawlers. In: Proceedings of WWW 2007, Brisbane Australia (1998)
3. Shoberg, J.: Building Search Applications with Lucine and Nutch. APress (2006); ISBN: 978-1590596876
4. Sigursson, K.: Incremental crawling with Heritrix. In: Proceedings of the 5th International Web Archiving Workshop (2005)
5. Marrs, T., Davis, S.: JBoss At Work: A Practical Guide. O'Reilly, Sebastopol (2005); ISBN: 0596007345
6. Alpert, J., Hajaj, N.: We knew the web was big... The Official Google Blog (2008)
7. Korus, P., Glowacz, A.: A system for automatic face indexing. Przeglad Telekomunikacyjny, Wiadomosci Telekomunikacyjne 81(8-9), 1304–1312 (2008); ISSN 1230-3496

Acoustic Events Detection Using MFCC and MPEG-7 Descriptors

Eva Vozáriková, Jozef Juhár, and Anton Čižmár

Technical university of Košice, Slovak Republic
Dept. of Electronics and Multimedia Communications, FEI TU Košice
Park Komenského 13, 041 20 Košice, Slovak Republic
{eva.vozarikova,jozef.juhar,anton.cizmar}@tuke.sk
http://kemt.fei.tuke.sk/

Abstract. This paper is focused on the acoustic events detection. Particularly two types of acoustic events (gun shot, breaking glass) were investigated. For any detection task the feature extraction methods play very important role. The feature extraction influences the recognition rate, therefore it is most important in any pattern recognition task. In this paper the impact of Mel-Frequency Cepstral Coefficients - MFCC and selected set of MPEG-7 low-level descriptors were examined. The best feature set contained MFCC and selected descriptors such as ASC, ASS, ASF. They were used to represent the sounds of acoustic events and background. We obtained the improvement of the detection rate using the mentioned set of features. In this task GMM classifiers are used to model the sound classes. This paper describes a basic aspect of our work.

Keywords: Acoustic events, feature extraction, MFCC, MPEG-7 low-level descriptors.

1 Introduction

Audio surveillance systems are highly requested systems nowadays. Over the past decades a great deal of work has been done and published in the area of detection and classification input pattern [1], [2]. The complex surveillance system [3] can be created by the fusion of sound and video information. The effectiveness of the surveillance system depends on environmental conditions. The visual part of detection system will probably fail in terms of the bad light condition or if the problematic situation is not in the visual field of surveillance camera. On the other hand, the audio part of system is very sensitive on the sound similarity. Generally, extreme weather conditions limit the performance of any surveillance system.

Surveillance systems are usually used at monitor public places, stadiums, vehicles and stations of public transport, etc. Some dangerous situations are more easily detectable via sound information than the video information for example calling for help, sound of gun shots, etc. Of course, there are many applications

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 191–197, 2011.

[4], [5], where the detection of specific sounds can be very helpful, e.g. in smart rooms, health care or industry applications, etc.

The goal of each surveillance system is help to protect life and property. Detection systems should generate the alert only if dangerous event is detected. It is very important to reduce false alarm as a result of incorrectly classified pattern. The proposed detection system is created to recognize potentially dangerous situations via sound information. Especially, the effort is to detect two types of acoustic events such as gun shot and breaking glass. These sounds represent abnormal behavior and they point to existence of some dangerous situation.

Acoustic signals have information redundancy and for this reason is necessary to specify effective feature extraction methods. The effective feature extraction should highlight the relevant information and reduce the number of input data by removing irrelevant information using various kinds of decorrelation methods.

One of the most popular approaches is based on the feature extraction methods that primary for speech signals were developed. The well known feature extraction methods such as MFCC (Mel-Frequency Cepstral Coefficients) [6], [7], or PLP (Perceptual Linear Prediction) coefficients [7] are used to extract the relevant information from the input speech signal.

Other effective approach exploits the low-level audio descriptors [8], [9], defined in MPEG-7 standard. MPEG-7 is focused on the describing of the multimedia content. It is oriented on the indexing, searching and retrieval of audio using the 17 low-level descriptors [10]. These descriptors can capture the nature of input acoustic signal.

This paper is focused on the feature extraction method that include the selected set of MPEG-7 low-level descriptors such as Audio Spectrum Spread (ASS), Audio Spectrum Centroid (ASC) and Audio Spectrum Flatness (ASF) and the advantages of speech parametrization like MFCC. In the classification stage Gaussian Mixture Models (GMMs) are used.

The rest of this paper has following structure. Section 2 describes the feature extraction methodology and Section 3 gives information about the proposed detection system. Section 4 includes experiments and results, finally the conclusion and future work proposal follows in Section 5.

2 Feature Extraction Methodology

As was mentioned in previous section the feature extraction method is very important part of the detection system. In general, a typical pattern recognition task can be divided into the feature extraction and classification task. The efficient feature extraction is a very important phase of overall process, because the recognition performance directly depends on the quality of extracted feature vectors.

In this paper we investigate the discriminative power of selected MPEG-7 descriptors in comparison of conventional speech parametrization MFCC.

2.1 MPEG-7 Low-Level Descriptors

MPEG-7 is an ISO/IEC standard developed in by the Moving Picture Experts Group (MPEG). It became an international standard in September 2001 [10] and includes the part dealing with audio information. This part of standard is MPEG-7 Audio. There are defined 17 low-level descriptors.

In our experiments basic spectral descriptors namely Audio Spectrum Centroid, Audio Spectrum Spread and Audio Spectrum Flatness were chosen according to the good results that were presented in the works [2], [8], [11].

Audio Spectrum Centroid - ASC

The Audio Spectrum Centroid (ASC) [10] gives the centre of gravity of a log-frequency power spectrum. All power coefficients below 62.5 Hz are summed and represented by a single coefficient, in order to prevent a non-zero DC component and /or very low-frequency components which can have a disproportionate weight. For a given frame of signal, ASC descriptor is computed from the modified power coefficients and their frequencies. In the Eq. (1) $P'(k')$ represents the power spectrum and $f'(k')$ represent corresponding frequencies.

$$ASC = \frac{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} log_2\left(\frac{f'(k')}{1000}\right) P'(k')}{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} P'(k')}. \quad (1)$$

ASC descriptor gives information on the shape of the power spectrum. It indicates whether in a power spectrum are dominated by low or high frequencies and can be regarded as an approximation of the perceptual sharpness of the signal.

Audio Spectrum Spread - ASS

The Audio Spectrum Spread (ASS) [10] is also called instantaneous bandwidth. It is measure of the spectral shape. In MPEG-7, it is defined as the second central moment of the log-frequency spectrum. For a given signal frame ASS is computed following way:

$$ASS = \frac{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} \left[log_2\left(\frac{f'(k')}{1000}\right) - ASC\right]^2 P'(k')}{\sum_{k'=0}^{(N_{FT}/2)-K_{low}} P'(k')}. \quad (2)$$

ASS descriptor is extracted by taking the root-mean-square (RMS) deviation of the spectrum from its centroid ASC. The ASS gives indications about how

the spectrum is distributed around its centroid. A low ASS value means that the spectrum may be concentrated around the centroid, whereas a high value reflects a distribution of power across a wider range of frequencies.

Audio Spectrum Flatness - ASF

The Audio Spectrum Flatness (ASF) [10] reflects the flatness properties of the power spectrum. More precisely, for a given signal frame, it consists of a series of values, each one expressing the deviation of the signal's power spectrum from a flat shape inside a predefined frequency band. In MPEG-7, the power coefficients are computed from non-overlapping frames where the spectrum B is divided into $1/4$ octave resolution logarithmically spaced overlapping frequency bands. For each band b, the spectral flatness descriptor is estimated as the ratio between the geometric mean and the arithmetic mean of the spectral power coefficients within this band:

$$ASF(b) = \frac{\sqrt[hiK'_b - loK'_b + 1]{\Pi_{k'=loK'_b}^{hiK'_b} P_g(k')}}{\frac{1}{hiK'_b - loK'_b + 1} \sum_{k'=loK'_b}^{hiK'_b} P_g(k')}, \qquad (1 \leq b \leq B). \tag{3}$$

For all bands under the edge of 1 kHz, the power coefficients are averaged in the normal way. For all bands above 1 kHz, power coefficients are grouped $P_g(k')$. The terms $hiK'b$ and $loK'b$ represent the high and low limit for band b. High values of ASF coefficients reflect noisiness, on the other hand, low values indicate a harmonic structure of the spectrum.

2.2 Mel-Frequency Cepstral Coefficients - MFCC

Ear's perception of the frequency components in the audio signal does not follow the linear scale, but rather the Mel-frequency scale, which should be understood as a linear frequency spacing below 1 kHz and logarithmic spacing above 1 kHz. So filters spaced linearly at a low frequency and logarithmic at high frequencies can be used to capture the phonetically important characteristics. The relation between the Mel-frequency and the frequency is given by the Eq.(4):

$$Mel(f) = 2595 \times log_{10}\left(1 + \frac{f}{700}\right), \tag{4}$$

where f is frequency in Hertz. MFCC coefficients are computed following way: a segment of signal is divided into the short frames, where the parameters of the signal are constant. The Hamming window method was applied on the frames. Then, they are transformed to the frequency domain via the Discrete Fast Fourier Transform (DFFT), and then the magnitude spectrum is passed through a bank of triangular shaped filters. The energy output from each filter is then log-compressed and transformed to the cepstral domain via the Discrete Cosine Transform (DCT) [8].

3 System Overview

Our system was developed to the purpose of the gun shots and breaking glass detection. Gaussian Mixture Models (GMMs) were used as a learning algorithm. For each acoustic event GMMs were trained up to 64 Gaussian Probability Density Functions (PDFs). Three types of feature sets (MFCC_E, ASS_ASC_ASF and MFCC_E+ASS_ASC_ASF) were compared.

Three types of low-level audio descriptors were used in our work. ASS, ASC and ASF descriptors that give promising results were computed from acoustic signal. ASF generated the vector with 24 coefficients for each signal frame and a scalar value was generated by ASS and ASC. The final MPEG-7 feature vector composed of 26 (24+1+1) coefficients per frame. The extraction of these descriptors was done in Matlab.

MFCC features were computed from the signal divided into the 30 ms frames with 50% overlapping between the neighboring frames. 29 triangular band filters were used. MFCC feature extraction algorithm generated 13 coefficients (12 static coefficients and log-energy coefficient _E). Extraction of MFCC was done using HCopy from HTK toolkit.

The coefficients of mentioned feature extraction algorithms were finally jointed to the one supervector with dimension 39 (MFCC_E+ASS+ASC+ASF = 13+1+1+24). Each feature extraction approach was performed and evaluated. The audio data of acoustic events used throughout the paper were recorded in relatively quiet environment. The sound data were recorded with sampling frequency 48 kHz with resolution of 16 bits per sample. Then, they were split into the training and the testing set. Recordings were cut and manually labeled using Transcriber. It is important to note that the definition of acoustic event and sound of background was specific for this application. Each sound that was not any acoustic event was considered as a background sound.

The GMM models were trained by the 40 recordings of shot, with the 40 recordings of breaking glass and 11 minutes of background sounds. Testing recordings were used to evaluate the detection system. The total duration of testing recordings was 40 seconds. They contained non-overlapping sounds of acoustic events and background sounds.

4 Experiments and Results

The performed experiments were focused on the detection of breaking glass and gun shot. HTK toolkit was used in this task. In the first step MFCC coefficients and log energy coefficients _E were computed from acoustic signal. Then ASS, ASC, ASF descriptors were computed and added in to the one MPEG-7 vector. Finally MFCC_E + MPEG-7 vector was used to describe the input acoustic signal. For evaluating the proposed detection system the measure accuracy [%] was used. Accuracy [%] is defined as:

$$ACC\ [\%] = \frac{N - D - S - I}{N} \times 100, \tag{5}$$

Table 1. Breaking glass detection - *ACC* [%]

Num. of PDFs	MFCC_E	MPEG-7	MFCC_E+MPEG-7
1	77.78	44.44	100
2	33.33	44.44	88.89
4	22.22	44.44	100
8	55.56	66.67	100
16	33.33	66.67	100
32	11.11	55.56	100
64	11.11	77.78	33.33

Table 2. Gun shot detection - *ACC* [%]

Num. of PDFs	MFCC_E	MPEG-7	MFCC_E+MPEG-7
1	87.50	37.50	87.50
2	62.50	50.00	87.50
4	87.50	50.00	87.50
8	87.50	50.00	87.50
16	62.50	50.00	87.50
32	37.50	62.50	62.50
64	12.50	62.50	37.50

where D is the number of deletion errors, S the number of substitution errors, I the number of insertion errors, and N is the total number of labels in the reference transcription files [12]. The results are depicted on the tables.

Table 1 presents the results of the breaking glass detection. The best recognition results were obtained with MFCC_E and MPEG-7 feature extraction. Every sound of testing recording was recognized correctly in case of GMMs (1, 4, 8, 16, 32 PDFs). Table 2 shows the results of the gun shot detection. In this case, the joint set of MFCC_E and MPEG-7 enabled to achieve comparable or better results against first two parametrization approaches. For the breaking glass and gun shot detection, the higher number of PDFs brought higher values of accuracies for sets of MPEG-7 descriptors in compare of MFCC_E, where better results were occurred with the lower number of PDFs.

5 Conclusion and Future Work Proposal

Presented results give us some base information about gun shot and breaking glass detection using of selected MPEG-7 descriptors and MFCC. We can notice that the results of the breaking glass detection was better against the results of the shot detection. It was probably caused by the location of acoustic events (gun shots) in the testing recordings. They were placed immediately behind one to another. In the future, we would like to evaluate the discrimination capability of each descriptor to the particular types of acoustic events. Also, we suppose that the combination of speech feature extraction algorithm, low-level descriptors and

feature reduction method can be very effective in this detection task, because the use of several descriptors and speech based parametrizations e.g. MFCC leading to the multidimensional feature vectors (supervectors). For this reason, Principal Component Analysis (PCA) can be apply to reduce the input feature supervectors. Extending of the sound corpus is also in progress.

Acknowledgments

This work has been performed partially in the framework of the EU ICT Project INDECT (FP7 - 218086) and by the Ministry of Education of Slovak Republic under research VEGA 1/0065/10.

References

1. Huang, W., Lau, S., Tan, T., Li, L., Wyse, L.: Audio events classification using hierarchical structure. In: ICICS-PCM, pp. 1299–1303 (2003)
2. Ghulam, M., Yousef, A.A., Mansour, A., Mohammad, N.H.: Environment recognition using selected MPEG-7 audio features and Mel-Frequency Cepstral Coefficients. In: International Conference on Digital Telecommunications, pp. 11–16 (2010)
3. Cristiani, M., Bicego, M., Murino, V.: Audio-visual event recognition in surveillance video sequences. IEEE Transactions on Multimedia 9/2, 257–266 (2007)
4. Temko, A., Malkin, R., Zieger, C., Macho, D., Nadeu, C., Omologo, M.: Acoustic Event Detection and Classification in Smart-Room Environment: Evaluation of CHIL Project Systems. In: The IV Biennial Workshop on Speech Technology (2006)
5. Rougui, J.E., Istrate, D., Souidene, W.: Audio sound event identification for distress situations and context awareness. In: International Conference of Engineering in Medicine and Biology Society, Minneapolis, September 3-6, pp. 3501–3504 (2009)
6. Zheng, F., Zhang, G., Song, Z.: Comparison of different implementations of MFCC. Journal of Computer Science and Technology 16/6, 582–589 (2001)
7. Psutka, J., Müller, L., Psutka, J.V.: Comparison of MFCC and PLP parametrizations in the speaker independent continuous speech recognition task. In: Eurospeech, Aalborg, September 3-7, pp. 1813–1816 (2001)
8. Mitrovic, D., Zeppelzauer, M., Eidenberger, H.: Analysis of the data quality of audio descriptions of environmental sounds. Journal of Digital Information Management 5/2, 48–55 (2007)
9. Casey, M.: General sound classification and similarity in MPEG-7, pp. 153–164. Cambridge University Press, Cambridge (2001)
10. Kim, H.G., Moreau, N., Sikora, T.: MPEG-7 audio and beyond: Audio content indexing and retrieval, p. 304. Wiley, Chichester (2005); ISBN: 978-0-470-09334-4
11. Ntalampiras, S., Potamitis, I., Fakotakis, N.: Automatic recognition of urban environmental sounds events. In: CIP, Santorini, June 9-10, pp. 110–113 (2008)
12. Young, S., et al.: The HTK Book. Cambridge University, Cambridge (2009)

Analysis of Particular Iris Recognition Stages*

Tomasz Marciniak, Adam Dąbrowski,
Agata Chmielewska, and Agnieszka Krzykowska

Poznań University of Technology, Chair of Control and System Engineering,
Division of Signal Processing and Electronic Systems,
ul. Piotrowo 3a, 60-965 Poznań, Poland
{tomasz.marciniak,adam.dabrowski,agata.chmielewska,
agnieszka krzykowska}@put.poznan.pl

Abstract. In this paper particular stages are analyzed present in the iris recognition process. First, we shortly describe available acquisition systems and databases of iris images, which can be used for tests. Next, we concentrate on features extraction and coding with the time analysis. Results of average time of loading the image, segmentation, normalization, features encoding, and also recognition accuracy for CASIA and IrisBath databases are presented.

Keywords: iris recognition, biometric, CASIA, IrisBath.

1 Introduction

Identification techniques based on iris analysis gained popularity and scientific interest since John Daugman introduced in 1993 the first algorithm for the identification of persons based on the iris of the eye [1]. Then many other researchers presented new solutions in this area [2-10].

The iris is an element, which already arose in early phase of the human life and remains unchanged for a long part of life. Construction of the iris is independent of the genetic relationships and each person in the world, even the twins possess different irises. However, there are also many problems to be faced when encoding the iris such as a change of opening angle in the pupil depending on lighting conditions, covering a portion of its regions by the eyelids and eyelashes, or the rotation of the iris due to the inclination of the head or eye movement.

2 Acquisition of Iris Image and Iris Databases

There are several conditions that must be met to perform a correct image of the iris, which could serve to identify the person.

First, the iris picture has to be taken with a good resolution, as the photographed iris region is small, with a diameter of about 1 cm. The higher the resolution the more

* This paper was prepared within the INDECT project.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 198–206, 2011.

details will be visible. The method should also be not invasive or causing discomfort to the photographed person. It has to be stressed that too much light entering the eye can cause pain.

Second, the shooting area should be properly cropped in order to include an interesting part of the eye only.

Finally, the acquired image should have no or small number of reflections in the iris, as it difficult to overcome them.

Acquisition of the iris should be implemented in accordance with the standards. The application interface has to be built using ANSI INCITS 358-2002 (known as BioAPI™ v1.1) recommendations. Additionally, the iris image should comply with ISO/IEC 19794-6 norm [11].

One of the first systems for iris acquisition were developed using concepts proposed by Daugman [1] and Wildes [2]. Daugman's system performs image of the iris with a diameter of typically between 100 and 200 pixels, taking pictures from a distance of 15–46 cm, using a 330 mm lens. In the case of the Wildes proposal the iris image has a diameter of about 256 pixels and the photo is taken with a distance of 20 cm using a 80 mm lens.

Currently, several iris acquisition devices as well as whole iris recognition systems can be found on the commercial market. Because now this technology is not yet widespread in use, the prices of these devices are high and the access to purchase them is limited. As examples of the iris devices the following products can serve:

- IrisGuard IG-H100 – a manual camera for iris acquiring and verification; this camera can be held in the hand or can be mounted on a tripod or a desk, 8 pictures of the eye can be registered in time less than 3 s, the system is equipped with RS170 connector composite video (NTSC), pictures are monochrome with resolution 470 TVL
- IrisGuard IG-AD100®multi-modal capture device, this system makes it possible to combine face and iris recognition; dual iris capture takes below 4 seconds (normal conditions), a person eyes can be acquired anywhere within a range from 21 cm to 37 cm away from the unit and be perfectly authenticated within the range
- OKI IrisPass-M – this camera is designed for mounting on the wall to control access to secured areas or can be used in border control, or to register travelers [12]; it takes a picture of the iris at the time of 1 s or less (depending on lighting conditions), identifies the iris at the time of 1 s or less (depending on the configuration of the PC as a controller and the size of the iris database); false acceptance rate (FAR): 1/(1.2 M); shooting from a distance of 30–60 cm
- Panasonic BM-ET330 – designed for the access control to buildings, mounted on the wall [13]; iris recognition time of about 1 s, acquisition from a distance of 30–40 cm, the camera field of view: 115 degrees horizontal, 85 degrees vertical; the maximum number of irises in the database is 1000 and FAR 1/(1.2 M).

Experimental studies were carried out using the databases containing photos of irises prepared by scientific institutions dealing with this issue. Two publicly available databases were used during our experiments, as shown in Section 4. The first database

was CASIA [14], coming from the Chinese Academy of Sciences, Institute of Automation, while the second IrisBath [15] was developed at the University of Bath. We have obtained also access to UBIRIS v.2.0 database [16] and the database prepared by Michael Dobeš and Libor Machala [17], which are currently undergoing testing by the contractors from WP7.

CASIA database is presented in three versions. All present photographs were taken in the near infrared. We used the first and third version of this database in our experimental research. Version 1.0 contains 756 iris images with dimensions 320×280 pixels carried out on 108 different eyes. The pictures in CASIA database were taken using the specialized camera and saved in BMP format. For each eye 7 photos were made, 3 in the first session and 4 in the second. Pupil area was uniformly covered with a dark color, thus eliminating the reflections occurring during the acquisition process.

The third version of CASIA database contains more than 22 000 images from more than 700 different objects. It consists of three sets of data in JPG 8-bit format. Section of CASI-IrisV3-Lamp contains photographs taken at the turned-on and off lamp close to the light source to vary the lighting conditions, while the CASIA-IrisV3 Twins includes images of irises of hundred pairs of twins.

Lately a new version of CASIA database has been created the CASIA-IrisV4. It is an extension of CASIA-IrisV3 and contains six subsets. Three subsets from CASIA-IrisV3 are: CASIA-Iris-Interval, CASIA-Iris-Lamp, and CASIA-Iris-Twins. Three new subsets are: CASIA-Iris-Distance, CASIA-Iris-Thousand, and CASIA-Iris-Syn.

CASIA-Iris-Distance contains iris images captured using self-developed long-range multi-modal biometric image acquisition and recognition system. The advanced biometric sensor can recognize users from 3 meters away. CASIA-Iris-Thousand contains 20 000 iris images from 1 000 subjects. CASIA-Iris-Syn contains 10 000 synthesized iris images of 1 000 classes. The iris textures of these images are synthesized automatically from a subset of CASIA-IrisV1.

CASIA-IrisV4 contains a total of 54 607 iris images from more than 1 800 genuine subjects and 1 000 virtual subjects. All iris images are 8 bit gray-level JPEG files, collected under near infrared illumination.

IrisBath database is created by a Signal and Image Processing Group (SIPG) at the University of Bath in the UK [15]. The project aimed to bring together 20 high resolution images from 800 objects. Most of the photos show the iris of students from over one hundred countries, who form a representative group. Photos were performed with resolution of 1280×960 pixels in 8-bit BMP, using a system with camera LightWise ISG. There are thousands of free of charge images that have been compressed into JPEG2000 format with a resolution of 0.5 bit per pixel.

3 Features Extraction and Coding

We can identify three successive phases in the process of creating the iris code [18]. They are determined respectively as: segmentation, normalization and features encoding as shown in Fig. 1.

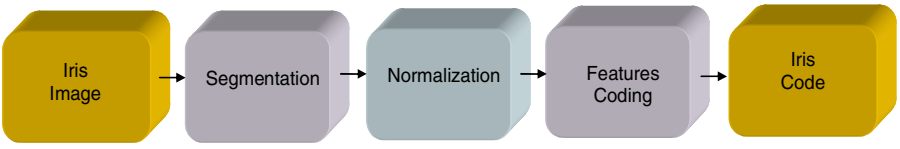


Fig. 1. Stages of creation of iris codes

3.1 Segmentation Process

Separation of the iris from the whole eye area is realized during the segmentation phase. At this stage it is crucial to determine the position of the upper and lower eyelids, as well as the exclusion of areas covered by the lashes. In addition, attention should be paid to the elimination of regions caused by light reflections from the cornea of the eye.

The first technique of iris location was proposed by the precursor in the field of iris recognition i.e. by John G. Daugman [1]. This technique uses the so-called integro-differential operator, which acts directly on the image of the iris, seeking the maximum normalized standard circle along the path, a partial derivative of blurred image relating to the increase of a circle radius. The current operator behaves like a circular edge detector in the picture, acting in the three-dimensional parameter space (x, y, r), i.e. the center coordinates and radius of the circle are looked for, which determine the edge of the iris. The algorithm first detects the outer edge of the iris, and then, limited to the area of the detected iris, is looking to get its inside edge. Using the same operator, but by changing the contour of the arc path, we can also look for the edges of the eyelids, which may in part overlap the photographed iris.

Another technique was proposed by R.P. Wildes [2]. In this case also the best fit circle is looked for but the difference (comparing to the Daugman method) consists in a way of searching the parameter space. Iris localization process takes place in this case in two stages. First, the image edge map is created then each detected edge point gives a vote to the respective values in the parameter space looking for the pattern. The edge map is created based on the gradient method. It relies on the assignment of a scalar bitmap vector field, defining the direction and strength increase in the pixel brightness. Then, the highest points of the gradient, which determine the edges, are left with an appropriately chosen threshold. The voting process is performed at the designated edge map using the Hough transform.

In our experimental program [19] we also used the Hough transform, and to designate the edge map we used a modified Kovesi algorithm [20] based on Canny edge detector. An illustration of the segmentation process with the time analysis is presented in Fig. 2.

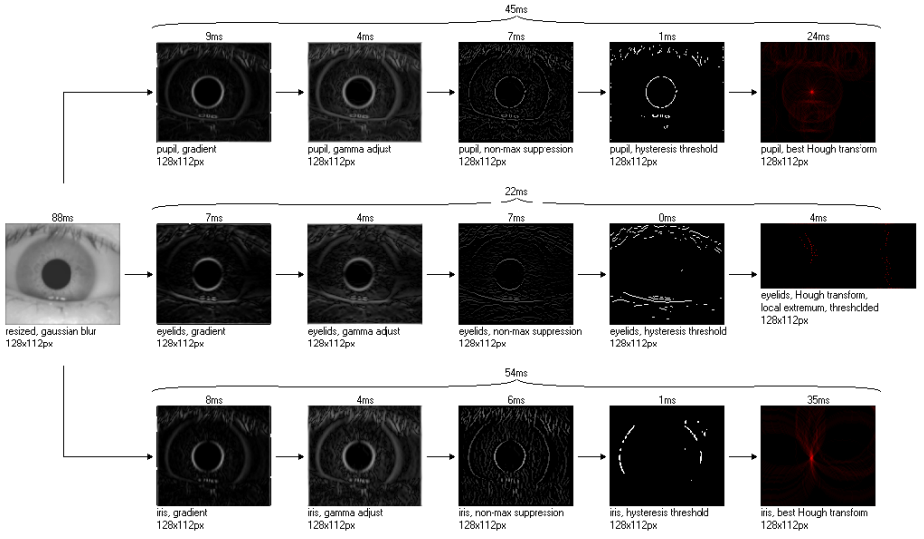


Fig. 2. Example time-consuming analysis of segmentation process

3.2 Normalization

The main aim of the normalization step is the transformation of the localized iris to a defined format in order to allow comparisons with other iris codes. This operation requires consideration of the specific characteristics of the iris like a variable pupil opening and non coordinated pupil and iris center points. A possibility of circulation of the iris by tilting the head or as a result of the eye movement in the orbit should be noticed.

Having successfully located the image area occupied by the iris, the normalization process has to ensure that the same areas at different iris images are represented in the same scale in the same place of the created code. Only with equal representations the comparing two iris codes can be correctly justified. Daugman suggested a standard transformation from Cartesian coordinates to the ring in this phase. This transformation eliminates the problem of the non-central position of the pupil relatively to the iris as well as the pupil opening variations with different lighting conditions. For further processing, points contained in the vicinity of a 90 and 270 degrees (i.e., at the top and at the bottom of the iris) can be omitted, This reduces errors caused by the presence of the eyelids and eyelashes in the iris area.

Poursaberi [21] proposed to normalize just only half of the iris (close to the pupil), thus by passing the problem of the eyelids and eyelashes. Pereira [22] showed, in the experiment, in which the iris region was divided into ten rings of equal width, that the potentially better decision can be made with only a part of the rings, namely those numbered as 2, 3, 4, 5, 7, with the ring numbered as the first one, being the closest to the pupil.

During our tests, the Daugman proposal and the model based on its implementation by Libor Masek in Matlab [18] was used in the step normalization stage. At the same time we can select an area of the iris, which is subject to normalization, using both angular distribution and the distribution along the radius. The division consists in determining the angular range of orientation, which is made with the normalization of the iris. This range is defined in two intervals: the first includes angles from –90 to 90 degrees and the second – angles from 90 to –90 degrees (i.e., angles opposite to clockwise).

3.3 Features Coding

The last stage of the feature extraction, which encode the characteristics, aims to extract the normalized iris distinctive features of the individual and to transform them into a binary code. In order to extract individual characteristics of the normalized iris various types of filtering can be applied. Daugman coded each point of the iris with two bits using two-dimensional Gabor filters and quadrature quantization.

Field suggested using a variety logarithmic Gabor filters, the so called Log-Gabor filters [23]. These filters have certain advantages over and above conventional Gabor filters, namely, by definition, they do not possess a DC component, which may occur in the real part of Gabor filters. Another advantage of the logarithmic variation is that it exposes high frequencies over low-frequencies. This mechanism approaches the nature of these filters to a typical frequency distribution in real images. Due to this feature the logarithmic Gabor filters better expose information contained in the image.

4 Time Analysis

During our research we used the program IrisCode_TK2007 [19]. A multi-processing was used in order to automatically create iris codes for multiple files. The study involved two databases described in Section 2, namely CASIA and IrisBath.

Test results are presented in Table 1. Section “Information” includes the total number of files and the number of classes of irises. Section “Results” contains the results of the processed images. These are average times of individual stages and the total processing time for all files. Figure 3 shows the times of individual stages, expressed in percentage of the overall time for all tested databases (processed with Intel Core i7 CPU; 2,93 GHz).

Our program contains also an option “Multithreading”, which a enables multithreaded processing on multiprocessor machines. Figure 4 presents the comparison of the processing times of various stages, when the option “Multithreading” was used or not (processed on Intel Core i7 CPU; 2,93 GHz) for IrisBath database. The total processing time for one processor was about 17 minutes. while for two processors was about 9 minutes.

Table 1. Processing times for two bases: IrisBath and CASIA

	Data base name	CASIA					IrisBath	All databases
		Iris Image Database (version 1.0)	IrisV3			Iris V4 Distance		
			Interval	Lamp	Twins			
Information	**Number of classes of irises**	217	498	822	400	142	22	2101
	Total number of files	756	2655	16213	3183	2572	432	25811
Results (times)	**The average time of image loading [ms]**	87	86	278	283	3956	1120	968
	The average time of segmentation [ms]	103	107	320	306	4404	1249	1081
	The average time of normalization [ms]	2	2	2	2	2	2	2
	The average time of features encoding [ms]	1	1	1	1	2	1	1
	The average total time coding [ms]	193	196	601	592	8364	2372	2053
	The average time of writing results on disc [ms]	6	5	6	6	8	5	6
	Total time database processing	**00:02:30**	**00:08:55**	**02:43:57**	**00:31:46**	**06:10:15**	**00:17:10**	**09:54:28**

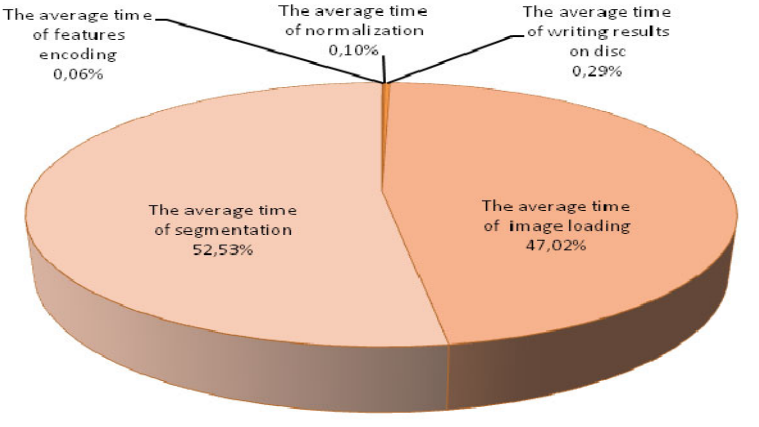


Fig. 3. Participation of individual stages, expressed in percentage, for all tested databases

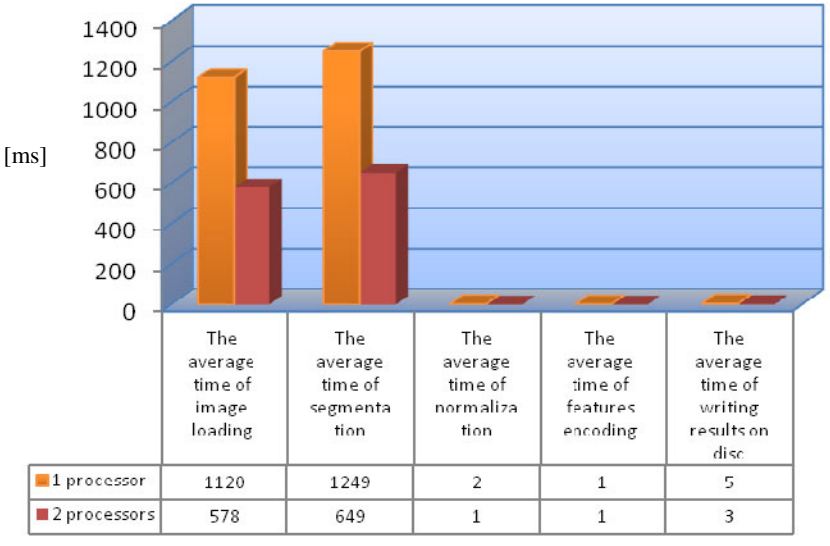


Fig. 4. Comparison of processing times [ms] of various stages (processed with Intel Core i7 CPU; 2,93GHz) for IrisBath database (time precision 1 ms)

5 Conclusions

The most important issue in the biometric identification process is recognition accuracy. The best result we received using the IrisBath database by means of the log-Gabor1D filter. The following results were obtained: FAR=0.351% (*false acceptance rate*) and FRR=0.572% (*false rejection rate*), which result in the overall factor of the iris verification correctness, equal to 99.5%. For the CASIA database v.1.0 the best result was obtained with the code size of 360×40 bits and the following results were obtained: FAR = 3.25%, FRR = 3.03%, and the ratio of correct verification of iris codes at the level of 97% [24].

It can be observed that the time of calculations is so short that the proposed iris recognition system can operate in real-time. However, an effective acquisition of the iris image remains a problem.

References

[1] Daugman, J.G.: High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. Pattern Anal.Mach. Intell. 15(11), 1148–1161 (1993)

[2] Wildes, R.P.: Iris recognition: An emerging biometric technology. Proc. IEEE 85(9), 1348–1363 (1997)

[3] Boles, W.W., Boashash, B.: A human identification technique using images of the iris and wavelet transform. IEEE Trans. Signal Process 46(4), 1185–1188 (1998)

[4] Ma, L., Tan, T., Wang, Y., Zhang, D.: Personal identification based on iris texture analysis. IEEE Trans. Pattern Anal. Mach. Intell. 25(12), 1519–1533 (2003)

[5] Ma, L., Tan, T., Wang, Y., Zhang, D.: Efficient iris recognition by characterizing key local variations. IEEE Trans. Image Process 13(6), 739–750 (2004)
[6] Sanchez-Avila, C., Sanchez-Reillo, R.: Two different approaches for iris recognition using Gabor filters and multiscale zero-crossing representation. Pattern Recognit. 38(2), 231–240 (2005)
[7] Vatsa, M., Singh, R., Noore, A.: Reducing the false rejection rate of iris recognition using textural and topological features. Int. J. Signal Process 2(1), 66–72 (2005)
[8] Yu, L., Zhang, D., Wang, K., Yang, W.: Coarse iris classification using box-counting to estimate fractal dimensions. Pattern Recognit. 38(11), 1791–1798 (2005)
[9] Monro, D.M., Rakshit, S., Zhang, D.: DCT-based iris recognition. IEEE Trans. Pattern Anal. Mach. Intell. 29(4), 586–596 (2007)
[10] Poursaberi, A., Araabi, B.N.: Iris recognition for partially occluded images: Methodology and sensitivity analysis. EURASIP J. Adv. Signal Process 2007(1) Article ID 36751, 20 (2007)
[11] Biometric data interchange formats – Part 6: Iris image data, ISO/IEC 19794-6 (2005)
[12] Oki IRISPASS®-M, http://www.oki.com/en/press/2005/z05049e-2.html
[13] Panasonic Iris Reader BM-ET330, http://panasonic.co.jp/pss/bmet330/en/
[14] Chinese Academy of Sciences' Institute of Automation, "CASIA-IrisV3", http://www.cbsr.ia.ac.cn/IrisDatabase.htm and Biometric Ideal Test website http://biometrics.idealtest.org/
[15] Signal and Image Processing Group (SIPG), University of Bath Iris Image Database, http://www.bath.ac.uk/elec-eng/research/sipg/irisweb/
[16] Proença, H., et al.: The UBIRIS.v2: A Database of Visible Wavelength Iris Images Captured On-The-Move and At-ADistance. IEEE Transactions on Pattern Analysis and Machine Intelligence, Digital Object Identifier 10.1016/j.imavis.2009.03.003 (2009)
[17] Dobeš, M., Machala, L.: Iris Database, http://www.inf.upol.cz/iris/
[18] Masek, L.: Recognition of Human Iris Patterns for Biometric Identification, M. Thesis, The University of Western Australia (2003)
[19] Kaminski, T.: Implementacja i analiza skuteczności identyfikacji osób na podstawie tęczówki (Implementation and efficiency analysis of person identification using iris recognition), M.Sc. Thesis, Supervisor: Tomasz Marciniak, Poznan University of Technology (2007)
[20] Kovesi, P.: Some of my MATLAB functions, http://www.csse.uwa.edu.au/~pk/
[21] Poursaberi, A., Araabi, B.N.: A Half-Eye Wavelet Based Method for Iris Recognition. In: Proceedings of the ISDA (2005)
[22] Pereira, M.B., Paschoarelli Veiga, A.C.: A method for improving the reliability of an iris recognition system, Department of Electrical Engineering – Federal University of Uberlandia(UFU) – Brazil (2005)
[23] Field, D.: Relations between the statistics of natural images and the response properties of cortical cells. Journal of the Optical Society of America (1987)
[24] Dąbrowski, A., et al.: D7.3 – Biometric features analysis component based on video and image information. INDECT Project FP7-218086 (2010)

Analysis of Malware Network Activity

Gilles Berger-Sabbatel and Andrzej Duda

Grenoble Institute of Technology, CNRS Grenoble Informatics Laboratory UMR 5217
681, rue de la Passerelle, BP 72
38402 Saint Martin d'Hères Cedex, France
Gilles.Berger-Sabbatel@imag.fr, Andrzej.Duda@imag.fr

Abstract. A botnet is a network of zombie computers compromised by some malware (virus, worm). Botnets are coordinated by a botmaster through a command and control channel (C&C) to which the malware connects to get instructions. A botmaster can use botnets to perform malicious activities. In this paper, we report on the development of a platform for analyzing malware and botnets.

1 Introduction

Our goal is to provide a platform with tools for capturing malware, running botnets in a controlled environment, analyzing their interactions with a botmaster, testing methods and techniques for mitigating botnet nuisance and eventually disrupting them.

2 Functional Architecture

As depicted in Figure 1, our platform is composed of honeypots, an online analysis environment, a Command and Control monitoring software, and an offline analysis environment.

2.1 Capture of Malware

We use Dionaea[1] to capture malware, a popular free software implementation of a low interaction honeypot running on Linux. Dionaea emulates a Windows system with known vulnerabilities. When an exploit of an emulated vulnerability is attempted, the emulation proceeds until the malware is downloaded by the honeypot. The downloaded binary is then stored in a file with a MD5 digest on its content as a file name to avoid storing multiple instances of the same binary malware. Events recorded by Dionaea are stored in an Sqlite database, which allows further statistical processing.

Different files does not mean a different malware, as most malware introduces a slight modification in its code when it propagates itself in an attempt to escape

[1] http://dionaea.carnivore.it/

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 207–215, 2011.

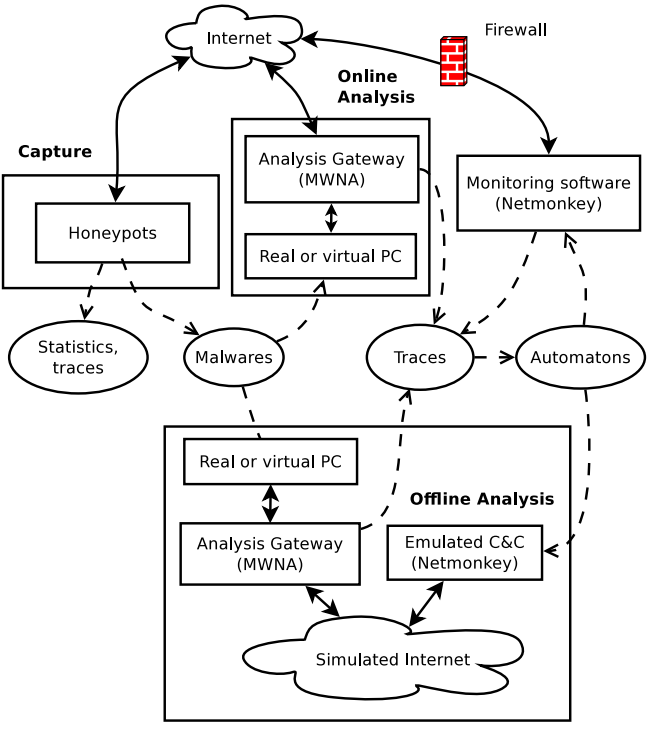


Fig. 1. Functional architecture of the platform

from the detection by an anti-virus. Hence, we use ClamAV[2], a free anti-virus Linux software to classify the capture malware. This is not very efficient, because almost 1/3 of malware is not even identified by ClamAV, and the same malware may correspond to different viral Clamav signatures. Even worse, the same signature may correspond to malware having different network activities.

As a result, we have currently captured about 1600 malware samples classified in 450 classes, while 700 samples are not classified at all. Hence, further work should focus on better classification methods.

After malware classification, we submit at least one sample of each class to three online sandboxed analysis tools: Norman SandBox Online Analyzer [3], CWSandbox analysis system[4], and Anubis[5]. These systems operate by tracing the execution of the processes in a real or emulated Windows environment, however they can be defeated by malware that contains anti-emulation or anti-tracing code.

[2] http://www.clamav.net/lang/en/
[3] Norman Company, http://www.norman.com/security_center/security_tools/
[4] University of Mannheim, https://mwanalysis.org/
[5] International Secure Systems Lab, http://anubis.iseclab.org/

2.2 Online Analysis

Then, we want to analyze network activities of a running malware code to determine if and how it connects to the C&C node (Command and Control), and thus, if it is actually enrolling in a botnet, and what kind of network activities it performs. We usually analyze one sample of each malware class unless we notice some important differences in the sizes, or dates of capture of samples in the same class.

The malware sample is then installed on a PC that acts as a victim machine. It runs an old version of Windows XP and connects to the Internet through a smart gateway that monitors all network traffic to detect and block harmful activities. We present the gateway in more details in Section 3. The Internet connection should not be filtered, but for the sake of security, a few restrictions are applied. We use virtual machines under Linux as victim PCs to speed up the cycle of experiments: the analysis of a sample takes only two minutes in the simplest cases.

2.3 Monitoring C&C

Once we have identified malware, we want to monitor the C&C traffic to detect some malicious activities. We could let the malware run and monitor its traffic, but it would require a dedicated machine for each monitored botnet and permanent monitoring by an operator to detect and remedy any unexpected harmful behavior.

To avoid these issues, we have developed Netmonkey, a tool for monitoring botnet C&Cs by reproducing their protocol through an automaton-based program. Netmonkey accepts rules generated from traces of C&C activities and is able to connect to the C&C and reproduce the behavior of the actual malware without presenting significant security risks. Interactions are recorded and unidentified interactions are notified to the operator, so that they can be replayed against the actual malware to determine their effect.

We have used Netmonkey to monitor the C&C of a botnet for a few weeks. It works fine for IRC-based protocols and should correctly work for other text-based unencrypted protocols. In most cases, monitoring of the C&C can be performed on a network protected by a firewall, as it only needs to connect to the C&C.

However, encrypted C&C communications may lead us to let the actual malware directly communicate the real C&C. In this way, even if we are not able to determine the meaning of interactions, we can detect and observe their effects. In this case, we have to carefully select the interactions that we can relay between the malware and the Internet.

2.4 Offline Analysis

Offline analysis can be used to replay unexpected interactions captured from the C&C against the real malware to determine its reaction. The environment is completely isolated, the C&C is simulated with Netmonkey, and a set of honeypots

emulate a real network environment. The result can then be used to augment the set of rules used by Netmonkey to monitor the real C&C.

3 Online Malware Analysis

The online malware analysis relies on a smart gateway allowing to intercept, monitor, and filter traffic. Its role is to present useful high level information to the user and filter out non significant transmissions, as well as detect and stop malicious activities. It is based on MWNA (MalWare Network Analysis), a tool first developed in C and then rewritten and extended in C++.

3.1 Principles of the Gateway

MWNA use the Linux packet filter mechanism that allows a program runing in the user space to intercept and process network packets flowing through the kernel. Processed packets can then be accepted (their regular processing by the protocol stack is resumed), dropped, and even modified. Packets are intercepted at the Ethernet level, so that all protocol layers have to be processed by MWNA. Rules can be specified to determine packet processing according to packet source or destination (IP address, port), or the fact that the address has been resolved or not.

In its current state, the following features are implemented in MWNA :

- Processing DNS packets of type A (queries and replies), so that a list of known (resolved) IP addresses is maintained.
- Tracking TCP connections and notification of opened connections.
- Identification and basic decoding of the IRC and HTTP protocols, even when not run on a standard port number (including ports reserved for other protocols such as).
- Detection of scans and denial of service attacks.

3.2 Malicious Activities and Their Prevention

We need to prevent analyzed malware from performing malicious activities, such as port scanning, Denial of Service attacks (DoS), sending spams, sending local private data (spyware), and click fraud. We have focused on the prevention of DoS and scans. Sending spam can be prevented by blocking connections to the port 25 (SMTP), or redirecting then to a local program accepting mails without transmitting them. Spyware activities would cause no harm, as there is no sensitive data on the victim systems: we are only interested in detecting them if they occurs. Click fraud is marginal.

Scanning. Almost all active malware scan the Internet to find vulnerable hosts for replication. Scanning must be stopped to prevent replication and a waste of network and computer resources. On the other hand, we want to observe a scan attempt. A scan is a repetition of connection attempts to the same port

on multiple targets. In the current implementation, scans on TCP ports are detected when the number of unanswered connection attempts to a port on a given interval of time exceeds some threshold value. The parameters of the detection (time interval and number of connection attempts) can be configured globally, or for specific port (such as 139 and 445, which correspond to services on Windows systems, and are the most frequent targets of scans). When a scan is detected, the destination port is recorded in a blacklist, and subsequent packets sent to the same port are dropped. This solution detected every scans on 139 and 445 ports. We have observed however some false positives due to repeated connection attempts to HTTP servers. Such false positives should be avoided, because they prevent the malware from regular operation.

Denial of Service. Denial of Service (DoS) are attacks targeted at a specific host. They can be defined as a series of connection attempts to the same port of the target host. Here again, detection is based on a threshold value of the number of connection attempts to the destination in a given interval of time. When a DoS is detected, the IP address of the destination is recorded in the blacklist and subsequent packets are dropped.

We have tested the method locally with well known DoS scenarios. As far as we can tell, we did not observe actual DoS attacks. However, we have observed false positives due to repeated connection attempts when malware tries to connect to the C&C. Such false positive should have little impact on the analysis, as they usually concern botnets whose C&C servers have been disrupted.

4 Results

We have analyzed about 460 samples of malware. 80 samples are active and participate in a botnet, but we can find several different samples participating to the same botnet. Other samples are inactive mostly because they fail to connect to their C&C: this is the case for most malware older that 6 months.

For 158 samples, including recently captured ones, there is no network activity. This is the case for all samples of the Palevo worm and almost all samples of the Allaple worm. This seems strange, as malware samples captured from the Internet should obviously present network activity and in some cases, such activity is reported by one or several of the sandbox analyzers to which the malware was submitted. We can conjecture that the malware lacks some functionalities from the system or the environment, or expect some event.

A few malware do not run at all, either because they do not find a library entry-point, or because they are not valid Win32 executable. In the last cases, either there was an error in the downloading of the malware, or the binary needed some postprocessing that the honeypot did not perform.

We have detected network activities in 302 samples, while Norman, MWanalysis, and Anubis detected respectively 39, 129, and 145 samples with network activities.

4.1 Life and Death of a Botnet

Previously, we presented statistics about a malware labeled "Trojan.SdBot-4763" by `clamav` [4]. Once started, the malware tries to connect to `botz.noretards.com`, port 65146 that appears to be an IRC server. This address corresponds to a pool of several machines, most probably compromised ones. Figure 2 plots the number of IP addresses corresponding to `botz.noretards.com` day by day on a period of 419 days ending on September 17th 2010. As we can see, from day 0 to day 206, the number of machines is mainly between 6 and 9. From day 207 to day 419, there are only two machines in the pool, but they seem to stay there for a much longer time and there are very few variations in the number of machines.

After September 17th 2010, the DNS returns "Host not found" when queried for `botz.noretards.com`. Hence, we considered this botnet as dead.

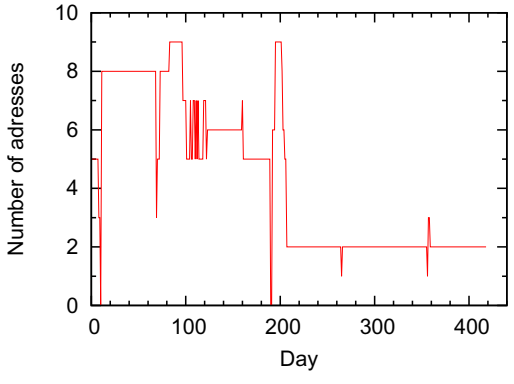


Fig. 2. Number of machines resolving to `botz.noretards.com`

Figure 3 presents the cumulative distribution function of the number of days each address remains in the pool. A majority of machines stays in the pool less than 20 days, but there is a long tail of machines that stays for a much longer time, up to 217 days.

Although we did several sessions of observation of this botnet, we never observed malicious activities other than scanning of ports 135 and 445. We used Netmonkey to monitor the C&C for a few weeks, and noticed a few unknown commands which could trigger other operations of the malware, but we had no mean by this time to replay them against the malware.

4.2 Monitored C&C Protocols

Most C&Cs we could observe are still based on IRC. The main advantage of IRC is probably that it can be supported by a pool of servers in a transparent way, allowing to easily build a robust and evolutive C&C. Some IRC-based C&C use what could be encrypted strings to transmit commands, but there is no

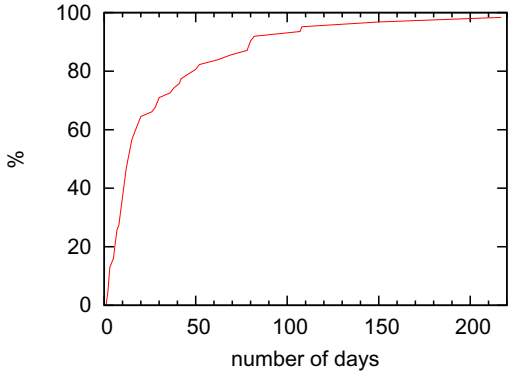


Fig. 3. Cumulative distribution function of the number of days each machine remains in the pool of compromised machines

evidence that strong cryptographic algorithm are used, as this would imply a complex protocol for the exchange of session keys.

Some botnets use another textual protocol, which seems to be a variant of IRC, with some keywords changed probably in an attempt to escape from detection.

We have also noticed the use of an unknown non-textual protocol. For a number of malware, we did not manage to clearly identify the C&C. Some of them may simply use HTTP to this end, as we observed malware connecting to HTTP servers before any other connection attempt.

4.3 Network Traffic

When the C&C has been contacted, the main observed malware traffic is to access HTTP servers, often multiple ones. They often download Windows code and some other data. On HTTP servers, we mainly observed GET methods, a few POST methods, and no PUT method. In one case, we observed an intense activity of downloading data from several HTTP servers, leading us to suspect a case of click fraud, but without evidence. We also observed one case of a malware trying to send an email. In most cases, HTTP servers are accessed through the standard 80 port, but we have a fair amount of accesses to other ports (81, 88, 888...).

4.4 Retrieving the C&C

Retrieving the address of the C&C is an important issue for a botnet and determine its resilience. Machines supporting the C&C are usually compromised ones, which can be recovered at any time by their administrators, so that they should be replaced by the botmaster.

Hardcoded addresses. A few malware try to connect to a host prior to any other network activity: it seems that they use a pool of hardcoded addresses.

This would require that the list of addresses be regularly refreshed. Hence, a zombie machine that has been down for a long time may become unable to retrieve the C&C and is lost for the botnet.

Such strategy present an advantage for botnets: a sample of the malware under investigation by security specialists may become obsolete, if it is not regularly refreshed.

Domain names. The majority of botnets use one or several domain names. Such names may correspond to the legitimate names of compromised machines or they may be obtained from providers with low security concerns and the addresses returned may be adapted to the available compromised machines.

This is particularly the case of the botnet presented in Section 4.1: its C&C was only accessed through the name `botz.noretards.com` and the botnet was alive for more than 400 days. However, this kind of botnets may become vulnerable, if the registrars become more concerned with security. Furthermore, when the name has been detected by security specialists, it becomes easy to monitor, or even block accesses to the C&C.

TXT DNS records. We have observed a few cases of malware connecting to IRC servers without other prior network interaction than a DNS TXT query. The reply is an apparently random string that probably encodes IP addresses of the C&C servers. This technique could help to escape from detection, as it is difficult to establish a relation between a TXT record and the address of the C&C, even with a moderately complex encoding.

5 Related Work

Several approaches to study botnets are possible. One of them is to passively monitor connection attempts on unused network address ranges to get a statistical view of malware activity (*telescopes*). Kumar et al. used telescopes for the analysis of the Internet worm propagation [5]. Anyway, a telescope only gives an external view of botnets and does not provide a real knowledge of complex internal botnet behavior.

One way of observing botnet activity is to monitor DNS lookups [3]. The problem with such a solution in a research context is that it requires cooperation of DNS administrators at a rather high level.

We can use reverse engineering, emulation, trace of execution, or automatic code analysis to analyze captured malware. Trinius et al. proposed for example an abstract instruction set for representing significant activities of malware [6]. Such representation may be derived from a trace of system calls and allows a higher level of analysis. However, some malware are able to detect the fact that their execution is traced and work around it.

Another way would be to let malware freely execute in a controlled environment. Alata et al. proposed a method based on connection redirection to monitor Internet attacks [2]. Rajab et al. described a platform for studying botnets in a confined environment [1].

6 Conclusions

In this paper, we report on the development of a platform for analyzing malware and botnets. It provides several tools for capturing malware, running botnets in a controlled environment, analyzing their interactions with a botmaster, testing methods and techniques for mitigating botnet nuisance and eventually disrupting them. We have used the platform for capturing a large number of malware samples and analyzing their behavior. We continue the development of the platform, especially the MWNA component, we plan to integrate a high-interaction honeypot, and develop better malware classification methods.

Acknowledgments

This work was partially supported by the EC FP7 project INDECT under contract 218086.

References

1. Abu Rajab, M., Zarfoss, J., Monrose, F., Terzis, A.: A multifaceted approach to understanding the botnet phenomenon. In: IMC 2006: Proceedings of the 6th ACM SIGCOMM Conference on Internet Measurement, pp. 41–52. ACM, New York (2006)
2. Alata, E., Alberdi, I., Nicomette, V., Owezarski, P., Kaaniche, M.: Internet attacks monitoring with dynamic connection redirection mechanisms. Journal on Internet Computer Virology 7(2) (2008)
3. Anirudh Ramachandran, D.D., Feamster, N.: Revealing botnet membership using dnsbl counter-intelligence. In: U. Association, editor SRUTI 2006: 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet, pp. 49–54 (2006)
4. Berger-Sabbatel, G., Korczyński, M., Duda, A.: Architecture of a Platform for Malware Analysis and Confinement. In: Proc. MCSS 2010: Multimedia Communications, Services and Security, Cracow (2010)
5. Kumar, A., Paxson, V., Weaver, N.: Exploiting underlying structure for detailed reconstruction of an internet-scale event. In: PROC. ACM IMC (2005)
6. Trinius, P., Willems, C., Holz, T., Rieck, K.: A malware instruction set for behavior-based analysis. Technical Report 2009-007, University of Mannheim (December 2009)

Software Implementation of New Symmetric Block Cipher

Jakub Dudek, Łukasz Machowski, Łukasz Romański, and Marcin Święty

AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow
jakub@dudek.in,
lukas.berkra@gmail.com,
{lukasz.romanski,marcin.swiety}@wp.eu

Abstract. Software implementation of new symmetric cryptography block algorithms - Indect Block Cipher - based on highly nonlinear substitution boxes is presented in this article. Innovative combination of key scheme and substitution-permutation phase of this algorithm provides promising level of robustness against cryptanalysis. Also results of first performance tests are satisfactory for now. At the beginning of the paper, application requirements and programming environment are briefly described. Functional description and structure of application are presented in the following sections. Graphical interface is described in detail as well as results of performance tests are provided. At the end, future development of IBC application is considered.

Keywords: security, symmetric cryptography, block ciphers, Indect Block Cipher.

1 Introduction

Cryptographic techniques fulfill an extremely important role in secure transmission and storage of sensitive private data. Hence, constant improvement of existing cryptographic algorithms or development of new ones is crucial for many applications. What is more, such algorithms are not only expected to provide the highest possible level of security and resistance to cryptanalysis [8] (which develops with similar rapidity) but proper performance and speed as well. As these two requirements are always mutually exclusive, new techniques ought to be designed with careful attention.

One of the main research areas in Work Package 8 of INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment* [4]) project is development of new symmetric block cipher [6]. The newly designed algorithm is called *Indect Block Cipher (IBC)*. It consists of several encryption/decryption algorithms based on substitution-permutation network [1][5] with different key length and different number of rounds. Hence, it can provide different levels of security in the sense of data confidentiality. This paper presents IBC application which is a software

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 216–224, 2011.

implementation of Indect Block Cipher [7]. The latest version of the application can be downloaded from the following website: `http://www.kt.agh.edu.pl/~niemiec/IBC`.

2 Requirements and Programming Environment

Before implementation, a few requirements which should be met by Indect Block Cipher application were defined. The most important of them are listed below.

- **Configurability.** The user of the application should have the possibility to adjust settings respectively to his/her needs with relative ease.
- **Portability.** Program should be as much independent of runtime environment (underlying operating system, hardware, etc.) as possible.
- **Modular structure.** This requirement was defined to assure ease of further development and modifications.
- **Speed and performance.** Generally, this requirement is usually met in the hardware implementations dedicated to encryption/decryption function. Even software-based encryption or decryption process should not take too much time and consume a lot of computational resources.

In order to meet listed requirements, C++, as a member of object-oriented programming languages group [2], was chosen for implementation of main algorithms. Graphical User Interface (GUI) was built by means of C++/CLI (Common Language Interface) language under .NET platform.

The application uses multithreading techniques for the separation of user interface and cipher engine. This allows user to cancel encryption or decryption operation in any moment (i.e. asynchronously). Delegating data processing to a different thread provides operation independent interface that is not locked during data encryption.

It should be noted, that this was the only purpose of using multithreading techniques. In some cases the algorithm requires that data have to be processed in determined order, which rules out the possibility of process parallelization.

3 Functional Description

In this section short description of IBC encryption algorithm is presented. The main idea of this symmetric algorithm is the unique approach to key scheme. It serves as a base for special AES [3] like matrices (S-boxes [9]) creation. Each of those matrices represents a unique nonlinear transform. They can be used both as substitution and permutation transforms. The sequence of 64 key bits is needed to create each of S-boxes. There are 4 key lengths chosen for practical use.

- **128 bits.** 2 S-boxes are coded: one for substitution transform and another for permutation transform. 8 rounds of algorithm are proposed in this case.
- **192 bits.** 3 S-boxes are coded: two for substitution transform and one for permutation transform. 10 rounds of algorithm are proposed in this case.

- **320 bits.** 5 S-boxes are coded: four for substitution transform and one for permutation transform. 12 rounds of algorithm are proposed in this case.
- **576 bits.** 9 S-boxes are coded: eight for substitution transform and one for permutation transform. 14 rounds of algorithm are proposed in this case.

Every of those key lengths and suiting number of rounds ensures unique level of protection. The IBC algorithm operates on 256 bit blocks of plaintext (the plaintext is divided into 256 bit blocks). Round of the algorithm is composed of substitution and permutation transforms in this particular order. Substitution operates on 8 bit parts of block and permutation operates on the whole block. If the length of the last block of plaintext is smaller than 256 bits it is padded with zeroes and then encrypted.

4 Structure of Application and GUI

This chapter presents the structure of Indect Block Cipher application. The application consists of many blocks with different functionalities. In order to clarify concept of Indect Block Cipher the flowchart of encryption process is introduced in Figure 1. Each functionality was developed independently to ensure ease of further development and modifications. At the end, all blocks were integrated in a single application. The block structure of IBC application is presented in Figure 2. The total number of source code lines is 2999.

According to configurability requirement, mentioned earlier, graphical user interface of Indect Block Cipher assures full control over application configuration without sacrificing its intuitiveness. It is shown in the Figure 3 with example encryption process configuration and result. In the further sections each control block of the interface will be described in detail.

4.1 Cipher Configuration

Configuration of the cipher properties is the first operation that should be carried out by the user before the actual encryption. Controls that set ciphers values and operation modes are grouped in a block named “Cipher configuration".

- *Security level.* That combo-box corresponds to the configuration set determining the security level of encryption process. Short after choosing security level the values of key size and number of rounds are automatically set to its respective values. The user can choose between 4 defined security levels and one custom level that allows user to manually define the properties of the cipher. These levels are:
  - *Low* – key size set to 128 bits with 8 rounds of encryption process,
  - *Medium* – key size set to 192 bits with 10 rounds of encryption process,
  - *High* – key size set to 320 bits with 12 rounds of encryption process,
  - *Very High* – key size set to 576 bits with 14 rounds of encryption process,
  - *Custom* – key size and rounds controls are unlocked, and the user is allowed to manually change each value.

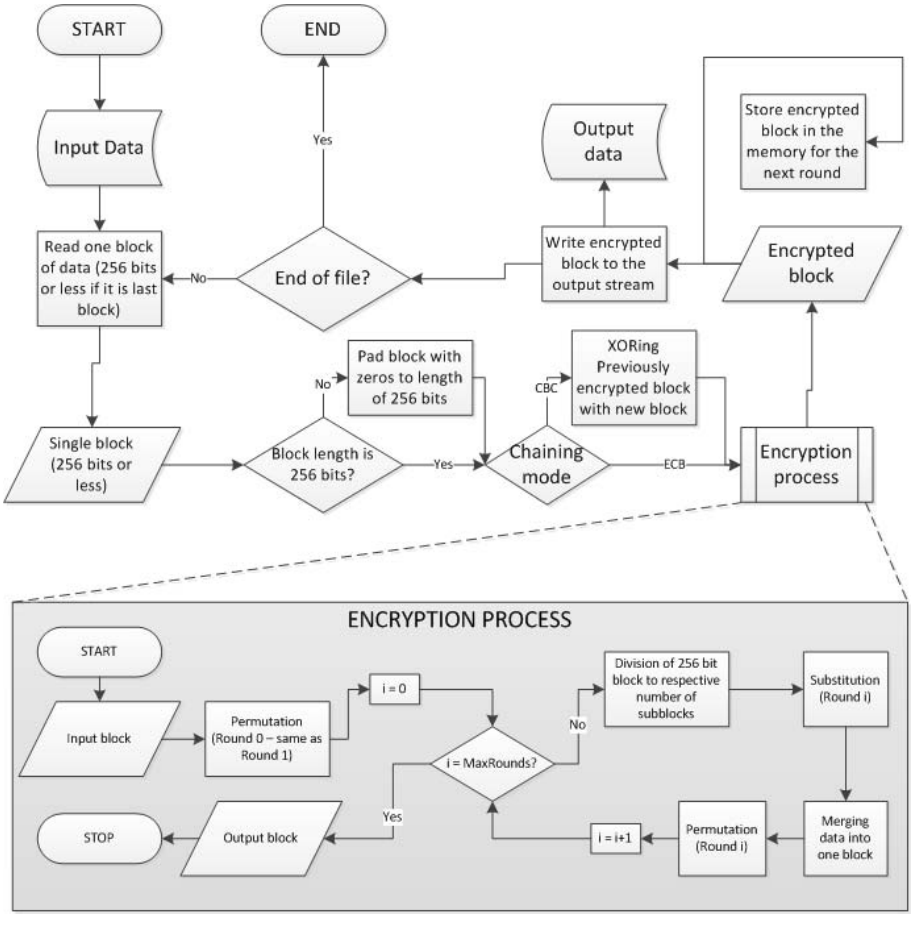


Fig. 1. Encryption process flowchart

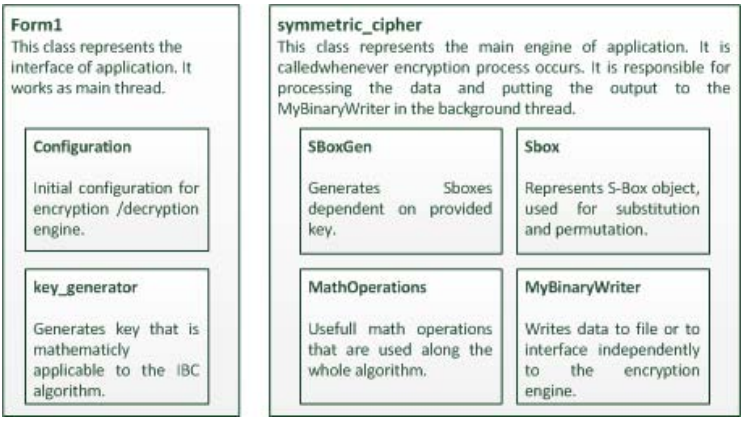


Fig. 2. The block structure of application

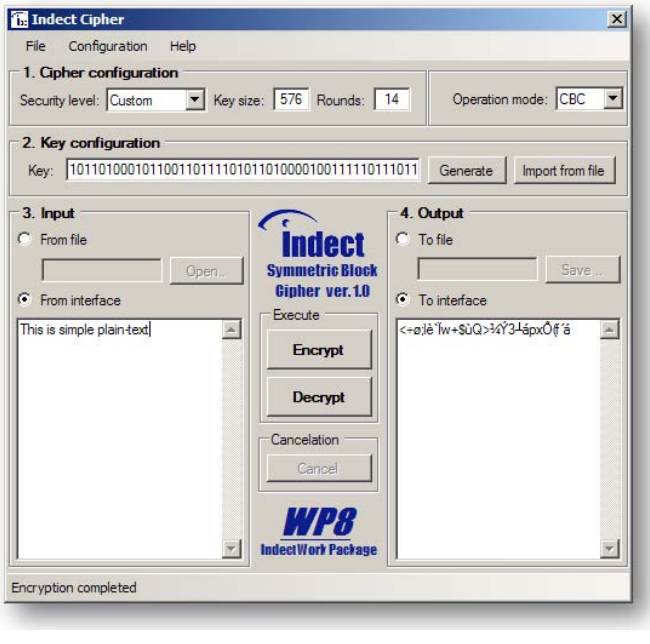


Fig. 3. IBC graphical user interface with example of encryption result

- *Key size.* Defines the size of the key in bits. Allowed values are: 128, 190, 320 and 576. This text box is automatically unlocked when the Custom security level is chosen.
- *Rounds.* Defines the number of rounds of encryption process. This text box is automatically unlocked when the Custom security level is chosen.
- *Operation mode.* Defines the method of encryption/decryption of data larger than one 256 bit block. The user can chose between the following modes.
  - *ECB* – Electronic Codebook. Each block of the data is encrypted separately and independently.
  - *CBC* – Cipher-block chaining. Provides better security but makes each block dependent of the previous one.

4.2 Key Configuration

This block consists of an input textbox that should contain a valid key in binary format. Key can be entered manually, generated by the application or imported from a text file. It contains the following controls.

- *Key* – a textbox that should contain a valid key - that is: with a correct length, corresponding to the one configured in block “cipher configuration", in binary format and mathematically applicable to the algorithm,
- *Generate* – button that invokes the generation of a valid key with length defined in block “cipher configuration",
- *Import from file* – button that allows user to choose a text file containing the key.

4.3 Input

This block represents input configuration. User is allowed to enter text manually or choose a file she/he wants to be encrypted or decrypted (Figure 3). It contains the following controls:

- *From file* – radio-button that sets container of input data to be a file,
- *File path* – text-box that should contain a valid path to file,
- *Open...* – button that allows user to choose a file,
- *From interface* – radio-button that sets the input textbox to be the data source,
- *Input* – text-box that is used to enter input data

4.4 Output

This block corresponds to the output configuration. Output can be set to a file or a text-box. It is highly recommended to use a file as an output to the encryption process, because of the encoding differences between the set of encrypted letters and those that can be printed out on the screen. Interface output is used only to show the process, but in case of real encryption a file should be used as the destination of encryption instead. It contains the following controls:

- *To file* – radio-button that sets the container of output data to be a file - highly recommended,
- *File path* – text-box that should contain a valid path to the output file,
- *Save...* – button that allows users to specify a file,
- *To interface* – radio-button that sets the textbox to be the data output,
- *Output* – text-box that is used to print output data - used only for demonstration purposes.

4.5 Execute

This block consists of two buttons that start the encryption or decryption process:

- *Encrypt* – button that when pressed starts the encryption process,
- *Decrypt* – manually starts the decryption process.

4.6 Cancellation

This is a block containing only one button and providing user a possibility to cancel the operation and control the interface again. *Cancel* button is unlocked when the encryption or decryption process occurs. By pressing it, user stops operation and goes back to the interface.

4.7 Status Bar

Status bar indicates the current state of the encryption engine. It also presents the progress of the encryption or decryption process. Possible messages are:

- *Symmetric Cipher Ready* – shown after starting the application or clearing configuration,
- *Encryption in progress* – shown when encryption is in progress,
- *Decryption in progress* – shown when decryption process occurs,
- *Encryption cancelled* – shown when encryption was interrupted by pressing cancel button,
- *Decryption cancelled* – shown when decryption process was canceled by user.

Progress bar shows percentage of completion of work, and it is only visible when actual encryption or decryption occurs.

5 Performance

Performance of certain aspects is essential for the suitability of a computer application in a real deployment. This is especially true in data processing applications. Since encryption is a special case of data processing, the performance of its realization, both software and hardware, should be always optimal, i.e. should

Table 1. Application performance

Security level	Key size	Rounds	Operation mode	Performance [Mbps]	
				Encryption	Decryption
Low	128 bit	8	ECB	2,160	2,166
			CBC	2,109	2,165
Medium	192 bit	10	ECB	1,758	1,748
			CBC	1,770	1,741
High	320 bit	12	ECB	1,477	1,470
			CBC	1,516	1,491
Very high	576 bit	14	ECB	1,274	1,291
			CBC	1,291	1,279

be as high as possible. Not only a simplification of the algorithm itself is critical, but its implementation is important as well. This includes the programming environment, used libraries, application design and programming techniques. In Table 1, some insights into the performance aspects of the designed application are presented.

Each test was taken in the following environment and configuration set:

- *Operating system:* Windows 7 Professional 64-bit
- *Source of the data:* File on local disk (about 3 MB)
- *Destination of the data:* File on local disk
- *Processor:* AMD Athlon^{TM} II X2 250 Processor 3.00 GHz
- *RAM:* 2.00 GB
- *Hard-drive:* Samsung HD252HJ, 250GB, 7200rpm, SATA2, 16MB Cache

As mentioned earlier, performance is a critical issue in the hardware implementations of encryption/decryption modules, which are solely devoted to cryptography. This software implementation was first of all created to show functionality of new ciphers, but it is possible to improve the implementation performance in the future (i.e. a migration of functions to low-level programming languages). Nevertheless, our software-based encryption/decryption process does not take too much time and can be used by end-users as it is.

6 Conclusions and Future Development

A new cryptographic application, which is a practical implementation of innovative block ciphers, was presented in the article. The structure of application, functionality and source components have been described. The graphical interface was presented in detail. Measured performance of implemented ciphers was also provided.

As far as future development of this application is considered, performance could be improved, as well as new features could be added. A key aspect is the optimization of the programming algorithms, which should allow user to encrypt files faster. To fulfill that goal, application rebuilding and migration of some of

the functions to low-level programming languages (e.g. assembly language) are required, since they are more effective in runtime. The creation of external library that could be used in other implementations (e.g. with other interface, on other operating system) is also anticipated. In the near future, this cipher is going to be used for network communication purposes. Finally, it is worth to mention that authors are going to implement IBC in hardware on FPGA devices.

Acknowledgments. This work has been performed in the framework of the EU Project INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment*)–grant agreement number: 218086.

Development of application's functionality and graphical interface have been co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme, INSIGMA project no. POIG.01.01.02-00-062/09.

References

1. Stinson, D.R.: Cryptography. Theory and Practice, 3rd edn. Chapman & Hall/CRC, Boca Raton (2006)
2. Eckel, B.: "Thinking in C++". Prentice Hall, New Jersey (2000)
3. FIPS Publication 197: The Advanced Encryption Standard (AES), U.S. DoC/NIST, November 26 (2001)
4. INDECT Project, `http://www.indect-project.eu`
5. Katz, J., Lindell, Y.: Introduction to Modern Cryptography. CRC Press, Boca Raton (2007)
6. Menezes, A.J., Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography, p. 21. CRC Press, Boca Raton (1997)
7. Niemiec, M., Machowski, Ł., Święty, M., Dudek, J., Romański, Ł.: New block ciphers (D9.13), Indect Project (December 2010)
8. Schneier, B.: Self-Study Course in Block Cipher Cryptanalysis. Cryptologia 24(1), 18–34 (2000)
9. Webster, A.F., Tavares, S.E.: On the design of S-boxes. In: Williams, H.C. (ed.) CRYPTO 1985. LNCS, vol. 218, pp. 523–534. Springer, Heidelberg (1986)

Multicriteria Metadata Mechanisms for Fast and Reliable Searching of People Using Databases with Unreliable Records*

Julian Balcerek and Paweł Pawłowski

Poznań University of Technology, Institute of Control and System Engineering,
Division of Signal Processing and Electronic Systems
Piotrowo 3a Street, 60-965 Poznań, Poland
{julian.balcerek,pawel.pawlowski}@put.poznan.pl

Abstract. In this paper an approach to multicriteria search for people is presented using data extracted from telephone calls to emergency services. In the considered application a procedure of searching the most similar object (objects) to the reference object using metadata mechanisms is presented. The proposed solution computes as much reliable result as possible using even unreliable records in the database. In the considered case the commonly used mechanism based on exact matching does not give the best result. Instead, the multicriteria metadata matching mechanisms exploiting weights, probabilities, distances, and correlations among database records are more reliable.

Keywords: database, metadata, multicriteria search, search reliability.

1 Introduction

Nowadays, search mechanisms for finding desired information have to operate on diverse types of fields, records and additionally very often on huge amount of data. Process of searching may be concentrated on exact matches or, if it is not possible, the best matches from the list of items, which only partially match.

Assume that the search mechanism looks for an object or group of objects that are the most similar to chosen object. If the object has rough or not precise description the searching mechanism must use description of the object only. Localization of objects in data set that are most similar to the query record is a similarity search [1, 4, 6, 8, 10]. An object is characterized by producing a possibly entire amount of description like: textual, quantitative (number based), qualitative (feature based), or binary (occurring or not occurring) description. Some of descriptions can be incorrect, but even in such a case, the proper object should be found. Additionally, searching process should also operate correctly for objects characterized in the past by using older values than the current data and such time change awareness mechanism should be included.

* This paper was prepared within the PPBW and INDECT projects.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 225–232, 2011.

In this paper a general multicriteria search mechanism, which collects different types of descriptions and correlations between data fields, is proposed.

2 Metadata Based Search

For querying the data based on their content, it is necessary to describe them using metadata, referred also to as the meta-information [4, 5, 7]. Metadata are generally textual descriptions. During looking for a given object a reference object is compared with other objects. Data searching mechanism gets metadata of the reference object and looks for metadata corresponding to other objects, then operates on all of them in order to reach the required results.

We have prepared the metadata based searching methods in order to develop a quite demanding application, namely the search in an emergency services telephone calls database. It can be seen as an example of a database that indexes real world events [3, 9]. Our specially prepared database should be organized in such a way that a particular conversation recording is stored together with its description as metadata. This is because the maintenance of such database by public emergency services like Police consists in searching for calls of the same person over a relatively long period of time like months or even years and a relatively large area such as the whole country. The described searching functionality may be very helpful to prevent unnecessary interventions under fake submissions or multi reporting of the same emergency case. Huge number of the recorded calls to be analyzed in a classic database of recordings of all and whole conversations hampers or even makes realization of this task impossible, not only in real-time, i.e., during a call, which typically lasts about no more than 1–2 minutes but also using resource unlimited off-line processing. It should be stressed that even in off-line processing there are strong time limits because of a demand for a quick reaction or a non-stop service with no free time slots. Using the classic database of the whole conversations it is usually not possible to find the required calls because it will be necessary to search in the entire database, thus to process a huge amount of data, but also because of a limited time of storage of the whole recordings. For example, just one year of recordings for a medium size city is counted in thousands of hours of the recordings. Thus just because of that the recordings are periodically deleted.

Analyzing all raw data in seconds or even minutes requires very large data stream and is rather unfeasible in a standard size computer center. Therefore we propose to define and store merely metadata: basic data about calls, metrics of the caller voice plus small but informative samples of his / her voice. Furthermore, even using this approach, recognition of the speaker and automated decision about relation of the present call with the already reported cases is not a trivial task [2].

In order to construct a reliable searching mechanism each call should be described using metadata stored in a specially prepared meta-database. These metadata are data about: event category, location of event, date, time, and duration of the call, telephone number, personal data about the caller (name, gender, age), but also acoustic and linguistic characteristics (i.e., caller voice samples together with samples of the acoustic background and together with the caller metrics like speed of speaking, vocabulary measure, occurrence of linguistic errors, foreign accent, stuttering,

repetition, hoarseness, speech intervals, weak logic, and additional comments). These records are partially computed automatically and partially put manually by the operator. The necessary human factor is an unreliable element of the meta-database.

3 Feature Conditions Based Search Classes

The best results occur if the search process operates on a variety of metadata and all metadata have a positive influence on the search. Regarding to the variety of types of features describing a particular call, various types of metadata can be distinguished. Accordingly to different types of metadata the respective searching strategies can also be divided into particular classes. They are analyzed below.

Some metadata values in the reference call and in the calls stored in the database should be the same. To obtain proper searching results with metadata of this type, plausibility or even exactness is a necessary condition. The exact matches of features mean that there are only two possibilities to choose: particular feature occurs or does not occur (binary match). For example, such binary match can be a prerequisite for occurrence of stuttering.

There are features, which in principle could be described quantitatively but due to difficulties in assigning a numerical value to them a quantitative, i.e., textual description is used instead (e.g., fast, medium, slow). In many cases it is possible to define similarity functions between different values of a feature. A reason of using the similarity function lies in a possibility of a mistake in the definition of the feature description. The feature description can be assigned subjectively by the telephone operator and the problem occurs if slight, difficult to distinguish differences between the feature values in two calls lead to different descriptions. As an example, the speed of speaking may be mentioned. The same speech for one operator may be felt as fast while for another one can be classified as e.g. the medium speed. Certainly, we assign high probability of match of two calls for the same speed of speaking is assigned for both but there should also be declared somehow lower (but substantially greater than zero) probability for slightly different descriptions of the speaking speed. In the contrary to the previously analyzed binary match, we cannot use an exact yes / no decision threshold in this case.

Distances between features can also be given quantitatively, i.e., with numerical data. During the matching process, the highest match score is assigned to the lowest difference between numerical values of a given feature in the reference object and that in the tested object. The score has the highest value for the lowest distance, because the most similar numerical values are the most probable ones. It is possible to define a dependence function with an exactness. The exactness depends on the spread of distances. In case of a telephone conversation a date is an example of a numerical value. We can assume that latter calls are more interesting than the older ones, thus their weights will be higher and also the probability of the same caller or of reporting the same emergency case will be higher.

To define dependences between features and to indicate how much one parameter changes its value according to changes of another parameter, we use correlations. The most natural are time and space correlations. In case of the time correlation we suppose how the respective feature can change in time. It should be an additional assumption

that the correlation is computed between a numerical value (like for example time feature) and a similarity function value (difference between feature values). Consider, for example, two fields: age and date. We compute how the age parameter changes in time. Developmental stages are interpreted as a human age range with the defined age at the beginning and at the end of each age range. If somebody calls, the operator can plausibly indicate the age range of the caller. If the same person called already many years ago, now is older, and can belong to another range of age. Thus the searching mechanism assigns, regarding to this dependence, higher probability of being the same person for the callers of these two conversations as a previously younger person is now respectively older.

For evaluating the searching results we propose a score system with the highest score for the best match. The global score is a sum of particular scores coming out from particular searching cases.

Summarizing the above analysis, it should be stressed that the developed search strategies are based on: exact matches, similarity matrices, ranges of differences, and correlations between features. A result of the searching process is the object with the highest score or a group of objects with the highest scores. They are the best candidates to be the same object with the reference object. In other words, records with the highest values of scores are possibly the most similar to the current record. A general schema of the proposed searching mechanism based on the metadata is presented in Fig. 1. The detailed formulas used for calculations of scores are given in [2].

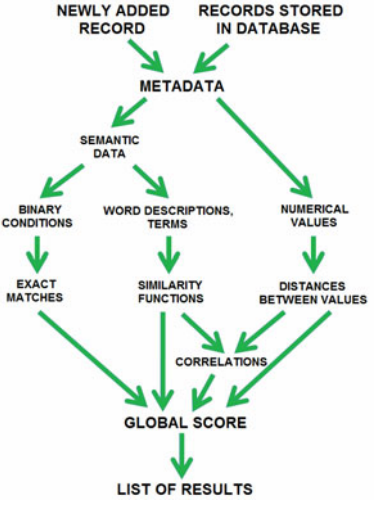


Fig. 1. Processing flow chart of searching mechanism using metadata

4 System Software Reliability

The described mechanisms are focused on the software reliability. First, the searching mechanism must be insensitive to incorrect feature values. Second, it must be aware of feature changes in time.

As a matter of fact, a wrongly given feature value lowers the probability of finding the proper object but it does not cause a total recognition failure. In other words, one or even more than one wrongly given or empty feature value reduces the probability of the correct choice but it does not exclude it.

In the developed multicriteria search mechanism similarity functions and / or other feature values are applied. In result the proper object may be chosen even if particular feature values are not exact. In our case of the emergency call system, more than 30 features describing an object (a telephone call) are included. The proposed mechanism allows to define weight or correlation for each feature. If we know, that a particular feature is often incorrect, we accordingly reduce value of the weight for this particular feature. Then the influence of this feature is lower on the search results and, in consequence, the final results are better or exactly correct. Otherwise, weights of more explicit features are higher. The system may also be protected against misprints by storing frequent misprints in similarity matrices. The problem of misprints may also be solved with the prompt system implemented within the user interface.

The proposed search mechanism learns to work properly in time. Due to the incorporated time correlations a precise monitoring of the feature changes in time is also possible.

5 Search Example

As an example let us consider checking if a person who is just calling to the emergency station has ever called before. A similarity matrix used in the calculations is given in Table 1 while the correlation matrix is shown in Table 2. Weight values of these matrices are given arbitrary, but in reference to the real world, only to illustrate the problem.

Table 1. Illustrative similarity matrix for city field

n.ad. \ in db	**Poznan**	**Buk**	**Warszawa**
Poznan	1.00	0.90	0.20
Buk	0.90	1.00	0.20
Warszawa	0.20	0.20	1.00

abbreviations: **n. ad.** = newly added, **in db** = in database

In Table 3 metadata of 10 telephone calls are presented. The current (the reference) call has ID equal to 10. A simplified assumption has been made that all features have the same importance value equal to 1. Adapting this value to each feature effect with better results of the search process. In Table 4 the global score is calculated and the results of the searching process are presented. A call of the same person was found (ID = 6) but also a call related to the same event (ID = 9) but reported by another person occurred to have a high value of the global score. The mentioned records are highlighted in Tables 3 and 4. Proper results were achieved even for simplified situations, for intuitively defined weight values, and with the same importance for all features.

Table 2. Illustrative time correlation matrix for changes of a human age from previous (past) call to the current call (columns with non zero weights are printed only)

No.	age / date change range [days]	n. ad.: child in db: child	n. ad.: young in db: child	n. ad.: young in db: young	n. ad.: adult in db: child	n. ad.: adult in db: young	n. ad.: adult in db: adult	n. ad.: senior in db: child	n. ad.: senior in db: young	n. ad.: senior in db: adult	n. ad.: senior in db: senior
1	0 – 10	1.0	0.0	1.0	0.0	0.0	1.0	0.0	0.0	0.0	1.0
2	10 – 3650	0.3	0.3	0.4	0.0	0.2	0.8	0.0	0.0	0.2	1.0
3	3650 – 7300	0.0	0.5	0.2	0.2	0.4	0.5	0.0	0.0	0.3	0.7
4	7300 – 10950	0.0	0.3	0.0	0.5	0.5	0.3	0.0	0.0	0.4	0.3
5	10950 – 14600	0.0	0.0	0.0	0.4	0.4	0.1	0.0	0.2	0.5	0.1
6	14600 – 18250	0.0	0.0	0.0	0.3	0.2	0	0.2	0.5	1.0	0.0
7	18250 and more	0.0	0.0	0.0	0.2	0.0	0	0.8	1.0	1.0	0.0

abbreviations: **n. ad.** = newly added, **in db** = in database

Table 3. Database record list

ID	Kind of the event	City	Date	Gender	Age	Speed of speaking	Linguistic errors	Foreign accent	Stuttering	Acoustic background
10	crime	Poznan	20110217	male	adult	slow	no	yes	no	street
9	crime	Poznan	20110217	female	young	quick	no	no	no	street
8	traffic	Warszawa	20100802	female	adult	medium	no	no	no	silence
7	inter-vention	Warszawa	20100717	male	adult	medium	no	no	yes	street
6	inter-vention	Buk	20020314	male	adult	slow	yes	yes	no	conver-sations
5	crime	Poznan	20000502	male	senior	medium	no	no	no	silence
4	inter-vention	Poznan	19991114	male	senior	slow	yes	yes	no	conver-sations
3	crime	Poznan	19980512	male	senior	medium	yes	no	yes	silence
2	crime	Poznan	19950721	male	child	medium	yes	no	yes	silence
1	traffic	Warszawa	19931009	male	adult	medium	yes	no	yes	silence

Table 4. Global score calculation for each record and results of searching process

ID	Kind of the event	City	Date	Gender	Age	Speed of speaking	Linguistic errors	Foreign accent	Stuttering	Acoustic background	Correlation	Global score	(Global score /max. global score)*100 %	Results
10	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	11.00	100.00	1
9	1.00	1.00	1.00	0.01	0.60	0.00	1.00	0.00	1.00	1.00	0.00	6.61	60.09	3
8	0.01	0.20	0.80	0.01	1.00	0.10	1.00	0.00	1.00	0.00	0.80	4.92	44.73	7
7	0.01	0.20	0.80	1.00	1.00	0.10	1.00	0.00	0.00	1.00	0.80	5.91	53.73	6
6	0.01	0.90	0.80	1.00	1.00	1.00	0.00	1.00	1.00	0.01	0.80	7.52	68.36	2
5	1.00	1.00	0.70	1.00	0.40	0.10	1.00	0.00	1.00	0.00	0.00	6.20	56.36	4
4	0.01	1.00	0.70	1.00	0.40	1.00	0.00	1.00	1.00	0.01	0.00	6.12	55.64	5
3	1.00	1.00	0.70	1.00	0.40	0.10	0.00	0.00	0.00	0.00	0.00	4.20	38.18	8
2	1.00	1.00	0.70	1.00	0.20	0.10	0.00	0.00	0.00	0.00	0.20	4.20	38.18	9
1	0.01	0.20	0.70	1.00	1.00	0.10	0.00	0.00	0.00	0.00	0.50	3.51	31.91	10

6 Concluding Remarks

The presented approach consisting in the development of a universal multicriteria mechanism for searching in a database for the most similar objects to the reference object joins various types of metadata together. Due to the proposed strategy the database search system is much less sensitive to incorrect feature values than standard search methods. Additionally our system is aware of feature changes in time.

The system was designed for a database of emergency telephone conversations in order to search for the most similar previous callers to the person who has just called.

The proposed methods can also be applied in other search engines replacing exact matching by multicriteria metadata matching mechanisms exploiting weights, probabilities, distances, and correlations of database records. A promising feature of our approach are reliable results even with somehow unreliable records in the database.

Testing of data search engines, which operate on automatically extracted data, is planned in future.

References

1. Allasia, W., Gallo, F., Chiariglione, F., Falchi, F.: An Innovative Approach for Indexing and Searching Digital Rights. In: Proceedings of the Third International Conference on Automated Production of Cross Media Content for Multi-Channel Distribution, pp. 147–154 (2007)
2. Balcerek, J., Drgas, S., Kmieciak, M., Pawłowski, P., Konieczka, A., Dąbrowski, A.: Database of emergency telephone calls – system tools for real-time registration and metadata searching. In: Proc. of IEEE Signal Processing SPA 2010, Poznań, September 23-25, pp. 89–94 (2010)
3. Pingali, G.S., Opalach, A., Jean, Y.D., Carlbom, I.B.: Instantly Indexed Multimedia Databases of Real World Events. IEEE Transations on Multimedia 4(2), 269–282 (2002)

4. Rawal, A., Kowar, M.K., Sharma, S., Sharma, H.R.: Automated Document Ranking Evaluation in Digital Libraries. In: 2010 International Conference on Advances in Recent Technologies in Communication and Computing (ARTCom), October 16-17, pp. 258–260 (2010)
5. Reveiu, A., Dardala, M., Smeureanu, I.: A MPEG-21 Based Architecture for Data Visualization in Multimedia Web Applications. In: International Conference Visualisation, pp. 84–89 (2008)
6. Samet, H.: Indexing Methods for Similarity Searching. In: Current Trends in Computer Science, ENC, Eighth Mexican International Conference, September 24-28, pp. xv–xv (2007)
7. Subramanya, S.R.: Multimedia databases – Issues and challenges. In: IEEE Computer, pp. 16–18 (December 1999/January 2000)
8. Li, T.: Ogihara, M.: Content-based music similarity search and emotion detection. In: Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, May 17-21, vol. 5.5, pp. V-705-8 (2004)
9. Xi, L., Tiyan, S., Jinjie, Z., Changmin, S.: A Spatial Technology Approach to Campus Security. In: IEEE International Conference on Networking, Sensing and Control, ICNSC, April 6-8, pp. 221–225 (2008)
10. Zezula, P., Amato, G., Dohnal, V., Batko, M.: Similarity Search - The Metric Space Approach. Springer, Heidelberg (2006)

Performance Measurements of Real Time Video Transmission from Car Patrol

Maciej Szczodrak, Andrzej Ciarkowski, Bartosz Głowacki, and Kamil Kacperski

Gdansk University of Technology, Multimedia Systems Department,
Narutowicza 11/12, 80-233 Gdansk, Poland
{szczodry,rabban}@multimed.org

Abstract. The HSUPA technology application to video streaming from moving vehicle to the central server is presented in the paper. A dedicated software for transmission control in case of non public IP address is employed. Quality of video streaming in urban area was measured. Several car routes were investigated in the area of the Polish Tricity. Measurements pointed out that the real time streaming quality during vehicle movement is sufficient within the prevailing time period.

Keywords: mobile camera, surveillance, transmission quality.

1 Introduction

Surveillance systems nowadays extend beyond the stationary, permanently installed cameras with wired media transmission. Often there is a need of real time transmission of the image from places where the infrastructure does not exist or where implementing a wireless network is not economically justified. Moreover, we can imagine monitoring of the interior of a moving vehicle where the operator of the vehicle is not able to watch the video. The security in public mass transport system provides an example of such an application [1]. The Moryne project aimed at increasing security in municipal public transport by real time transmission of video and data acquired by additional sensors from a bus to a traffic control center.

The aim of this work is to evaluate practically the quality of video transmitted from moving vehicle. The available channels of wireless transmission in the urban area of Polish Tricity (Gdansk, Sopot, Gdynia) can be either WiFi or UMTS. The first is practically limited to the public hotspots deployed in the selected city areas, thus using it in video transmission would be impractical. Therefore, cellular telephony is the most reasonable choice.

The extensive development of broadband wireless connections is observed in last time. Cellular telephony evolution has resulted in the appearance of data transmission services. In Europe the most popular standards are GSM and UMTS. The increasing number of mobile telephony users and research of new solutions effected in development of UMTS [2]. High speed Internet access is provided by implementation of HSPA (High Speed Packet Access), which is the integral part of the system. Two elements are distinguished: HSDPA and HSUPA [3] [4]. The first was introduced in the Third Generation Partnership Project Release 5 standards to reach the theoretical

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 233–241, 2011.

data rate up to 14.4 Mbit/s. HSUPA was introduced in the 3GPP Release 6 standards to achieve theoretical data rate up to 5.76 Mbit/s. With both HSDPA on the downlink and HSUPA on the uplink, UMTS users get improved data services throughput and Quality of Service. HSUPA introduced improved physical layer features to the uplink: smaller transmission time interval, hybrid automatic repeat request, and fast mechanism to request uplink resources [5]. This technology is widely available in most of the cities.

LTE technology is even more appropriate for streaming video. However, at present only a few 3G transmitters have been adopted to HSPA+ and have LTE ready status.

Therefore we decided to check the quality of streaming video from the mobile terminal using HSUPA. The quality is evaluated on the base of comparison of the received image and recorded locally in the mobile terminal. The synchronization of selected sequences was made in order to use comparative methods.

The paper is organized as follows. Section 2 describes the design of the transmission system. Next section presents the methodology of measurements. Following section depicts the results of the experiments. The last section contains conclusions.

2 Overview of Technology

The choice of HSPA technology as a transmission medium has a serious impact on the ability to conduct video streaming sessions from the mobile terminal. One of the most important properties of this technology is the relatively small area covered by a single base station, which often causes breaks in the connectivity as the vehicle travels between areas located in different “cells”. Switching the base stations results in the change of network-assigned IP address of the mobile terminal and causes connections based on session-oriented protocols to be terminated. Moreover, various operators deploy different strategies related to the availability of the public IP addressing, NAT filtering and firewalling in their networks. Such hostile environment makes it particularly difficult to achieve stable and reliable video transmission without resorting to specialized tools.

The most straightforward method of video streaming relies on the use of RTSP protocol to establish and to supervise a session. [6] There exist a number of powerful RTSP tools which could be used to achieve the above-set goal, however due to the aforementioned connectivity breaks and IP number changes, RTSP session would be dropped and needed to be reestablished very often. Moreover, mobile terminal acting as an RTSP server would need a way to notify the recording “client” of its new IP address after each network reconnect. Last but not least, the presence of NAT devices or firewalls would effectively prevent the client from reaching the mobile server, making the establishment of video stream impossible.

To combat the above-mentioned limitations another approach was chosen, based on “raw” RTP protocol, without the use of session control protocol. [7] RTP protocol is a standard tool for the streaming of multimedia data, however typically it is used in conjunction with some higher-level protocol used for controlling of the session

("signaling"), e.g. RTSP or SIP. It is possible to use RTP alone, yet the terminals must be configured by some other means to be able to communicate. The proposed setup consisted of two RTP terminals, one acting as a "sink", the other one as a "source". The "sink" was deployed on a machine with public IP address, which allowed the "source" to send RTP datagrams to it regardless of its conditions (public/private IP address, behind a NAT or firewall). As RTP is based on connectionless UDP protocol, there was no "danger" of terminating the connection between "source" and "sink" when network connectivity was broken. In fact, the "source" terminal was producing the RTP packets during the connection breaks the same way as when the connection was active. The packets were silently ignored by the terminals TCP/IP stack, however such behavior is consistent with the definition of real-time data, which need not being retransmitted, as they become outdated. This of course incurred high packet loss rate, however the streaming was automatically resumed as soon as the connectivity was reestablished.

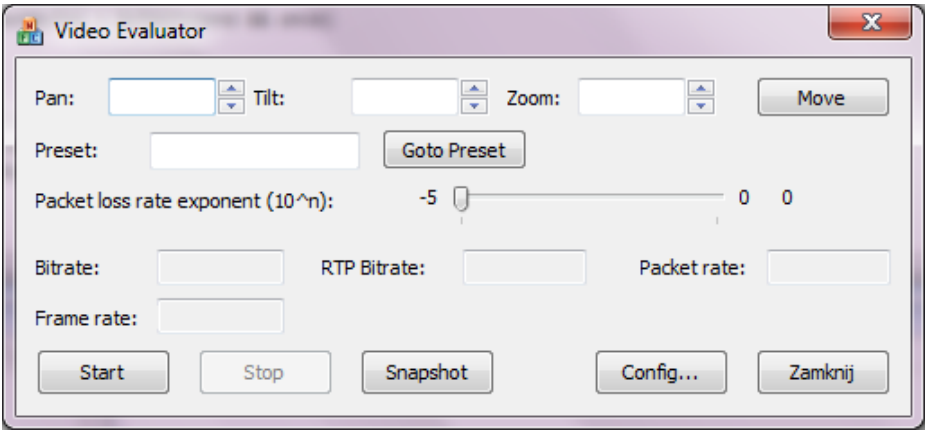


Fig. 1. RTP video streaming terminal (video preview window not included)

In order to adopt the above-described methodology a dedicated software system was developed consisting of two applications, namely the "source" and the "sink". Both programs were designed as Win32 GUI applications sharing about 80% of common source code. The applications were based on VoIP framework developed previously by the authors, therefore most of the work related to the integration and configuration of the components. The "source" application implemented the following feature set: video capture from local USB or remote IP camera, local video preview, recording of video stream to local file for evaluation reference and streaming RTP packets with H.263+ or Theora payload to preconfigured "sink", which included the following features: receiving RTP stream from remote endpoint, jitter-buffering, depacketization and decoding the video stream followed by recording it to local file as well as previewing it. Both applications used similar user interface, which was depicted in Fig. 1.

3 Measurement Methodology

The experiments were conducted in the urban area of Gdansk, Sopot and Gdynia (the Polish Tricity). The mean of transport was a car with IP camera Axis 223M, laptop, HSPA modem iCon 401 and GPS receiver GPSmap 60Cx installed. The camera was deployed in such way that allowed for observing the scene in front of the car. The sender application was installed on laptop, and receiver application on the remote server. The devices were working in setup presented in Fig. 2.

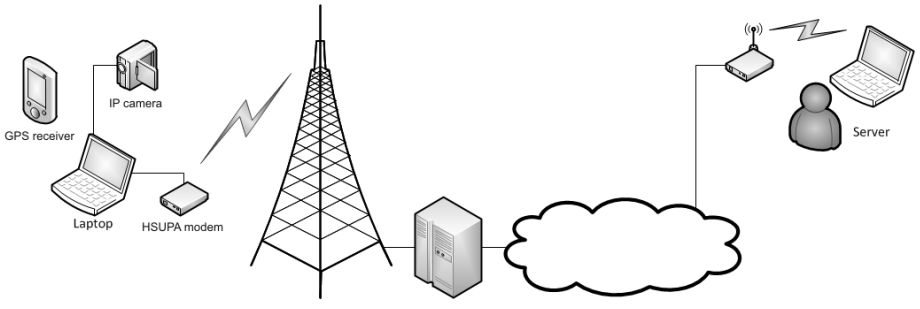


Fig. 2. The test setup scheme

The laptop remained behind NAT and firewall, thus the access by IP address was not possible. The server had the public IP address. The video was streamed at resolution of 352×288 at 10 fps (which is the maximum framerate of the camera). The encoding of the camera image was done by H.263+ codec and the output bitrate was about 152 kbit/s.

The image acquired by the camera was recorded locally on the laptop and treated later as the reference signal. The received stream was recorded on the remote server. The distortion of video signal in this case is caused by compression, packet data transmission and the Internet access. The compression of image is the source of such distortions as: appearing of pixel blocks, changes of saturation, sharpness, loose of fluency in dynamic scenes. The amount of distortions depends on the compression level and the codec. Packet transmission of data induces total or partial loose of image during transmission interferences or still image. These distortions are caused by the network.

The quality of received stream was evaluated with objective measures. The program MSU Video Quality Measurement Tool (MSU–VQMT), available as free software, was employed to investigate the video quality [8]. Relative and non relative metrics was used to evaluate image quality: PSNR, VQM, SSIM, Blurring Metric [9] [10].

4 Evaluation of Measurement Results

The experiment session resulted in gathering the video recordings. The total length of video, comprised of 7 fragments, recorded on the mobile terminal, equaled 2 hours 55 minutes 40 seconds. For the remote server the time was 15 minutes and 1 second shorter. The overall distance of the route was about 120 km.

Table 1. Summary of lost frames in transmitted video

Seq. number	Transmitted [mm:ss]	Received [mm:ss]	Difference [mm:ss]	Lost frames
1	10:34	10:02	00:32	326
2	15:13	14:22	00:51	511
3	52:09	47:17	04:52	2929
4	43:01	37:18	05:43	3427
5	04:36	04:22	00:14	132
6	32:39	30:56	01:43	1029
7	17:25	16:19	01:06	656

The duration differences between send and received video are caused by transmission distortion on the particular areas of telecommunication infrastructure. The observed disruptions in transmission was from few milliseconds (difficult to notice, lose of individual frames) to about 30 seconds. The time of video is presented in Tab. 1. All sections of route referenced geographically are presented in Fig. 3 (left).

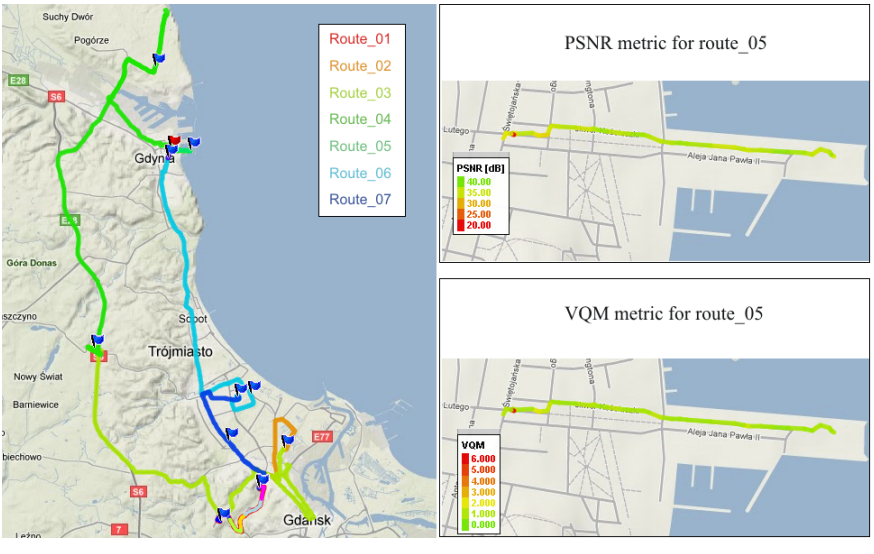


Fig. 3. Route map (left) and metrics (right)

The objective comparative analysis of video using the MSU Video Quality Measurement Tool requires that each inquired frame has to be compared with the reference frame. The number of frames in both sequences has to be equal in order to achieve reliable and correct result. Sequences gathered during experiment were characterized by different duration of transmitted and received video. In consequence reliable comparison of the obtained recordings becomes difficult. One

of the solution is manual synchronization of sequences. The empty frames are inserted in the moments of transmission breaks. The process is very time consuming, so for the purpose of analysis and presentation of outcomes, only the sequence 5 was synchronized. The image from camera send and received for distorted signal was presented in Fig. 4. In the best case the received image is the same as original.

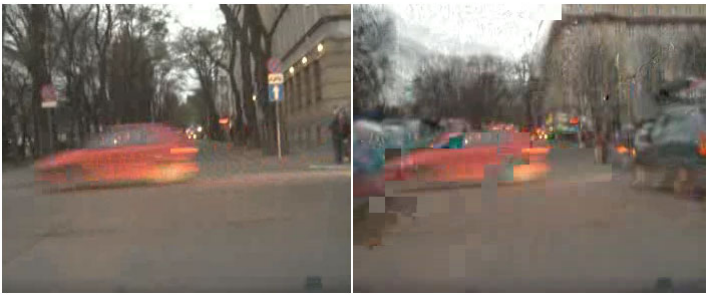


Fig. 4. Send and received frame 2170 in sequence 5

The received frame is visibly distorted in comparison to send frame. Remains of previous frames caused by breaks in transmission and packet loss are visible.

Synchronized video was compared with described metrics. The results of PSNR are presented in Fig. 5.

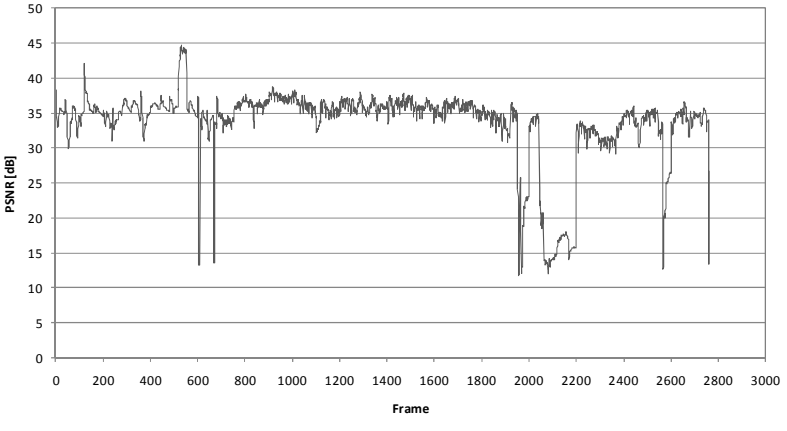


Fig. 5. PSNR metric for sequence 5

The higher value of PSNR metric, the better quality. For sequence 5 average of PSNR is 24.9 dB. According to the chart, the value of this metric for the prevailing number of frames was about 35 dB. Such a level provides subjectively unaffected

quality of received image in comparison to original image. The locations where the transmission were lost are characterized by rapid decrease of PSNR level. The VQM metric is presented in Fig. 6. Lower value means better image quality. The average VQM value is 1.66. Localizations of transmission loss are characterized by rapid increase of VQM. Similarly as in case of PSNR, the small distortions were observed in the first set of images. The SSIM metric is presented in Fig. 7. The average value of this metric is about 0.94 what denotes good quality of video transmission. Fig. 8 presents Blurring Metric. Apart from previous metrics it shows the level of blurring and sharpening of analyzed image in the reference to the original image. Continuous line denotes sharpening, dashed line blurring. In case of investigated recording the highest level of blurring was observed in localizations of partial or total transmission loss. The average value of blurring is 8.76 and sharpening 9.65. Very low value of both sharpening and blurring appears between frame 516 and 552. During experiment in this instant a piece of paper was put in front of camera. The following minima of blurring symbolize dropped frames. The longest period of such distortion lasted 107 frames (between frames 2060 and 2167), which is also most noticeable in the other metric charts. The quality map of VQM and PSNR metrics for sequence 5 is presented in the right part of Fig. 3.

The transmission quality on the selected route is satisfactory. The problem in comparative objective analysis is in synchronization of video sequences. In order to present the quality of transmission on the other routes, the simple metric which describes the proportion of send and received frame number on the route section can be utilized. The definition is given by equation:

$$Q_F = \left(1 - \frac{N_{Send} + N_{\mathrm{Re}cv}}{N_{Send} \cdot L_F}\right) \cdot 100 \tag{1}$$

where:
N_{Send} – number of frames in original sequence,
N_{Recv} – number of frames in received sequence,
L_F – length of the route fragment divided by 10^3 m.

The results of presented above quality measure for fragments of sequence 1 are presented in Tab. 2.

Table 2. Introduced Q_F metric for fragments of sequence 1

Fragment	L [m]	N_{Send}	N_{Recv}	Q_F
1	979	1421	1411	99.28
2	1000	1266	1204	95.10
3	932	889	878	98.67
4	964	693	693	100.00
5	1001	1093	851	77.88

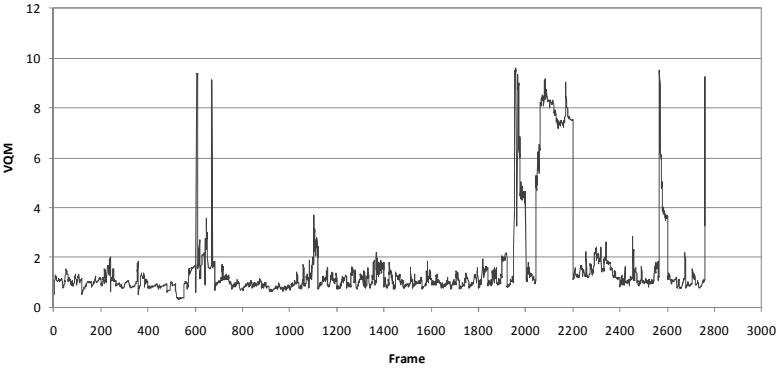


Fig. 6. VQM metric for sequence 5

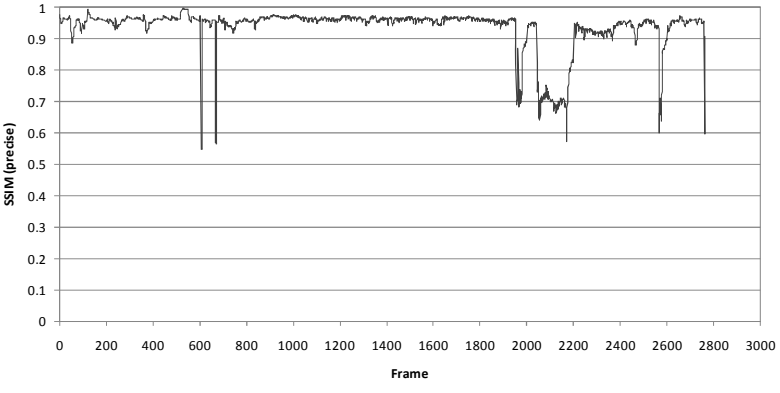


Fig. 7. SSIM metric for sequence 5

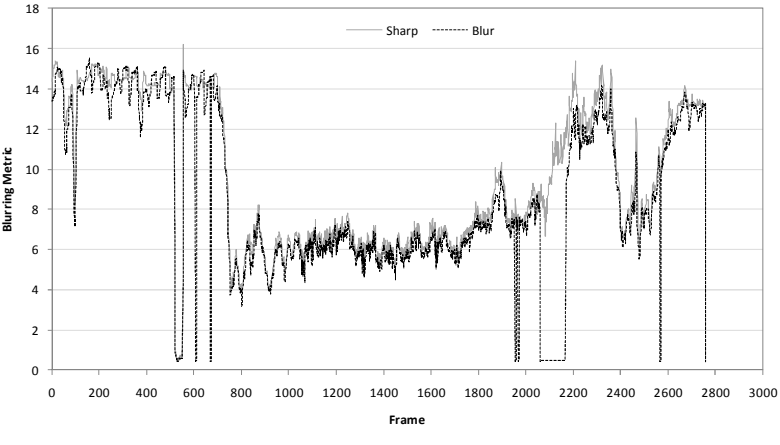


Fig. 8. Blurring metric for sequence 5

5 Conclusions

The measurements of real time video streaming were conducted. The implemented software for RTP video streaming was applied in order to record both the reference and transmitted data. The presented outcomes indicate that quality of video transmission from a moving car is sufficient. The distributed mobile monitoring can be applied in practice after integration of hardware components utilized in the experiment into a single device. The extension of the system functionality is foreseen with event detection or number plate recognition algorithms. Moreover, the popularization of LTE technology would provide the sufficient base for a high resolution streaming.

Acknowledgements

Research is subsidized by the European Commission within FP7 project "INDECT" (Grant Agreement No. 218086).

References

[1] EU-IST FP6 Moryne Project, http://www.fp6-moryne.org/
[2] Holma, H., Toskala, A.: HSDPA/HSUPA for UMTS. John Wiley & Sons Ltd., Chichester (2006)
[3] Dahlman, E., Parkvall, S., Skold, J., Beming, P.: 3G Evolution: HSPA and LTE for Mobile Broadband. Academic Press Ltd., London (2007)
[4] Liu, J., Tapia, P., Kwok, P., Karimli, Y.: Performance and Capacity of HSUPA in lab environment. In: Proc. Vehicular Technology Conference, May 11-14, pp. 1906–1909. IEEE, Los Alamitos (2008)
[5] 3GPP TS 25.309 V6.6.0, Group Radio Access Networks; FDD Enhance Uplink; Overall Description (2006)
[6] Schulzrinne, H., Rao, A., Lanphier, R.: Real Time Streaming Protocol (RTSP), RFC 2326 (April 1998)
[7] Schulzrinne, H.: RTP profile for audio and video conferences with minimal control, RFC 1890 (January 1996)
[8] MSU Video Quality Measurement Tool, http://compression.ru/video/quality_measure/video_measurement_tool_en.html
[9] Wang, Z., Bovic, A.C.: A universal image quality index. IEEE Signal Processing Letters 9, 81–84 (2002)
[10] Pinson, M., Wolf, S.: A New Standardized Method for Objectively Measuring Video Quality. IEEE Transactions on Broadcasting 50(3), 312–322 (2004)

Fast Face Localisation Using AdaBoost Algorithm and Identification with Matrix Decomposition Methods*

Tomasz Marciniak, Szymon Drgas, and Damian Cetnarowicz

Poznan University of Technology, Chair of Control and Systems Engineering, Piotrowo 3A, Poznan, Poland
{tomasz.marciniak,szymon.drgas,damian.cetnarowicz}@put.poznan.pl
http://dsp.put.poznan.pl

Abstract. In this paper automatic recognition of faces in images is considered. This problem can be divided into to stages: localization and identification. Despite satisfactory accuracy of many existing face localisation algorithms, it is important to perform this operation efficiently, therefore a tool for fast face localisation with AdaBoost algorithm is presented. The face recognition is a difficult problem due to variations of illumination, facial expression, etc. In order to achieve good accuracy, proper features must be selected. A tool for face feature extraction (both holistic and localized) is presented.

Keywords: face localisation, face recognition, AdaBoost, ICA, NMF, eigenfaces.

1 Introduction

The face analysis algorithms gather faces from the images (still pictures or video streams from cameras) and search the database of known people faces in order to identify the particular person. Two processes are involved. First, face detection selects parts of the image that contain faces, using a cascade of Haar classifiers (Section 2). In the next stage, face identification is performed by utilization of holistic analysis or local features analysis (Section 3) [15].

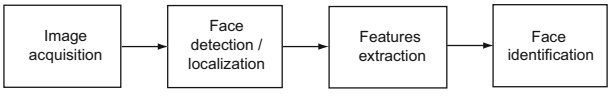


Fig. 1. General block diagram of typical face recognition stages

In holistic approach a face is treated as a no dividable image and compared with other images in the database. This comparison can be performed by

* This paper was prepared within the INDECT project.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 242–250, 2011.

means of various statistical multidimensional analyses. In local features analysis, the face is first split into distinctive regions (or points), then a mathematic description of those elements is calculated, and finally the obtained values are compared to the ones from the database. Hybrid approach is also possible, combining both holistic and local approaches.

The following sections focus on face recognition methods for biometric feature analysis. Section 2 contains a brief description of localization algorithms, and Section 3 an overall comparison of existing and established methods for face identification.

2 Fast Face Localisation

The first stage of face recognition is its localization. During that process position and scale of face is estimated and these values can be used to scale and position face at the input of indentification algorithm. Many face localization methods are compared in [12]. An interesting solution is the algorithm, which can select proper set of Haar features [10]. It enables not only fast picture analysis but also training of fast classifiers. An application of extended set of Haar features gives a possibility of moving object analysis in limited range. The next important feature of this method is a possibility of state classifiers learning using the samples from real objects. Due to this the method is more universal and is not limited to specific person, in strictly defined conditions.

After face localization, the next step in face recognition is position determination of characteristic face elements such as: eyes, nose or lips. It can be done by means of hierarchical segmentation [6] and Bezier deformation model [8].

Assuming, that elements of interest are eyes, detection of double, identical objects in bounded area defined as face image is considered. Eye localization is based on iris position determination. Iris is approximately round and very contrastive at eye sclera. It is most visible for red component.

The algorithm consists of the following steps:

- restriction of searching area to eye sockets surrounding
- red component extraction
- enlargement of area
- filtering and contrast enhancement
- final contrast enhancement
- gradient determination
- adaptive thresholding
- Hough transform.

The restriction of searching area to eye sockets surrounding is essential for computational cost reduction. Dimensions of this area was determined experimentally.

Extraction of the red component also reduces the contrast for between iris and sclera. After component extraction, it is still difficult to localize eyes because of low resolution and high level of noise. For resolution 320×240 pixels, the eyes

area can be represented by merely few of them. In order to detect circle shape an increment of resolution of the analyzed image is needed.

Unfortunately, with increment of the sample areas, noise is also enlarged. In order to reduce grain noise coming from camera, the area is filtered with median filter. Filtering parameters should be set in a proper way to reduce the noise but to avoid sharpness reduction. After filtering contrast increment is performed. It is done by shifting and scaling. For example for 8 bit image after scaling and shifting pixel values should be between 0 and 255. An additional contrast enhancement operation is thresholding of values below 3 to 0 and above 252 to 255. The image processed in such a way is ready to start a proper stage of iris localization. First, the gradient of the image is computed. An approximation of the gradient is Sobel function.

Many accidental lines still hinder iris localization. It seems that appropriate threshold is the solution, however it is difficult to find the threshold value. Incorrect threshold may lead to iris edge loss. In such cases iterative threshold can be employed. However, there are cases where 70 iterations are needed to obtain satisfactory result, what disqualifies this silution. A better way is adaptive threshold. An image is thresholded locally and the threshold is determined from values of pixels in a close neighborhood. The size of neighborhood and precision should be set, to obtain iris edge but with not to high computational cost. The result of the threshold is in Fig. 2.

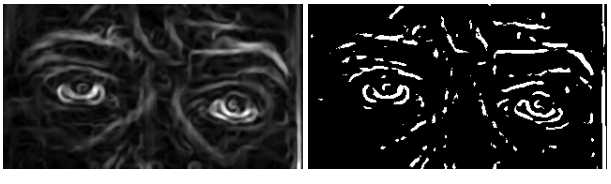


Fig. 2. Adaptive threshold result

After thresholding a binary image with oval iris contours is obtained. It can be said that the threshold helps in situations when the face is unequally lit. Next, the Hough transform is performed in order to find circular shapes (Fig. 3) [5].

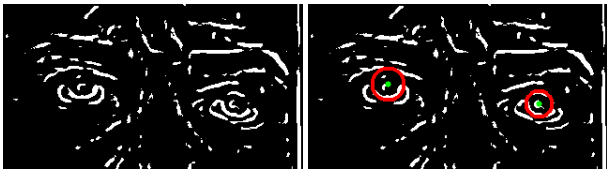


Fig. 3. Iris localization using Hough transform

In Figure 4 the main application screen with fps counter is showed. The software [3] was prepared with C++Builder by Borland and OpenCV library.

Typically, the prepared application can process from 11 to 13 fps. It should be noted, that face localization without iris searching gives about 40–42 fps. This difference is mainly caused by interpolation. The interpolation is necessary if low quality video is analyzed. Application uses about. 12 MB of the system memory.

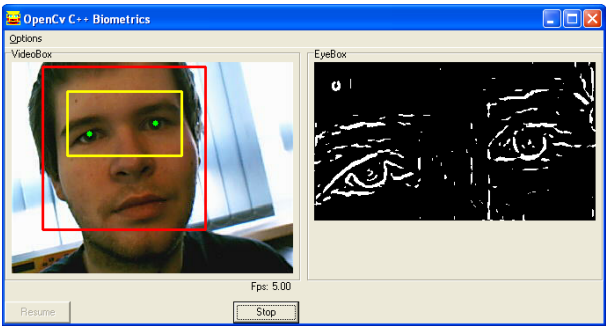


Fig. 4. Application window

A critical aspect of this application is sensitivity of the algorithm to varying light conditions. Four lightening tests were performed: scattered lightening, weak lateral lightening combined with scattered lightening, relatively strong lateral lightening, and low light. It turns out, that correct eyes localization is possible in all four cases. However weak lightening makes localization more difficult (Fig. 5).

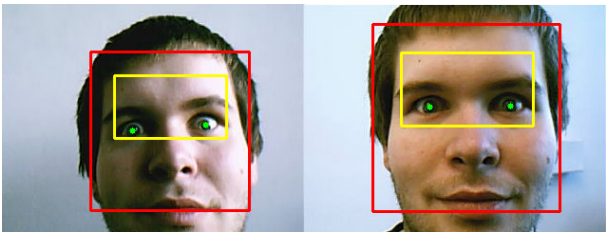


Fig. 5. Algorithm results for: (a) weak lateral illumination, (b) and strong lateral lightening

If the only light source is monitor (Figure 6) a weak frontal light program can still manage to localize the face. Only a small deviation from eye can be observed.

The most common face detector / localization method is based on Viola et. al. [11]. It analyses the presence of particular contrasts in the image. The contrasts are described as the so called Haar-like features. The localization, orientation,

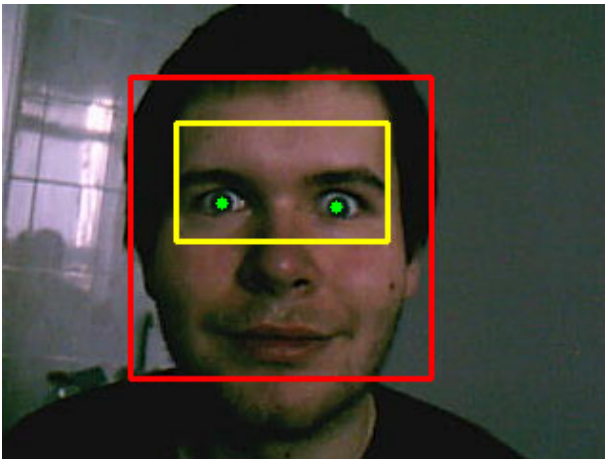


Fig. 6. Localization result without lightening

and size of those contrasting features are related to particular elements of the face, such as eyes, nose, forehead, etc. (Fig. 7). If most of those contrasts are found in particular regions of the image, it is assumed that the face was detected there.

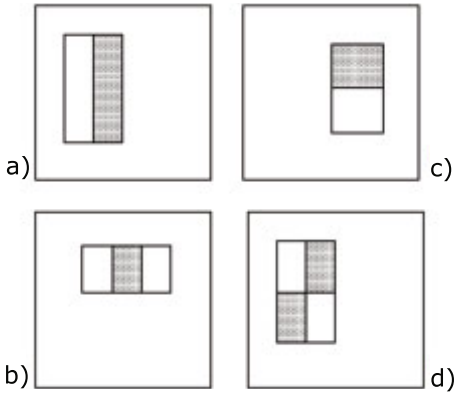


Fig. 7. Sample of rectangle features: a) and b) are two rectangle features, c) and d) are four rectangle features [11]

The best set of features for description of a face was trained on thousands of positive samples (images containing faces) and negative samples (images without faces), for assuring the best distinction between face and non-face. Next, all Haar-like features are used to create a cascade of classifiers. This approach is

called boosting. There exist numerous methods for boosted classifiers, with most popular AdaBoost. The cascade is performed as follows: first feature is tested, and if the result is positive the algorithm goes to the next one, otherwise it ends and no face is detected. The process is repeated for all features. They are sorted from the most distinctive to less distinctive; therefore classification of non-faces is very fast, because the tested region is excluded in first steps of the cascade.

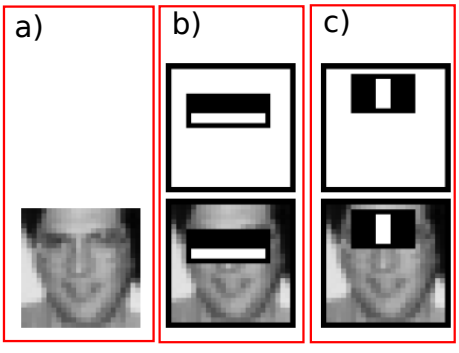


Fig. 8. First two steps of boosted classifier: a) input face, b) first feature for testing eyebrows presence, c) second feature for testing eyes presence [11]

A well known and established image processing tool set, the programming library OpenCV (*open computer vision library*) [2] comes with the trained classifiers for frontal and profile faces detection.

3 Analysis of Face Identification Algorithms

A face recognition system should automatically identify faces present in images. This problem is hard because of variations due to illumination, viewing direction, pose or facial expression. They are often higher than the variation related to various persons. A face recognition can be performed in the following steps:

- face localization and alignment (described in the previous section)
- feature extraction
- classification.

After face localization and alignment the feature extraction is needed in order to transform the image into a set (often smaller than the image dimensionality) of values of parameters that are maximally identity dependent and noise independent. In this section feature extraction methods for face recognition are analyzed.

There are two main classes of face feature extraction methods: appearance based and model based. In the appearance based methods the face image is

represented as a combination of basis images. Linear and nonlinear appearance based methods can be distinguished. The model based methods consist in detection of particular face elements and representing face by shape, size, and relative position of these elements.

In our experiments we used linear appearance based methods due to their performance, simplicity, and speed. In these methods each facial image is represented as a linear combination of basis images

$$x \approx Bh \ , \tag{1}$$

where x is a vector with pixel values of the analyzed image, B is a matrix, which in columns contains the basis images, and h is a vector of weights. A set of face examples X is used to learn the basis B, thus

$$X \approx BH \ . \tag{2}$$

The feature extraction methods differ according to the learning criteria as well as basis and weights constraints.

Methods based on the minimal reconstruction error select such matrices B and H that the reconstructed matrix X minimally differs (according to some metric) from BH. The most popular method, i.e., eigenfaces uses the squared Frobenius norm with the orthonormality constraint on the basis:

$$\min_{B,H} ||X - BH|| \tag{3}$$

subject to

$$B^T B = I \ . \tag{4}$$

Another decomposition method that can be used for face recognition is a nonnegative matrix factorization (NMF), in which there is no orthonormality constraint but nonegativity constraint of elements of matrices B and H. There are variants that differ according to the reconstruction metric: Frobenius norm NMF and Kullback-Leibler NMF, in which the criterion is expressed as

$$\min_{B,H} \sum_{i=1}^{D} \sum_{j=1}^{N} X_{ij} \log(BH)_{ij} - (BH)_{ij} \ . \tag{5}$$

Another criterion for the basis learning is statistical independence. In these methods matrices B and H are selected in such a way that they will give statistically independent features.

In Fig. 9 the basis images obtained using the eigenfaces method are shown. All of these basis images cover the whole image. Thus this is a holistic decomposition.

In Fig. 10 NMF basis images are shown. This algorithm learned the basis that corresponds to face elements such as nose, eyes, etc.

ICA basis images are also shown in Fig. 10.

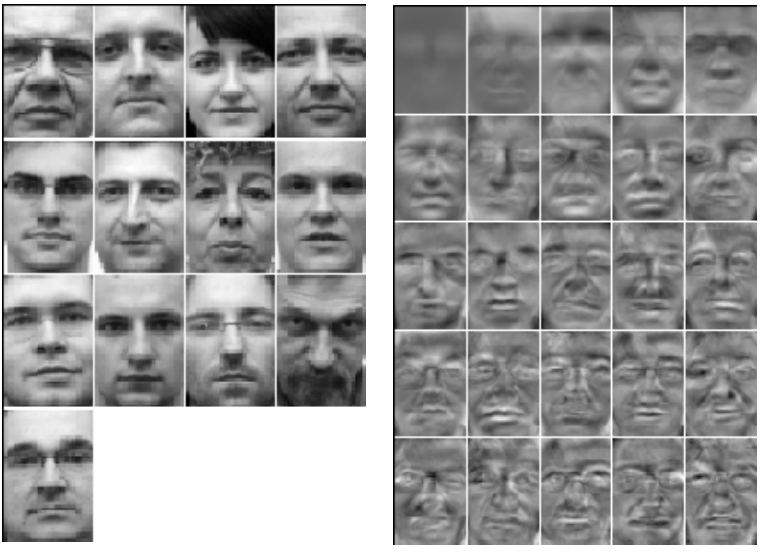


Fig. 9. Original faces (left) and eigenfaces (right)

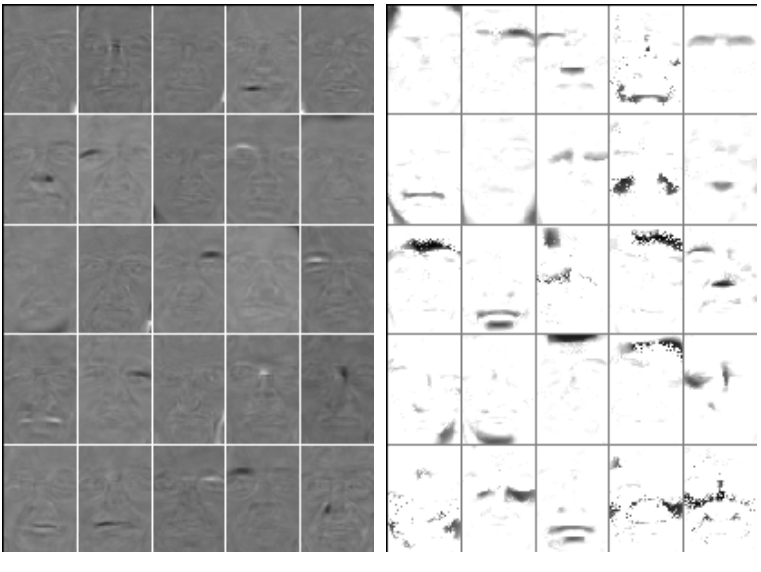


Fig. 10. ICA basis (left), NMF basis (right)

4 Conclusions

The experiments undertaken by means of the presented software tools showed that:

1. the algorithm based on Haar like features and AdaBoost can give satisfactory results in varying lightening conditions.
2. NMF and ICA methods in contrary to eigenfaces give localized features. It suggests that by this means higher recognition accuracy can be achieved, particularly in the presence noise.

References

1. Beumier, C.: 3D Face Recognition. In: IEEE International Conference on Industrial Technology, ICIT 2006, pp. 369–374 (2006)
2. Bradski, G., Kaehler, A.: Learning OpenCV, Computer Vision with the OpenCV Library. O'Reilly Media, Sebastopol (2008), sourceforge.net/projects/opencvlibrary/
3. Ciesielski, P.: Wyszukiwanie punktów charakterystycznych twarzy, (Extraction of feature points from faces), MSc. Thesis Supervisor: Tomasz Marciniak, Poznan University of Technology (2008)
4. Heseltine, T.D.: Face Recognition: Two-Dimensional and Three-Dimensional Techniques, PhD Thesis, The University of York, Department of Computer Science (September 2005)
5. Information about Hough transform, http://homepages.inf.ed.ac.uk/rbf/HIPR2/hough.htm
6. Lievin, M., Luthon, F.: A hierarchical segmentation algorithm for face analysis, Application to lipreading. In: IEEE ICME, vol. 2, pp. 1085–1088 (2000)
7. Russ, T., Boehnen, C., Peters, T.: 3D Face Recognition Using 3D Alignment for PCA. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1391–1398 (2006)
8. Tao, H., Huang, T.S.: A piecewise Bezier volume deformation model and its applications in facial motion capture. In: Advances in Image Processing and Understanding (2002)
9. Uchida, N., Shibahara, T., Aoki, T., Nakajima, H., Kobayashi, K.: 3D Face Recognition Using Passive Stereo Vision. In: IEEE International Conference (2005)
10. Viola, J.: Robust Real-time Object Detection. In: IJCV 2001(2001)
11. Viola, P., Jones, M.: Rapid Object Detection using a Boosted Cascade of Simple Features. In: Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii (2001)
12. Yang, M.H., Kriegman, D., Ahuja, N.: Detecting faces in images: a survey. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(1) (2002)
13. Yong-Li, H., Bao-Cai, Y., Shi-Quan, C., Chun-Liang, G.: An improved morphable model for 3D face synthesis. In: Proc. of the Third International Conference on Machine Learning and Cybernetics, Shanghai, vol. 7, pp. 4362–4367 (2004)
14. Yuan, X., Lu, J., Yahagi, T.: A method of 3D face recognition based on principal component analysis algorithm. In: IEEE International Symposium on Circuits and Systems, ISCAS, vol. 4, pp. 3211–3214 (2005)
15. Zhao, W., Chellappa, R., Phillips, P.J., Rosenfeld, A.: Face Recognition: A literature survey. ACM Computing Surveys 35(4), 399–458 (2003)

WSN-Based Fire Detection and Escape System with Multi-modal Feedback

Zahra Nauman, Sohaiba Iqbal, Majid Iqbal Khan, and Muhammad Tahir

COMSATS Institute of Information Technology, Park Road, Islamabad, Pakistan
{zee_nom,duatazeem}@yahoo.com,
{majid_iqbal,muhammad_tahir}@comsats.edu.pk

Abstract. Disasters do happen in our lives and we cannot avoid them but we can prepare ourselves so that when a disaster occurs, damage can be minimized. One such disaster can be the event of fire. In this paper, we have proposed a wireless sensor network based solution that can be used to detect fire, set-off a fire alarm and identify safe evacuation routes from the building. In order to accomplish this goal, we have designed a distributed algorithm for systematically deployed sensor nodes in a building. Then the obtained information about the safe evacuation routes is communicated to the occupants of the building using multimodal (visual + audio) feedback. In order to evaluate proposed system we have developed a simulation using Active Tcl and ns2. Front end of the simulation is designed using Active Tcl that can be used to sketch the blue-print of a building, place sensor nodes and set fire. At the back-end safe route identification algorithm is implemented in ns2 that is feed with the deployed network topology to identify safe routes for the occupants of the building.

Keywords: Wireless Sensor networks, Fire detection and escape system, Multimodal feedback.

1 Introduction

During past few decades urban building architecture has seen revolutionary changes all around the globe. Because of the significantly increasing urban population major focus remained on efficient utilization of the space that resulted in multi-storey and complex building architectures. However, till-date ensuring occupant's safety during a disaster in these buildings still remains an open issue.

One such disaster can be the event of *fire*. Current state of the art fire alarm systems usually provide only two types of services: fire detection – implemented using smoke / temperature sensors; fire alarm – implemented using an alarm which is triggered through smoke/temperature sensors. Usually when a fire alarm starts to ring occupants rush towards nearest exit using static escape signs shown in Figure 1. However, it is possible that the nearest escape route pointed by the escape sign is blocked due to fire. After physically experiencing this harsh fact residents will then have to rush towards alternate exits to escape from the building as shown in Figure 2. This can cause panic especially among those occupants who are not very well familiar

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 251–260, 2011.

with the building because now they are not getting any help from deployed escape signs. As a result, crucial time gets lost which can cause loss of precious lives.

Fig. 1. Building escape sign [1]

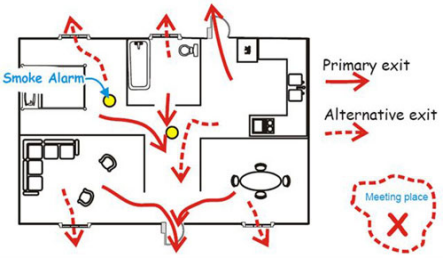


Fig. 2. Possible fire scenario in a building [2]

We argue that just ringing an alarm and guiding residents using static escape signs is not sufficient. Instead fire detection and escape system should be intelligent that is capable of detecting fire, identifying safe evacuation routes from the building and reporting these routes to the occupants of the building. In this paper we have addressed this issue by proposing a WSN-based fire detection and escape system. WSNs are result of recent advances in the development of low cost sensing devices that have found their use in diverse types of applications such as environmental monitoring, structure monitoring etc. Low cost and reliable indoor communication mechanism are two characteristics of WSN that motivated us to use them as the basic building block in our fire detection and escape system.

Proposed solution talks about the deployment of customized sensor nodes that are programmed with a distributed algorithm to compute safe route from the building in a decentralized manner. Accurate and clear indication of the identified safe passage for the occupants of the building is also very crucial. To accomplish this goal we have proposed to attach a combination of visual (directional lights) and audio (alarm) feedback (multimodal feedback) to the sensor nodes. Although we have not provided any real world implementation of the proposed model but a simulation based analysis of the proposed evacuation route identification algorithm is discussed.

Rest of the paper is organized as follows, Section 2 provides a literature review, Section 3 gives the problem statement, Section 4 presents proposed fire detection and escape system, Section 5 evaluates the performance of proposed safe route identification algorithm and Section 6 concludes the paper.

2 Literature Review

This section presents prevailing standards regarding residential fire alarm and escape systems, also few experimental and commercial WSN-based fire detection and escape systems are discussed.

During past few years a great deal of effort has been put in to utilize sensors for fire detection in buildings. Early fire detection systems were mainly utilizing smoke in the environment to detect fire. Soon it was realized that only smoke sensing is not

sufficient because it often triggers false alarms due to smoke from cigarettes, toaster etc. As a result many commercial vendors jump in to develop efficient sensing units that should utilize multiple parameters along with fuzzy logic schemes for correct fire detection. Also many standards for fire detection and alarm systems were defined e.g. European EN 54 standard and the Dutch NEN 2575 standard. EN 54 is actually a complete suite comprising of many standards for alarm devices, call points, power supplies etc. On the other hand, NEN 2575 not only defines standard for products but also specify installation and cabling process [3].

Based on these standards various WSN-based fire detection and alarm systems were proposed. In [4], a WSN-based fire detection and alarm system for large buildings is proposed. It is based on dense deployment of the sensor nodes in the building that are responsible for periodically checking temperature/smoke concentration and reporting these values to a surveillance centre through self-organizing hierarchal wireless sensor network. The central server is then used to coordinate post fire event activities. A major drawback of such a system is its centralized nature because when fire breaks out in a building the first action item of the fireman is to disconnect the electricity. As a result the central control station shuts down causing failure of the complete system.

In [7], a WSN-based fire detection system was proposed to ensure safety of the people working in mines. In this paper authors have analyzed various network topologies and communication protocol for optimal performance of the proposed solution.

To aid the fire fighters in [8] authors have proposed a fire alarm application that is based on TelosB motes. For decision making activity of the occupants and firefighters authors have utilized three parameters; humidity, temperature and light. Moreover, they have collected information regarding failed sensor nodes (due to fire). In [5, 6] authors have proposed the use of machine learning techniques for WSN-based fire detection.

In [9] authors talks about coupling the exit sign units with the heat/smoke sensors for input, speech synthesizer and strobe light for output and a communication unit to provide communication mechanism between exit sign units and central monitoring unit. They argue that the flashing strobe light can be used to attract occupants and speech synthesizers can be used to provide instructions to them. In [10] inventors have also emphasized the importance of having efficient strobe lights that can provide clear guidance to the occupants under low visibility because of thick smoke cover. We argue that even this proposed solution in [9] have a major limitations that is it depends on speech synthesizer to communicate information about the escape route to the occupants. However, in panic situations such as a fire, occupants will be running for their lives and there is a great probability that nobody would be interested in responding on speech synthesizer. Therefore the proposed system must be robust that require minimal/no human intervention.

Apart from residential fire detection and alarm systems a great deal of effort has been put in to utilize WSN for forest fire detection. In [11, 12] authors argue that WSN-based fire detection system can report a forest fire event much more effectively compared to traditional satellite based detection approach. They propose dense deployment of the sensor nodes in the forest that are responsible of gathering crucial information such as temperature, humidity, fire propagation pattern etc. and reporting

it to a central base station where fire fighting team can take necessary actions. Other work regarding the use of WSN for forest fire detection can be found at [13, 14]. However, in this paper we have focused only on indoor fire detection and escape system.

3 Problem Statement

It can be inferred from our discussion in Section 2 that most of the current state-of-the-art fire detection and escape systems suffer from following limitations:

Firstly, although most of the current state-of-the-art fire alarm systems implement extremely sophisticated mechanism to detect fire and trigger an alarm. However, these systems lack a highly important component that is safe evacuation route identification incase of fire.
Secondly, an efficient mechanism is required to convey safe evacuation route information to the occupants of the building. Static escape signs (as shown in Figure 1) can be useful but incase of fire it is quit possible that these signs guide occupants directly into the fire. Furthermore, static escape signs are often attached on the wall above the person's average height while, incase of fire people generally walk bend forward close to floor in order to avoid smoke through lungs. Thus intelligent guiding system for the occupants along with the correctly placed escape signs is required.

Another feature of the existing fire alarm systems is the steady sound of ringing alarm that only indicates a fire event has occurred. We believe that the sound intensity at fire alarms can be varied to point towards safe exits.

4 WSN-Based Fire Detection and Escape System

In this section we have proposed a WSN-based fire detection and escape system where custom designed sensor nodes are strategically deployed inside the building. These nodes are responsible of detecting a fire event, triggering an alarm, identifying safe evacuation routes and helping occupants in exiting the building through safe evacuation routes. Theoretically, proposed model can be divided into two major parts that are discussed in the following,

Safe evacuation route identification: This sub-section presents proposed WSN-based fire detection and safe route identification algorithm. In order to accomplish this goal we propose dense but systematic deployment of sensor nodes, where each node knows its neighbor as well as its direction e.g. *node-y* knows that *node-x* and *node-z* are positioned on its left and right simultaneously. Furthermore, based on their functionality and region of deployment sensor nodes can be divided into two types.

Indoor_nodes - deployed inside the building along the pathways and are responsible of communicating with their neighboring nodes to identify safe evacuation routes.
Exit_nodes – deployed at building exits and are responsible of initiating the safe route identification algorithm.

Once nodes are deployed then the algorithm executes in following steps:

Initialization phase: During this phase each *exit_node* floods the network with an initialization message containing three fields, *exitNodeId, senderNodeId* and *hopCount*. Each *indoor_node* on receiving this message performs following tasks:

```
if hopCount is greater than storedHopCount then   // storedHopCount is currently know shortest
    storedHopCount = hopCount                     // distance to exit node
    nxtNode = senderNodeId
    exitNode = exitNodeId  // When exit node broadcast the initialization msg senderNodeId=exitNodeId
    hopCount ++
    senderNodeId = Id      // Before forwarding increment hopCount and set senderNodeId equals own Id
    Broadcast initialization message
else
    discard message
```

Thus at the end of the initialization phase each node in the field knows the identity of neighbouring node through which the nearest exit can be reached. For example, if we consider *node-c* in Figure 3(a) its *nxtNode=d* through which *EXIT-I* can be reached (closest safe exit from *node-c*). Similarly for *node-b, nxtNode=a* through which *EXIT-II* can be reached (closest safe exit from *node-b*).

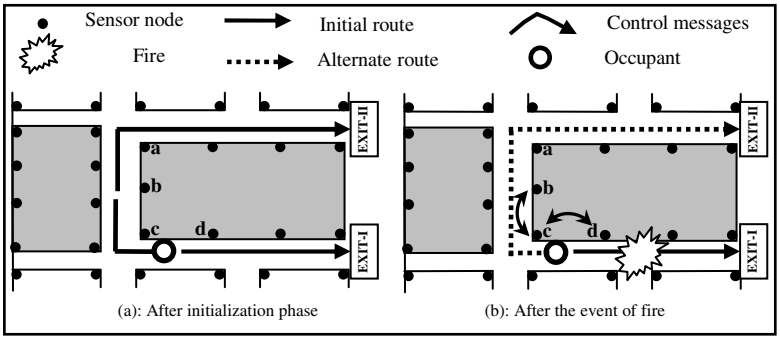


Fig. 3. Route identification algorithm

Fire detection phase: When a node senses a fire it broadcast a message *(Id, fire)* to its neighbouring nodes regarding this new development. The receiver node then checks whether the sender node is my *nxtNode* to reach exit. If the answer is yes, it means the route to exit is now broken and a request for the identification of an alternate route to exit is broadcasted as shown below,

```
if senderNodeId == nxtNode then
    route_to_exit = broken
    BroadcastMsg(Id, requestForAlternateRouteToExit)
else
    reply_safeRoute(Id, hopCount, exitNodeId)
```

Each node on receiving alternate route request will also execute above piece of code until a node with safe route *(senderNodeId* != *nxtNode)* is reached which will initiate a reply message. Each node on receiving *replay* message responds in following manner,

```
if route_to_exit == broken then
    nxtNode = senderNodeId
    storedHopCount = hopCount
    exitNode = exitNodeId
    hopCount ++
    Broadcase [reply_safeRoute(Id, hopCount, exitNodeId)]
else
     discard message
```

Thus on receiving *reply_safeRoute* message each node with broken route to exit will update its safe evacuation route. For example Figure 3(b) shows the execution of fire detection phase. When *node-d* senses a fire it broadcasts a fire message that is received at *node-c* which checks whether *nxtNode==d*. Since the answer is yes therefore *node-c* broadcast an alternate route request. When *node-b* receives an alternate route request it checks whether *nxtNode==c* the answer is no thus *node-b* broadcast a *reply_safeRoute* message and *node-c* updates its safe evacuation route to exit without the use of any central authority as shown in Figure 3(b).

Thus utilizing proposed algorithm deployed WSN successfully computes the safe evacuation routes from the building. In order to deliver this information regarding safe evacuation routes to the occupants of the building we propose a multimodal feedback mechanism.

Multimodal feedback: A user-friendly perspective: In life critical situations (such as, fire), providing correct and timely information to people is extremely important. In order to address this particular issue we have proposed a solution that is based on the philosophy of multimodal feedback.

We propose the use of custom designed senor nodes that are equipped with illuminating directional lights (to guide people visually), a speaker (for alarm and perhaps instructions), a mic (for possible communication between fire-fighters and occupants) and temperature plus smoke sensor to detect fire, as shown in Figure 4.

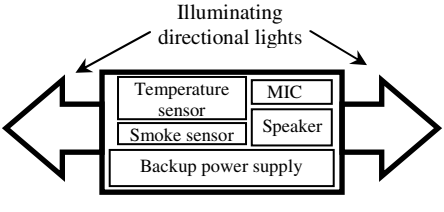


Fig. 4. Sensor guiding people towards exits

The design of directional lights also matters. For example, circular or square design of lights does not provide valuable information. Therefore the illuminating lights are designed in the shape of “arrows” where arrow-head indicates the direction of safe passage. These newly designed lights when illuminated with proper wavelength of light e.g. red colour, they become more beneficial (visible under thick smoke). Thus good design of the illuminating lights along with proper light intensity and wavelength can help a lot in providing clear and timely information to the occupants.

Now these sensor nodes should be placed low on the wall so that they are clearly visible, for people walking bend forward, even under thick smoke cover (that mostly fills a room's upper section).

The audio feedback can be used in two different ways. The first is to provide "speech" output indicating the direction of exit. For example, "move right", "move left", or "down stairs" etc. Other option is to vary the pitch and loudness of sound at different nodes of the WSN. For example, nodes along the wall, far from exit, can have "low" level of a sound but still "audible" for the people passing nearby. This level of sound gradually increases while moving towards exit. Setting gradual increase in the level of "same sound", towards the exit, can provide reliable information to people i.e. indicating that *you are close to exit*. This combination of visual and audio feedback when used wisely can help prevent disastrous situations effectively.

In order to keep proposed multimodal feedback systems running under a power failure we have also proposed to equip each node with a backup power supply that can be used during main power failures.

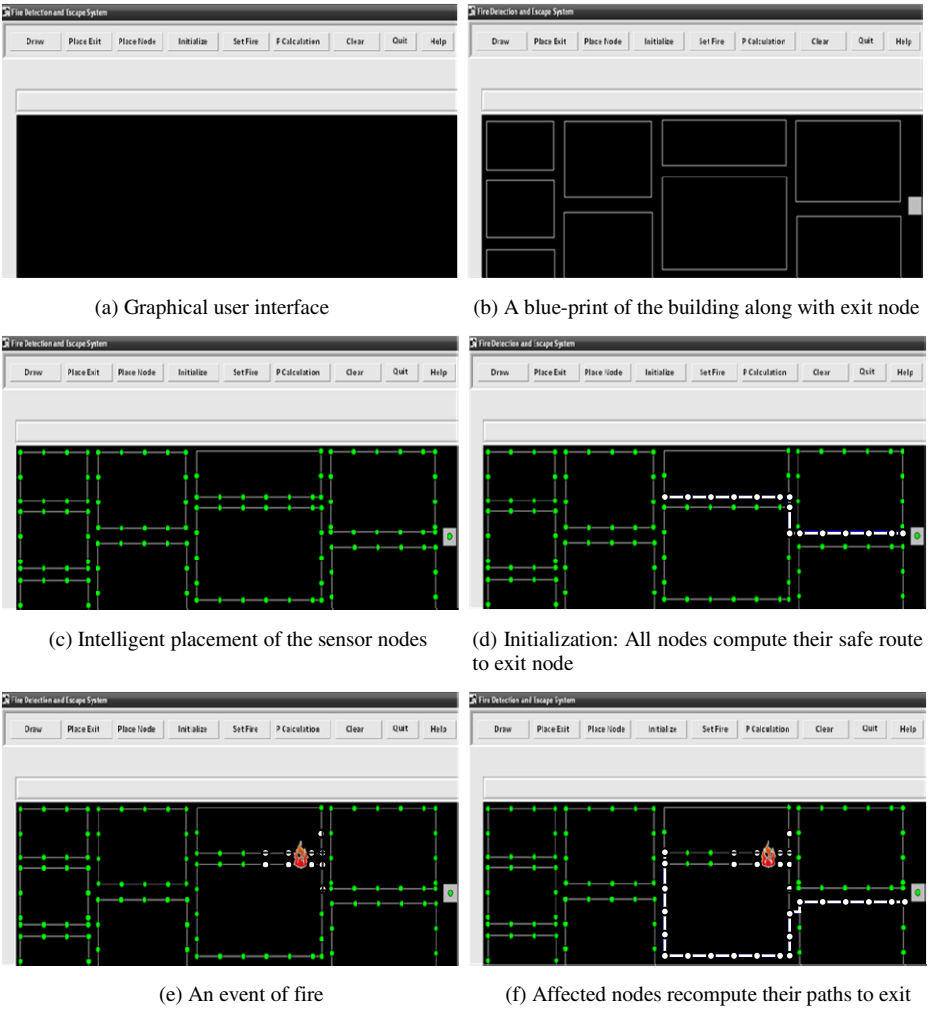
5 Testing and Evaluation

We have used simulations to validate the correctness of route identification algorithm that is a major building block of our proposed fire detection and escape system. A GUI is provided to the user for drawing blue-print of a building (for now only single-storey), identifying exits of the building and placement of sensor nodes. Then the route identification algorithm (that is implemented in ns2) can be invoked from the GUI to identify safe evacuation routes from the building (initialization phase of our algorithm discussed in Section 4). GUI also provides a provision to initiate the event of fire in any section of the building then *fire detection phase* can be triggered to compute alternate safe evacuation route to exit as discussed in Section 4. The GUI is developed using Active Tcl and its salient features are as following:

1) Draw: After clicking Draw button user can draw rectangles that represent rooms as shown in Figure 5(b). In order to ensure connected network topology we have ensured that the distance between two rooms cannot be greater than the preset communication range of the sensor node. Furthermore, user can also delete a wrongly drawn room with the right click of mouse.

2) Place Exits: It can be used to place multiple exits of the building by a click of the mouse as shown in Figure 5(b). Users can also delete a wrongly placed exit with a right click of the mouse. Again in order to ensure connected topology we have restricted that exit node must be within the communication range of at least one sensor node.

3) Place Nodes: Clicking this button leads to automatic placement of the sensor nodes along the pathways in the building. To accomplish this goal we have designed and implemented a robust scheme that takes into account the drawn building blue-print and preconfigured communication range of the nodes. In order to ensure connected topology in adverse situations we opted for dense node deployment because during an event of fire nodes can get destroyed or malfunction.



(a) Graphical user interface

(b) A blue-print of the building along with exit node

(c) Intelligent placement of the sensor nodes

(d) Initialization: All nodes compute their safe route to exit node

(e) An event of fire

(f) Affected nodes recompute their paths to exit

Fig. 5. Graphical user interface showing the working of proposed Fire Detection and Escape System

4) Initialization: After clicking the Initialization button shortest routes gets calculated using initialization phase of our algorithm discussed in Section 4. A user can view computed safe route to exit by clicking a node on the canvas as shown in Figure 5(d). Furthermore, each node also stores this route information and takes necessary actions to guide occupants towards safe exits.

5) Set Fire: After clicking the Set Fire button user can place fire in the building by single click of the mouse as shown in Figure 5(e). The nodes that detect fire turn red and broadcast fire message. Then the *BroadcastMsg* and *reply_safeRoute* messages will be exchanges among the nodes to re-compute safe route to exit as discussed in

Section 4. User can analyse the path calculation algorithm by placing fire at different location on the map. At the moment our simulator allows the placement of only one fire at a time.
6) P Calculation: Once a user changes the location of the fire in the building then by clicking the P Calculation button safe evacuation routes gets recalculated using route identification algorithm discussed in Section 4. User can see the safe paths towards exit by clicking on any node as shown in Figure 5(f).
7) Clear Canvas: It removes all items (e.g. map, nodes, exits, fire) from the canvas.
8) Quit: It closes the application.

Developed GUI based application can be of great help for diverse types of users. Deployment team can use it for pre-deployment analysis and cost estimation. Architects can use it to asses and improve the design of the building by ensuring congestion free escape routes for the occupants in the case of any disaster such as fire.

6 Conclusion and Future Work

In this paper we have proposed efficient fire detection and escape system that not only detects fire and triggers an alarm but also guides occupants to safe evacuation routes, in a user-friendly way. We have also presented a simulation model that can be used by (fire detection and escape system) deployment team for pre-deployment planning and analysis. Furthermore, proposed simulation model can also be used by the architects to improve building designs by providing congestion free routes.

Currently, we are working on a real world implementation of proposed system. During our testing we plan to analyze how occupants of the building respond to such fire detection and escape system in a real world scenario.

References

1. http://stfire-and-security.com/Escape-Route-Signs.html (accessed on 28.03.2011)
2. The City of Yorkton, http://www.yorkton.ca/dept/fire/householdfireescapeplans.asp (accessed on 28.03.2011)
3. Bahrepour, M., Meratnia, N., Havinga, P.J.M.: Automatic Fire Detection: A Survey from Wireless Sensor Network Perspective, Centre for Telematics and Information Technology (CTIT), Technical Report TR-CTIT-08-73, Pages 13 (December 2008); ISSN 1381-3625
4. Zhang, L., Wang, G.: Design and Implementation of Automatic Fire Alarm System based on Wireless Sensor Networks. In: Proceedings of the International Symposium on Information Processing (ISIP 2009), August 21-23, pp. 410–413 (2009)
5. Bahrepour, M., Meratnia, N., Poel, M., Taghikhaki, Z., Havinga, P.J.M.: Distributed Event Detection in Wireless Sensor Networks for Disaster Management. In: International Conference on Intelligent Networking and Collaborative Systems, INCOS, pp. 507–512 (2010)
6. Lim, Y.-s., Lim, S., Choi, J., Cho, S., Kim, C.-k., Lee, Y.-W.: Fire Detection and Rescue Support Framework with Wireless Sensor Networks. In: International Conference on Convergence Information Technology (ICCIT 2007), pp. 135–138 (2007)

7. Tan, W., Wang, Q., et al.: Mine Fire Detection System Based on Wireless Sensor Network. In: International Conference on Information Acquisition, ICIA 2007 (2007)
8. Bernardo, L., Oliveira, R., et al.: A Fire Monitoring Application for Scattered Wireless Sensor Networks: A peer-to-peer cross-layering approach. In: International Conference on Wireless Information Networks and Systems (WINSYS 2007), Barcelona, Spain (2007)
9. Patent:4531114, Intelligent fire safety system, Topol, Peter (Tahoe City, CA), Slater, Michael (Palo Alto, CA) (July 1985), http://www.freepatentsonline.com/4531114.html
10. Patent application title: Tactile Fire Escape System, Inventors: Daniel J. Halberg John Halberg Agents: Neustel law offices, Ltd., http://www.faqs.org/patents/app/20080282961
11. Akyildiz, I.F., Su, W., Sankarasubramaniam, Y., Cayirci, E.: Wireless sensor networks: a survey. Computer Networks (Amsterdam, Netherlands: 1999) 38(4), 393–422 (2002)
12. Yu, L., Wang, N., Meng, X.: Real-time Forest Fire detection with wireless sensor networks, pp. 1214–1217. IEEE, Los Alamitos (2005)
13. Nasipuri, A., Li, K.: A Directionality Based Location Discovery Scheme for Wireless Sensor Networks. In: Proceedings of the 1st ACM International Workshop on Wireless Sensor Networks and Applications, Atlanta, Georgia, USA. ACM, New York (2002)
14. Bagheri, M.: Efficient K-Coverage Algorithms for Wireless Sensor Networks and Their Applications to Early Detection of Forest Fires. Computing Science, SIMON FRASER UNIVERSITY. MSc: 7 (2007)

Building Domain-Specific Architecture Framework for Critical Information Infrastructure

Norbert Rapacz[1], Piotr Pacyna[1], and Grzegorz Sowa[2]

[1] AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland
[2] ComArch S.A., al. Jana Pawła II 39a, 31-864 Kraków, Poland
{rapacz,pacyna}@kt.agh.edu.pl,
grzegorz.sowa@comarch.com

Abstract. Critical Infrastructures (CI) are large and complex enterprises, often composed of multiple systems. Organizations seek to establish common frameworks that will allow them to manage the development and evolution of CIs. The enterprise architecture frameworks are templates for the development of enterprise architectures that aim to describe the enterprise structure, goals, processes, resources and organization. We argue that an architecture framework is required for critical infrastructures and that it will facilitate the critical infrastructure protection. In this article we show an approach to the task of deriving an abstracted, domain-specific Enterprise Architecture Framework for Critical Infrastructures from generic Architecture Frameworks and from the emerging standards for security management. We call it CrAF. In CrAF, the emphasis is put on Critical Information Infrastructure that is inherently coupled with Critical Infrastructure. We provide both a review of recent advances in the area of enterprise architectures and the new material resulting from our own research. The work involved in the development of such an AF is underway within the EU FP7 NI2S3 Project and National Research Grant for Polish Ministry of Science and Education.

Keywords: architecture framework, critical infrastructure, CI, CII, security.

1 Introduction

Critical Infrastructure (CI) is an infrastructure, which, when disrupted or destructed, threatens the safety of citizens, undermines national wellbeing and wealth, undermines the security of the country or region, and hinders or degrades its potential for economic growth. Critical Infrastructure may include installations and services in the area of transportation, energy and water supply, communications, and also in other sectors including, for example, public administration, financial system and healthcare [1]. CI is often multi-domain, multi-platform and multi-technology construct, geographically scattered and interconnected. CI is heavily dependent on Information Infrastructure, which is developed and operated with the use of standard components like data communication networks, thus making the CI particularly vulnerable to accidental or intentional security incidents. Therefore, federal governments and large organizations have started their work to establish guidance for CI operators in the

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 261–267, 2011.

form of best practices, recommendations and regulations to help them ensure that such systems are adequately protected. Among the prospective approaches, now under consideration, are the efforts to define common frameworks that would assist in the development and evolution of systems, in which business continuity and safety of operation is of primary concern. To this end, Enterprise Architecture Frameworks (EAF) are established templates, and methods, for the development of Enterprise Architectures (EA), which capture and represent all the relevant aspects of a business-driven enterprise.

In this paper we postulate that a domain-specific EAF is required for the particular business of critical infrastructure. The framework would serve as a guide to support enterprise architects in the design and specification of architecture of a CI. Its role would be manifold: (i) to ensure that the enterprise architecture description addresses all the essential concerns inherent to a CI, (ii) to ensure that the description is understandable and approved by the stakeholders of this business venture, (iii) to ensure that the architecture design is maintained, updated and kept in sync with the Solution Architectures throughout the lifetime of the CI system, (iv) to ensure, that transformations of the CI are orchestrated and agreed upon prior to deployment. The main idea behind the creation of a domain-specific AF for the domain of Critical Infrastructure is the preparation of a common space, in which all the stakeholders including top-level managers, executives, architecture designers, and system designers can express all major concerns pertaining to protection of CI.

The rest of the paper is organized as follows. First, we provide rationale for taking the Critical Infrastructure as an enterprise and evaluate the feasibility of applying the enterprise architecting approach for modelling the CI. Then, we highlight some trends in enterprise architecting. Next, we discuss the proposed domain-specific architecture framework. A few distinct and general elements are identified and sample solutions are provided. Last, but not least, we focus on the methodology. This part is based on enterprise architecture development method originating from TOGAF framework.

2 Critical Infrastructure Protection

The Critical Infrastructure protection activity is an analytical template, made of a CI model and a collection of methods that guide the design of a systematic protection for CI. The use of the methods results in a sequence of decisions, and associated actions, that should assists CIO (Chief Information Officer) working close with the top-level management, in ultimately determining what needs protection, as well as what level of security protection is required in an enterprise. CIP is not a product or system, but an enterprise governance activity aiming to systematically deploy the effective protection of CIs. Being a time-efficient and resource-constrained activity, it ensures the protection of only those infrastructures upon which survivability, continuity of operations, and mission success depend. While it may be impossible to prevent all attacks against a CI, CIP can reduce the chances of success of an attack and mitigate the impact in the event they are successful. For the reasons given above, the CIP activity should have the following steps: (i) identifying critical systems within CI,

essential for mission accomplishment, (ii) determining threats against systems by analyzing resources, on which an enterprise is dependent, business processes and their execution, (iii) analyzing the vulnerabilities, (iv) assessing the risks of the CI degradation, (v) determining countermeasures, for where the risk can get beyond accepted levels, (vi) defining security controls, (vii) monitoring for change, and re-evaluating the CIP process as well as re-applying security controls in an infrastructure.

The areas of risk assessment and management, business continuity and contingency planning require deeper analysis which falls beyond the scope of this paper; these areas needs to be integrated in the further work on the proposed framework incorporating the results of recent developments, such as the special publications from NIST [2].

The definition of the CI protection activity should be carried out with the objective, that CI is a fully fledged enterprise.

3 Architecture Frameworks for Large and Complex Enterprises

Enterprise Architectures (EA) help create a consolidated picture of a structure and operations of an enterprise. EA is a set of models that help to show, how various business and technical elements work together as a whole. EA defines the vocabulary, describes the composition of enterprise components, shows their internal relationships, and their ties with the external environment. What is important, it also lists guiding principles for eliciting the requirements for the design and for the evolution of an enterprise [3]. Multiple EA are available [4][5][6]. The Department of Defence Architecture Framework (DODAF) is an example of enterprise architecture framework used by the DoD, while Federal Enterprise Architecture (FEA) is in use in the US federal domain. NATO NAF, UK Ministry of Defence MODAF and DNDAF in Canada represent similar developments. Zachman Framework and TOGAF are known enterprise architecture frameworks often used in non-military domain. They follow similar strategy and vision of enabling complex systems architectures to evolve towards uniform description and to model them in a compact and homogeneous way, starting from the mission, through the guiding principles and rules of operation, down to organizational structure and procedures for all major forms of business activity in a manner understandable to all stakeholders in a particular enterprise. The efforts are undertaken to standardize the architecture description [7].

To understand better the relative position of the architecture framework proposed in this paper, a hierarchical categorization of architecture frameworks is shown in Figure 1 [8]. The five levels position the domain-specific Architecture Framework between an organization-specific AF and the generic AF. Its usefulness results from that it inherits all the features of a generic Architecture Framework, but it also accommodates the methods that allow addressing systematically the domain-specific concerns. ‘Domain specific’ means that terminology is related with the domain of critical infrastructures, that typical processes and procedures found in CI and CII are identified, and that the catalogue of good practices and guidelines is prepared together with a most common problems and their potential solutions.

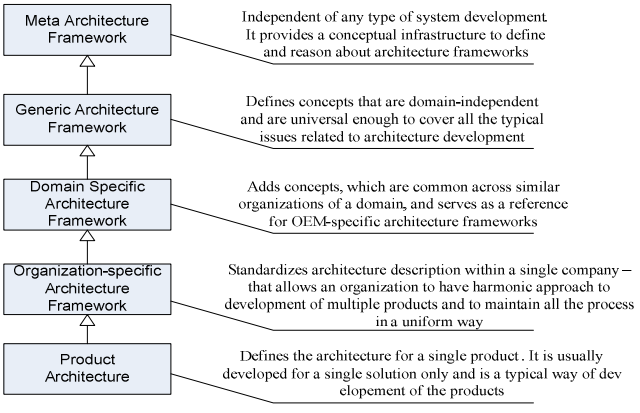


Fig. 1. A hierarchy of Architecture Frameworks

4 Towards AF Tailored for CII - CrAF

Complex interactions between the elements of an enterprise in daily operations are captured by the business and operational processes. In the course of developing a domain-specific framework we will focus on uniform description that allows expressing different means to secure these essential processes in the critical infrastructure.

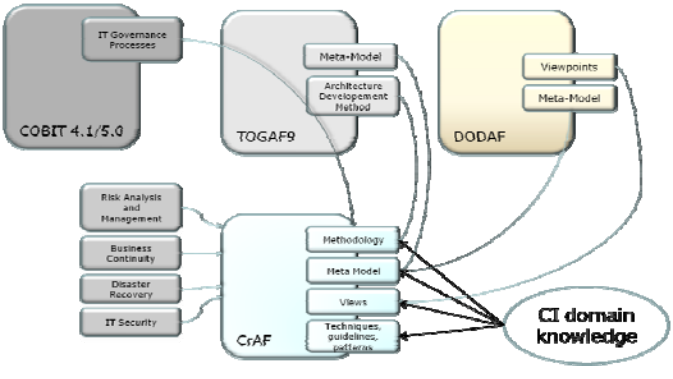


Fig. 2. General composition of Critical Infrastructure Architecture Framework

It is necessary that the framework makes provision for the presentation layer (Views) that shows how the procedures and security features affect these processes. Figure 2 shows along with Views, other general components of the Critical Infrastructure Architecture Framework (CrAF) under development in NI2S3 FP7 Project [9].

4.1 CrAF Meta-model

A meta-model is a repository of terms, used to represent the domain of interest. It is required during the AF development to name and to refer all the important elements

of architecture. Being a common, generic vocabulary, it needs to be complete, consistent, and minimal. A meta-model can be based on a generalized meta-model, such as, for example, DM2 from DODAF2 or Content Framework from TOGAF with appropriate extensions from work done in fields of CI/CII ontology [10].

The domain-specific meta-model, understandable and widely accepted by all the stakeholders, allows for a compact representation of concepts, recommended solutions, procedures and techniques applicable in the field of CII protection.

Current effort in the NI2S3 Project is focused on the development of meta-model in the field of Risk Assessment and Management, Risks in Business Processes, Vulnerability Assessment, Data Fusion and Correlation, Data Acquisition and Control.

Another applications of the meta-model in NI2S3 project include task of environmental data correlation and fusion. When implemented, such data correlation subsystem shall improve among the others the situation awareness in the enterprise.

4.2 CrAF Viewpoints

The purpose of viewpoints is to enable human engineers to comprehend very complex systems. Viewpoints are collections of views that represent architecture data in the human readable way, sufficient to describe solutions to domain-specific issues brought up by stakeholders involved in the architecture development – they're usually collections of useful views. The views are concrete instances of the architecture data captured in a model. CrAF framework inherits viewpoints from the DoDAF framework with guidelines described in [11]. DODAF similarly to MODAF and NAF defines a set of viewpoints: All Viewpoint, Capability Viewpoint, Data and Information Viewpoint, Operational Viewpoint, Project Viewpoint, Services Viewpoint, Standard Viewpoint and Systems Viewpoint.

4.3 CrAF Techniques, Guidelines, Best Practices and Design Patterns

On the basis of the defined viewpoints, consisting of multiple selective models and selective views of an CI enterprise, it becomes possible to incorporate into the framework the domain specific knowledge with the use of the domain-specific concepts from the meta-model. During that activity, the set of models with predefined processes, enterprise design patterns [12], and structures library are created. The proposed domain-specific concerns in NI2S3 project include issues of systems dependability, environmental data acquisition, aggregation, fusion and correlation in order to improve situational awareness.

4.4 CrAF Methodology

The CrAF inherits the baseline methodology from TOGAF. The most important part of TOGAF is the Architecture Development Method (ADM) - a prescriptive, step-by-step instruction guide for how to architect an enterprise. It is presented in a series of phases that guide a team of architects through the lifecycle of architecting.

There are four main iterations within ADM: Architecture Context iterations that allow initial mobilization of architecture activity by establishing the architecture approach, principles, scope, and vision. Architecture Definition iterations allow for the creation of architecture content by cycling through the Business, Information Systems,

and Technology Architecture phases. These iterations also allow viability and feasibility tests to be carried out by looking at 'opportunities and migration planning'. Transition Planning iterations support the creation of formal change roadmaps for the defined architecture. Architecture Governance iterations support change management towards the defined Target Architecture.

5 Problems and Opportunities

A vast number of AFs is readily available. These AFs are large and they require both experience and effort to use them. A particularly big effort is need to master or adapt them for a certain area of uses. This effort investment shall be one-time. The possible advantage of having a domain specific AF for CI is that the most of typical concerns in the domain are pre-elaborated within the AF, and the most appropriate generic meta-model, a template model, viewpoints and methods are selected in the AF as design patterns, with the required capabilities to represent concerns and attributes to capture essentials, are already included. Additionally, AF advantages are (i) common terminology, validated by its use for the description of the characteristic problems, which are well known to the experts, (ii) common understanding by stakeholders of different origin and background; (iii) easy introduction of newcomers in the universe of domain, (iv) productive collaboration between organizations using the domain-specific AF within an industry sector.

What is more, the domain-tailored AF can set out some new paradigm regarding the enterprise architecture modelling. An example of such an introduction is the net-centric architecting paradigm. DODAF and similar frameworks support net-centricity.

6 Conclusions

The goal of this paper was to present a CI as a certain kind of enterprise and to show that CI protection should be carried out in a systematic way with the support for enterprise architectures. In order to allow for the protection of a CI, the enterprise mission, its critical business processes, the organization structure as well as internal and external relationships must be precisely identified. In contrast to well-established but generic EAs, security concerns: confidentiality, integrity, availability, dependability, reliability, continuity and assurance are central in the architecting process and need to be represented with particular care. A resulting domain-specific AF shall become a structured collection of domain knowledge in the form of best practices, design patterns, and predefined, customized security viewpoints and models. Its special value is in that once completed, it can be used in a repeatable fashion, thus improving the efficiency of the architecting process. The guidelines, embedded in the AF shall push the architects to consider the important security aspects, while the embedded constraints should prevent them from making mistakes.

The article focused on the approach to derive an architecture framework for CI protection, through the CII protection, and explained the roles of the major AF elements. Some extensions have been proposed including a domain meta-model, sub-methodologies for architecture development, design patterns and predefined models.

The work described in this paper is not finalized and is currently underway in the FP7 project NI2S3 and in National Research Grant for Polish Ministry of Science and Education.

References

1. Brunner, E.M., Suter, M.: International CIIP Handbook 2008/2009, Center for Security Studies, ETZ Zurich (2008)
2. Integrated Enterprise-Wide Risk Management, Special publication NIST 800-39 (2010)
3. Wolthusen, S.D.: Modeling critical infrastructure requirements. In: Proceedings from the Fifth Annual IEEE SMC Information Assurance Workshop (2004)
4. Tang, A., Han, J., Chen, P.: A Comparative Analysis of Architecture Frameworks. In: 11th Asia-Pacific Software Engineering Conference (APSEC 2004)
5. Lim, N., Lee, T., Park, S.: A Comparative Analysis of Enterprise Architecture Frameworks based on EA Quality Attributes. In: 10th ACIS, SNPD 2009 (2009)
6. Sessions, R.: A comparison of the Top Four Enterprise-Architecture Methodologies (2007), http://msdn.microsoft.com/en-us/library/bb466232.aspx
7. Every Architecture Description Needs a Framework: Expressing Architecture Frameworks Using ISO/IEC 42010, D.Emery, R. Hillard, IEEE/IFIP WICSA/ECSA (2009)
8. Broy, M., Gleirscher, M., Merenda, S., Wild, D., Kluge, P., Krenzer, W.: Towards a Holistic and Standardized Automotive Architecture Description. In: IEEE Computer (December 2009)
9. Network Information and Integration Services for Security Systems FP7 EU Project, http://ni2s3-project.eu
10. Castorini, E., Palazzari, P., Tofani, A., Servillo, P.: Ontological Framework to Model Critical Infrastructures and their Interdependencies. In: Complexity in Engineering, COMPENG 2010, February 22-24, pp. 91–93 (2010)
11. TOGAF 9 and DODAF 2.0, The Open Group White paper
12. Patterns for e-business, https://www.ibm.com/developerworks/patterns/map.html

Prototypes of a Web System for Citizen Provided Information, Automatic Knowledge Extraction, Knowledge Management and GIS Integration*

Antoni Ligęza, Weronika T. Adrian, Sebastian Ernst, Grzegorz J. Nalepa, Marcin Szpyrka, Michał Czapko, Paweł Grzesiak, and Marcin Krzych

Institute of Automatics,
AGH University of Science and Technology,
Al. Mickiewicza 30, 30-059 Kraków, Poland
{ligeza,wta,ernst,gjn,mszpyrka}@agh.edu.pl

Abstract. This paper presents preliminary results of work conducted within the INDECT project, Deliverable D4.14. The main goal of this deliverable is to develop a web system for acquisition of information from citizens, automatic knowledge extraction and knowledge management with GIS features. The paper presents three prototype systems and outlines their functionality. Social features have been implemented to encourage users to actively participate in data acquisition. Each of the presented systems offers a slightly different functionality and is based on a somewhat different conceptual model. The prototypes have been developed with the use of various software development technologies and tools for practical evaluation of each of the possible approaches. The results presented here constitute the basis for the implementation of the final system.

Keywords: security, citizens, GIS, knowledge management, INDECT.

1 Introduction

Local safety of citizens in urban environments is an issue of great importance to authorities, police and citizens themselves. While information systems, such as Crime Mapping[1], by means of which the police inform people about potential threats exist, there seems to be a niche for a social networking service of a similar functionality. Examples of existing social network systems show that people are willing to cooperate and share information when the subject is important to them and the application user interface is easily understandable. Therefore, it is desirable to investigate how modern technologies, including Geographic Informaiton Systems (GIS) can be used to develop such a system.

* The research presented in this paper is carried out within the EU FP7 INDECT Project: "Intelligent information system supporting observation, searching and detection for security of citizens in urban environment" (http://indect-project.eu).

[1] See http://www.crimemapping.com

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 268–276, 2011.

This paper presents the results of work aimed at design and implementation of a web system capable of collaborative acquisition, storage, sharing and analysis of knowledge regarding threats to public security in urban environments. The adopted concept assumes that the main source of the knowledge is citizen-provided information that can be submitted, retrieved and evaluated via a web interface. In order to improve information exchange and collaboration between individuals using the system, the prototypes have been supplemented with social features. This approach encourages citizens to actively cooperate, by showing that their contribution is of great significance and actually helps in increasing safety of their families and friends. However, this is only possible if the public services are able to make use of the information and provide users with useful and reliable feedback.

The basic assumptions and requirements for the system have been presented in [4]. These include: (i) willingness of citizens to share information about dangers with others, (ii) collaborative rating of the credibility, usefulness and importance of the threats entered into the system, (iii) possibility of entering information in various formats (text, photos and geographical coordinates), and (iv) geographic data presentation independent of the mapping system (Google Maps, Yahoo Maps and others). The general idea of the system can be observed in Figure 1.

The input data, in general, may be composed of: a text description of a threat, its spatial (2D) location, and visual (multimedia) documentation. The data, stored in a relational database equipped with spatial features should be presented to the audience in a combined visual and textual form. The system should provide means for searching, filtering, aggregation and grouping of data for final users, according to their preferred form and level of detail. The primary use case is based on the assumption that the user can perceive and analyze threats located within an area of interest. The threats should be presented in a convenient and transparent way as icons located on the map.

The functional requirements defined for the system can be summarised as follows: (i) threat data submission and management (ii) map-based information sharing and visualisation, (iii) data analysis using predefined reports, (iv) advanced search capabilities, (v) multiple categories of users, (vi) user evaluation framework to assess credibility of submitted threats, (vii) a notification system, including alerts for significant dangers submitted by credible users, (viii) a newsletter, with attribute- and location-based customisation, (ix) a rule-based engine to facilitate data administration and report generation, (x) integration with existing social network applications, (xi) user authentication with open authentication frameworks, (xii) support for mobile platforms.

Three independent prototype systems have been implemented and deployed within the project:

- INDECT Threat Monitor [2],
- Threat Radar [3],
- Safety Protect [1].

The systems offer slightly different functionalities, present a somewhat different conceptual model, and have been developed with use of different software

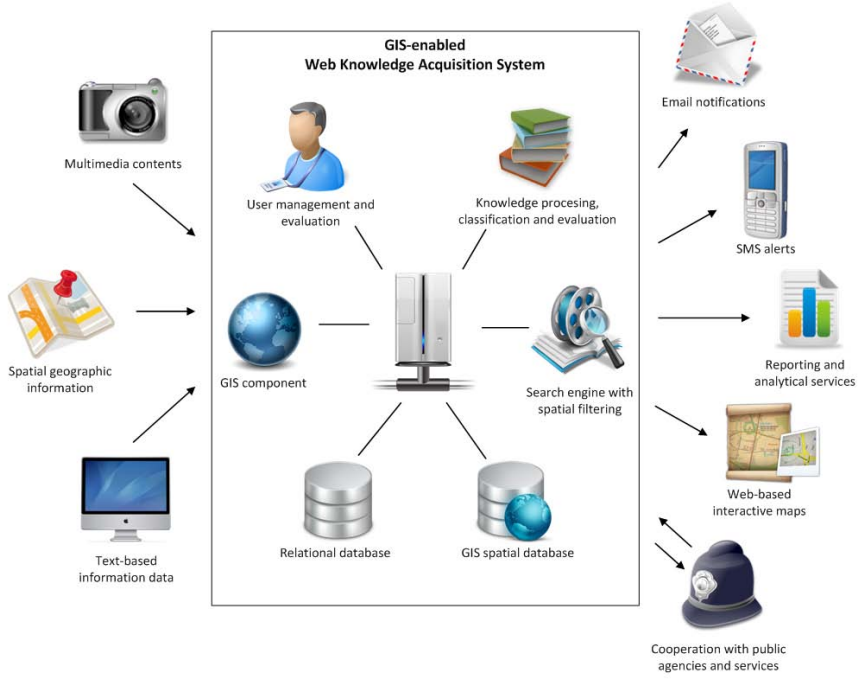


Fig. 1. Conceptual model of the system

development technologies and tools. These prototypes are aimed at further testing before development of the final system. The systems are described in detail in the following sections.

2 Functionality of the Prototype Systems

In this Section, two of the three developed prototypes are described. The systems constitute publicly accessible social portals presenting potential threats and their location on a map. The threats are stored in a spatially-enabled relational database; each threat has spatial coordinates (point, line, or polygon). A registered user can add new threats and their characteristics.

2.1 INDECT Threat Monitor

INDECT Threat Monitor [2] is available at the following location: `http://marvin.ia.agh.edu.pl/indect/pgr/`. Its interface is in English, but it is possible to implement additional language versions. A screenshot of the system, with example threats, may be observed in Fig. 2. The system fulfils the crucial functional requirements defined in Section 1.

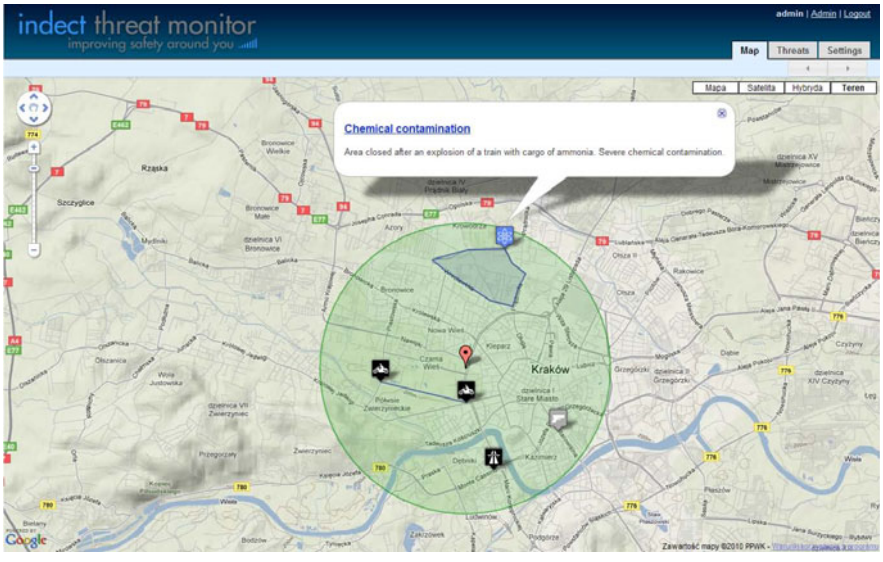


Fig. 2. Screen shot of the INDECT Threat Monitor system

User accounts and account management. The end user of the system is an authenticated citizen with a system account and access to the Internet. Each user is able to manage their own account: set the language and time zone preferences, add an avatar image or fill the profile with personal information such as address, phone number, gender or date of birth.

System users categories. The users are divided into the following categories: *guest*, *member*, *editor* and *administrator*. A *guest* is an anonymous user with restricted system access. A *member* is a registered user who can view the map, create and manage own threats, search and view threats added by others, edit the profile, upload photos, comment and vote on existing threats. An *editor* can create emergency threats and controls threat records, while an *administrator* has full control over the system.

User evaluation. The credibility and contribution rating for each user is determined using several criteria, such as the number and relevance of submitted threats and votes received from other users.

Open authentication. Aside from the standard authentication method with login and password, the system provides full support for decentralised open authentication standards such as OpenID[2]. It allows users to log onto different web platforms and services with the same digital identity, stored and provided by the so-called OpenID providers.

Visual threat representation and threat management. The system provides a complete solution for data representation, submission and management. The focus is on the visual presentation of threats with the use of a map service.

[2] See `http://www.openid.net/`

Threat classification. Further analysis and interpretation of the gathered knowledge requires application of a unified classification scheme. Users are supposed to assign one or more predefined categories (tags) to newly created threats.
GIS component. Every piece of knowledge gathered in the system is supplemented by geographic coordinates, representing the shape and the location of the danger. This enables more precise analysis and extremely effective, map-based knowledge sharing and visualisation.
Notification system Users can define custom notification rules for receiving e-mail and SMS alerts. The rule editor allows creating and joining multiple conditions based on threat geographical location, severity and category.
Voting system. Threats submitted into the system can be assessed by other registered users in terms of usefulness, importance and credibility. This can be performed by a voting scheme for registered threats.
Commenting system. Registered users can discuss threats submitted into the system by writing comments.
Tagging system. Any threat can be assigned multiple tags. A tag is a term assigned to a piece of information or data. This kind of metadata helps to describe and classify items and facilitates searching and browsing.
Search engine. It is crucial to design an advanced and efficient search engine, which allows users to find threats given standard criteria (such as date, name, description, category or creator) and provide GIS features to enable spatial searching and filtering based on a geographic area selected on a visual map component. At the moment, searching by name, credibility points and date is supported, while location-based search has not yet been implemented.
Data storage. The knowledge is stored in a high-performance spatially-enabled relational database. It guarantees rapid data access and processing, thanks to implemented GIS features such as geographic querying and indexing.
Internationalization and localization. The system is designed and implemented in a way that allows itself to adapt to various languages and regions without engineering changes. These requirements are fulfilled by means of configurable drop-down lists of available languages and time zones.
Social network integration. Integration with existing social web networks such as Facebook (http://www.facebook.com) or Twitter (http://www.twitter.com) would result in worldwide and fast propagation of the idea that stands behind the INDECT project and, what is probably even more important, increased number of active users and therefore increased effectiveness and performance of the system.
Site management. Individuals responsible for system administration and maintenance are equipped with a powerful and flexible tool for control and management of the website contents.

2.2 Threat Radar

Threat Radar [3] is another system for threat registration and monitoring, installed at: `http://marvin.ia.agh.edu.pl/indect/mkr/`. The system interface

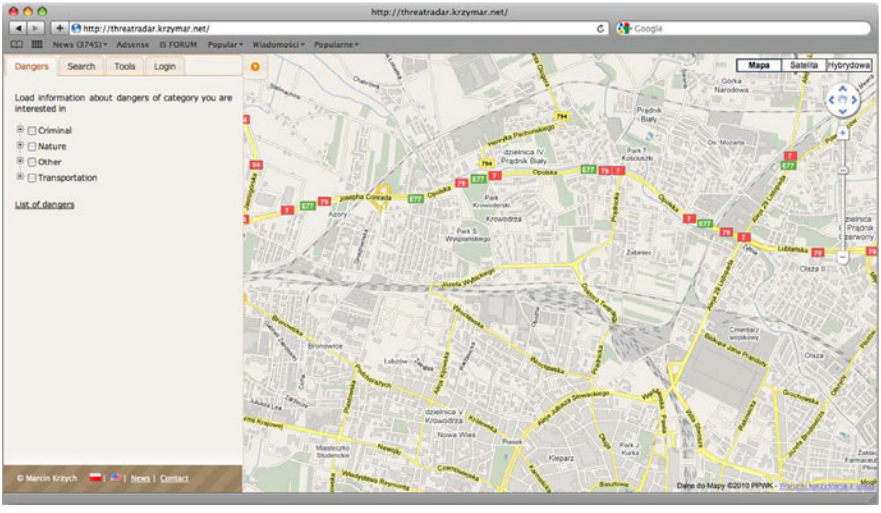


Fig. 3. Screen shot of the Threat Radar system

is in Polish and English; it is possible for a user to switch between these two languages. An example screenshot can be observed in Figure 3.

Three user categories have been defined: *user*, *system administrator* and *staff member*. Based on the category, a user has access to particular functionalities. For unregistered users (Guests), the following features are available:

Data browsing. The main page consist of the map and the function panel placed on the left. The function panel can be hidden so that the map can be as big as screen resolution allows it to be. This makes it easier to browse the map, navigate to other places or adapt zoom.

Data filtering. All threat information can be filtered by category. This can be done with use of the category tree in the function panel. Both category groups and single categories can be selected. Requested information is presented on the map. When user navigates to the other places, application presents dangers which are located in the currently chosen area.

Threat detail display. User can click on each icon on the map at the main page to see the details of a threat. An info cloud will show up with the threat name and its category. User can see the following details of the threat: description, category, comments and rates from users and photo gallery.

Data searching. The search method which is vital is the search for an address. There is no need for searching for dangers, because while browsing the map users can easily filter dangers by category, so that they can see only those threat information they want. Searching for an address allows users for quick navigation from one place to another without time consuming map operations. In order to search for an address, the user has to fill in a dedicated form.

User registration. Apart from standard registration procedure, it is possible to create a profile by submitting information about danger and providing an

e-mail address. In this case, upon e-mail validation, a new user is registered and the entered information is linked to this new created profile.
Map position saving. User can save the current position of the map if they know they may need it in the future.

Registered users (Members) have access to these additional functionalities:
Adding new threat information. In order to add information about new threat a user has to be logged in or should provide their e-mail address in the form with threat data. In order to add a point or a space, the user clicks a proper icon and then choose a place on the map to locate the danger. Afterwards, the user can specify the details of the submitted threat.
Commenting and rating. Each threat can be commented and rated. To add a comment and rate the information, user must fill a dedicated form. Comments and rating are added to the threat information details.
Photo submission. Photos presenting the place connected with the area of threat can be added to each entry. Up to four photographs can be submitted at a time.

As for the staff members, the system offers the following functions:
Alert Panel. Alerts are used to mark important events in the system. Their main objective is to signal that a severe threat from a reliable user has been submitted. Alerts are generated by the rule engine according to rules.
Rule Engine Management. Rule engine management is performed in the data grid component (see [3] for details). Authorised users may build rules out of possible components with use of a special form. Afterwards, they define options for conditions and actions chosen in the previous step.
Reports. Reports are designed to present the gathered data in an aggregated form. The application provides some predefined reports available for the user. After choosing the type of the report, the user has to configure the report criteria.

Administrators have access to all application functions available for members and guests. Moreover, the are some Application management functionalities. The main administrative areas include: data import, export and management, category management, user accounts management, news management, and internationalization.

3 Implementation and Deployment

The prototypes have been implemented with the use of different software technologies and tools.

The INDECT Threat Monitor system combines the features provided by the *Django Web Framework*[3], *Google Maps*[4] and *PostGIS*[5] . The major advantage of the proposed set of tools and technologies is that it allows developers to build reliable, high-performance applications fast and lets them focus on the

[3] `http://www.djangoproject.com/`
[4] `http://maps.google.com`
[5] `http://postgis.refractions.net/`

INDECT-specific parts of the project. The other important conclusion is that no problems or difficulties have occurred during system development and component integration.

Threat Radar uses a number of technologies working on the client and server side. The former include XHTML, CSS, JavaScript libraries: *JQuery*[6] and *JQuery UI*[7] and *Google Maps*[8], chosen due to its internationalization and popularity. The latter include the Apache web server[9], PHP-based Zend framework[10] and Maestro – a set of components for web applications. PostgreSQL[11] with the PostGIS extension has been chosen as the data storage backend.

The Safety Project is based on the Ruby on Rails[12] framework and uses the Phusion Passenger™[13] Rails deployment technology. The web server used here is also Apache, via the `mod_passenger` module. The installation of three prototypes has been completed. A dedicated server has been set up to host the developed system prototypes. It is set up as a virtual machine using the Oracle VirtualBox[14] virtualisation technology. The host machine configuration is as follows: Eight (8) Intel(R) Xeon(R) E5430 CPUs @ 2.66GHz, 24GB of ECC RAM, 293.3 GB of storage using a hardware RAID controller, Debian GNU/Linux 5.0.7 ("Lenny") with 2.6.32-bpo.5 kernel, Oracle VirtualBox 3.2.6.

The host machine for prototype deployment platform features a stringent security policy: Only HTTP/S (TCP ports 80, 443) is publicly available, Secure Shell (sshd; TCP port 22) is available for specific users logging in from specific remote addresses.

As the host machine features no graphical UI nor dedicated displays, the dedicated guest machine has been set up in headless mode, i.e. the system console is available remotely via the network, using the VirtualBox Remote Display Protocol (VRDP), available on port 3389 of the host machine. Low-level security of the guest machine is guaranteed by the stringent security policy of the host machine – the users' machine has to be entitled to access the host machine via SSH, and a SSH tunnel has to be set up to access the TCP port 3389 of the host machine.

The INDECT server is running the Ubuntu Server Ubuntu 10.04.1 LTS (Lucid Lynx) operating system (supported until April 2015), as the package repository features most software required by the prototypes, this facilitating maintenance and management.

The server is available via a separate, public IP address/hostname, with an external firewall allowing only connection on TCP ports 22, 80 and 443. The following server software components are available: Apache 2.2.14 HTTP Server, PHP

[6] `http://jquery.com/`
[7] `http://jqueryui.com/`
[8] `http://maps.google.com/`
[9] `http://httpd.apache.org/`
[10] `http://framework.zend.com/`
[11] `http://www.postgresql.org/`
[12] `http://rubyonrails.org/`
[13] `http://www.modrails.com/`
[14] `http://www.virtualbox.org/`

5.3.2 interpreter, MySQL 5.1.41 database server, PostgreSQL 8.4.5 database server, PostGIS 1.4.0 geo-referenced data extension for PostgreSQL, Postfix 2.7.0 mail transport agent, QNotifier 1.0.1 remote monitoring software, Python 2.6.5 interpreter, OpenJDK Java Runtime Environment, database management tools.

4 Summary

The prototypes presented in the paper constitute a novel approach to the process of acquisition, management, processing and visualisation of threat-related data. The novelty of the approach mainly consists in application of a social network concept and providing users with a friendly yet comprehensive information environment. To the best of our knowledge no similar threat information systems exist. The prototype implementations serve as a proof-of-concept and are the basis for the development of the final release of the system.

The future improvements and development works will focus on advanced threat categorisation, automatic evaluation of submitted data, data warehousing and statistic services for public law enforcement and regular users, and support for mobile devices.

References

1. Czapko, M.: Design and Implementation of a Web-based Acquisition and Knowledge Management System for the Safety and Security Needs with the GIS Module. Master's thesis, AGH University of Science and Technology (2010)
2. Grzesiak, P.: Analysis of Available GIS Technologies and Tools and Project and Implementation of a Prototype Public Security Support System. Master's thesis, AGH University of Science and Technology (2010)
3. Krzych, M.: Project and Implementation of an Internet Danger Information Acquisition and Knowledge Management System with GIS Module. Master's thesis, AGH University of Science and Technology (2010)
4. Ligęza, A., Ernst, S., Nowaczyk, S., Nalepa, G.J., Furmańska, W.T., Czapko, M., Grzesiak, P., Kałuża, M., Krzych, M.: Towards enregistration of threats in urban environments: practical consideration for a GIS-enabled web knowledge acquisition system. In: Dańda, J., Jan Derkacz, A.G. (eds.) IEEE International Conference on Multimedia Communications, Services and Security, MCSS 2010, Kraków, May 6-7, pp. 152–158 (2010)

Human Tracking in Multi-camera Visual Surveillance System

Piotr Marcinkowski, Adam Korzeniewski, and Andrzej Czyzewski

Multimedia Systems Department, Gdansk University of Technology
Narutowicza 11/12, 80-233, Gdansk, Poland
{pmarcin,adamkorz,andcz}@sound.eti.pg.gda.pl

Abstract. A short survey of visual human tracking technologies used in intelligent surveillance systems is presented. Face recognition algorithms combined with human tracking systems are not meant to identify human face and personality. There is no database with persons' biometric features employed, thus in this case there is no problem with violating privacy policy. The concept of combining human tracking technology with face recognition techniques, in order to increase efficiency, has been described. The paper also includes the description of KASKADA - hardware and software supercomputer platform for development of multimodal (audio and video) algorithms, including object and person tracking video monitoring systems. Face recognition algorithm on the KASKADA platform was proposed. Method of implementation of the proposed algorithm was described.

Keywords: visual surveillance, multi-camera systems, human tracking, face recognition.

1 Introduction

Nowadays, visual surveillance is an important tool not only for fighting crime and ensuring citizens' safety, but also for monitoring traffic, crowd behavior, etc. Management of visual surveillance systems can be difficult (especially with multiple cameras), thus it requires from operating person an ability to work with constant attention and focus. Application of computer vision analysis techniques provides automation which increases effectiveness and reduces workload of human operators.

This paper focuses on the human tracking technology using some specific Cloud computing methods. The Context Analysis of the Camera Data Streams for Alert Defining Applications platform (Polish abbreviation: KASKADA) is a technology platform of MAYDAY EURO 2012 project and is used to analyze media streams of different types. This paper reveals work on multi-camera vision streams. Authors in their work are using the KASKADA platform to provide automation in designing surveillance systems.

2 Human Tracking Technology

The main purpose of human tracking technology is detection of the person appearance in the monitored area of interest (AOI) and determining its movements. This

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 277–285, 2011.

information can be used to control access to certain zones, count people entering or leaving the area, detect loitering on a parking lot and many more.

Single-camera human tracking methods provide basic data which can be used to develop multi-camera systems. Lee *et al.* [1] proposed fusion-based tracking algorithm using optical flow under the non-prior training active feature model. Li *et al.* [2] used improved HOG (Histogram of Oriented Gradient) and Kalman filter for detection and tracking. HOG is also used in [3] combined with color histogram. Single-camera human tracking has been described in many others works in the literature [4], [5], [6]. Increasing number of cameras in surveillance systems and overall technical progress caused evolution of these methods. Single-camera human tracking methods were adapted to meet new requirements.

More sophisticated human tracking methods are used in multi-camera visual surveillance systems in which the human tracking algorithms transmit information on movement direction to locate the same person in different cameras. Multi-camera visual surveillance systems can be divided into overlapping or non-overlapping field of view (FOV) systems. The concept of overlapping fields of view (FOV) in multi-camera methods can be understood in two ways. From visual surveillance's point of view, overlapping FOV means that a tracked human is visible in at least one camera (Fig. 1). Khan *et al.* [7] presented system that is able to discover spatial relationships between cameras FOV. Using this information, when a person appears in one camera, the system is able to predict all other cameras in which this person will be visible. In another work [8], spatial and appearance features for object description and recognition were utilized. For a spatial description, an approximated object position in world coordinates is estimated and evaluated by an inconsistency detector before associated to a Kalman filter. Appearance similarity calculation using an appearance model and a

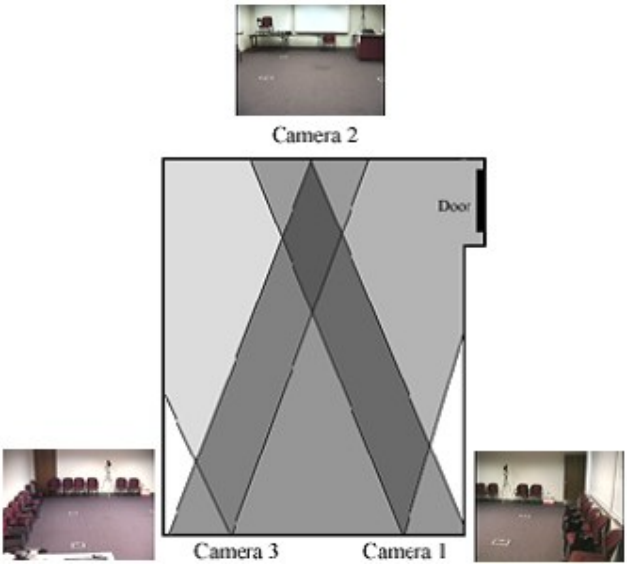


Fig. 1. Human tracking system with overlapping cameras' set FOV [7]

similarity metric is based on the Earth Mover's Distance. The paper referred as [9] studies a method to determine multi-cameras' common relay area (RA) and FOV lines and presents an adaptive blocked rule of the tracking target based on target features. It also proposes a target gradual handoff and a target relay tracking strategy in a common RA based on multisubblock features correlation matching.

Overlapping FOVs means also that the area of interest (AOI) is seen from many perspectives (Fig. 2). In this case, 2D images are processed for the purpose of extracting 3D information of a tracked human. Mohedano *et al.* [10] presented a 3D people positioning and tracking system. People detection is carried out through a template-based correlation strategy. Detected people are tracked independently in each camera view by means of a graph-based matching strategy. 3D tracking and positioning of people is achieved by geometrical consistency analysis over the tracked 2D candidates, using head position to increase robustness to foreground occlusions. A multi-view multi-hypothesis approach to segmenting and tracking multiple persons on a ground plane was proposed in the paper [11]. In this approach, during tracking, several iterations of segmentation are performed using information from human appearance models and ground plane homography.

In case of a large AOI, its fully coverage is difficult and expensive. Usually, in such situations less cameras are applied, covering only the most important subareas. This solution is the cause of blind regions in AOI and tracking discontinuity. In order to solve this problem, many approaches of non-overlapping FOV (Fig. 3) have been developed. An automated surveillance system, performing human detection and tracking across multiple non-overlapping cameras, was proposed by Mehmood *et al.* [12]. In this approach, emphasis is put on a single camera level where motion-based segmentation is achieved using optical flow estimation. Feature matching and region-based shape descriptors are used for tracking. Color Brightness Transfer Function is used in the mentioned paper [13]. Wang *et al.* [14] presented real-time distributed system using multiple features (color histogram, height, speed) in which cameras communicate with each other in a peer-to-peer manner. A new method, that uses multiple features that are dynamically weighed for matching moving people across cameras, was proposed by Montcalm *et al.* [15]. In particular, the Zernike moment shape descriptor has been used together with blob histogram and other features to describe a moving object. A higher weight is given to more suitable features, based on

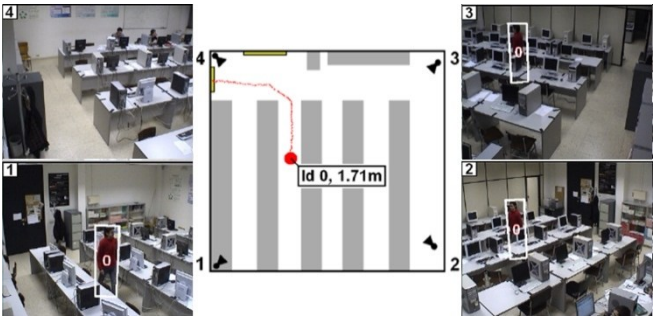


Fig. 2. Human tracking system where AOI is seen from many views [10]

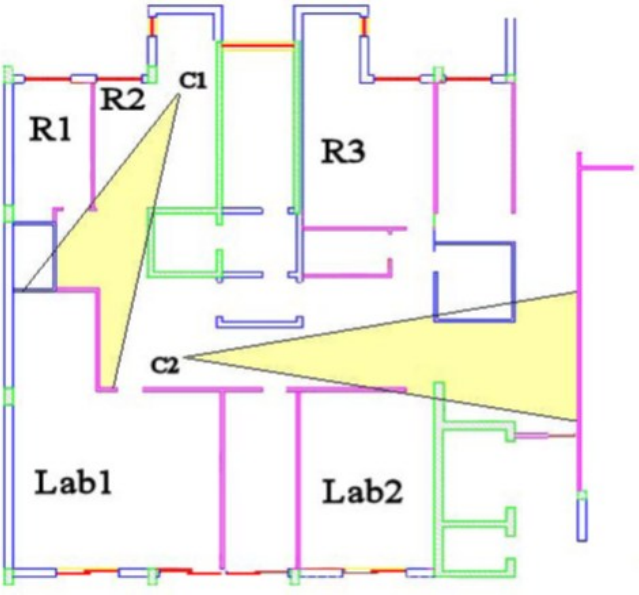


Fig. 3. Human tracking without overlapping cameras' set FOV [13]

their stability, reliability and their life-time in the system. This weighting is used both during appearance aggregation and object comparison. Other approaches have been also described in the literature [16][17][18].

3 Human Tracking Supported by Face Recognition Techniques

Face recognition techniques, similarly to human tracking, are dynamically evolving. Every year some new, more sophisticated and efficient face recognition algorithms emerge to meet higher requirements. There is a need for such algorithms in human tracking systems to increase their efficiency.

First it should be stressed that face recognition algorithms combined with human tracking systems are not meant to identify human face and personality. There is no database with persons' biometric features, thus in this case there is no problem with violating privacy policy. The system uses only biometric features of persons in its FOV and there is no connection with any personal information.

The main purpose of using recognition algorithms is to keep track of the same person in multi-camera visual surveillance system. The problem of accurate face recognition is specific. To meet the algorithm's requirements regarding data acquisition a precise camera positioning is needed. There is a useful technique proposed by Shen *et al.* [19] to measure Quality-Of-View (QOV) of the cameras' set to adjust them properly. Human tracking systems utilizing face recognition algorithms might be appropriate when standard tracking techniques fail. A good example of combining those methods is human tracking in an underpass, hall, airport etc. Other suitable situations are those with big crowd density in which typical tracking algorithms encounter problems with discrimination of a single person.

4 KASKADA Supercomputer Platform

Creating and executing more and more sophisticated algorithms requires flexible and powerful platform. KASKADA is an actively developed supercomputer platform of context-depended analysis of multimedia data streams. Its task is to identify specified objects or safety threats. Algorithms and services based on the platform are developed to create three main applications: protection of intellectual property (plagiarism checking application), support of medical examination (diseases detecting application) and recognition of people and events (visual surveillance application). The platform is developed in order to increase safety and public trust. Future end-users of the platform and its applications could be: scientists, students, IT software developers, medical facilities, local governments, entertainment organizers, security services.

The hardware layer of KASKADA platform is a computing cluster built of 192 processing nodes (reaching performance of 20 TFlops) and 12 mass storage nodes of 512TB capacity. The platform consists of:

- database server (12 nodes),
- parallel-processing management server (189 nodes),
- notifications server (1 node),
- multimedia streams management server (2 nodes),

The development framework of the platform includes tools for easy implementing and testing various algorithms on the computing cluster. These tools are represented in C++ libraries. The KASKADA also provides a GUI to:

- merge algorithms into services,
- manage streams on platform side,
- create own benchmark streams,
- access and filter events generated by algorithms,
- monitor and manage running services.

Additionally a part of the KASKADA platform forms the network of multimodal servers featuring digital video cameras and audio sensors that transmit multimedia streams to the platform for analyzing.

The architecture of KASKADA platform and the schema of launching services are shown in Fig 4.

5 Implementation of Face Recognition Algorithm on KASKADA Platform

In order to implement a basic face recognition algorithm, the modified Elastic Bound Graph Matching (EBGM) [20] has been implemented on the KASKADA platform. Original EBGM has been altered by changing decomposition method and face detection method. The detection of face is carried out by trained Haar cascade classifier. The decomposition is carried out by LogGabor transform (implementation of Kovesi's function [21] in C++). The algorithm is presented in Fig. 5.

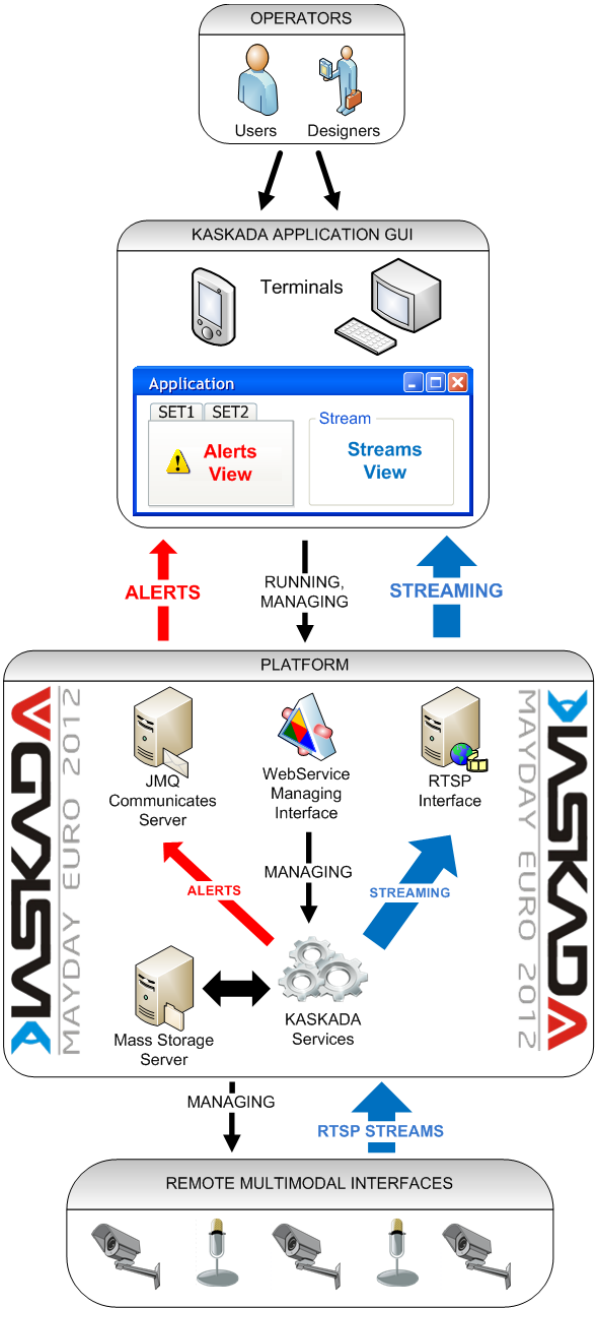


Fig. 4. Architecture and services schema on the KASKADA platform

Every algorithm's step is designed to work in multithread environment and divided into tasks. Tasks are organized in a queue and processed by thread pool. As soon as a thread completes its task, it will request the next task from the queue.

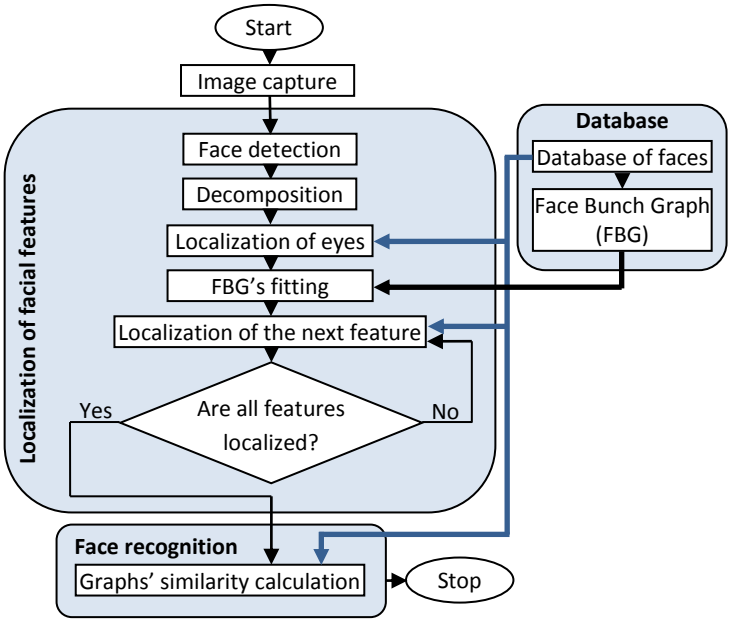


Fig. 5. Modified Elastic Bound Graph Matching algorithm

In the first step of the algorithm a face is detected in captured image. As previously mentioned, Haar cascade classifier is used. The next step involves face's image decomposition using LogGabor transform. The transform performs convolution of the input image and 40 complex wavelets (1). After the decomposition, the position of eyes is calculated. Next, Face Bunch Graph (the faces representation containing information on average of facial features position and graph's edges length) is fitted to the localized eyes. Using FGB data, the algorithm is able to compute the region of interest (ROI) of a localized feature. Localization of the next node is performed by similarity calculation (2). The last step involves calculation of similarity between localized features and biometric templates in a database (3).

$$J_j(\vec{x}) = \int I(\vec{x}')\psi_j(\vec{x}-\vec{x}')d^2\overrightarrow{x'}, \vec{x} = (x,y), j = 0,1,\ldots,39\ , \tag{1}$$

where *jet* $J_j(\vec{x})$ is description of gray values in an image $I(\vec{x}')$ around a given pixel $\vec{x} = (x,y)$ and $\psi_j(\vec{x})$ is LogGabor wavelet.

$$S(J,J')_\varphi = \frac{\sum_j a_j a'_j \cos(\varphi_j - \varphi'_j)}{\sqrt{\sum_j a_j \sum_j a'_j}}, \tag{2}$$

where similarity of two jets $S(J,J')_{\varphi}$ is calculated using magnitude $a_j$and phase φ_j.

$$S_G(G^I,G^M)=\frac{1}{N}\sum_n S_a(J_n^I,J_n^M),\ S(J,J')_a=\frac{\sum_j a_j a'_j}{\sqrt{\sum_j a_j \sum_j a'_j}}, \tag{3}$$

where similarity of two graphs $S_G(G^I,G^M)$ is calculated using similarity function $S(J,J')_a$ for every graph's node N.

6 Future Work

Future work will be focused on implementation of selected human tracking algorithms and on development of existing face recognition algorithms on the KASKADA supercomputer platform. Because of the high performance of the platforms' hardware, more complex algorithms can be implemented for reliable results. The variety of implemented face recognition and tracking methods will allow to increase efficiency of human tracking in multi-camera visual surveillance system.

Another task is to create the dedicated benchmark framework to test and compare performance of all services. Multimodal servers network will be expanded to assure visual data acquisition. It will be used in face recognition and tracking algorithms.

Acknowledgments

Research funded within the project No. POIG.02.03.03-00-008/08, entitled "MAYDAY EURO 2012 - the supercomputer platform of context-depended analysis of multimedia data streams for identifying specified objects or safety threads". The project is subsidized by the European regional development fund and by the Polish State budget.

References

1. Lee, J., Kim, S., Kim, D., Shin, J., Paik, J.: Feature Fusion-Based Multiple People Tracking. In: Ho, Y.-S., Kim, H.-J. (eds.) PCM 2005. LNCS, vol. 3767, pp. 843–853. Springer, Heidelberg (2005)
2. Li, C., Guo, L., Hu, Y.: A New Method Combining HOG and Kalman Filter for Video-based Human Detection and Tracking. In: 3rd International Congress on Image and Signal Processing (2010)
3. Jin, L., Cheng, J., Huang, H.: Human tracking in the complicated background by Particle Filter using color-histogram and HOG. In: International Symposium on Intelligent Signal Processing and Communication Systems (2010)
4. Benezeth, Y., Emile, B., Laurent, H., Rosenberger, C.: Vision-Based System for Human Detection and Tracking in Indoor Environment. International Journal of Social Robotics 2(1), 41–52 (2010)
5. Zhou, J., Hoang, J.: Real Time Robust Human Detection and Tracking System. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops (2005)

6. Rowe, D., Reid, I., Gonzàlez, J., Villanueva, J.J.: Unconstrained Multiple-People Tracking. In: Franke, K., Müller, K.-R., Nickolay, B., Schäfer, R. (eds.) DAGM 2006. LNCS, vol. 4174, pp. 505–514. Springer, Heidelberg (2006)
7. Khan, S., Javed, O., Rasheed, Z., Shah, M.: Human Tracking in Multiple Cameras. In: 8th International Conference on Computer Vision (2001)
8. Monari, E., Maerker, J., Kroschel, K.: A Robust and Efficient Approach for Human Tracking in Multi-camera Systems. In: 6th IEEE International Conference on Advanced Video and Signal Based Surveillance, September 2-4, pp. 134–138 (2009)
9. Chang, F., Zhang, G., Zhang, P., Li, J.: Multi-camera Target Relay Tracking Strategy Based on Multi-Subblock Feature Matching. In: 8th World Congress on Intelligent Control and Automation, pp. 6229–6233 (2010)
10. Mohedano, R., del-Bianco, C.R., Jaureguizar, E., Salgado, L., Garcia, N.: Robust 3D People Tracking and Positioning System in a Semi-overlapped Multi-camera Environment. In: 15th IEEE International Conference on Image Processing, pp. 2656–2659 (2008)
11. Kim, K., Davis, L.S.: Multi-camera Tracking and Segmentation of Occluded People on Ground Plane Using Search-Guided Particle Filtering. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 98–109. Springer, Heidelberg (2006)
12. Mehmood, M.O., Khawaja, A.: Multi-camera Based Human Tracking with Non-overlapping Fields of View. In: 5th International Conference on Image and Graphics, pp. 313–318 (2009)
13. D'Orazio, T., Mazzeo, P.L., Spagnolo, P.: Color Brightness Transfer Function Evaluation for Non overlapping Multi Camera Tracking. In: 3rd ACM/IEEE International Conference on Distributed Smart Cameras (2009)
14. Wang, Y., He, L., Velipasalar, S.: Real-time Distributed Tracking with Non-overlapping Cameras. In: 17th IEEE International Conference on Image Processing, pp. 697–700 (2010)
15. Montcalm, T., Montcalm, T.: Object Inter-camera Tracking with Non-overlapping Views: A New Dynamic Approach. In: Canadian Conference on Computer and Robot Visio (2010)
16. Zhu, L.J., Hwang, J.N., Cheng, H.-Y.: Tracking of Multiple Objects Across Multiple Cameras with Overlapping and Non-overlapping Views. In: IEEE International Symposium on Circuits and Systems, pp. 1056–1060 (2009)
17. Lim, F.L., Leoputra, W., Tan, T.: Non-overlapping Distributed Tracking System Utilizing Particle Filter. The Journal of VLSI Signal Processing 49(3), 343–362 (2007)
18. Pflugfelder, R., Bischof, H.: Tracking Across Non-overlapping Views via Geometry. In: 19th International Conference on Pattern Recognition (2008)
19. Shen, C., Zhang, C., Fels, S.: A Multi-Camera Surveillance System that Estimates Quality-of-View Measurement. In: IEEE International Conference on Image Processing, vol. 3, pp. 193–196 (2007)
20. Wiskott, L., Fellous, J.M., Krüger, N., von der Malsburg, C.: Face Recognition by Elastic Bunch Graph Matching. In: Intelligent Biometric Techniques in Fingerprint and Face Recognition, pp. 355–396. CRC Press, Boca Raton (1999)
21. Kovesi's MATLAB and Octave Functions for Computer Vision and Image Processing, (03/10/2011), `http://www.csse.uwa.edu.au/~pk/Research/MatlabFns/index.html`

Quantum Cryptography Protocol Simulator

Marcin Niemiec, Łukasz Romański, and Marcin Święty

AGH University of Science and Technology
Al. Mickiewicza 30, 30-059 Krakow
niemiec@kt.agh.edu.pl,
{lukasz.romanski,marcin.swiety}@wp.eu

Abstract. This paper presents newly developed application for evaluation and testing of quantum cryptography protocols. At the beginning the reason of creation as well as basics of the quantum mechanics are provided. Successive sections of this article show the design and building blocks of application. Two use cases are proposed as well. At the end, reader is provided with short demonstration of simulator's work in the means of exemplary results obtained through quantum protocol simulation execution.

Keywords: security, quantum cryptography, quantum key distribution, simulation.

1 Introduction

Quantum cryptography (or more precisely *Quantum Key Distribution*—QKD) was introduced in 1984 by Bennett and Brassard [4] and independently by Ekert in 1990 but from a little different perspective [1]. Its most important and unique feature is, that based on laws of nature (physics), it can always assure faultless detection of eavesdropping. Thanks to that, we are able to make sure that no one knows anything about transmitted message, which similarly to asymmetric cryptography is usually the key for symmetric ciphers.

However, asymmetric cryptography is rather slow and it is not determined that it is robust to technology changes [7], i.e. improvement in computational performance of computers. Thus, it is very inconvenient to use it with one-time-pad [5] cipher for *perfect secrecy*, knowing that in some time our data could be read regardless of our effort in securing that content. Quantum Cryptography does not share this disadvantage which opens the possibility to use this cipher and finally obtain *perfect secrecy* property in practically realizable and cost-effective cryptographic scheme.The main advantages of quantum cryptography are presented below.

- It assures certain detection of eavesdropping.
- As based on physics laws, it is completely secure against any technological improvements (in means of computational performance).
- It is faster than most popular asymmetric cryptographic systems, thus we are finally able to use one-time pad cipher.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 286–292, 2011.

– Constant improvements are being made in the field of practical realizations and equipment to overcome technological vulnerabilities and make the security of QKD schemes device-independent.

The security of quantum cryptography is based on two principles: *Heisenberg uncertainty principle* and *no-cloning theorem* [3][9].

Heisenberg uncertainty principle states that it is impossible to measure two uncorrelated physical quantities with the same, assumed in advance precision. It simply tells us that having measured a position of a particle in some dimension, we are able to measure its momentum along this dimension, with only limited precision.

Another theory, which plays an important role in quantum cryptography, is a theory about the impossibility of cloning the quantum states. It says that it is impossible to create a device that would be able to clone state of *qubit* (quantum bit), without any change to its original state. This is a big advantage of quantum systems, because it prevents from passive eavesdropping.

More detailed information about the quantum key distribution schemes can be read in [7], [4]

2 Reasoning

Methods of assessing the correctness and accuracy of work associated with the development of quantum cryptography protocols are highly important during the designing phase. They provide results that are essential in the analysis of current work. The development process is based on a repeated cycle consisting of three stages. These are: implementation, testing, and improving. Thanks to the repetition of the cycle, the designer is able to continuously verify project progress and assess the significance of changes introduced in the previous iteration. For reliable evaluation of quantum cryptography protocols, a new simulator was developed.

This application simulates operation of quantum key distribution protocol. The main tasks of simulator are validation of quantum cryptography protocol's efficiency in the means of security it is offering as well as delivery of the detailed protocol's results at each step of its operation. Simulator is expected to provide information regarding the transmission of key that is determined by many parameters which define transmission channel, eavesdropping type etc.

In order to clarify the purpose of new application creation, the following criteria were fixed.

– **Deterministic processing.** Application should operate rather on deterministic than probabilistic conditions. It is due to the fact, that in analyzing some conditions that have been assumed, the results should correspond directly to these conditions;
– **Configurability.** Interface should allow user to model every step of the protocol with different values in order to provide a clear and multi-scenario analysis means;

- **High level of simulation.** For development purpose of novel QKD protocol, what basically boils down to new reconciliation scheme, the binary level was assumed. Thus, simulation occurs on binary values. Each channel is modeled also as binary channel;
- **Generation of detailed reports.** That is reports containing data gained on every step of algorithm. Assumed format was textual for the report, and *Comma Separated Values* (CSV) for series of simulations;
- **Two modes of operation.** Application should provide way to simulate single and specific variant of configuration. Moreover, it should allow user to simulate protocol through continuous iteration of one parameter;
- **Professional programming techniques.** That is modularity, multithreading and many other for ease of further development and stability of operation;
- **User friendly interface.** That is graphical user interface with intuitive labels and input elements;

Moreover, proposed solution introduced other properties and features. These are listed below:

- Modular design with correlation to successive phases of QKD protocol,
- C++ programming language with object oriented approach,
- Graphical interface build in C++/CLI (*Common Language Interface*) under the .NET platform (*Windows Forms API*),
- Multi-threaded operation that allows independent use of interface against processing engine,
- Microsoft Windows family working environment.

Application described in this paper was anticipated to be helpful in development of novel quantum protocols. Creation of simulation tool for evaluation and testing of QKD protocols proves to be essential for deep analysis of quantum protocol's properties.

3 Design and Construction

Design of application was assumed to be both modular and highly hermetic. QKD protocol itself operates in modular manner. Therefore it was proposed that building blocks of created application should directly correspond to successive phases of QKD protocol's work. That concept was complemented with functional classes in order to fully take advantage of object oriented approach [8]. Project was divided into two main areas.

- **Engine of application.** Simulation core that executes single simulation or series of simulations with given initial configuration. Engine collects computed results and groups them in convenient results object. After this work is done, results are passed to the second part of application—the Interface.

– **Interface.** Communication layer between user and engine. It is responsible for provision of clear and convenient way of defining initial configuration. Interface also presents textual reports based on gathered results.

Figure 1 presents the *Graphical User Interface* (GUI). Configuration tab, shown on the left side, allows user to define parameters and simulation conditions. Results tab, presented on the right side, is place where user is able to view and save obtained data after successful simulation.

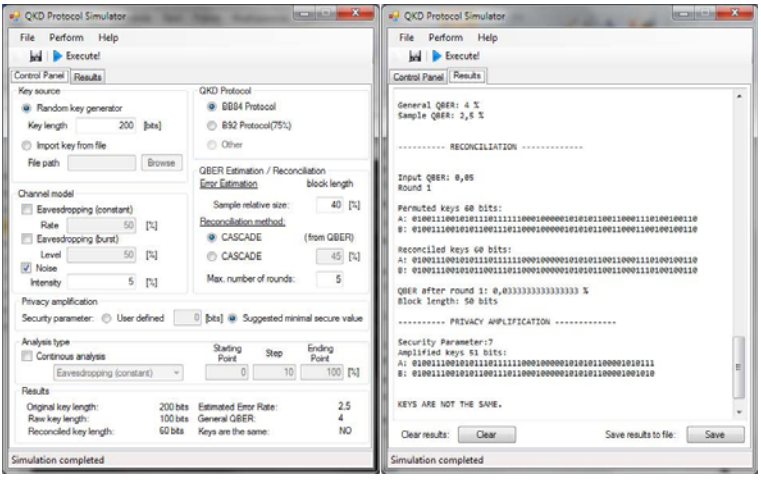


Fig. 1. Graphical User Interface. (*left*) Configuration tab; (*right*) Results tab.

Both of mentioned areas, interface and engine, are totally independent, thanks to the multi-threaded operation. Therefore user is able to break simulation any time, and redefine the initial configuration in order to simulate another scenario. Detailed flow-chart of application is presented in Figure 2.

Processing engine consists of several modules that models successive phases of QKD protocols. Simulation execution is controlled by dedicated Control module. It is also responsible for results data collection and raport generation. Due to the serialized architecture, the connection type between each module had to be defined. Because the only data type that flows through QKD protocol is set of two keys (Sender's and Receiver's), the two_keys class was designed. Therefore, each module takes input set of keys and processes it. The output set of keys is then passed to the next module. In the same time the results obtained by module operation are saved in the control module for raport generation purpose.

It should be noted, that this application operates by means of two independent threads. Interface is responsible for building initial configuration and passing it to the processing engine. After successful simulation, output data, grouping all gathered results, are passed back to the interface thread and presented to the user. Due to the order of data processing, parallelization in engine thread proves to be pointless.

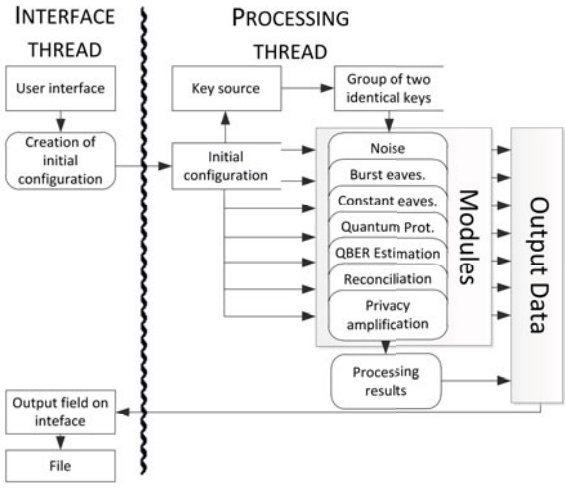


Fig. 2. Data flow-chart

4 Use Case and Exemplary Results

Two possible simulation scenarios were anticipated. Firstly, user should be able to simulate single execution of QKD protocol with parameters of his choosing. This way, user is provided with possibility of performing accurate simulations and gathering detailed and specific results. Other mode of operation is multiple rounds analysis—analysis of protocol's work with iterative changes of one parameter. User can determine which parameter should be changed during series of simulation, set starting and ending point of simulation as well as define percentage step that corresponds to the continuous change. In this section both scenarios with exemplary results are presented.

4.1 Single Simulation

Single simulation provides detailed report generated by the main engine of application. Each of the operating modules provides essential data for further analysis of QKD protocol. That data is complemented with adequate description and grouped by QKD protocol's phase. Due to the extensive amount of data stored in report, in this paper we present only a small and exemplary portion of obtained results—the output of QBER Estimation module.

In this particular example, the textual report consists of three different types of data:

- **Key sequences as well as key's size.** Groups of keys that represent sample keys and raw keys.
- **Internal processing data.** In this particular example, it would be mask that is used for key sample extraction.

– **Values calculated during operation of module.** In this example, these are general level of QBER and corresponding error rate of sample keys.

Output of QBER Estimation module

```
Mask for extracting sample key:
M: 00100110110110001000000010011110010
Sample keys 14 bits:
A: 00010101000100
B: 00011111001000
Raw keys (without sample) 21 bits:
A: 100011000101101000110
B: 100010000011001000010
General QBER: 25,7142857142857 %
Sample QBER: 28,5714285714286 %
```

4.2 Series of Simulations

Design of application introduces other approach for evaluating and testing QKD protocols—the possibility of performing a series of simulations. User can define each parameter as in case of single simulation, but in addition, he or she can choose which parameter (eavesdropping rate or noise level) should be tested through series of values. Starting and ending point, as well as step of change can be set through the interface. This mode of operation generates data formatted in CSV format for simplification of chart generation procedure. Here, small portion of that data is presented.

Series of simulation output

```
    Eavesdropping Rate;Sample QBER;General QBER;Reconciled Keys
    ->Length;Amplified Keys Length;Keys are the same
(1) 0;0;0;308;299;True
(2) 0,5;0,00490196078431373;0,001953125;308;297;True
(3) 1;0,00980392156862745;0,0078125;308;295;True
(4) 1,5;0;0,005859375;308;299;False
(5) 2;0,00980392156862745;0,0078125;308;295;True
(6) 2,5;0,0147058823529412;0,0078125;308;294;True
```

Output procured after series of simulations consists of following values:

– **Eavesdropping rate.** That is, rate introduced to the key or noise level that disturbed transmission of the key.
– **Sample and General QBER.** Represents estimated and general value of QBER in percent.
– **Reconciled keys length, Amplified keys length.** Length of the keys after reconciliation phase and privacy amplification provided in bits.
– **"Keys are the same" field.** Notice whether keys were the same at the end of QKD protocol.

As reader can see, presented results shows instability of simulated protocol. For example row of eavesdropping rate equal to 1.5% (row number 4) demonstrates that key distribution failed despite of low level of introduced disturbance.

5 Conclusions

Presented quantum cryptography protocol simulator meets all defined criteria. Moreover, application was built with use of modular approach in object oriented manner. Convenient graphical user interface as well as extensive manual were provided. The multithreaded operation gives user possibility of canceling operation and redefining initial configuration. It should be noted that main engine of application is programmed with prediction of further use under different operating systems. Created application is currently used in scientific work that concerns novel high-level security protocols [6].

Acknowledgments. This work has been performed in the framework of the EU Project INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment*)—grant agreement number: 218086.

References

1. Ekert, A.: Quantum cryptography based on Bell's theorem. Physical Review Letters 67(6), 661–663 (1991)
2. INDECT Project, http://www.indect-project.eu
3. Lindblad, G.: A general no-cloning theorem. Royal Institute of Technology, Stockholm (1998)
4. Lomonaco Jr., S.: A Quick Glance at Quantum Cryptography. University of Maryland Baltimore County, Baltimore (1998)
5. Menezes, A.J., Oorschot, P.C., Vanstone, S.A.: Handbook of Applied Cryptography, p. 21. CRC Press, Boca Raton (1997)
6. Niemiec, M., et al.: D8.2 Evaluation of components, INDECT Project Deliverable (2010)
7. Romański, Ł., Święty, M., Niemiec, M.: Current status and future directions of quantum cryptography. In: Proc. MCSS 2010: Multimedia Communications, Services and Security: IEEE International Conference, Poland, Krakow, pp. 176–181 (2010)
8. Święty, M.: Design and implementation of a simulation tool for testing and evaluating the quantum cryptography protocols, BSc dissertation, AGH University of Science and Technology in Krakow, Krakow (2011)
9. Wooters, W.K., Zurek, W.H.: A single quantum cannot be cloned. Nature 299(5886), 802–803 (1982)

Human Re-identification System on Highly Parallel GPU and CPU Architectures

Sławomir Bąk[1], Krzysztof Kurowski[2], and Krystyna Napierała[3]

[1] INRIA Sophia Antipolis, PULSAR group, France
slawomir.bak@inria.fr
[2] Poznań Supercomputing and Networking Center, Poland
krzysztof.kurowski@man.poznan.pl
[3] Institute of Computing Science, Poznań University of Technology, Poland
krystyna.napierala@cs.put.poznan.pl

Abstract. The paper presents a new approach to the human reidentification problem using covariance features. In many cases a distance operator between signatures based on generalized eigenvalues has to be computed efficiently, especially once the real-time response is expected from the system. This is a challenging problem as many procedures are computationally intensive tasks and must be repeated constantly. To deal with this problem we have successfully designed and tested a new video surveillance system. To obtain the required high efficiency we took the advantage of highly parallel computing architectures such as FPGA, GPU and CPU units to perform calculations. However, we had to propose a new GPU-based implementation of the distance operator for querying the example database. In this paper we present experimental evaluation of the proposed solution in the light of the database response time depending on its size.

Keywords: Re-identification, Covariance Matrix, Generalized Eigenvalues, High Performance Computing, GPU.

1 Introduction

Human re-identification is one of the most challenging and important problems in computer vision and pattern recognition. The re-identification problem can be defined as a determination whether a given person of interest has already been observed over a network of cameras. This issue (also called the *person re-identification problem*) can be considered on different levels depending on the information cues currently available in the system. Biometrics such as *face*, *iris* or *gait* can be used to recognize identities. Nevertheless, in most video surveillance scenarios such detailed information is not available due to a low video resolution or a difficult segmentation (crowded environments, *e.g.* airports, metro stations). Therefore a robust modeling of a global appearance of an individual is necessary to re-identify a given person of interest. In these identification techniques (named *appearance-based approaches*) clothing is the most reliable information about the identity of an individual (there is an assumption that individuals wear

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 293–300, 2011.

the same clothes between different sightings). A model of appearance has to handle differences in illumination, pose and camera parameters to allow matching appearances of the same individual observed in different cameras. High accuracy of re-idenfication approaches can only be achieved using an appearance representation based on descriptors which are invariant across different camera views. Recently, a covariance descriptor [9] has proved its effectiveness in recognition [1] and classification approaches [8]. It has been shown that the performance of the covariance features is superior to other methods, as rotation and illumination changes are absorbed by the covariance matrix [1]. Moreover, *integral images* [10] used for fast covariance computation make this descriptor very efficient concerning extraction of the covariance.

However, as covariance matrices do not lay on Euclidean space, it is necessary to apply complex differential geometry to compute a distance between two covariances. The distance operator is computationally heavy as it requires solving *the generalized eigenvalues problem*. As a consequence, matching of the covariance descriptors is slower than matching of other computer vision descriptors, which are usually represented by vectors laying on Euclidean space. This often makes covariance-based approaches difficult to apply in real-time systems in spite of their effectiveness. Hence, we propose a new hybrid architecture based on Graphics Processing Units (GPU) and Field Programmable Gate Arrays (FPGA) accelerators to take advantage of high performance computing and make covariance-based approaches applicable to large-scale databases.

This paper makes two contributions. First, we describe a new GPU- and FPGA-based architecture for the person re-identification problem, which can be easily adjusted to other video surveillance problems, such as object recognition or classification (Section 3). Second, we propose an implementation for finding generalized eigenvalues and eigenvectors for distance operator, using NVIDIA GPU architecture (Section 5). We evaluate our approach in Section 6 before concluding the paper.

2 Motivations and Related Work

There is an increasing demand for effective surveillance systems, *i.e.* systems that can perform low-cost, low-power and high-speed operations. Recently, recognition problems have become one of the most important tasks in video surveillance. As recognition is an extremely difficult task, the existing approaches are computationally heavy. Thus, a new high performance architectures are necessary to apply these approaches to real-time systems. In [7] biologically-inspired algorithms are adjusted to GPU to perform large-scale object recognition. Similarly, in [6] GPU-based neural network is presented to recognize human faces.

We introduce a surveillance system based on FPGA and GPU architectures giving much better computing facilities in comparison with traditional CPU-based systems. The most demanding part of the system – the generalized eigenvalues calculations – is performed using the GPU architecture. There are already a few approaches of finding eigenvalues of symmetric matrices using GPUs [5],

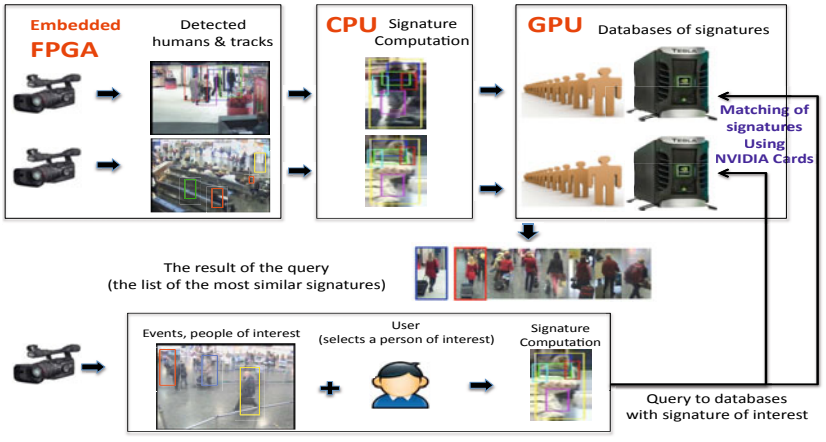


Fig. 1. FPGA- and GPU-based architecture for the person re-identification

but these implementations are focused on computation of the eigenvalues of large matrices (the implementations are optimized for matrices larger than 1024). In contrary to these approaches, we concern a domain-specific problem where it is necessary to solve the generalized eigenvalues problem for a large number of small matrices (the covariance descriptor is mostly represented by square matrix of size between 8 and 16). To the best of our knowledge, we are the first to propose an implementation for finding the generalized eigenvalues and eigenvectors of a large amount of small matrices using NVIDIA GPU cards.

3 System Architecture

Our GPU- and FPGA-based surveillance system for the person re-identification problem (Fig. 1) is designed to assign the tasks to the most suitable architectures. The system consists of a network of cameras with embedded FPGA units, which preprocess a video stream (image denoising, object detection and classification). FPGA units are well suited for video processing tasks which usually expose a high level of parallelism. Therefore, only the necessary information (*blobs* of interest) is sent in the network, without the need of transfering the whole video stream to the central unit. The partially transformed data is collected on a central unit with a CPU and GPU processor. In our approach a *human appearance* is represented by a set of covariance matrices extracted on different resolutions from detected body parts [1]. In total, the human signature is represented by 26 covariance matrices of the size 11. Covariance signature can be computed on CPU efficiently using *integral images*. Signature is then stored on a GPU unit which serves as a database of signatures. The advantage of such a solution is that, first, CPU unit is offloaded from storing this information, and second, when the distance is calculated, the data is already stored on a proper unit. We use a Tesla S1070 with four GPU units, on which there is 4GB of available global memory

for each unit. Taking into account the free space needed for calculations for a query to a database, about 200,000 signatures can be stored in the database on one unit. The user (administrator of a surveillance system) can select an object of interest and query the database with a new signature. The new signature is compared using the distance operator to all signatures in the database. Distance operator is calculated in parallel directly on the GPU unit. The result is a list of the most similar signatures. The most important part of the system is the database stored on GPU and the calculation of distance on this architecture. Preprocessing on FPGA is not that crucial for making the system real-time, therefore we show only how to calculate the distance on GPU (Sec. 5), and we evaluate experimentally the database response time (Sec. 6).

4 Covariance Descriptor

In [9] the covariance of d-features has been proposed to characterize a region of interest. Below we describe only *the geodesic distance* definition proposed in [3], as its computation is the main topic of our work. The distance between two covariance matrices is defined as

$$\rho(C_i, C_j) = \sqrt{\sum_{k=1}^{d} \ln^2 \lambda_k(C_i, C_j)} \tag{1}$$

where $\lambda_k(C_i, C_j)_{k=1\ldots d}$ are the generalized eigenvalues of C_i and C_j, determined by $\lambda_k C_i x_k - C_j x_k = 0,\ k = 1 \ldots d$ and $x_k \neq 0$ are the generalized eigenvectors. Traditional methods work efficiently for dimension d below 5, so often a dimension of covariance matrix has to be decreased to conform the requirements of real-time systems at the expense of accuracy.

5 GPU Implementation

5.1 GPU Architecture

GPU is a high performance architecture in which the emphasis is put on the computing units instead of data caching and control flow units in contrast with CPU. Threads work in a SIMD model and are grouped into blocks. All threads of a block reside on the same processor core and are split into basic scheduling units called warps, consisting of 32 threads. We use the NVIDIA Tesla S1070 with 4 graphic cards, each consisting of 240 processors.

5.2 Generalized Eigenvalues Problem

The calculation of generalized eigenvalues of positive definite symmetric matrices A and B (two corresponding covariance matrices from the compared signatures) is defined by the equation $\mathbf{A}x = \lambda \mathbf{B}x$, where λ denotes a vector of eigenvalues and x is the eigenvector. This equation can be decomposed to the equation $(\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-T})\mathbf{L}^T x = \lambda(\mathbf{L}^T x)$, where $\mathbf{L}$ is the upper triangular matrix calculated as $\mathbf{B} = \mathbf{L}\mathbf{L}^T$. The decomposed equation corresponds to original eigenvalues problem. We solve the generalized eigenvalues problem in 4 steps:

1. Calculate the Cholesky Decomposition to solve the equation $\mathbf{B} = \mathbf{L}\mathbf{L}^T$
2. Calculate twice the Forward Substitution to solve $\mathbf{C} = (\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-T})\mathbf{L}^T x$
3. Tridiagonalize the symmetric matrix $\mathbf{C}$ to facilitate finding the eigenvalues
4. Find the eigenvalues of $\mathbf{C}$ with the Bisection Algorithm

Let us note that there are many methods to calculate the eigenvalues of a symmetric matrix. On CPU we use a Jacobi algorithm, which does not need to perform the tridiagonalization first, however this algorithm is not easy to parallelize. For the GPU implementation we first tridiagonalize the matrix and then use the bisection algorithm, which can be parallelized much more efficiently.

5.3 Porting the Algorithm to the GPU Architecture

Two sources of parallelism can be used in the algorithm. The first one is obvious – a comparison of two signatures involves comparing n pairs of covariance matrices, which can be naturally processed in parallel. The second one involves the parallelism extracted from each step of the equation performed on a given pair of covariance matrices. We will now briefly describe how we use the parallelism of each procedure to efficiently port it to the GPU architecture.

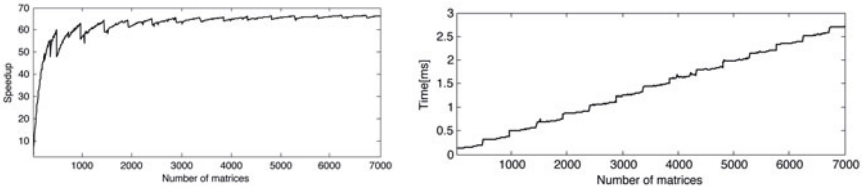
In **Cholesky Decomposition of n matrices**, the formula to calculate each element of the $\mathbf{L}$ upper triangular matrix is given as

$$L_{j,j} = \sqrt{B_{j,j} - \sum_{k=1}^{j-1} L_{j,k}^2}, \qquad L_{i,j} = \frac{1}{L_{j,j}}(A_{i,j} - \sum_{k=1}^{j-1} L_{i,k}L_{j,k}) \qquad for \;\; i > j \quad (2)$$

All calculations are performed using fast shared memory. We use the Cholesky-Crout algorithm which calculates the matrix column by column, as processing the data in column order ensures the memory coalescence. To calculate an element of a column, only the elements from the columns to the left are needed. So, all the elements in one column can be processed in parallel. For our matrices of size 11, we can use 11 threads to calculate each column in parallel, using 11 iterations to calculate the whole matrix. The natural decomposition of a problem would be to assign one matrix to a block. However, 11 threads is much less than the size of a warp, which is the smallest scheduling unit. Assigning two matrices to a block enables to process two matrices without the increase of processing time, as 22 threads still form only one warp scheduled in one operation.

Forward Substitution solves the equation $\mathbf{L}\mathbf{x} = \mathbf{A}$. Two forward substitutions solve the equation $\mathbf{C} = (\mathbf{L}^{-1}\mathbf{A}\mathbf{L}^{-T})\mathbf{L}^T\mathbf{X}$. We process them together to avoid copying the intermediate data to the global memory. An element $x_{m,k}$ of a matrix $\mathbf{x}$ is calculated as $x_{m,k} = (A_m - \sum_{i=1}^{m-1} L_{m,i}x_{i,k})/L_{m,m}$. Columns of $\mathbf{x}$ are calculated independently. As a result, for one matrix we need 11 threads, and following our observations from Cholesky Distance procedure, we assign 22 threads to one block.

Tridiagonalization is a part of the generalized eigenvalues algorithm which has the lowest level of paralellism. We use the Householder transformation [4] in which $n - 2$ iterations need to be performed sequentially; in each iteration



(a) Speedup - CPU vs GPU (b) Time (ms) of Bisection algorithm

Fig. 2. Speedup for finding generalized eigenvalues and time of Bisection algorithm

the appropriate elements in the *k-th* row and column are zeroed. Some computations such as matrix multiplication and vector-vector multiplications can be parallelised using $n \times n$ threads for each matrix. As 121 threads is more than a size of a warp, we can assign one matrix per block without loosing the efficiency.

Bisection Algorithm is used to calculate the eigenvalues of a symmetric tridiagonal matrix **C** with a given approximation (for details see e.g. [2]). The core function of this algorithm is the `Count()` procedure returning the number of eigenvalues present in a given interval. The algorithm starts with the initial interval constructed using Gerschgorin's theorem. Then, it is divided in two and `Count()` procedure returns a number of eigenvalues in each subset. The unempty nodes are further subdivided into two subsets, until each subset contains only one eigenvalue and the size of a subset is not bigger than the assumed approximation (for our purposes 10e-6 is enough). The `Count()` function can be calculated independently in each node. As only 11 eigenvalues can be found, on each level of the binary tree there will be only up to 11 unempty nodes. We therefore use 11 threads to calculate one matrix and again we assign 2 matrices to a block.

6 Experimental Results

To evaluate the database response time, we calculated the performance for matrices number ranging from 10 to 3000. The database of signatures is stored directly on the GPU (see Section 3), therefore we do not take into account the time of the data transfer, because the reference signatures already reside in the device memory, and the time of transferring the query signature is negligible.

In Fig. 2(a) we present the speedup obtained with comparison to the optimal CPU implementation (Jacobi algorithm from LTI library, see discussion in 5.2). The speedup grows with the number of matrices, and reaches its maximum (66) from about 1500 matrices (corresponding to about 50 signatures). Distributing the database of signatures between 4 GPU cards of Tesla, and performing the calculations for a query signature on all cards in parallel, would result in further speedup improvement. Table 1 presents the time of the component procedures for 5 numbers of matrices. The Cholesky and Forward Substitution times are the lowest. The biggest impact on the total time comes from tridiagonalization procedure, which has the lowest degree of parallelism.

Table 1. Time[ms] of component procedures

N	Cholesky	Forward.S	Tridiag.	Bisect.	Total.GPU	Total.CPU
200	0.037	0.035	0.140	0.147	0.359	16
400	0.049	0.071	0.272	0.187	0.579	32
600	0.082	0.078	0.389	0.325	0.874	48
800	0.093	0.112	0.522	0.352	1.08	64
1000	0.126	0.120	0.657	0.505	1.41	80

In Fig. 2(a) one can notice the *"stairs"* (decrease of speedup) which occur in regular intervals every 480 matrices. This is a result of a similar effect observed in component functions (see e.g. Bisection Algorithm in Fig. 2(b)) and is correlated to the number of processors. The architecture is the most efficiently used when all the processors (240 for Tesla S1070) have some blocks assigned, i.e. when k*480 matrices are processed (each block calculates two matrices). The worst case is for k*480+1 matrices – then, the time is almost equal to processing (k+1)*480 matrices, which results in sudden decrease of speedup in these points.

7 Conclusions

In this paper we demonstrate that it is possible to improve significantly the performance of examplary video surveillance procedures, in particular the distance operator for querying the database. We observe that the speedup grows with the number of matrices. Moreover, we can easily distribute the demanding calculations on many GPU units and obtain very good scalability of the system. Although GPU-based approach of four routines of generalized eigenvalues problem have been already proposed, they are optimized for solving only one large matrix. In our approach we use many very small matrices, which can be efficiently stored in a shared memory on many GPUs. This requires different memory alignments, memory access optimization and simplification of some subprocedures, but as tested experimentally could yield much better performance.

Acknowledgement

The presented work was sponsored by the UCoMS project under award number MNiSW (Polish Ministry of Science and Higher Education) Nr 469 1 N - USA/ 2009 in close collaboration with U.S. research institutions involved in the U.S. Department of Energy (DOE) funded grant under award number DE-FG02-04ER46136 and the Board of Regents, State of Louisiana, under contract no. DOE/LEQSF(2004-07).

References

1. Bak, S., Corvee, E., Bremond, F., Thonnat, M.: Person re-identification using spatial covariance regions of human body parts. In: AVSS (2010)
2. Demmel, J.W., Heath, M.T.: Applied numerical linear algebra. In: Society for Industrial and Applied Mathematics. SIAM, Philadelphia (1997)
3. Förstner, W., Moonen, B.: A metric for covariance matrices. In: Quo vadis geodesia..? TR Dept. of Geodesy and Geoinformatics, Stuttgart University (1999)
4. Householder, A.S.: Unitary triangularization of a nonsymmetric matrix. Journal of the ACM 5 (1958)
5. Lessig, C.: Eigenvalue computation with CUDA. In: NVIDIA techreport (2007)
6. Poli, G., et al.: Processing neocognitron of face recognition on high performance environment based on GPU with CUDA architecture. In: SBAC-PAD, pp. 81–88. IEEE Computer Society, Los Alamitos (2008)
7. Sriram, V.: Design-space exploration of biologically-inspired visual object recognition algorithms using CPUs, GPUs, and FPGAs. In: MRSC (2010)
8. Tosato, D., Farenzena, M., Spera, M., Murino, V., Cristani, M.: Multi-class classification on riemannian manifolds for video surveillance. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010. LNCS, vol. 6312, pp. 378–391. Springer, Heidelberg (2010)
9. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3952, pp. 589–600. Springer, Heidelberg (2006)
10. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: CVPR 2001, pp. 511–518 (2001)

Face Occurrence Verification Using Haar Cascades - Comparison of Two Approaches

Piotr Boryło, Andrzej Matiolański, and Tomasz M. Orzechowski

Department of Telecommunications,
AGH University of Science and Technology,
Cracow, Poland
pborylo@pluton.kt.agh.edu.pl, matiolanski@kt.agh.edu.pl,
tomeko@agh.edu.pl
http://www.kt.agh.edu.pl/

Abstract. Face occurrence verification is mandatory stage for color based face detection systems. Facial features extraction using Haar cascades is one of the possible way to classify regions as faces. The purpose of this paper is to provide novel and valuable comparison of two approaches for using Haar cascades in face occurrence verification stage. Performance and accuracy tests have been carried out to decide which of approaches described in the article is more suitable. Results might be crucial while implementing similar face detection system based on skin detection and facial features localization.

Keywords: Haar cascades, skin detection, face occurrence verification, facial features extraction.

1 Introduction

Face detection is one of the most important issues in the computer vision. In the recent years many researches have been conducted to find effective and efficient solutions. Human face detection is important because it is the first step for many useful and powerful applications, e.g. security systems, face recognition (including mug shot matching), face gesture recognition, face tracing etc. Due to differences in illumination, scale, angles, orientation, pose, camera calibration, contrast etc. creating an universal and robust face detection system is a complex and demanding issue.

Face detection based on skin color regions localization is one of the possible approaches. Figure 1 presents stages of the algorithm. Using skin color model in RGB color space, *pixel by pixel* processing, morphological operations and segmentation skin color regions are extracted. Those areas should be treated as *face candidates* and face occurrence verification must be conducted against them. Due to the fact that verification process is crucial for system effectiveness it must be deeply analyzed.

Many methods for face occurrence verification have been proposed in the recent years. Some examples are: shape verification [5], wavelet packet analysis

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 301–309, 2011.

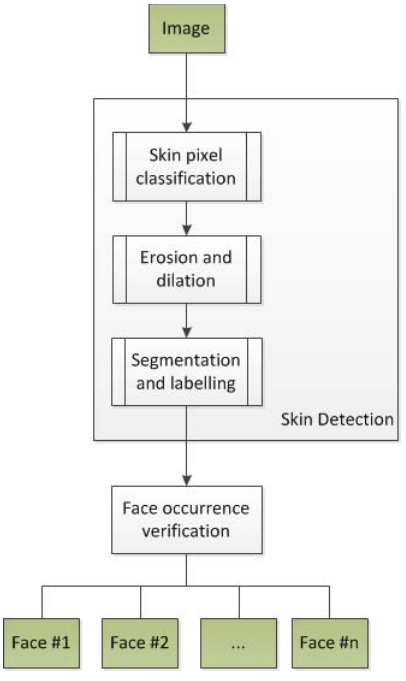


Fig. 1. Process of face detection

[3], facial features extraction using greylevel information [14], hair and skin connectivity analysis [2] and neural networks [8]. Our system uses facial features extraction using Haar cascades to verify face occurrence. This paper provides far-reaching comparison of two approaches for Haar cascades usage. Both, accuracy and efficiency are significant to face detection system.

The rest of the paper is organized as follows. Color based skin detection process is presented in the following section. Section 3 sheds light on a problem of Haar cascades. In section 4 two approaches for face occurrence verification using Haar cascades are described. Experimental results are presented and discussed in section 5. In section 6 conclusions are provided. All images presented in this paper are real output of the system. Black rectangles in images have been added after detection process to hide personalities.

2 Color Based Skin Detection

Skin color regions detection starts with classification of each pixel as skin or non-skin. Different color spaces can be used for skin localization e.g. RGB, HSV, YCrCb etc. [5,12,14,8]. Some additional information about color spaces using in skin detection can be found at [16,13]. Model proposed in [16] has been chosen as the most accurate one with test image set.

Next, erosion and dilation operations take place. Those morphological operations use structuring element of user defined shape and size to operate on images. For binary image input erosion results in deletion of regions smaller than structuring element and perimeter of larger objects. However, dilation connects regions separated by spaces smaller than structuring element and adds perimeter to each region. Erosion followed by dilation is defined as opening and is used in the system to delete small regions and to make big regions smoother and continuous.

Segmentation and labeling transform skin color regions into objects called *face candidates*. Due to the fact that binary image is an input of this stage, the simplest variant of segmentation can be used. Skin pixels are grouped using structuring element defined to examine regions connectivity. An output of segmentation and labeling is set of *face candidates* coordinates. Figure 2 presents example image with *face candidates* borders. Some false-positive detections can be observed what proves that face occurrence verification process is indispensable.

Fig. 2. Example image after skin detection stage

3 Haar Cascades

Usage of Haar cascades classifier is a machine learning approach for object detection. Described by P. Viola and M. Jones algorithm was presented in [10].

Cascades are training with set of positive and negative samples. For each sample there is a set of weak classifiers works on differences between sample images. Learning process based on AdaBoost[11] algorithm combines results of weak classifiers and create a good one.

Because Haar detection based on difference between the images far better cope with the detection of objects containing significant number of details (e.g. faces - the effectiveness of about 93.7%[17]). Unfortunately, for face features detection (e.g. mouth, nose, eyes), the results are not such good. This is influenced by the small size of the objects and weak highlighting face features in the image. An additional problem is high false-positive detections rate. For instance, the eyes in an image are often fuzzy so represent only as two circles. They can be identified as a nose hole or worse as a background - outside the face.

The problems shows that extraction of face features using Haar cascades is not enough efficient and powerful for face detection.

4 Two Approaches for Face Occurrence Verification Using Haar Cascades

The main difference between two approaches for face occurrence verification using Haar cascades concerns the area on which cascades operate. An order of skin detection, facial features extraction and face occurrence verification is crucial to distinguish both methods.

4.1 Haar Cascades Detect Facial Features in Original Input Image

Figure 3 illustrates the first approach where Haar cascades operate on original input image. An output of skin detection and facial features extraction is set of *face candidates* coordinates and set of facial features coordinates respectively. The task of face occurrence verification is to check which *face candidates* enclose any facial features and classify them as faces.

4.2 Haar Cascades Detect Facial Features Only in Face Candidates Regions

Figure 4 presents the second approach. Facial features are extracted using Haar cascades only within *face candidates* regions. If any facial feature is detected, the area is classified as a face.

Example image after face occurrence verification using the second approach has been presented in the figure 5. Blue rectangles outline facial features detected within *face candidates* regions. False positives have been correctly rejected.

5 Experimental Results

The purpose of conducted tests was to compare two approaches for face occurrence verification described in section 4. Test data set consists of twenty photos

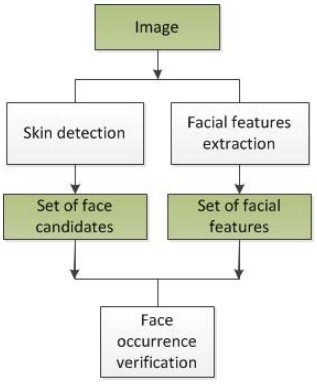


Fig. 3. Haar cascades detect facial features in original input image

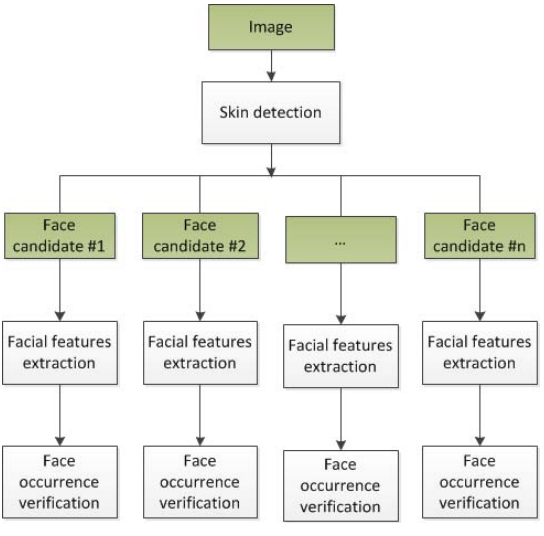


Fig. 4. Haar cascades detect facial features only in face candidates regions

which contain total eighty faces. Images were selected to be not only various and representative but also demanding and problematic for the system. Test script was written using libraries: OpenCV[15] and Scipy[9].Ten cascades trained to extract different facial features (e.g. eyes, nose, mouth) have been tested (with default parameters) for each approach and results for all of them have been provided. Additionally, the average outcomes have been counted and presented.

Table 1 provides results for accuracy tests of facial features extraction using Haar cascades. Indicators of positive detections are on unacceptable low level for both approaches. When cascades process original image only 16,8% of facial

Fig. 5. Example image after face occurrence verification

features are properly detected on average. For the second approach positives detection rate is only 3,7%. Also a number of false positives is unsatisfactory high: 40,8% and 32,3% respectively for both methods. What should be mentioned is the fact that cascades are trained to detect concrete feature. Thus, when cascade is made for e.g. nose extraction and detects eye, false positive case is noticed.

The most important indicators for our system concerns face verifications using facial features extraction. The main difference in comparison with the previous test is that when cascade trained for nose detection extracts eye, accurate face verification is noticed. We do not have to detect exact facial feature because if any is extracted then *face candidate* is classified as face. Table 2 presents percentage values of correctly verified faces and false positives for both approaches. When cascades process full image 19,6% of all faces were correctly verified and 5,4% of *face candidates* were wrongly classified as faces. Rate of positive detections is significantly improved for the second approach and reaches the level of 68,2% on average. A number of false positives comes to 6,8%. Analyzing results for each cascade individually, outcomes for cascade number six cannot be omitted – 98% and 7% for true and false positives respectively.

Performance is critical indicator for the proposed system so appropriate tests have been conducted. Absolute values are depend on image resolution, number of *face candidates*, number of extracted facial features etc. Because the aim of

Table 1. Facial features extraction accuracy tests

	Cascades process an original input image		Cascades process face candidates regions	
Cascade number	Detected facial features	False positives	Detected facial features	False positives
1	15%	40%	2%	47%
2	11%	25%	2%	0%
3	12%	0%	4%	17%
4	20%	67%	5%	40%
5	16%	54%	5%	29%
6	28%	76%	4%	67%
7	18%	9%	2%	17%
8	14%	67%	4%	53%
9	20%	70%	5%	39%
10	14%	0%	4%	14%
Average	**16,8%**	**40,8%**	**3,7%**	**32,3%**

Table 2. Face occurrence verification accuracy tests

	Cascades process an original input image		Cascades process face candidates regions	
Cascade number	Correctly verified faces	False positives	Correctly verified faces	False positives
1	18%	11%	36%	4%
2	11%	5%	53%	0%
3	9%	0%	56%	0%
4	29%	11%	87%	14%
5	18%	5%	71%	11%
6	31%	4%	**98%**	**7%**
7	24%	0%	47%	0%
8	18%	9%	87%	16%
9	29%	9%	89%	16%
10	9%	0%	58%	0%
Average	**19,6%**	**5,4%**	**68,2%**	**6,8%**

Table 3. Performance tests results

Method	Average execution time
Cascades process an original input image	1595 ms
Cascades process face candidates regions	477 ms

test was to compare two approaches, results have been averaged for the whole set of data. Otherwise numerous figures may hide differences. Table 3[1] presents that the second approach is 3,34 times faster. This is due to a smaller amount of information which is processed.

6 Conclusion and Future Work

Some far-reaching finding can be based on the provided tests results. None of tested cascade in the proposed approaches is appropriate for systems where accurate facial features extraction is mandatory. In those cases other techniques for facial features extraction should be engaged [1,7,4,6]. Nevertheless, the average percentage accuracy rates for face occurrence verification are quite satisfactory when cascades process only *face candidates* regions. For that approach outcomes for some individual, well prepared, cascades are entirely acceptable. Performance tests interchangeably points second approach as more efficient.

Comparison of two approaches for face occurrence verification using Haar cascades has been conducted. Accuracy and performance aspects have been brought up. Processing only *face candidates* regions by cascades gives more effective and scalable solution. Additional advances can be reached by training cascades to detect any of facial features, thus it will be aim of the further work.

Acknowledgment

This work has been performed in the framework of the EU Project INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment*) – grant agreement number: 218086 and co-financed by the European Regional Development Fund under the Innovative Economy Operational Programme, INSIGMA project no. POIG.01.01.02-00-062/09. Development of algorithm and implementation have been fund by EU Project INDECT. Development of system, tests and result analysis have been fund by INSIGMA project.

References

1. Campadelli, P., Lanzarotti, R., Lipori, G.: Automatic facial feature extraction for face recognition. In: Face Recognition, pp. 31–58. I-Tech Education and Publishing (2007)
2. Chen, Y., Lin, Y.: Simple face-detection algorithm based on minimum facial features. In: IECON 2007, 33rd Annual Conference of the IEEE, Taipei, pp. 455–460. Industrial Electronics Society (November 2007)
3. Garcia, C., Tziritas, G.: Face detection using quantized skin color regions merging and wavelet packet analysis. IEEE Transactions on Multimedia 1(3), 264–277 (1999)

[1] Intel Core 2 DUO T7100 1.8 GHz, 2 GB RAM, OS Windows 7 32-bit.

4. Gunduz, A., Krim, H.: Facial feature extraction using topological methods. In: Proceedings of Image Processing, ICIP 2003, vol. 1, pp. 673–676 (September 2003)
5. Hadid, A., Pietikainen, M., Martinkauppi, B.: Color-based face detection using skin locus model and hierarchical filtering. In: Proceedings of 16th International Conference Pattern Recognition, vol. 4, pp. 196–200 (2002)
6. Hassan, M., Osman, I., Yahia, M.: Walsh-hadamard transform for facial feature extraction in face recognition (2007)
7. Kam-art, R., Raicharoen, T., Khera, V.: Face recognition using feature extraction based on descriptive statistics of a face image. In: 2009 International Conference on Machine Learning and Cybernetics, Baoding, vol. 1, pp. 193–197 (July 2009)
8. Mostafa, L., Abdelazeem, S.: Face detection based on skin color using neural networks. In: GVIP 2005 Conference, Cairo, Egypt (December 2005), CICC
9. Open Source Library of Scientific Tools (March 03, 2011), http://www.scipy.org/
10. Jones, M., Viola, P.: Rapid object detection using boosted cascade of simple features. In: Computer Vision and Pattern Recognition, vol. (I), pp. 511–518 (2001)
11. Jones, M., Viola, P.: Robust real-time object detection. In: Second International Workshop on Statistical and Computational Theories of Vision - Modeling, Learning, Computing and Sampling, Vancouver, Canada (July 2001)
12. Sandeep, K., Rajagopalan, A.N.: Human face detection in cluttered color images using skin color and edge information
13. Singh, S., Chauhan, D.S., Vatsa, M., Singh, R.: A robust skin color based face detection algorithm, tamkang. Journal of Science and Engineering 6(4), 227–234 (2003)
14. Sobottka, K., Pitas, I.: Segmentation and tracking of faces in color images. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, pp. 236–241 (October 1996)
15. Open source computer vision library (March 03, 2011), http://opencv.willowgarage.com
16. Vezhnevets, V., Sazonov, V., Andreeva, A.: A survey on pixel-based skin color detection techniques. In: PROC. GRAPHICON 2003, pp. 85–92 (2003)
17. Schapire, R.E., Freund, Y.: A decision-theoretic generalization of on-line learning and an application to boosting. In: Vitányi, P.M.B. (ed.) EuroCOLT 1995. LNCS, vol. 904, pp. 23–37. Springer, Heidelberg (1995)

Implementation of the New Integration Model of Security and QoS for MANET to the OPNET

Ján Papaj, Anton Čižmár, and Ľubomír Doboš

Department of Electronics and Multimedia Communications, Technical University of Košice, Letná 9, 042 00 Košice, Slovak Republic
{jan.papaj,anton.cizmar,lubomir.dobos}@tuke.sk
http://www.tuke.sk

Abstract. The implementation of the new designed model used to integrating security and Quality of Service (QoS) as one parameter in mobile ad-hoc network (MANET) is introduced and studied in this paper. Security and QoS represent a highly important field of research in MANET and they are still being considered separately with no mechanisms used to establish cooperation between them. This new model provides alternative to cooperation between QoS and Security via Cross Layer Design (CLD) and modified Security Service Vector (SSV). Main motivation of this paper is indicating how could be security integrated as a QoS parameter to the MANET via this model. The performance analysis of the new designed model in the well-know simulator OPNET Modeler is also provides.

Keywords: OPNET Modeler, MANET, QoS, Security, modified Security Service Vector.

1 Introduction

OPNET Modeler is very popular commercial network simulation and analytical tool, which is used for network design, modeling and simulation [1]. OPNET Modeler also provide very useful tools for simulations of the mobile ad-hoc network (MANET). MANET represents a set of mobile devices and nodes with self-configuring features and with the ability to mutually communicate. MANET nodes can establish and maintain connections as needed without fixed infrastructure and central management [2]. MANET is characterized as a dynamic network with ability of the nodes to join or leave the network at randomly set times and ways. Current research trends in MANET are oriented to following categories: QoS, security and cross layer design. The field of QoS provides a wide space for research. The notion of QoS is a guarantee provided by the network to satisfy a set of predetermined service performance constraints for the user in terms of the end-to-end delay statistics, available bandwidth, probability of packet loss, etc. [2]. In MANET, QoS is essential to satisfy the communication constraints [8]. Another very important research field is security. Security solves problem of protected communication between mobile nodes in a hostile environment. Security issues are detected in many different areas. In MANET, there

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 310–316, 2011.

are solved problems of physical security, key management, secure routing and intrusion detection. Due to their architecture, MANETs are more easily attacked than wired network [3]. The cross-layer design (CLD) approach is a new dynamic area of research into MANET networks. This approach provides new possibilities to increase the performance and adaptability of MANET and research of cross-layer networking is still at the beginning [4].

1.1 Main Motivation

Basic ideas of the integration process are to provide QoS and security mechanisms at the same time, and that user or services had the possibilities to interact with system via CLD. Integration itself is necessary for proper functioning of both mechanisms in terms of QoS and security. Moreover, users can specify requirements for new services in MANET.

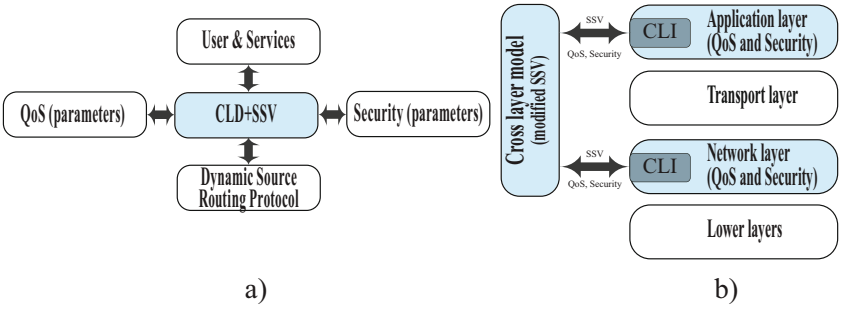


Fig. 1. Model of integrating QoS and security in MANET (a) and CLD model (b)

In this article, we provide a new model indicating how could be security integrated as a QoS parameters to the MANET via modified SSV and Cross Layer Interface (CLI). CLI enables the interaction between user/system and is also used to collect the relevant information and to cooperate between application and network layer of the MANET layer model. Based on this information the system can evaluate and choose the optimum algorithms for achieving required parameters and guidelines [8]. The modified SSV is used for cooperation between several blocks of the new model and also provides the decision algorithms for the selection of routes. Model enables to specify requested parameters and the user has the ability to participate in the routing process. The advantage of this model is that it can be used for different services and not only for QoS and security.

2 Introduction of New Integration Model QoS and Security as One Parameter in MANET

We have designed a new model, which allows integrating security and QoS as one parameter via modified SSV and CLD in MANET [5], [8]. This model also

enables cooperation of security and QoS mechanisms in MANET (Fig.1a). Model consists of the main components necessary for interactions between the user and system in order to provide the requested security and QoS related services [5]. The main function block of the model is block CLD+SSV.

2.1 Concept of Modified SSV for MANET

The modified SSV is based on security service vector designed specially for wired IP networks [6], [7]. Modification of the SSV can be defined by two ideological parts: user and system parts. The user part deals with process of collecting the relevant data about requested services. In our case, these data are created by QoS and security parameters. Parameters can represent different QoS and security parameters or mechanisms for providing QoS and security processes [8]. In this model, users can specify the required parameters and using this approach can actively affect the system (routing) processes. The system part of our modification represents the new method of processing collected data and also deals with routing processes of the routing protocol. Each MANET node has implemented algorithm to process the routing packet (RP). Algorithms analyze the routing information stored in RP and analyze the information about requested parameters, QoS and security (rSSV). A main idea of modified SSV is shown in Fig. 2.

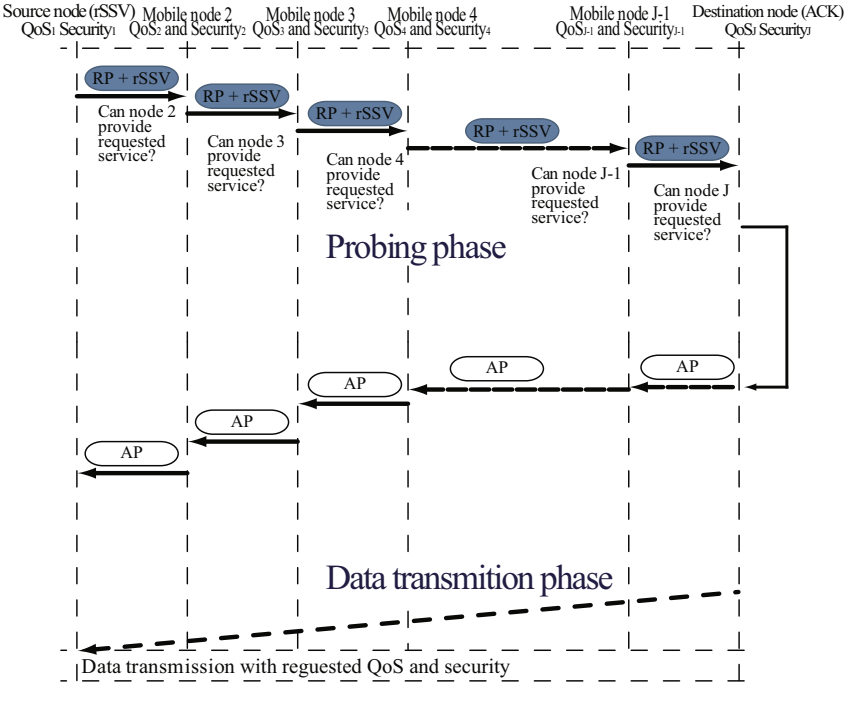


Fig. 2. Modified SSV in MANET

2.2 CLD in OPNET Modeler

The CLD was designed for collecting of the security and QoS related information. These entries are necessary for processing of the modified SSV and the dynamic source routing protocol (DSR). The CLD is used for the bidirectional transferring and collecting data from application layer/network layer. The collected SSV attributes consist of information about security and QoS parameters that a node is able to provide. Main idea of CLD is depicted in Fig. 1b. In the case of source node, the user defines SSV attributes via CLI interface located on application layer and CLD interface sends these data to network layer, where are stored to the modified route cache. In the case of routing node, the CLI analyzes the incoming packet and reads information about SSV stored in packet. If the modified route caches don't includes the information about security and QoS from application layer, the CLD interface activates re-collecting process of these data from application layer. In the case of destination node, the CLD collects data about requested QoS and security from routing packet and from modified route cache.

In OPNET, the CLI is designed as finite state machine (FSM). The CLI process model consists of four states (Fig. 2a). State $INIT$ consists of the initialization of the process model and this block is necessary for OPNET modeler implementation. This state initializes every variable, statistic, table, and user parameter that is used by the CLI process model. State $IDLE$ is used to analyzing the type of requests and this is the default state where the process waits for an event. State $From\_APPL$ is activated when $CLI\_REQUEST\_APPL$ request from application layer is obtained. This request is used in the case of source node, when user specifies the QoS and security requirements. State $From\_IP$ indicates, that network layer activates request for collecting information about QoS and security from the application layer on the node. That means, the CLI interface located on network layer, sends the request called $CLI\_REQUEST\_IP$ to process of collecting SSV attributes from application layer. In network and

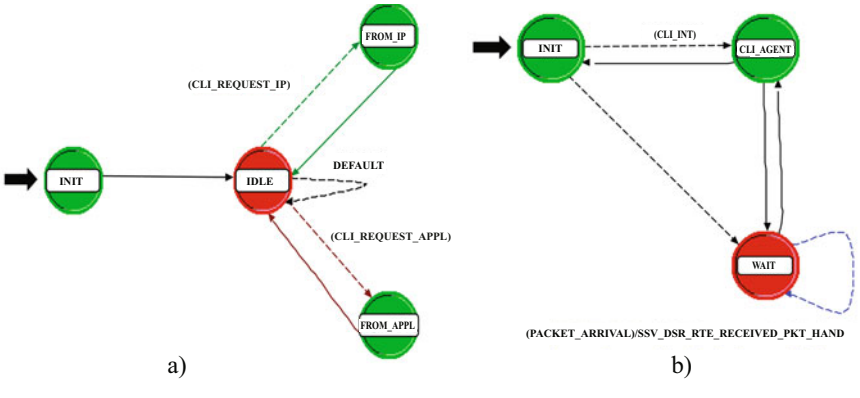


Fig. 3. Cross layer model in OPNET: (a) CLI model, (b) CLI interface

Table 1. Delay of MANET [s] analysis depending on the number of nodes generated traffics

Number of nodes	Model	20%	40%	60%	80%	100%
20	DSR	0, 00239	0, 00292	0, 00306	0, 00422	0, 00512
	DSR+SSV	0, 00297	0, 00382	0, 00393	0, 00431	0, 00612
	DSR+SSV_CLD	0, 00253	0, 00295	0, 00338	0, 00425	0, 00513
40	DSR	0, 00191	0, 00227	0, 00264	0, 00222	0, 00268
	DSR+SSV	0, 00243	0, 00363	0, 00284	0, 00295	0, 00302
	DSR+SSV_CLD	0, 00198	0, 00242	0, 00267	0, 00279	0, 00283
60	DSR	0, 00165	0, 00303	0, 00702	0, 00921	0, 00985
	DSR+SSV	0, 00184	0, 00340	0, 00807	0, 00987	0, 01102
	DSR+SSV_CLD	0, 00171	0, 00320	0, 00785	0, 00954	0, 01078
80	DSR	0, 00105	0, 00204	0, 00245	0, 00254	0, 00281
	DSR+SSV	0, 00132	0, 00236	0, 00260	0, 00275	0, 03240
	DSR+SSV_CLD	0, 00110	0, 00226	0, 00246	0, 00261	0, 00291
100	DSR	0, 00289	0, 00300	0, 00335	0, 00352	0, 00391
	DSR+SSV	0, 00341	0, 00332	0, 00374	0, 00382	0, 04010
	DSR+SSV_CLD	0, 00303	0, 00317	0, 00349	0, 00362	0, 00381

application layer are implemented CLI agents (Fig. 2b). These agents interact with main CLI interface in order to transmit the relevant SSV attributes. In application layer CLI agent is used for collection. In network layer, CLI agent is activated by modified SSV in the case, that the modified route cache does not contain the node's information about SSV.

3 Simulation and Results

The main ideas of the simulations were to verify the effectiveness of implementing a new designed model in MANET terminals. All behavior of the proposed model was simulated in OPNET Modeler 16.0 simulator. Three types of simulations were used to evaluate effectiveness of integrating a new model with CLD and modified SSV. A first simulation represents model where the nodes used routing protocol DSR without modified SSV and CLD (DSR). Second simulation scenario is model where the nodes used modified routing protocol DSR with implemented the modified SSV (DSR+SSV). Third simulation scenario is model where the nodes used modified routing protocol DSR with implemented modified SSV and CLD (DSR+SSV_CLD). To check functionality of the proposed model were used parameters delay of MANET and total packet processing delay. Delay of MANET represents the value of the average end-to-end delay measured from the network layer on the source node, where the MANET packet is created, to the delivery of the packet to the destination node and also taken into account is the processing time for SSV of information layers in source-target transport. Total packet processing delay parameter represents the average delay in MANET

Table 2. Total packet processing delay MANET [s] analysis depending on the number of nodes generated traffics

Number of nodes	Model	20%	40%	60%	80%	100%
20	DSR	0, 00200	0, 00206	0, 00210	0, 00201	0, 00203
	DSR+SSV	0, 00213	0, 00233	0, 00221	0, 00213	0, 00224
	DSR+SSV_CLD	0, 00207	0, 00212	0, 00216	0, 00204	0, 00211
40	DSR	0, 00129	0, 00184	0, 00217	0, 00245	0, 00316
	DSR+SSV	0, 00170	0, 00193	0, 00262	0, 00284	0, 00332
	DSR+SSV_CLD	0, 00130	0, 00173	0, 00216	0, 00275	0, 00324
60	DSR	0, 00183	0, 00262	0, 00357	0, 00395	0, 00378
	DSR+SSV	0, 00219	0, 00290	0, 00397	0, 00414	0, 00418
	DSR+SSV_CLD	0, 00190	0, 00273	0, 00373	0, 00400	0, 00396
80	DSR	0, 00150	0, 00230	0, 00321	0, 00409	0, 00322
	DSR+SSV	0, 00169	0, 00293	0, 00272	0, 00234	0, 00265
	DSR+SSV_CLD	0, 00167	0, 00282	0, 00247	0, 00204	0, 00245
100	DSR	0, 00224	0, 00263	0, 00269	0, 00266	0, 00325
	DSR+SSV	0, 00280	0, 00309	0, 00312	0, 00295	0, 00376
	DSR+SSV_CLD	0, 00266	0, 00282	0, 00276	0, 00280	0, 00339

networks from sending a packet to the adoption of the packet on the IP layer of the target node. The parameter does not reflect the time needed to processing information SSV.

The 5 separated simulation scenarios that were formed of 20, 40, 60, 80 and 100 nodes were created to check the effectiveness of operation of the modified SSV and CLD in MANET. Transmit power was set up to 1mW. The random mobility model was used to simulate the mobility of nodes. Simulation period was in all cases 1000 seconds. The changing parameter was the initial value of movement, which gives a different initial position of individual nodes in the simulated project. The result of each simulation was a set of values that were then statistically processed and evaluated. Each sample was made up of a set of 100 values from each simulation (10000 values were recorded).

In the experiment was monitored how the traffic increases can affect behavior of the network by applying the new designed model with modified SSV and CLD. The burden in this case is seen as the number of nodes generating traffic (packets), thus becoming simultaneously the source, routing and destination nodes. In Table 1 and Table 2 are shown the comparative delay of MANET and the total processing delay analyses when the numbers of nodes that generated the traffic (%) for different networks were changed. Based on collected results, it can be concluded that the integration of modified SSV (DSR+SSV) into MANET layer model represented an increase in the values as compared with standard layer model (DSR). After applying the cross layer model to MANET the delay was reduced, as compared with DSR+SSV. These situations could be caused by following factors: i) density distribution of nodes and their mobility - the values depended on the distribution and movement of nodes and ii) activity modified SSV and CLD - the delay would increase mainly decision algorithms at routing nodes.

4 Conclusions

This article presents an implementation of our designed model, which can be used to integrate QoS and security as one parameter in MANET, to the well-know simulator called OPNET Modeler. The performance analysis is introduced and tested too. The results obtained in delay and total packet processing delay indicate that to integrate the modified SSV with CLD resulted in insignificant increase of delays of the MANET network and of total processing delay. When performance of implemented modified SSV and CLD model was simulated in MANET, comparable results were achieved in the DSR model.

Acknowledgments. This work has been performed partially in the framework of the EU ICT Project INDECT (FP7- 218086) and by the Ministry of Education of Slovak Republic under research VEGA 1/0065/10.

References

1. OPNET Modeler Simulation Software, http://www.opnet.com
2. Gerla, M.: Ad hoc networks: Emerging applications, design challenges and future opportunities. Ad Hoc Networks: Technologies and Protocols, 1–45 (2004)
3. Patwardhan, A., Parker, J., Joshi, A., Karygiannis, A., Iorga, M.: Secure Routing and Intrusion Detection in Ad Hoc Networks. In: Third IEEE International Conference on Pervasive Computing and Communications, pp. 8–12 (2005)
4. Srivastava, S., Motani, M.: The road ahead for cross-layer design. In: Proceedings of 2005 2nd International Conference on Broadband Networks, pp. 551–556 (2005)
5. Papaj, J.: Proposal for a integration model of Security and Quality of Service in mobile ad-hoc network. Doctoral dissertation, Technical University of Košice (2010)
6. Sakarindr, P., Ansari, N., Rojas-Cessa, N., Papavassiliou, S.: Security-enhanced Quality of Service (SQoS) networks: a network analysis. In: Military Communications Conference MILCOM, vol. 4, pp. 2165–2171. IEEE, Los Alamitos (2005)
7. Sakarindr, P., Ansari, N., Rojas-Cessa, N., Papavassiliou, S.: Security-enhanced Quality of Service (SQoS) networks. In: IEEE Sarnoff Symposium on Advanced in Wired and Wireless Communications, pp. 129–132 (2005)
8. Papaj, J., Doboš, Ľ., Čižmár, A.: New cross layer model to integration QoS and security as one parameter in mobile ad hoc network. In: IEEE International Conference on Multimedia Communications, Services and Security, MCSS 2010, Krakow, May 6-7, pp. 1–6 (2010); ISBN 978-83-88309-92-2

One Approach of Using Key-Dependent S-BOXes in AES

Nikolai Stoianov

Technical University of Sofia, INDECT project team
8, Kliment Ohridski Str., 1000, Sofia, Bulgaria
nkl_stnv@tu-sofia.bg

Abstract. The use of key-dependent substitution matrices (S-BOXes) is one of the commonly used methods to change the characteristics of a cryptographic algorithm. In this report is presented one approach for changing the S-BOXes used in the algorithm AES. By this approach are used substitution matrices which depend on the key while the parameters of the created new S-BOXes have characteristics equal to those in the original algorithm AES. The proposed new matrices were tested for Balancing, Nonlinearity (Additivity and Completeness), Strict Avalanche Criterion (SAC), Low XOR Table, Diffusion Order and Invertability using the software simulator developed by Work Package 8 (WP8) in the project INDECT. The new S-BOXes were also tested for static and dynamic independence between the input and output data.

Keywords: cryptography, substitution, S-BOX, AES, Rijndael.

1 Introduction

Developments in information technology and in particular the multiplication of the speed of processor devices necessitated the need to review the used cryptographic algorithms. The National Institute of Standards and Technology of USA (National Institute of Standards and Technology - NIST) jointly with the industry and cryptographic communities worked together to create a new cryptographic standard. The main objective was a federal standard (Federal Information Processing Standard - FIPS) to be created, which specifies a cryptographic algorithm (s) with the possibility of better protection of the sensitive government information. It was expected that the algorithm can be used both in the governmental structures of the United States and in the corporate and private sectors. [1], [2], [3]. After thorough analysis (mathematical, cryptographic, statistical, engineering, etc.) of the algorithms, NIST announced that the new standard would use the Rijndael algorithm and since 2001 it is the basis of the new cryptographic standard AES.

2 Defining the Problem

AES is based on the well-known practice "substitution-permutation network". It is relatively fast in both types of implementation (software and hardware), unlike DES

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 317–323, 2011.

and 3DES. Also, Feistel network is not used in AES. In the algorithm, the block size is fixed and is 128 bits and key length is 128, 192 or 256 bits. Considering, that one byte is equal to 8 bits, the size of the data block is 128 / 8 = 16 bytes. AES operates with an array of size 4x4 bytes in each cycle (period) of encryption / decryption. Most of the calculations performed in the algorithm are carried out on finite fields. In general it can be said that AES is based on a repetition of transforming cycles that convert the incoming explicit text to an output encrypted text. Each cycle consists of several steps and always includes one which depends on the cryptographic key. Multiple reverse cycles determine the transformation of the encrypted text in the original, using the same cryptographic key [2], [3], [4].

One of the main functions used in AES is the function “Substitution of bytes (SubBytes)". This function performs a non-linear substitution, which is carried out independently on each input byte. The matrix which gives the relationship between input and output bytes, so called (S-BOX) in the AES algorithm is invertible.

Each such a matrix must meet the following criteria: Balancing, Nonlinearity, Completeness, The Strict Avalanche Criterion (SAC), Low XOR Table, Diffusion Order, Invertability, Static criteria (Independence between the input and output data, Independence between the output and input data, Independence between the output and output data), Dynamic criteria (Dynamic Independence between the input and output data, Dynamic Independence between the output and input data, Dynamic Independence between the output and output data), Private criteria (Completeness of S-BOX and Non-contradiction). These criteria are defined and described in details in [5], [6], [7].

So the defined requirements that each S-BOX must meet are determined by the algorithm necessity to be stable to both linear and differential cryptanalysis.

Therefore, to meet the requirements specified above it, is necessary to be found new substitution matrices to be applied in the algorithm depending on the parameters or values of key and at the same time these S-BOXes to be such that their characteristics be the same or better than those used in the standard AES.

3 Determination of the Usable S-BOXes

According to the published by NIST standard AES [2] is offered one S-BOX for the operation encryption and one (the opposite of it) S-BOX for the operation decryption.

These S-BOXes were tested by the software simulator developed in Work Package 8 (WP8) in the Project INDECT.

The results obtained for the substitution matrix used in encryption are as follows:

```
Probability of changing output bit if one input bit is
changed (SAC) equals: 50%
Completeness is ensured for 100% possible inputs
Faultily XOR distribution
Diffusion order equals: 0
```

```
Function is balanced
Minimum distance to affine function equals (bent
function has 128):
For 1 bit 112  (most significant)
For 2 bit 112
For 3 bit 112
For 4 bit 112
For 5 bit 112
For 6 bit 112
For 7 bit 112
For 8 bit 112
```

The results obtained in testing the inverted S-BOX (used to decrypt) are as follows:

```
Probability of changing output bit if one input bit is
changed (SAC) equals: 50%
Completeness is ensured for 100% possible inputs
Faultily XOR distribution
Diffusion order equals: 0
Function is balanced
Minimum distance to affine function equals (bent
function has 128):
For 1 bit 112  (most significant)
For 2 bit 112
For 3 bit 112
For 4 bit 112
For 5 bit 112
For 6 bit 112
For 7 bit 112
For 8 bit 112
```

Analyzing the data shows that the results are identical. It is therefore logical to assume that both S-BOXes are interchangeable, i.e. so called Inv S-BOX can be used for encryption and S-BOX to be used for decryption operation.

4 Proposal for Two New S-BOXes

Taking as a basis the substitution matrix used in encryption to AES it is necessary other such matrices with the same or better features to be found. The main S-BOX, used in the encryption operation is considered as a basal. On the basis of the left diagonal a change of the locations of the corresponding bytes in the matrix is performed. Thus gets a new S-BOX that will call it $S\text{-}BOX_{Left}$. Figure 1 schematically shows the way of the change of the cells by using the left diagonal as symmetrical axis.

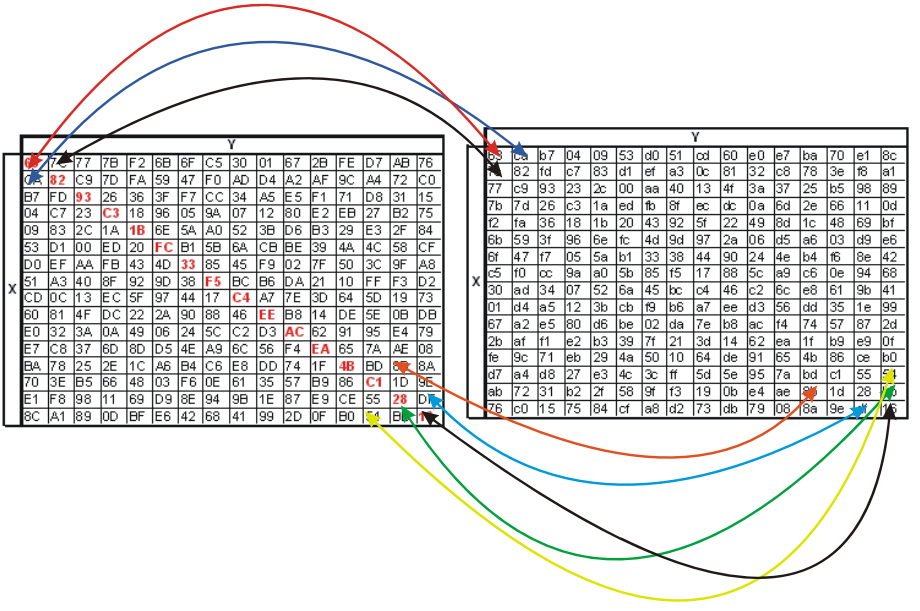


Fig. 1. Scheme of obtaining S-BOXLeft

Analogically, a new S-BOX$_{\text{Right}}$ is obtained, using the right diagonal as symmetry axis. Figure 2 shows the scheme to compute S-BOX$_{\text{Right}}$.

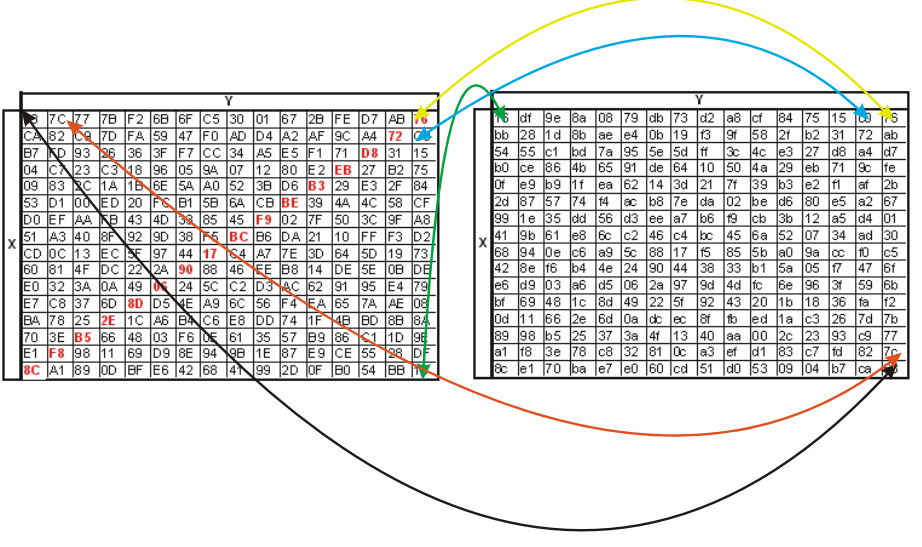


Fig. 2. Scheme of obtaining S-BOXRight

Two new S-BOXes are obtained by such presented transformations. They are shown respectively in Table 1 and Table 2.

Table 1. S-BOX$_{Left}$

63	ca	b7	04	09	53	d0	51	cd	60	e0	e7	ba	70	e1	8c
7c	82	fd	c7	83	d1	ef	a3	0c	81	32	c8	78	3e	f8	a1
77	c9	93	23	2c	00	aa	40	13	4f	3a	37	25	b5	98	89
7b	7d	26	c3	1a	ed	fb	8f	ec	dc	0a	6d	2e	66	11	0d
f2	fa	36	18	1b	20	43	92	5f	22	49	8d	1c	48	69	bf
6b	59	3f	96	6e	fc	4d	9d	97	2a	06	d5	a6	03	d9	e6
6f	47	f7	05	5a	b1	33	38	44	90	24	4e	b4	f6	8e	42
c5	f0	cc	9a	a0	5b	85	f5	17	88	5c	a9	c6	0e	94	68
30	ad	34	07	52	6a	45	bc	c4	46	c2	6c	e8	61	9b	41
01	d4	a5	12	3b	cb	f9	b6	a7	ee	d3	56	dd	35	1e	99
67	a2	e5	80	d6	be	02	da	7e	b8	ac	f4	74	57	87	2d
2b	af	f1	e2	b3	39	7f	21	3d	14	62	ea	1f	b9	e9	0f
fe	9c	71	eb	29	4a	50	10	64	de	91	65	4b	86	ce	b0
d7	a4	d8	27	e3	4c	3c	ff	5d	5e	95	7a	bd	c1	55	54
ab	72	31	b2	2f	58	9f	f3	19	0b	e4	ae	8b	1d	28	bb
76	c0	15	75	84	cf	a8	d2	73	db	79	08	8a	9e	df	16

Table 2. S-BOX$_{Right}$

16	df	9e	8a	08	79	db	73	d2	a8	cf	84	75	15	c0	76
bb	28	1d	8b	ae	e4	0b	19	f3	9f	58	2f	b2	31	72	ab
54	55	c1	bd	7a	95	5e	5d	ff	3c	4c	e3	27	d8	a4	d7
b0	ce	86	4b	65	91	de	64	10	50	4a	29	eb	71	9c	fe
0f	e9	b9	1f	ea	62	14	3d	21	7f	39	b3	e2	f1	af	2b
2d	87	57	74	f4	ac	b8	7e	da	02	be	d6	80	e5	a2	67
99	1e	35	dd	56	d3	ee	a7	b6	f9	cb	3b	12	a5	d4	01
41	9b	61	e8	6c	c2	46	c4	bc	45	6a	52	07	34	ad	30
68	94	0e	c6	a9	5c	88	17	f5	85	5b	a0	9a	cc	f0	c5
42	8e	f6	b4	4e	24	90	44	38	33	b1	5a	05	f7	47	6f
e6	d9	03	a6	d5	06	2a	97	9d	4d	fc	6e	96	3f	59	6b
bf	69	48	1c	8d	49	22	5f	92	43	20	1b	18	36	fa	f2
0d	11	66	2e	6d	0a	dc	ec	8f	fb	ed	1a	c3	26	7d	7b
89	98	b5	25	37	3a	4f	13	40	aa	00	2c	23	93	c9	77
a1	f8	3e	78	c8	32	81	0c	a3	ef	d1	83	c7	fd	82	7c
8c	e1	70	ba	e7	e0	60	cd	51	d0	53	09	04	b7	ca	63

To meet the necessary performance the resulting new matrices are tested by software simulation. The results for S-BOX$_{Left}$ are:

```
Probability of changing output bit if one input bit is
changed (SAC) equals: 50%
Completeness is ensured for 100% possible inputs
Faultily XOR distribution
Diffusion order equals: 0
Function is balanced
Minimum distance to affine function equals (bent
function has 128):
For 1 bit 112  (most significant)
```

```
For 2 bit 112
For 3 bit 112
For 4 bit 112
For 5 bit 112
For 6 bit 112
For 7 bit 112
For 8 bit 112
```

The results obtained for S-BOX_{Right} are:

```
Probability of changing output bit if one input bit is
changed (SAC) equals: 50%
Completeness is ensured for 100% possible inputs
Faultily XOR distribution
Diffusion order equals: 0
Function is balanced
Minimum distance to affine function equals (bent
function has 128):
For 1 bit 112  (most significant)
For 2 bit 112
For 3 bit 112
For 4 bit 112
For 5 bit 112
For 6 bit 112
For 7 bit 112
For 8 bit 112
```

It is clear that the characteristics of the resulting two S-BOXes (S-BOX_{Left} and S-BOX_{Right}) are identical to those of their prototype. It can therefore be assumed that irrespectively which of the proposed matrices will be used, the stability of the algorithm to linear and differential cryptanalysis will be the same. This can be claim because this stability of AES depends on the characteristics of the used substitution matrices.

5 Algorithm Using the S-BOXes Depending on the AES Key

On the basis of the four surveyed substitution matrices and depending on the particular value of the encryption key used in AES the following algorithm is offering:

1. Select a key for AES;
2. The first byte of the Key (it could be anyone else) is divided into 4 integers (mod 4).
3. Depending on the remainder in the integer division one of four S-BOXes is selected i.e.:
    - In a remainder of 0 for encryption is selected S-BOX;
    - In a remainder 1 for encryption is selected Inv S-BOX;
    - In a remainder 2 for encryption is selected S-BOX_{Left};
    - In a remainder 3 for encryption is selected S-BOX_{Right};
4. Continue according to the algorithm set out in AES.

6 Conclusion

In this paper a research of the characteristics (Balancing, Nonlinearity, Completeness, The Strict Avalanche Criterion (SAC), Low XOR Table, Diffusion Order, Invertability, Static criteria (Independence between the input and output data, Independence between the output and input data, Independence between the output and output data), Dynamic criteria (Dynamic Independence between the input and output data, Dynamic Independence between the output and input data, Dynamic Independence between the output and output data), Private criteria (Completeness of S-BOX and Non -contradiction) of the proposed in the standard AES S-BOXes was carried out. Two new substitution matrices have been developed by using the left and right diagonal as axis of symmetry. These matrices were tested with the software simulator developed by WP8 in the Project INDECT for the same characteristics. Analyzing the results shows that the characteristics of the four S-BOXes are identical from which a conclusion is drawn: it is possible to use each of them for an encryption. This will not lead to a deterioration of stability of AES to linear and differential cryptanalysis. An algorithm for using these matrices is proposed, as it is based on a pre-selected byte of the used key and depending on the result in the integer division with it to 4, one of the four S-BOXes is selected.

Acknowledgments. This work has been funded by the EU Project INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment*) — grant agreement number: 218086.

References

1. Schneier, B.: AES Announced, October 15 (2000)
2. FIPS PUB 197, Advanced Encryption Standard (AES), National Institute of Standards and Technology, U.S. Department of Commerce (November 2001), http://csrc.nist.gov/publications/fips/fips197/fips-197.pdf
3. Advanced Encryption Standard (AES), National Institute of Standards and Technology, http://csrc.nist.gov/archive/aes/index.html
4. Schneier, B.: AES News, Crypto-Gram Newsletter, September 15 (2002), http://www.schneier.com/crypto-gram-0209.html (retrieved July 27, 2007)
5. Dawson, M.H., Tavares, S.: An expanded set of S-box design criteria based on information theory and its relation to differential-like attacks. In: Davies, D.W. (ed.) EUROCRYPT 1991. LNCS, vol. 547, pp. 352–367. Springer, Heidelberg (1991)
6. Stoianov, N.: AES S-BOX generator: analysis of requirements. In: International Science Conference 2009, Communication and Information Systems, Shoumen, Bulgaria (2010)
7. INDECT Consortium, D8.2: Evaluation of Components (June 2010), http://www.indect-project.eu/files/deliverables/public/deliverable-8.2

Scalability Study of Wireless Mesh Networks with Dynamic Stream Merging Capability

Jun Ye and Kien A. Hua

School of Electrical Engineering and Computer Science, University of Central Florida.
Orlando, Florida 32816, U.S.A.
{jye,kienhua}@eecs.ucf.edu

Abstract. In recent years, there has been a dramatic increase in the number of users who access online videos from wireless access networks. It is highly desirable that such wireless networks are robust in handling sudden spurts in demand for some videos due to special events. An abrupt increase in the network usage should not significantly impact normal access to other applications. In this work, we tackle this problem in wireless mesh access networks by applying a distributed video sharing technique called Dynamic Stream Merging (DSM). DSM improves the robustness of the access network without directives from the video servers. We perform comprehensive simulation study based on a grid network topology. The experimental results, based on NS2 simulation, indicate that DSM is highly scalable.

Keywords: Wireless Mesh Network, Access Network, Dynamic Stream Merging, Scalability Study.

1 Introduction

Rapid advances in networking technology have enabled large-scale deployment of online video streaming services in today's Internet. In particular, wireless technology is expected to dominate much of the broadband access and a significant share of the distribution markets. Unlike wired environment, wireless networks have very limited bandwidth. It is desirable to have a robust wireless access environment that can sustain sudden spurts of interests for specific videos due to, say current events.

By taking advantage of broadband wireless technologies such as 802.11 and WiMax, *wireless mesh network* (WMN), a special case of *mobile ad hoc network* (MANET) with mesh topology, has been used as an edge technology to provide broadband Internet access in residential, business, and even city-wide networks. In such a network, the wireless mesh access routers form a mesh-like wireless backbone network that allows users to access services in the Internet through mesh gateways. We refer to such a network environment as a *Wireless Mesh Access* (WMA) network. In this paper, we study the scalability of a robust WMA network design based on a technique called *Dynamic Stream Merging* (DSM). We first discuss related work in Section 2. DSM is then introduced in Section 3. In Section 4, we present our simulation study. Finally, we conclude this paper in Section 5.

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 324–330, 2011.

2 Related Work

A promising approach to sustain spurts of demand for certain videos is to share video streams in the video delivery. Techniques such as periodic broadcast [1], overlay multicast [2][3] and Patching [4] are designed to facilitate video stream sharing. However, these multicast techniques would require the continuous coordination between the video servers in the Internet and various WMA networks at the edges. This is generally difficult to achieve efficiently in practice.

On the other hand, several video multicast techniques [5][6] have been proposed for stand-alone WMN's to facilitate data sharing. In this environment, the video server is a node on the WMN, and therefore may participate in the mesh network protocols. In practice, video servers are generally separated from the WMA networks by the Internet. To support this environment, we apply in this paper a method called *Dynamic Stream Merging* (DSM) [7]. It does not require directives from the video servers located in the Internet. Instead, DSM is a light-weight distributed solution which can improve the robustness of a WMA environment without imposing significant overhead on the network. The DSM technique was originally proposed in our previous work [7], where we present a prototype, with a 3-node linear topology, to demonstrate the operational capability of the system and the benefits of the technique. The purpose of the current paper is to investigate the scalability of DSM through simulation study. We target a much larger mesh network environment with a more complex topology.

3 Dynamic Stream Merging

We briefly present the DSM technique in this section to make the paper self-contained. More details of the DSM technique can be found in [7]. Generally speaking, the main idea of DSM is the opportunistic reuse of data at each mesh node in order to save bandwidth for other applications. DSM consists of three basic functions: stream merging, buffer management, and merge cancelling. We discuss them in the rest of the section.

A mesh network can be modeled as a directed graph G(V, E) where V is the set of mesh nodes and E is the set of links. A link <i, j> is a pair of two mesh nodes and who are in the communication range of each other, nodes are labeled with i, j where i, j = 1, 2, …, N, and N = | V |. We model a video stream as a sequence of segments and assume that the segment ID of a video stream starts from 1 and grows in increasing order. We use t_i^x to represent the highest segment ID of all the video segments that have arrived at node i from stream x.

We explain the merging process by giving a simple example as shown in Fig. 1(a). Two identical streams, s_1 and s_2, are passing through link <1, 2>. After node 1 receives segment 5 from stream s_1 (i.e., $t_1^{s_1} = 5$) and segment 2 from stream s_2 (i.e., $t_1^{s_2} = 2$), node 1 notices that both streams s_1 and s_2 are the same video stream and they share the same next hop (i.e., node 2). Node 1 then notifies node 2 of its intension to merge these two streams (denoted as "$s_1 \rightarrow s_2$") and soon block stream s_2 at node 1, after segment 5. We refer to this segment as $\tau_1^{s_2}$, the last segment from stream s_2 before the stream merging occurs. In response, node 2 treats data packets arriving from s_1 as data for

both streams. That is, after node 2 finishes forwarding segments 2, 3, 4, and 5 from stream s_2 to the next hop, node 2 continues to forward the subsequent segments by reusing these segments received earlier from stream s_1. This can be achieved by spoofing the packet header with the header information of s_2. The desirable effect of this merger is that we have one less stream transmission between nodes 1 and 2.

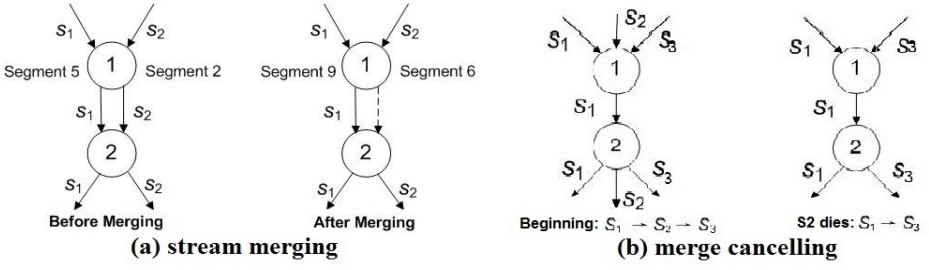


Fig. 1. Merging and cancelling process

To facilitate reuse of the data segments for the other stream later, we need to cache the current segments to a buffer. By buffering, DSM ensures that data packets are relayed in the same order they arrive at a mesh node. Suppose two streams x and y of the same video are forwarded at link <*i, j*>; and stream y begins to merge stream x (i.e., $y \rightarrow x$) at τ_i^x. The buffering can be done as follows. Node j caches y's segments after segment τ_i^x into a buffer while it is relaying the segments between t_i^x and τ_i^x from stream x to the next hop in the downstream. When node j does not see any more data arriving in stream x from node i (i.e., i has blocked the stream), node j fixes the size of the buffer and starts to use it as a FIFO queue as follows. Each subsequent data segment from stream y is appended to the end of the queue while node j removes the segment at the head of the queue and forwards it to the next hop. By using this buffering scheme, node j maintains the order of stream x while it also reuses the data from y for x. The buffer is released at the end of the video session.

Due to the dynamics of the network, we may have to cancel a merger in a link. For example, consider the merger $s_1 \rightarrow s_2$ at link <1, 2> as illustrated in Fig. 1(a). If stream s_2 is terminated by the end user or the routing protocol diverts s_2 from node 1, it will experience the discontinuation of stream s_2. In a different scenario, if the routing protocol changes the next hop of s_2 at node 1, node 1 can recognize this change from the routing table. In any of these cases, we say that stream s_2 is dead at node 1; and it needs to inform node 2 to stop caching data for stream s_2 and release the FIFO buffer.

Fig. 1(b) illustrates the process of cancelling a merged stream. In this example, we consider 3 streams s_1, s_2 and s_3 with the same video ID. We label the subscripts of the streams according to the order of their start time, in which s_1 starts first. At the beginning, only s_1, s_2, and s_3 are passing on link <1, 2>. In this scenario, we have the merging chain $s_1 \rightarrow s_2 \rightarrow s_3$. That is, we first have $s_1 \rightarrow s_2$, and this merger then merges s_3. As a result, node 1 needs only send s_1 to node 2 while node 2 reuses the data from s_1 for both s_2 and s_3. Sometime later, s_2 is dead at node 1. The affected mergers $s_1 \rightarrow s_2$ and $s_2 \rightarrow s_3$ are cancelled at link <1, 2>. Since s_3 is now set as unmerged, it is re-merged by s_1 (i.e., $s_1 \rightarrow s_3$).

4 Performance Study

Our current prototype [7] performs reliably with seven mesh nodes (using seven small netbook computers). However, DSM can efficiently support much larger mesh networks. To study its scalability, we conducted simulations using NS-2.34 [8] due to limited hardware resources.

4.1 Experiment Setup

In the simulation study, the transmission range and carrier sensing range are both set to 140 meters. RTS/CTS is disabled to reduce the overhead. The MAC parameters are set according to IEEE 802.11g standard, where basic data rate is set to 54Mbps. AODV is employed as the unicast routing protocol. Evalvid [9] framework is employed as an extension of video transmission in NS2. DSM is set to allow merging of two video streams if the temporal difference in their start times is less than 10 seconds. An mp4 video clip, encoded by an MPEG-4 encoder with 100 seconds in length, is used. Since we repeat each simulation run 100 times for each network scenario to obtain the average results, a shorter video than typical ones was used in our study for the sake of reasonable simulation times.

A 5×5 grid topology is applied in our study. The distance between any two adjacent nodes is 100 meters. The 25 nodes cover an area of 700x700 square meters. Node 0 is the gateway and is also the video source. Node contention is significantly more severe in a grid topology as compared to simpler topologies such as a linear chain. To increase the network capacity, we further implemented multi-channel multi-interface (MCMI) extension of NS2 [10] for both DSM and non-DSM techniques in our simulation. Since channel assignment techniques are beyond the scope of our study, we simply adopted the technique presented in [10] for our implementation.

In this study, we assume that video requests are equally likely initiated by clients at any one of the 24 non-gateway mesh nodes. The arrival of these requests follows a Poisson process. We varied the average inter-arrival time to investigate the different degrees of stress on the mesh network. We note that our mesh network design is intended for a wireless access network. However, since this work focuses on the scalability of DSM in the wireless environment, we perform our simulation based on a stand-alone wireless mesh network.

4.2 Experimental Results

Three performance metrics are used in this study: (1) *Packet loss rate* is the number of packets received at the destination divided by the total number of packets for the video clip. In our experiment, the UDP packets are set with a maximum size of 1052 bytes. (2) *End-to-end delay* is calculated as the difference of time elapsed between the time a packet is initiated at the gateway and the time it is successfully received at the destination. Since packet can be dropped during the transmission, we only take into account the received packets in the end-to-end delay study. (3) *Node workload* is the amount of data forwarding activity measured at each mesh node in terms of the total number of data packets forwarded. All the performance data presented herein is the average over 100 simulation runs under the same setting.

We performed stress test to validate the scalability of DSM with a grid topology under heavy traffic. In each test setting, a request for the same video is initiated with a constant arrival rate λ at any one of the 24 mesh nodes with equal probability. This random process repeated until the total number of video streams reaches a set value. For a given number of concurrent video streams, a more scalable design should place less stress on the network and experience less average packet loss rate.

The results for the stress test are in plotted in Fig. 2(a). It reveals that the stress on the DSM network is significantly less (i.e., the packet loss rate is much lower) compared to that of the non-DSM network. As the number of concurrent streams increases, the packet loss rate gradually rises to 0.15 when there are 30 concurrent streams in the DSM network; whereas it soars all the way to 0.7 for the non-DSM network. From a different perspective, if the packet loss rate is not allowed to exceed 0.15, then non-DSM can serve no more than 7 concurrent users while DSM can sustain more than four times (i.e., 30) as many users. This observation confirms that DSM is a very robust design that can absorb a much bigger spurt in the video demand.

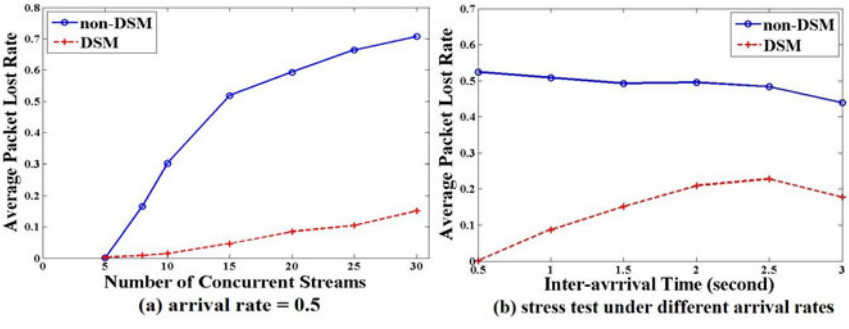


Fig. 2. Stress test under concurrent video streams

To investigate the impact of arrival rate λ on DSM, we repeat the experiments of stress test under different inter-arrival times. The number of concurrent streams is fixed to 15 and inter-arrival times of the video requests vary from 0.5 to 3.0 seconds. These inter-arrival times are randomly generated as follows: $t_i = -log(R_V)/\lambda$, where t_i is the inter-arrival time between stream i and stream i+1, and R_V is a uniformly distributed random variable ranging from 0 to 1. The simulation results are plotted in Fig. 2(b). As the t_i increases, the traffic reduces slightly for non-DSM resulting in a slightly lower average packet loss rate. In contrast, the advantage of DSM reduces with a higher t_i due to the decreases in the opportunity for stream merging. Nevertheless, the overall performance gaps between the two techniques are still very large throughout. We can expect that this performance gap disappears when the inter-arrival time is increased beyond 10 seconds. This is due to the fact that we set the merging criterion as 10 seconds (i.e., 1/10 of the video length), and very few streams can be merged when the inter-arrival time exceeds 10 seconds. We note that this does not mean that DSM is beneficial only when inter-arrival time is noticeably less than 10 seconds. This is only due to the very short video (100 seconds) used in our simulation study and we have to set the merging criterion very small accordingly. In practice, we often have much longer videos and a 10% of the video length would provide a very significant temporal window for merging video streams.

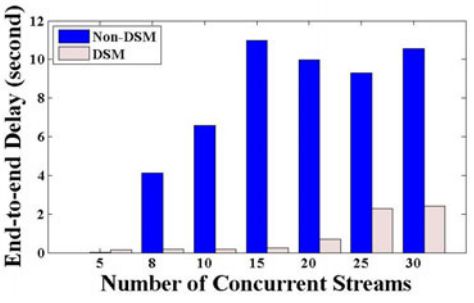


Fig. 3. Average end to end delay with arrival rate is 0.5

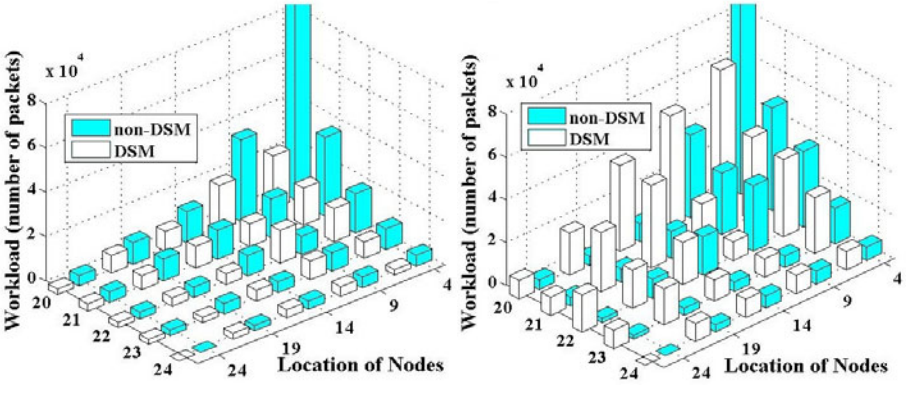


Fig. 4. Workload under different traffic conditions (left:10 streams, right:25 streams)

We further investigated the end-to-end delay. We stressed the network by gradually increasing the number of concurrent streams. The result is shown in Fig. 3. Again, we observe that non-DSM suffers significantly higher end-to-end delay, whereas DSM can sustain the spurts in demand for the video much better.

Node workload is a good measure to get insight into the performance characteristics of DSM. We plot the workload for each of the 25 mesh nodes under both DSM and non-DSM in Fig. 4. This study considers both a light traffic condition (10 streams) and a heavy traffic condition (25 streams). We observe the same behaviour in both scenarios – the gateway has the highest traffic and nodes closer to the gateway experience higher workloads. Under the light traffic condition (10 streams), DSM, in comparison, is much less demanding on the gateway bandwidth leaving much of the bandwidth for other applications and users. This is evident under the heavy traffic condition (25 streams). Since DSM, in this scenario, does not exhaust the gateway capacity, it allows every mesh node in the network to participate in data forwarding. In other words, users farther away from the gateway can also use the network and experience good performance. In contrast, non-DSM quickly exhausts the gateway. Heavy congestion renders many packet drops at the gateway. Consequently, mesh nodes farther away from the gateway have little data to forward despite there are a large number of active streams in the network.

5 Conclusion

In this work, we performed simulation to study the scalability of a technique, called *Dynamic Stream Merging* (DSM), in providing a robust wireless mesh access environment. The simulation results based on a 25-node grid topology indicate that DSM is effective in mitigating the stress on the network when it experiences spurts of demand in certain videos due to situations such as special events. Under such conditions, a DSM mesh network does not overload the network leaving much of its bandwidth available for other applications and users. DSM also lessens the demand on the gateway. This allows a given gateway to support a larger mesh access network to service a larger user community.

Our future work includes study of handoff techniques to support mobile users. Furthermore, we will consider optimization techniques in multi-channel multi-interface allocation to further enhance DSM performance.

Acknowledgements

The author would like to thank the reviewers for their insightful comments and feedback. This work was partially supported by National Science Foundations under Grant Number CNS-0917082.

References

1. Hua, K.A., Sheu, S.: Skyscraper Broadcasting: A New Broadcasting Scheme For Metropolitan Video-On-Demand Systems. In: ACM SIGCOMM, pp. 89–100 (1997)
2. Banerjee, S., Kommareddy, C., Kar, K., Bhattacharjee, B., Khuller, S.: Construction ofAn Efficient Overlay Multicast Infrastructure for Realtime Applications. In: IEEE INFOCOM (2003)
3. Cui, Y., Li, B., Nahrstedt, K.: oStream: Asynchronous Streaming Multicast in Application-Layer Overlay Networks. J. IEEE Journal on Selected Areas in Communications, Special Issue on Recent Advances in Service Overlay Networks 22(1), 91–106 (2004)
4. Hua, K.A., Cai, Y., Sheu, S.: Patching: A Multicast Technique for The Video-on-Demand Services. In: ACM Multimedia, Bristol, UK, pp. 191–200 (1998)
5. Xie, F., Hua, K.A., Jiang, N.: Optimizing Patching-based Multicast for Video-on-Demand in Wireless Mesh Networks. J. International Journal of Comm. Sys. (2009)
6. Zhu, X., Schierl, T., Wiegand, T., Girod, B.: Video Multicast over Wireless Mesh Networks with Scalable Video Coding (SVC). In: Visual Communication and Image Processing, San Jose, CA (2008)
7. Hua, K.A., Fei, X.: A Dynamic Stream Merging Technique for Video-on-Demand Services over Wireless Mesh Access Networks. In: IEEE SECON, pp. 1–9 (2010)
8. The Network Simulator – NS2, `http://www.isi.edu/nsnam/ns/`
9. EvalVid, `http://www.tkn.tu-berlin.de/research/evalvid/`
10. Calvo, R.A., Campo, J.P.: Adding Multiple Interface Support in NS-2 (2007), `http://personales.unican.es/aguerocr/files/ucMultiIfacesSupport.pdf`

Overview of the Security Components of INDECT Project

Nikolai Stoianov[1], Manuel Urueña[2], Marcin Niemiec[3], Petr Machník[4], and Gema Maestro[5]

[1] Technical University of Sofia, INDECT Project Team, 8, Kliment Ohridski St., 1000 Sofia, Bulgaria
nkl_stnv@tu-sofia.bg
[2] Universidad Carlos III de Madrid, Departamento de Ingeniería Telemática, Avda. de la Universidad, 30 E-28911 Leganés (Madrid), Spain
muruenya@it.uc3m.es
[3] AGH University of Science and Technology, Department of Telecommunications, Mickiewicza 30 Ave., 30-059 Krakow, Poland
niemiec@kt.agh.edu.pl
[4] VSB-Technical University of Ostrava, Department of Telecommunications, 17. listopadu 15, 708 33, Ostrava, Czech Republic
petr.machnik@vsb.cz
[5] APIF Moviquity SA Madrid, Madrid, Spain
gmm@moviquity.com

Abstract. In this paper an overview of the security components developed by the INDECT project is presented. This paper is focused on creating a multilevel security architecture, and describes the five basics areas for which new algorithms, components and tools are been created. In particular Virtual Private Networks, Symmetric Cryptography Block Ciphers, Quantum Cryptography, Federated ID Management and Secure Ad hoc Multipath Routing are described in detail.

Keywords: Virtual Private Networks, Symmetric Cryptography Block Ciphers, Quantum Cryptography, Federated ID Management, Secure Ad hoc Routing Protocol.

1 Introduction

INDECT (Intelligent information system supporting observation, searching and detection for security of citizens in urban environment) [1] is a Collaborative Research Project funded by the 7th EU Framework Program. Its main aim is to develop cost-efficient tools for helping European Police services to enforce the law and guarantee the protection of European citizens. These tools must comply with both country-level laws as well as European-level directives including, among many others, the European Declaration on Human Rights. Information security and data confidentiality are two of the main topics of the project. This paper is focused on new security algorithms, protocols, tools and techniques being developed by Work

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 331–337, 2011.

Package 8 (WP8) of the INDECT project. The objective of this paper is to present the security-related issues required to cover all aspects of data protection by means of one holistic multilevel security platform.

2 Security Components of INDECT Project

Virtual Private Networks

Because the INDECT system has a distributed structure with multiple remote nodes and servers interconnected over public networks, it is necessary to ensure the secure communication among these remote nodes and servers. For this purpose, the best solution seems to be the implementation of a virtual private network (VPN). By using a VPN, it is possible to secure data transmission over insecure public networks, which is a very important feature in the INDECT system where sensitive information is exchanged among its different subsystems. The security of the communications in the INDECT system is based on two types of VPN technologies – IPsec (Internet Protocol Security) VPN and SSL (Secure Socket Layer) VPN – which are the most widespread VPN technologies nowadays. Each of these technologies is more suitable for different scenario. It is also possible to use multiple VPN solutions in the same network.

The most important tasks of the VPN implementation that have been carried out are: Analysis of the INDECT system communication networks – their topologies, number of terminals and network nodes, application types, amount of traffic transmitted, transmission media (wired or wireless), ownership of constituent network segments (public or private networks), possible future growth of these networks, etc. Various security levels will be defined for individual VPNs with diverse security requirements, since different VPNs require differently strong encryption algorithms or authentication mechanisms. On the basis of the network topology, the most suitable types of VPN connections will be chosen - a remote access or a site-to-site connection, a point-to-point or a hub-and-spokes connection. For each virtual private network, the most convenient technology will be employed – IPsec or SSL VPN.

SSL and IPsec VPNs are well-known technologies so testing their cryptographic security is not necessary. On the other hand, it is desirable to evaluate the operation of particular VPN implementations for the INDECT system because otherwise a secure VPN technology that is badly implemented does not guarantee to be secure. To achieve this goal, the following evaluation tools can be used: The correct creation of secure tunnels should be tested, which means to evaluate whether end nodes and users are authenticated, all the transmitted traffic is encrypted and the data integrity guaranteed, and the anti-replay protection works. It should be also tested whether transmission parameters like delay, jitter, throughput, or packet loss are not negatively influenced as a consequence of the VPN usage. Finally, it should be evaluated whether end users are satisfied with the chosen VPN.

The most important benefits of IPsec and SSL virtual private networks are as follows: Secure communication over insecure public networks which are ensured by data confidentiality, integrity protection, anti-replay protection, and authentication. In the case of IPsec VPNs, all types of unicast IP traffic can be secured. In the case of IPsec/GRE VPNs, all types of traffic can be secured. IPsec VPNs support site-to-site

and remote access communication. SSL VPNs support remote access communication. In a SSL VPN, the secure remote access can be carried out from anywhere, only a terminal with a web browser is needed. SSL VPNs are very easy to use by end users. In the case of clientless or thin client SSL VPNs, users do not even need administrative rights on the used terminals.

Software simulator for testing symmetric cryptography algorithms

Encryption is the process of transforming information into unreadable to anyone except users who possess special knowledge. In this case information/data is called *plaintext* and its unreadable form is called *ciphertext*. If we want to encrypt plaintext (or decrypt ciphertext) we need the key, which must be used with an appropriate cryptography algorithm, called cipher. One of the mainly used functions in symmetric crypto algorithms is substitution. An S-box (substitution box) is a matrix consisting of `n` rows and `m` columns. This matrix produces (substitutes) input bits into output bits. Usually an S-box realizes a random and non-linear transformation. The defined requirements, to which every S-box shall comply, are dictated by the need of the algorithm to be reliable both to differential and linear crypto analysis and to be resistant against interpolating attacks. Considering those conditions, the following functions have been implemented in a newly developed software simulator [2]: Balancing, Nonlinearity, Completeness, SAC, Low XOR Table, Diffusion Order, Invertibility, Static criteria, Dynamic criteria and Private criteria.

During the work performed by WP8 we have created a simulator which checks some mathematical properties of cipher algorithms and S-boxes. These properties influence directly the security level of the new algorithm. The most important requirement for the simulator is flexibility, so we decide to check just the relationship between input and output bits of cipher (or element of cipher). This relationship is tested with application-defined functions that operate on arrays containing all combinations of inputs and outputs. WP8 has performed an in-depth research to verify security of cryptographic algorithms based on already proposed S-boxes. The analysis of properties of these S-boxes has motivated the creation of new ones and their comparison with the existing S-boxes in the AES standard. The simulator is able to test the properties of any substitution box for both AES and DES algorithms. The simulator is able to evaluate the full block cipher as well as a single S-box. Functions for testing statistical, dynamic and private criteria for each S-box and inverse S-boxes have been created.

The functions developed in the “Checks” module shall identify if the data provided for the S-box are correct. Having in mind that the statistic and dynamic tests take time, it is reasonable to check first each generated (delivered) S-box for correctness. Other checks include: *S-box completeness* - to check if the check subject S-box is complete i.e. if all the cells of the matrix have values. *Non-contradiction* avoids the duplication of values in the cells of the matrix. *Independence between input/output and output/input data* - For testing the independency between the input and output data, the substituted text and the number of the known input/output bytes are compared. *Dynamic independence between input/output and output/input data* - For testing the independence between the input and output data, different texts processed by the S-box and the number of the known input/output bytes are compared.

Quantum Cryptography

Quantum Key Distribution (QKD) was introduced in 1984 by Bennett and Brassard [3]. The most important and unique feature is that, based on laws of nature (physics), it can always assure faultless detection of eavesdropping. One of the main objectives of WP8 focuses on quantum cryptography [2]. The main work concentrates on the quantum cryptography methods for security and privacy assurance. During this work we need to define the methods of assessing the correctness and accuracy of the work associated with the development of a new quantum cryptography protocol. Such methods are highly important during the designing phase, and they provide results that are essential for the analysis of current research work. The developing process is based on a repeated cycle consisting in three stages: implementation, testing and improving. Thanks to the repetition of the cycle, the designer is able to continuously verify project progress and assess the significance of changes introduced in the previous iteration. For reliable evaluation of quantum cryptography protocols, another simulator has been developed. The main task of the simulations is to validate the quantum cryptography protocol's efficiency in the means of security. The application delivers the detailed protocol's results at each step of its operation. The simulator is expected to provide information regarding the transmission of key, which is determined by many parameters that define transmission channel, eavesdropping type, etc. The main requirements placed on development of simulator are: assuring the modular structure, ease of further development, and the best possible performance.

The main components of this application are: *Key source* block is responsible for generating the cryptographic keys (usually a symmetric/secret key). *QKD protocol* block enables the user to choose the proper QKD protocol and it is the responsible for simulating the protocol behavior. *Channel model* block contains the methods responsible for simulating quantum channel behavior. The user can choose a combination of three options: constant eavesdropping, burst eavesdropping and noise generation, as well as to set their parameters. *QBER estimation and reconciliation* block focuses on Quantum Bit Error Rate (QBER) estimation and the procedure for key reconciliation. QBER is similar to the classical Bit Error Rate (BER) concept, but in quantum protocols it is usually computed by sampling the raw key (the key obtained after its transmission though a quantum channel and then extracting the bits where the two sides used the same quantum bases). Reconciliation is based on parity check operations, performed on each block, and the results are compared via a public channel. If parity does not agree on both sides, a binary search algorithm is applied to blocks and the parity is checked every search round until the error is finally localized and can be eliminated. In this way, we obtain the reconciled quantum key, free of errors. The GUI provides the ability to set direct simulation parameters. Unless we use the "Continuous analysis" mode, a single execution of the simulation algorithm is carried out. Otherwise, we have to define other parameters: simulated feature of quantum transmission (constant eavesdropping rate, burst eavesdropping level or noise intensity), starting point, stop value and simulation step — where the last three values determine number of simulation rounds. Finally, the most important element in the program is the "control" block, which provides a bridge between the GUI and the described blocks. It collects parameters, adjusted by the user, from the interface and passes them to the functions of the appropriate blocks.

European ID authentication and identity management

The INDECT System has a distributed structure with many services, and it is necessary to implement a common authentication process for the final users, taking care of web services provisioning, and ensuring accessibility rights and anonymity via secure communications. This way the user is able to access all the services without having to repeat the identification procedures for each individual service, while ensuring that her/his access rights are recognized by the different services. The tasks to achieve this goal are: The integration of complete certification methodologies and the modeling of INDECT user interactions through the usage of X.509 certificates. To research the main platforms for Single Sign–On and Federated ID, in order to provide a unique platform for user authentication, and to enable the basic functionalities for certification management. The objectives of the ID Management are: to share an identity among different services and providers => Single Sign On, to keep a unified authentication provided by a trusted provider => Federated ID, and to keep a unique profile managed by a trusted provider => Circle of Trust (COT).

Single Sign On (SSO) is a means by which a Service Provider (SP) or Identity Provider (IdP) may convey to another Service Provider or Identity Provider that the user has been correctly authenticated. SSO solutions are designed to manage the credentials and passwords of users, providing a single credential and a single password at the beginning of the session, and then establishing a security context in order to achieve a secure and transparent access to all applications [2]. *Federated Identity* enables a single identity to be housed in an Identity Provider and to be accepted by one or more Service Providers. It is characterized by a relationship of trust among its members, allowing the communication and validation of user data in a safe manner without sharing directories as with SSO. The main initiatives and standards for Identity Federation are [2]: Liberty Alliance, WS-Federation and SAML (Security Assertion Markup Language). A User can be integrated in an Identity Federation when different Liberty entities know who this user is. This User is accepted, when accessing via this Identity Federation, into a *Circle of Trust (CoT).* Privacy is enhanced when an Identity Federation is given to the User so these entities don't need to know who the User is or her/his characteristics but they rely on an Identity Provider that has signed certain business agreements and contracts to be part of the CoT.

A set of specific elements are being developed and evaluated for the Project based on the X.509 Certificates standard [2] including the architecture of modules needed for managing INDECT certificates and the set of tools able to create a secure transmission. This additionally includes the keys (public and private key pair) required by the end user. There will also be a specific set of smart cards and their readers in order to demonstrate and evaluate the management and access of the certificate information stored within those cards.

Secure Ad hoc Multipath Distance Vector (SAOMDV) protocol

A new routing protocol for Mobile Ad hoc Networks (MANET) called SAOMDV is being designed in order to provide a secure communication infrastructure for security forces or other first responders in an emergency scenario. In particular, SAOMDV is a

secure routing protocol based on AODV [4], the most popular reactive ad hoc routing protocol, which is standardized by the IETF. Unlike other secure ad hoc protocols, which only consider logical attacks [5], SAOMDV also deals with physical attacks (i.e. jamming). For that purpose SAOMDV is able to employ several disjoint paths between each source and destination pairs. Therefore, even if an attacker is able to disrupt one of the paths by means of jamming, the communication will be still possible by using the remaining paths. Essentially, when the source node A wants to communicate with the destination node B, it sends a Multipath Route Request (MRREQ) that floods the network. These MRREQ messages will arrive to B though multiple paths. Then B sends a Route Reply (RREP) message back to the source using the different disjoint paths that have been found. In order to know which paths are fully disjoint it is only necessary to compare the first hop of each path (i.e. A's neighbor), since intermediate nodes only forward the first received MMREQ to avoid loops. Furthermore, intermediate nodes use these MRREQ and RREP messages to learn, respectively, where the source and destination nodes are. This simple multipath routing is inspired in the AOMDV [6] multipath routing protocol, although SAOMDV adds several security mechanisms to avoid logical attacks. First of all, SAOMDV performs access control by means of digital certificates issued to all Police officers by a trusted Certificate Authority (CA). This way, a SAOMDV node will only accept as its neighbor another node that has a valid certificate. However, since in most cases ad hoc networks are not connected to Internet, it is not possible to use a Certificate Revocation List (CRL) from the CA to check the validity of a certificate.

Therefore it is necessary to design a new certificate revocation mechanism for isolated ad hoc networks, such as some kind of distributed node reputation system. However those mechanisms are quite complex and should be designed with great care or otherwise they could be abused by attackers. Instead, thanks to the existing hierarchy of Police forces, any high rank officer can just periodically broadcast a message that revokes the (otherwise valid) certificate of a low rank officer featuring a malicious behavior or whose private key has been compromised. Therefore SAOMDV nodes will (after some verification) relay those revocation messages and build a CLR cache to validate current and future neighbors' certificates against it. Moreover, the SAOMDV access control protocol employs random nonces and timestamps to avoid replay attacks and to be sure that there is a bidirectional connection in order to thwart certain kinds of "Rush" attacks [5]. Also, all SAOMDV messages are signed by the source node, as well as the previous hop, in order to avoid impersonation and injection attacks. For mutable fields in SAOMDV control packets like the hop count, a Hash Chain mechanism is employed to prevent an insider attacker to advertise (invalid) shorter paths. Finally, SAOMDV data packets are encrypted end-to-end with a symmetric session key, securely exchanged during the path setup. However even encrypted information may suffer from statistical analysis attacks. Therefore for extremely sensible and confidential information, SAODVM may choose to only use paths composed by high rank officers. This information is also obtained during the path setup phase by means of a hierarchy of hash-based keys that enables a high rank officer to generate all the lower-rank keys, whereas the reverse action is impossible.

3 Conclusion

This paper presents five areas where the WP8 of INDECT project is developing and testing elements, components and tools for creating a multilevel security platform for European Police forces. In particular IPsec and SSL Virtual Private Networks are employed to secure communication over public networks. A new symmetric encryption algorithm and new substitution boxes (S-box) are being developed and tested by means of a newly developed simulator. This way a user can check different symmetric algorithms and compare their results in order to choose the most appropriate one. The main benefit for INDECT end-users related to the work carried out in the Quantum Cryptography research area is the improvement of security in Quantum Cryptography systems, so the potential eavesdropper will be uncovered much easier and end-users could be able to define and choose an adequate security level of Quantum Key Distribution. Moreover this paper offers an overview about the feasibility of deploying a Federated Identity platform within INDECT, and it identifies tools and modules required for integrating authentication mechanisms based on Smart Cards (i.e. identity cards) containing X.509 Certificates. The proposed Secure Ad hoc On-demand Multipath Distance Vector (SAOMDV) routing protocol will provide a robust and secure communication network for early emergency responders. The ability to find and employ several paths in parallel is a novel feature that increases both the robustness of the network in case of terminal mobility, as well as its security, because it avoids low-level attacks that are overlooked by most ad hoc security researchers.

Acknowledgment. This work has been funded by the EU Project INDECT (*Intelligent information system supporting observation, searching and detection for security of citizens in urban environment*) — grant agreement number: 218086.

References

1. INDECT project web site, `http://www.indect-project.eu`
2. INDECT Consortium. D8.2: Evaluation of Components (June 2010), `http://www.indect-project.eu/files/deliverables/public/deliverable-8.2`
3. Bennett, C.H., Brassard, G.: Public key distribution and coin tossing. In: Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing, pp. 175–179 (1984)
4. Perkins, C., Belding-Royer, E., Das, S.: Ad hoc On-Demand Distance Vector (AODV) Routing. RFC 3561 (July 2003)
5. Djenouri, D., Khelladi, L., Badache, A.N.: A Survey of Security Issues in Mobile Ad Hoc and Sensor Networks. IEEE Communications Surveys & Tutorials 7(4), 2–28 (2005)
6. Marina, M.K., Das, S.R.: 2002, On-demand Multipath Distance Vector Routing in Ad hoc Networks. In: Proceedings of IEEE International Conference on Network Protocols (ICNP), pp. 14–23 (2001)

Interactive VoiceXML Module into SIP-Based Warning Distribution System

Karel Tomala, Jan Rozhon, Filip Rezac, Jiri Vychodil,
Miroslav Voznak, and Jaroslav Zdralek

VSB - Technical University of Ostrava, 17. listopadu 15,
70800 Ostrava, Czech Republic
{karel.tomala,jan.rozhon,filip.rezac,jiri.vychodil,
miroslav.voznak,jaroslav.zdralek}@vsb.cz

Abstract. This article discusses the use of the Voice Extensible Markup Language (VoiceXML, VXML) to create a complex voice menu in danger alert communication system. The system was created as a part of research at Department of Telecommunications at the VSB – Technical University of Ostrava. Creating a voice menu provides end-users more information about the impending danger as well as instructions on how to behave in a given situation. If users receive a pre-recorded warning message in the form of a phone call, it will provide a telephone number on which they can obtain more information. In order to achieve the desired functionality, we had to use open-source PBX Asterisk, the VoiceGlue package which features both the VoiceXML interpreter and the Text-to-Speech (TTS) module.

Keywords: Voice Extensible Markup Language, VoiceGlue, Text-To-Speech, Hypertext Preprocessor.

1 Introduction

Today, more than ever before, information plays an important role, in particular information received at the right moment. In order for a person to be able to make the right choice, s/he should have access to information when it is most needed. People are increasingly aware of this, as today information can be accessed not only through a computer or laptop but also through various PDAs, mobile phones and other communications devices. However, a situation may arise in which the only link to the outside world and the only way to access information will be the voice communication via a mobile phone. Speech is a fundamental and most important means to transfer information between human beings, so naturally there is an effort to use it in modern automated systems that surround us in everyday life.

Our interactive voice menu uses the VoiceXML markup language combined with some open-source tools. It was created as an extension on our own system to alert the public to danger by means of sending voice messages over the communication system directly to their terminal device (mobile phone, fixed line, etc.) [1]. Besides the warning, the delivered voice message also includes a phone number to dial to reach

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 338–344, 2011.

the voice menu which is located directly in the system. This is where the end-user obtains more information about the emergency as well as more instructions on how to address the problem occurred.

2 Used Technology

The comprehensive voice menu was designed using open-source software components Linux, VoIP Asterisk PBX [2], VoiceGlue (VXML language interpreter OpenVXI) [3] and Cepstral Swift text-to-speech module.

VoiceXML [4] enables to easily create complex voice menus that are considered an equivalent of web pages in the area of voice communications. VXML, as the name suggests, is a subset of Extensible Markup Language (XML) [5] and uses a precisely specified set of tags. Through these tags, the type and nature of voice conversation between the end-user and the voice automat can be unambiguously determined. The source code in the VXML format is processed by the VXML interpreter language to a user-friendly voice conversation format. Subsequently, the interpreter has to options to generate the voice message for transmission: to play pre-recorded messages in a given sequence or to feed the text input into the module processing text to speech (Text-To-Speech). The Text-to-Speech module is one of many VXML complements. Another example is a module that enables processing speech to text (Speech-To-Text). The VXML interpreter is perfectly capable of interpreting the VXML language. However, where necessary, advanced features such as Text-To-Speech module should be implemented, in particular if it is not clear in advance whether the system will need to interpret sound.

2.1 VoiceXML Document Structure

A VXML document consists of blocks linked by pairs of tags. In addition, it also contains one stand-alone VXML document in which all other blocks are nested. Figure 1 provides an overview of the fixed VXML document structure. Below the element level, any of the above listed types of blocks, such as a ‘prompt’ block, can be added.

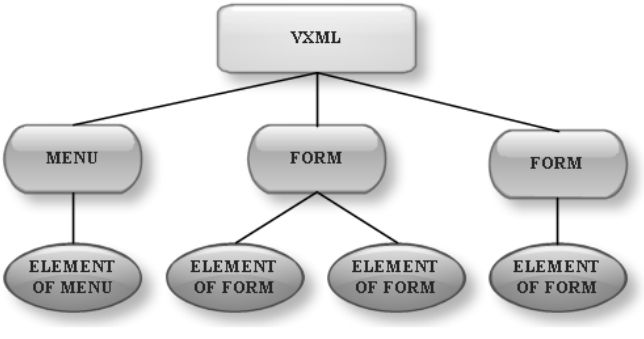


Fig. 1. Fixed structure of VXML document

3 Implementing VXML into Asterisk PBX

To be able to launch the interactive voice menu, PBX Asterisk needs to be configured. At the end-user's request, information about the emergency is read in the voice menu. In other words, a Hypertext Preprocessor (PHP) script that enables translating information entered into the web form into the VoiceXML and play it as a voice message had to be created.

The VoiceGlue package was applied to achieve the required functionality. It features both the VXML language interpreter and the TTS module. Voice output from the integrated TTS module was not sufficiently intelligible, and a separate TTS module had to be installed. The platform consists of the following open-source software components:

- OS Ubuntu x64 10.10,
- Asterisk PBX 1.6.2.15,
- VoiceGlue 0.12,
- Cepstral Swift.

The VoiceXML interpreter OpenVXI, Text-To-Speech module Cepstral Swift and the developed PHP script [6] to process information from a Web server to VXML were implemented directly into the Softswitch Asterisk [7]. This script is described in Chapter 4. The platform mentioned, including its communication with the danger alert system, is sketched in Figure 2.

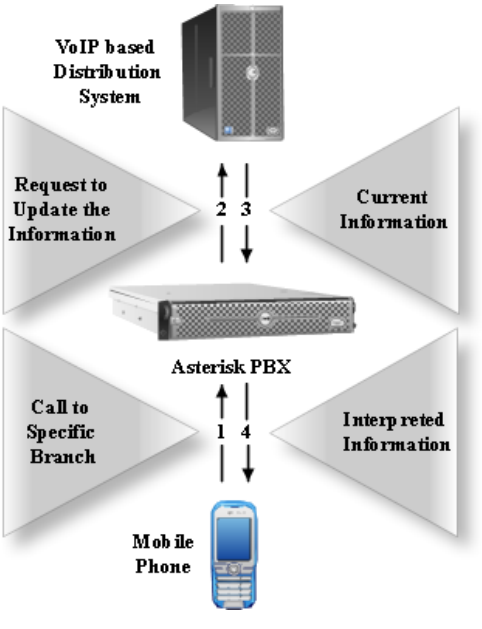


Fig. 2. Voice menu platform with the VoiceGlue system

As VoiceGlue requires a flawless and stable version of Asterisk to be able to function properly, the operating system's packages of Asterisk should be avoided and most recent version of the source code should be compiled and used. Before installing VoiceGlue, all dependencies for this application had to be installed first [3]. Once all dependencies have been installed, the VoiceGlue source code can be downloaded and the installation script launched. Subsequently, *dynlog, voiceglue* and *phoneglue* applications were installed. The first is a dynamic draw lots system, an analogy to Syslog. The second application serves as an interface between the system and VoiceGlue. *Phoneglue* then serves as an interface between the system, VoiceGlue and Asterisk.

3.1 Configuring Asterisk to Work with VoiceGlue

First, it was necessary to create a user 'phoneglue' to manage Asterisk's interface. This can be done in the */etc/asterisk/manager.conf* file by adding the following lines:

```
read = system,call,log,verbose,command,agent,user,
       originate
write = system,call,log,verbose,command,agent,user,
        originate
```

Subsequently, it was necessary to create a text string that will contact VoiceGlue via the Asterisk Gateway Interface (AGI). The basic functionality was achieved using the following dialplan in the Asterisk Extension Language (AEL). The dialplan below should be inserted into the */etc/asterisk/extensions.ael* file. Subsequently, Asterisk has to be restarted.

```
contextphoneglue {
    1 => {
Answer();
Agi(agi://158.196.244.148);
Hangup();
    };
};
```

3.2 Configuring VoiceGlue

Configuring VoiceGlue requires two steps. The first step is mandatory. A dialplan similar to that of Asterisk is implemented into the */etc/voiceglue.conf* file. This line should define which file will be interpreted at the VoiceGlue's launch. The following line has to be entered:

```
* http://158.196.244.148/katastrofix.vxml
```

This line tells the system to execute the *katastrofix.vxml* file whenever a call is placed on any extension. This file can be downloaded from the local web server through Hypertext Transfer Protocol (HTTP). Another possible configuration is to modify the Text-to-Speech module. This can be done in the */usr/bin/voiceglue_tts_gen* file. To use the Cepstral Swift voice generator, the following configuration was used:

```
#!/usr/bin/perl --      -*-CPerl-*-
$file = $::ARGV[2];
system ("/usr/local/bin/swift",  "-m", "text", "-o",
$file, $ARGV[1]);
```

When all changes have been made, you need to restart all three applications related to VoiceGlue. Applications should be shut in the following order: *VoiceGlue*, *Phoneglue* and *Dynlog*. Subsequently, the applications need to be launched again, this time in reverse order. To ensure that the system functions within a pre-defined system specification, it is also necessary to add a PHP script that retrieves the latest information from the communication system's web form and stores as a VXML document.

4 PHP Script for Conversion into VXML

This section describes the script which was used to build the interactive voice menu and which ensures the desired system functionality. The script is written in PHP. The script stores the information collected from the web form into the VoiceXML document format. The first part of the script is used to store the header of the output VXML file in the variable *vxmlstart* for the voice menu [4]:

```
<?php
$vxmlstart=
'<?xml version="1.0" encoding="UTF-8"\?\>
<vxml version = "2.0" xmlns="http://www.w3.org/2001/
vxml">
  <menu id="top" dtmf="true">
  <property name="inputmodes" value="dtmf"/>
  <prompt>
    <enumerate>
      For <value expr="_prompt"/>, press <value expr="
      _dtmf"/>
    </enumerate>
  </prompt>
';
```

The second part of the script is used to convert the information received through the web form in the VXML format. This is the information that you want the voice automat to play. Individual items for *menu* and *form* variables are created here:

```
echo "<html> <body>";
mysql_connect('localhost','voicexml', 'voice.*');
mysql_select_db('voicexml');
if($_POST)
{
  $choices=$_POST;
  for($i=1;$i<10;$i++)
  {
    if($choices[choice][$i])
```

```
    {
      $menu.='
        <choice next="#'.$i.'">
            '.$choices[choice][$i].'
        </choice>';
      $form.= '
        <form id="'.$i.'">
          <block>
            <prompt>
              '.$choices[text][$i].'
            </prompt>
            <goto next="#top" />
          </block>
        </form>
      ';
      $sql="UPDATE voicexml set choice='".$choices
      [choice][$i]."', text='".$choices[text][$i]."'
      where id=$i;";
      mysql_query($sql);
    }
  }
  $file=fopen("katastrofix.vxml", "w+");
  fwrite($file, $vxmlstart);
  fwrite($file, $menu."</menu>");
  fwrite($file, $form."</vxml>");
  fclose($file);
}
else
{
  $choices=mysql_query('select * from voicexml');
}
```

5 Conclusion

We developed an application in the VXML language that uses an interactive voice menu to obtain detailed information on an emergency. Using the VXML language, the Asterisk dialplan was outsourced into an external application – we opted for VoiceGlue. The system enables creating a voice menu without the need to record all voice messages in advance. This is ensured by incorporating the TTS module. The PHP script then ensures that the content collated through a web form is converted into VXML rather than HTML format.

Acknowledgments. This post leading to these results has received funding from G1 1485/2011 titled Modernization of laboratory exercises in the field of New Generation Information Systems in Switching course and from the Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 218086.

References

1. Voznak, M., Rezac, F., Zdralek, J.: Danger Alert Communication System. In: IWSSIP 2010 - 17th International Conference on Systems, Signals and Image Processing, Rio de Janeiro, Brazil, June 17-19 (2010) ISBN 978-85-228-0565-5
2. Meggelen, J.P., Smith, J., Madsen, L.: Asterisk: The Future of Telephony. O'Reilly Media, Sebastopol (2005)
3. VoiceGlue, http://www.voiceglue.org
4. McGlashan, S., Burnet, D., Carter, J., Danielsen, P., Ferrans, J., Hunt, A., Lucas, B., Porter, B., Rehor, K., Tryphonas, S.: Voice Extensible Markup Language (VoiceXML) Version 2.0, W3C - Recommendation (2004)
5. Bray, T., Paoli, J., Sperberg-McQeen, C.M., Maler, E., Yergeau, F.: Extensible Markup Language (XML) Fifth Edition, W3C - Technical Report (2008)
6. Lerdorf, R., Tatroe, K., MacIntyre, P.: Programming PHP, 2nd edn. O'Reilly Media, Sebastopol (2006)
7. Kwon, H.-J., Hong, K.-S.: A Design of User-Initiative Voice Web Using RSS and VoiceXML. In: 5th ACIS International Conference on Software Engineering Research, Management & Applications, SERA 2007, August 20-22, pp. 281–288 (2007)

A Solution for IPTV Services Using P2P

Le Bich Thuy and Yoshiyori Urano

Global Information and Telecommunication Studies, Waseda University
Tokyo, Japan
lebichthuy@asagi.waseda.jp

Abstract. Next Generation Network (NGN) based on IP Multimedia Subsystem (IMS) platform is an open architecture which provides multimedia services. IPTV is also considered as one of the most promising services for service providers. Unfortunately, from the service provider perspective, IPTV, given its complicated architecture and overwhelming bandwidth requirements, is not yet popular. In Internet environment, using Peer to Peer (P2P) overlay to deliver IPTV services has gained lots of attention. However, in order to use P2P technology to provide IPTV, there are many issues such as the effects of peer churn which join and leave the overlay unpredictable and QoS problems to be dealt with. Thus to tackle those matters, We propose an architecture for IPTV service using P2P overlay between NGN/IMS clients. In this architecture, We design a hybrid tree/mesh topology of the application layer multicast (AML) as a content distribution method. In this hybrid topology, stable peers play an important role as the parent peer while other peers will be the child peers. The main characteristics of the proposed approach are: (1) benefits from QoS mechanism of NGN, and (2) minimizing the effects of peer churn problems in P2P network.

Keywords: IMS, IPTV, P2P, NGN, AML, stable peer.

1 Introduction

IPTV service architecture is a very promising technology with many multimedia services such as Live streaming services; video on demand, etc has been a very hot topic for years. Nevertheless the IPTV architecture is very complicated and requires a great investment for the whole system.

On Internet environment, Peer to peer network has become more and more popular in multimedia streaming given its various advantages such as scalability, robustness, and cost effectiveness. But still, there remains QoS problem in in P2P media streaming service. Up to date, there have been many studies about how to improve the quality of IPTV service, but the best-effort in the Internet is the major barrier towards the design and deployment of P2P video system [2]. Another problem in P2P, where peers rely on other peers, is peer churn. Peers can join and leave at any time will cause problems in time delay, reducing the quality of service (QoS) and increasing overheard in network.

For the telecommunication companies, the Next Generation Network (NGN) based on IMS is an ideal architecture for multimedia services because NGN/IMS network

A. Dziech and A. Czyżewski (Eds.): MCSS 2011, CCIS 149, pp. 345–351, 2011.

has function elements support end to end QoS mechanism which is very important to provide multimedia services.

This paper presents an IPTV solution using P2P network overlay on top of NGN network. This solution can provide IPTV services with guaranteed QoS and also can take advantages of P2P network. In this solution, We implement hybrid topology (mesh and tree) for P2P overlay network because this topology has a small start up time (tree topology) and reduces packet loss or delay caused by leaving peers (mesh topology). In order to reduce the effects of peer churn problem, We introduce the stable peers' selection and arrangement into hybrid topology.

Section 2 briefly discusses about the IPTV architecture based on NGN/IMS solution which was proposed by TISPAN group. In the section 3 focuses on the idea of using P2P network to delivery IPTV service is presented. Section 4 presents in details about the current studies of combining the P2P technology and IPTV service over NGN/IMS network. Section 5 describes the P2P IPTV architecture and the basic operation mechanisms of the proposed architecture. Finally, section 6 will provide a conclusion for the paper as well as the future work of the research.

2 IPTV Architecture Over NGN/IMS Network

IP multimedia subsystem (IMS) which was first developed by the Third Generation Partnership Project (3GPP) has been considered as key solution for telecom operators to increase revenue from multimedia services while the traditional services like voice services dramatically reduce. However, the IMS without having proper and sufficient service standardization gradually becomes less attractive. The researchers in TISPAN organization want to merge IPTV services as an application in NGN architecture. They propose two architectures to provide IPTV service over NGN network.

NGN-based IPTV architecture: In this architecture, IPTV architecture is a subsystem in NGN network. It enables interaction and interworking over specified reference point between IPTV application and some existing common NGN components. These components include transport control elements for Resource administration and Control subsystem (RACS) or the TISPAN Network Attachment subsystem (NASS). This approach uses delicate IPTV subsystem within NGN to provide all necessary IPTV required functionalities [7].

IMS-based IPTV architecture: It replaces the IPTV control by IMS based subsystems and employs these functions to make services initiation and control based on SIP protocol. The advantage of this architecture is that IMS can act as an unified service control subsystem for all NGN services instead of establishing an additional specialized subsystem (case of NGN based IPTV architecture) [8].

3 IPTV and P2P Overlay Network

IPTV services have merged as the new trend for the service providers. The traditional architecture is use client-server model, where a single server served multi-clients. The client-server approach has advantages of small delay and easy management but the server can be burden due to a large number of client requests.

The idea of using P2P where the users having the dual roles of both clients and servers can remove the pressure from media servers as well as enhance the scalability of the system. Therefore, P2P technology has been used in multimedia streaming services to enhance the scalability and performance of the system.

Basically, P2P multimedia streaming technology can be classified into two categories namely tree-based and mesh-based [11]. Tree-based topology has good real-time transmission and is easily constructed, but it has some shortcomings. Once an intermediate peer leaves, downstream peers may not receive the streaming data. This situation does not occur with mesh-based architectures; however, the transmission latency of mesh-based architectures is higher than that of tree-based architectures. Mesh-based peers must exchange a considerable number of messages, including associated packets and the locations of peers that hold the next packets. Therefore, transmission time is longer and much bandwidth is wasted during message exchange.

The proposed architecture combines tree topology and ring topology [4] can solve the problems of both topologies. In [4], the novel Hierarchical Ring Tree (HRT) architectures where peers is contracted flow the tree topology, but all the children of node form a ring called DDR (Distributed/Decentralized Dual Role), each node contains the information of the two neighbor nodes (the right sibling and the left sibling). In each ring, a selected peer will contact to its parent's right sibling node to set up another backup link. When a parent peer departs, the backup link is immediately switched to regain data and the DDR ring will be enabled to carry data stream to other peers belongs to the ring. Then the recovery procedure is started. The departed peer will be replaced by the latest joining peer, the corresponding DDR ring and backup link that are established to preserve the stability of the architecture. This solution significantly reduces the time required by a peer to regain data and the related overhead. The HRT topology can be constructed and maintained efficiently without tree splitting or merging.

4 IPTV Over NGN/IMS Architecture Using P2P Delivery

The idea of using the P2P overlay to provide IPTV service for IMS client has been proposed for recently. Those proposals do not modify the IMS architecture and reuses SIP session control, SDP media negotiation and QoS guarantee mechanism of NGN-IMS network. A number of proposals introduce some new functions for IPTV architecture based on IMS to support providing P2P - IPTV services [4], [6] over NGN/IMS.

Nozzilla architecture is the solution of providing IPTV service by maintaining a P2P overlay between Residential Gateways [2] in NGN based IMS. The AML overlay uses either mesh or tree topologies to distribute the media stream from the Media servers to all connected RGWs. The RGW becomes source of media stream for UEs.

The P2P IPTV service-based IMS [1] proposes service architecture for IPTV service in TISPAN NGN using Application Layer Multicast to provide live Television services. They choose tree topology where Media server for each channel is the root of the tree and peers are tree nodes. The most significance in [1] is that they introduce the foster peers-the group of peers with pre-configured, inactive SIP sessions to

reduce the signaling steps and the reserved network resources. These foster peers reduce the latency when peer fails or changes the channel unexpectedly, the orphan peers can be connected to the foster peers without having to wait for the signaling procedure.

Nevertheless, there remain many issues related to the design of P2P multimedia streaming systems such as peer churn, packet loss, etc.

5 Proposed Architecture

My proposal is a solution to provide IPTV for fixed network so We assume that the uplink bandwidth of all peers is the same and each peer can be a parent of a certain number of peers-Cmax. Inheriting the idea from [4], I implement the Hybrid topology (tree and mesh) P2P overlay to distribute IPTV over NGN infrastructure. In addition, We will also present the idea of implementing the stable peer into IPTV AML model to stabilize the network architecture.

Stable peers are peers whose watching time is longer than the normal peers [1], [5]. There are only 30% of peers have the watching time more than one min [10], so in this paper, We simply define that stable peer are those having a watching time longer than 1min.

5.1 The Architecture Overview

The media server plays a role as the root in the tree topology. Stable peers will serve as parent peers or intermediate peers in the tree. The newly joined peers or the peers that have watching time less than 1 min will be considered the child peers.

Each peer has a data link connection with its parent peer to receive data stream. On the other hand, all the children of a peer form mesh overlay as a backup link. There is a candidate to replace the parent position in tree topology incase the parent peers depart. The candidate peer establishes a back up link connection with its grandfather peer. When a parent peer left (or crashed), the back up between candidate peer and grandfather are active to regain data; at the same time the backup link of mesh overlay switches to data link to ensure that all the children of the departed peer will receive data.

5.2 Registration and Join Procedure

Registration to IMS core is the first step for an IMS client to use the IPTV services. The signaling flows and the stream transmission network inherit the session control based on SIP [11]. A peer needs to register to IMS system to be authentication and get QoS guarantee with reserved resource, UE sends an INVITE message to the Core IMS. Once IMS core received the request; it will send a QoS message to the Resource Access Control Function (RACF) in the transport layer before forwarding the request to the IPTV AS. The IPTV AS will check whether the user has the rights to access the service. If authentication is successful then the IPTV AS will start looking for a peer to stream media to the new peer.

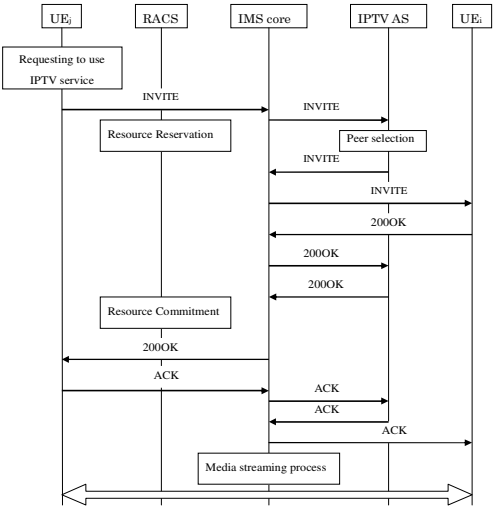


Fig. 1. Peers join procedure

The IPTV AS will search for a peer which is a stable and having the number of child peers smaller than Cmax. In case there are several qualified peers are, priority will be given to the one located in the same network domain or near to the Media server as in order.

Once the available peer is found, the IPTV AS will send the address information (IP address, port number to transfer media stream...) to the IMS Core, IMS core will be responsible for sending an INVITE message to the selected peer to ask him to be the parent of the new peer. Since the selected peer send an OK message, the two peers can start communicating and a media stream will be active with the reserved bandwidth.

In case the ITPV AS cannot find any peer that full satisfies all the requirements or there is no peer watching that content, IPTV AS will connect the new peer to the Media server.

In the worst situation when AS cannot find any peer to be the parent of the new peer and the IPTV Media server exceeds bandwidth, the stable peer definition is extended to the longest watching time in the same domain. Such situations always happen in specific events such as the Olympic, sport events and Oscar, etc., when peer has tendency to be stable. By choosing the longest watching time peer in that domain, the new peer will be connected to the peer nearest to the media server and the media part will be transported within domain.

5.3 Peer Reconstruction Procedure

When a peer departs, its children need to be connected to another peer to receive media stream. The back up between the candidate peer and grandfather becomes active to regain data; at the same time the backup link of mesh overlay switches to data link to ensure that all the children of the departed peer will receive data.

The mesh topology of the departed peer layer rebuilds. The selection of candidate peer and the connection between the candidate peer and its grandfather will be set up again if necessary.

In case of failure, the children of the departed peer have to detect the loss of video stream and after waiting time, they will send a NOTIFY messages to the IPTV AS. After receiving the information of collapsed peer, IPTV AS starts *Peer reconstruction procedure*. If the departed peer is a leaf in the tree topology, its parents must detect its missing and send the NOTIFY message to the IPTV AS. In both situations, the mesh overlay has to recognize whether the departure peer is the parent or the child peer.

This mechanism reduces the effect of leaving peers in the topology. Their children immediately receive media stream with minimized delay.

5.4 Recovery Procedure

After *Peer reconstruction procedure*, the mesh topology is temporary transmitting media stream. To reorganize the topology, the *Recovery procedure* is activated.

The candidate peer now becomes a child peer of its grandfather. If this peer has enough capacity, it will become a parent of other peers in the mesh. If not, the selection process for the parent peer, which is similar to the joining procedure, will be active to select new parent peers for other peers in that mesh overlay.

6 Conclusion

This paper presents a solution using AML hybrid tree/mesh topology to provide IPTV service over NGN network. The hybrid topology with backup links can significantly reduce the time delay for peers in their attempts to regain data and rapidly recover when parent peers depart. The stable peers that form tree topology take a responsible to stabilize network topology and reduce the overhead in the network. In addition, using this P2P overlay between NGN/IMS clients, on one hand can exploit the QoS mechanism proving by NGN/IMS network; on the other hand can increase the scalability and reliability for the system.

This issue is still under research. The next step is evaluating the proposal by using simulation. The simulation results will prove how hybrid topology as well as stable peers can reduce the effects of leaving peers (how many time the recovery process need to be active) and how less the time delay to obtain data of child peers.

References

1. Bikfalvi, A., Garcia-Resinoso, J., Vidal, I., Valera, F.: A peer-to-peer IPTV service architecture for the IP Multimedia Subsystem. International Journal of Communication Systems (2010)
2. Bikfalvi, A., Garcia-Resinoso, J., Vidal, I.: Nozzilla: A peer -to-peer IPTV service Distribution service for an IMS-based NGN. In: Proceedings of the 2009 Fifth International Conference on Networking and Services, pp. 450–455. IEEE Computer Society, Valencia (2009)
3. Huang, N.-F., Chu, Y.-M., Chen, Y.-R.: Design of a P2P Live Multimedia Streaming System with Hybrid Push and Pull Mechanisms

4. Huang, N.-F., Tzang, Y.-J., Chang, H.-Y., Ho, C.-W.: Enhancing P2P overlay network architecture for live multimedia streaming. Information Sciences 180 (May 2010)
5. Wang, F., Liu, J., Xiong, Y.: Stable Peers: Existence, Importance, and Application in Peer-to-Peer Live Video Streaming. In: Proceedings of the IEEE INFOCOM 2008 (2008)
6. Lin, L., Yu, D., Xiao, S., Yu, B.: IMS-based P2P Streaming Service System. In: International Conference on Computer Application and System Modeling (2010)
7. ETSI TS 182 028 V2.0.0 (2008-01), TISPAN; IPTV Architecture; Dedicated subsystem for IPTV functions (2008)
8. ETSI TS 182 027 V2.0.0 (2008-02), TISPAN; IPTV Architecture; IPTV functions supported by the IMS subsystem (2008)
9. ETSI TR 182 010 V.0.0.9 (2009-11), TISPAN; Peer to peer for content delivery for IPTV services: analysis of mechanisms and NGN impacts (2009)
10. Cha, M., Rodriguez, P., Crowcroft, J., Moon, S., Amatriain, X.: Watching television over an IP network. In: Proceedings of the 8th ACM SIGCOMM conference on Internet Measurement Conference, pp. 71–84. ACM, New York (2008)
11. Camarilla, G., Garcia-Martin, M.A.: The 3G IP Multimedia Subsystem (IMS): Merging the Internet and the Cellular Worlds, 2nd edn. John Wiley & Sons Ltd., Chichester (2006)
12. Banerjee, S., Bhattacharijee, B., Kommareddy, C.: Scalable Application Layer Multicast. In: SIGCOMM 2002, Pennsylvania, USA, August 19-23 (2002)

Author Index

Adrian, Weronika T. 165, 268
Anzel, Wojciech 46
Aved, Alexander J. 120

Bąk, Sławomir 293
Balcerek, Julian 64, 225
Berger-Sabbatel, Gilles 207
Białas, Jarosław 1
Blinkiewicz, Michał 73
Boryło, Piotr 301
Bryk, Damian 28

Cetnarowicz, Damian 242
Chmiel, Wojciech 174
Chmielewska, Agata 198
Ciarkowski, Andrzej 37, 233
Čižmár, Anton 191, 310
Czapko, Michał 268
Czyżewski, Andrzej 55, 277

Dąbrowski, Adam 64, 198
Dalka, Piotr 37
Dębski, Roman 19
Doboš, Ľubomír 310
Drgas, Szymon 242
Duda, Andrzej 207
Dudek, Jakub 216
Dziech, Andrzej 1, 144

Ernst, Sebastian 174, 268

Figaj, Andrzej 73

Głowacki, Bartosz 233
Głowacz, Andrzej 10, 46
Grega, Michał 28, 46
Grzesiak, Paweł 268
Gurappa, Varalakshmi 120
Gusta, Marcin 28
Guzik, Piotr 158

Haselhoff, Anselm 113
Hoehmann, Lars 113
Hua, Kien A. 120, 324

Iqbal Khan, Majid 251
Iqbal, Sohaiba 251

Jachnik, Arkadiusz 73
Janowski, Lucjan 10, 137
Juhár, Jozef 191

Kacperski, Kamil 233
Kadłuczka, Piotr 174
Kisiel-Dorohinicki, Marek 183
Klapaftis, Ioannis P. 100
Konieczka, Adam 64
Korus, Paweł 1
Korzeniewski, Adam 277
Kotus, Józef 55
Krzych, Marcin 268
Krzykowska, Agnieszka 64, 198
Kummert, Anton 113
Kurowski, Krzysztof 293

Lach, Seweryn 46
Le, Bich Thuy 345
Leszczuk, Mikołaj 10, 91
Ligęza, Antoni 165, 268
Łopatka, Kuba 55

Machník, Petr 331
Machowski, Łukasz 216
Maestro, Gema 331
Manandhar, Suresh 100
Marciniak, Tomasz 198, 242
Marcinkowski, Piotr 277
Matiolański, Andrzej 158, 301
Meuter, Mirko 113
Mirek, Ryszard 10
Musiał, Jan 46

Nalepa, Grzegorz J. 268
Napierała, Krystyna 293
Napora, Maciej 28
Nauman, Zahra 251
Niemiec, Marcin 286, 331
Nunn, Christian 113

Olech, Piotr 1
Opaliński, Andrzej 183
Orzechowski, Tomasz M. 144, 301

Pacyna, Piotr 261
Pandey, Suraj 100
Papaj, Ján 310
Papir, Zdzisław 137
Pawłowski, Paweł 225

Rapacz, Norbert 261
Rezac, Filip 338
Romaniak, Piotr 10, 137
Romański, Łukasz 216, 286
Rozhon, Jan 338
Rusek, Krzysztof 144

Śnieżyński, Bartłomiej 19
Sowa, Grzegorz 261
Stankiewicz, Mateusz 64
Stoianov, Nikolai 317, 331
Świerczek, Arkadiusz 19
Święty, Marcin 216, 286
Szczodrak, Maciej 233
Szczuko, Piotr 82
Szpyrka, Marcin 268
Szwabe, Andrzej 73
Szwoch, Grzegorz 37, 149

Tahir, Muhammad 251
Tomala, Karel 338
Turek, Wojciech 183

Urano, Yoshiyori 345
Urueña, Manuel 331

Vozáriková, Eva 191
Voznak, Miroslav 338
Vychodil, Jiri 338

Waliszko, Jarosław 165
Wassermann, Jakob 129
Włodek, Piotr 19

Ye, Jun 324

Zdralek, Jaroslav 338