

Guest Editorial: Intelligent Processing for Citizen Security

Andrzej Dziech¹ · Remigiusz Baran² · Mikołaj Leszczuk¹

Published online: 5 June 2015

© The Author(s) 2015. This article is published with open access at Springerlink.com

1 Introduction

The contemporary world presents us with numerous dangers and challenges. We hear about traffic accidents, crimes, confrontations between football hooligans, building disasters, etc. on the news almost every day. In addition, terrorism has become a growing threat. In order to prevent and combat such hazards, new intelligent solutions are required. Traditional urban monitoring and surveillance systems, with operators overseeing any activities taking place, are no longer sufficient, mainly because of the petabytes of data per minute that need to be analyzed. Such a vast volume of data in current urban monitoring systems is the result of widespread and growing use of various sensors, mainly digital cameras and microphones. Therefore, new intelligent approaches for image and audio processing and recognition are vital for modern security systems; in turn, they would not be effective without innovative data encryption and protection techniques, including those relying on digital watermarking.

The aims of this special issue are three-fold: (1) introducing innovative research in intelligent processing techniques for security systems; (2) presenting new ways of applying them; and (3) discussing different aspects of security architectures for law enforcement agencies.

2 Review process

The special issue of “Intelligent Processing for Citizen Security” collects reports of scientific research conducted on a wide range of topics. In general, issues addressed by the submissions

✉ Andrzej Dziech
dziech@kt.agh.edu.pl

Remigiusz Baran
r.baran@tu.kielce.pl

Mikołaj Leszczuk
leszczuk@kt.agh.edu.pl

¹ AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland

² Kielce University of Technology, al. 1000-lecia P.P. 7, 25-314 Kielce, Poland

concern various fields of digital signal processing, ranging from digital watermarking techniques to face detection in thermal images. There are certain characteristics which are common to all the works reported; as stressed in Section 1 of this editorial, all the approaches aim to improve citizen security.

Taking this into account, the Guest Editors represent expertise which covers digital signal processing and its security applications. Prof. Andrzej Dziech is an expert on digital communication, data analysis and compression, information and coding theory, and watermarking technology. He has also coordinated several security-oriented projects, such as the major European Union FP7 integrated project INDECT. Remigiusz Baran is a specialist in image analysis, feature extraction and pattern recognition. His expertise is also closely related to different security aspects, as he has been a researcher on a number of relevant projects including INSIGMA. Mikołaj Leszczuk's expertise is mainly in quality assessment (especially for recognition tasks), video processing (particularly video summarization, indexing and compression) and in image retrieval. As the other editors, he has also been involved in security-oriented projects, such as acting as a steering committee member in the INDECT project.

Each submission was reviewed by at least three experts, both during the first and the second round. Ultimately, a total of 15 papers were accepted for this special issue.

3 Guide to included papers

This special issue includes 15 papers. They fall into the following five main categories:

- i. speech and other acoustic signal analysis,
- ii. object (including face) detection, recognition and classification,
- iii. image and video indexing and retrieval,
- iv. digital watermarking,
- v. integrated security architectures.

3.1 Speech and other acoustic signal analysis

This category includes three papers. In “Speaker recognition based on multilevel speech signal analysis on Polish corpus” ([10.1007/s11042-013-1502-0](https://doi.org/10.1007/s11042-013-1502-0)), the authors present an approach in which spectral and high-level features (prosodic, articulatory, and lexical) are combined to text-independent speaker verification. Instead of using support vector machines (SVM), as in other successful speaker discrimination approaches, the authors propose a cosine similarity system with scoring methods based on a modified z-norm (zero normalization) technique. The features are combined by the AdaBoost algorithm. Since the aim of this paper was to show advantages of high-level features in Polish language, the authors created a corpus consisting of semi-spontaneous telephone conversations to be used as the test set.

The next item in this category, titled “Feature selection for acoustic event detection” ([10.1007/s11042-013-1529-2](https://doi.org/10.1007/s11042-013-1529-2)), discusses an effective framework for the recognition of acoustic events such as breaking glass and gunshots. When such

sounds are detected in a public space, they are generally treated as being anomalous and indicating a potentially dangerous situation (e.g., a robbery). As such, their automatic recognition is highly desirable as a part of modern, intelligent security systems. The paper presents superior feature sets, representative of these sounds and selected following their processing using different feature extraction algorithms, as well as the entire framework for effective recognition of acoustic events. As the framework is based on minimum redundancy maximum relevance and joint mutual information algorithms, and it uses hidden Markov model-based classification, these methods are also discussed.

The third paper, titled “Multiple sound source localization in a free field using an acoustic vector sensor” ([10.1007/s11042-013-1549-y](#)), covers a somewhat different area in the field. Instead of recognition, the paper focuses on localization of sound sources using an array of dedicated sensors. The localization is achieved by finding out the direction of arrival (DOA) of the sound sources under consideration. The paper discusses the methods used to achieve this, regardless of the frequency of sounds being analyzed, as well as presenting a test set of synthetic and real acoustic signals. Results obtained using synthetic and real acoustic signals are discussed with reference to the localization accuracy of the methods presented and distribution of spectral energy.

3.2 Object (including face) detection, recognition and classification

The second category includes a total of six papers; it is divided into two sub-categories related to objects (e.g., knives) and faces, respectively. The first object-related paper is titled “Visual detection of knives in security applications using active appearance models” ([10.1007/s11042-013-1537-2](#)). The detection scheme presented in the paper is especially well suited to applications such as luggage-scanning systems. Although it is based on the well-known technique of active appearance models (AAMs), the paper is undeniably innovative, since the approach is the first successful application of AAMs in general class object detection. In addition, according to the authors’ best knowledge, it is the first example of research focusing on knife detection.

The next object-related paper is titled “Efficient real- and non-real-time make and model recognition of cars” ([10.1007/s11042-013-1545-2](#)). It presents two different MMR approaches; the first is capable of recognizing cars within real-time constraints with a satisfactory classification accuracy (approx. 92 %), while the other, known as a visual content classification (VCC) approach, gives a high accuracy (approx. 97 %) while maintaining reasonable, non-real parameters of processing time. The real-time approach combines speeded-up robust features (SURF) and support vector machines, while the VCC approach combines selected MPEG-7 visual content descriptors and local features, including the scale invariant feature transform (SIFT) and SURF.

The third paper in this sub-category, “A method for counting people attending large public events” ([10.1007/s11042-013-1628-0](#)), concerns the problem of analyzing video sequences and detecting objects defined as “having dimensions similar to the size of an average human body”. The aim of the system proposed in the paper is to count people in a crowd and estimate the number of incomers as they are entering a public space (such as a large sports arena) through the gates. The solution, known as virtual gate, is based on the modified dense optical flow method and consists of two parts: the main module

which performs image processing, and the calibration module where the optimal counting threshold, quantifying the human silhouette at a given camera view, is determined.

The title of the first face recognition-related paper is “Face detection and facial expression recognition using simultaneous clustering and feature selection via an expectation propagation statistical learning framework” (10.1007/s11042-013-1548-z). The statistical learning framework is proposed in the paper as being dedicated to face detection and facial expression recognition. It is based on a nonparametric Bayesian analysis technique known as the Dirichlet process, which has been used to extend the finite generalized Dirichlet (GD) mixture model (successfully utilized in previously reported relevant works) into an infinite case. Nonparametric Bayesian analysis makes it possible to avoid the problems of data size over- and under-fitting. Therefore, different feature subsets, chosen using the localized feature selection scheme, can be combined with different mixture components.

“Influence of low resolution of images on the reliability of face detection and recognition” (10.1007/s11042-013-1568-8) addresses the problem of face detection and recognition in low-resolution images (e.g., from video monitoring images) and its efficiency in real-time constraints. The paper presents various face detection techniques and test-sets composed of images selected from different databases in accordance to biometric standards. The results, in particular those related to minimum resolution requirements and their influence on face detection and recognition, show that face recognition can be performed accurately even when the pixel dimensions of the image are small, e.g., 21 by 21 pixels. Moreover, the approach presented in the paper ensures high recognition accuracy when the training dataset is small.

The next face recognition-related paper, titled “Automatic method for the detection of characteristic areas in thermal face images” (10.1007/s11042-013-1745-9), presents a novel algorithm for thermogram analysis, enabling the automatic localization of characteristic areas of the face. The approach can be applied at locations such as airports in order to detect people suffering from fever as a potential symptom of an infectious disease. The main advantages of the approach are its robustness to variation in the subject’s appearance, changes in head position and orientation, and background clutter (such as caused by hair and the hairline). Each variable has been confirmed by experiments carried out on a large test set which includes images registered using several thermal cameras, each with a different sensitivity. In addition, the test images have been collected in real conditions, maintaining the principles related to taking thermal images for the purposes of medical thermography. They have been registered at the Department of Pediatrics and Child and Adolescent Neurology in Katowice, Poland, courtesy of its authorities and patients. The approach assures a high localization accuracy for the centers of the eye sockets (approx. 87 %) and nostrils (approx. 93 %).

3.3 Image and video indexing and retrieval

The next category – Image and video indexing and retrieval – includes two papers. “Urban photograph localization using the INSTREET application – accuracy and performance analysis” (10.1007/s11042-013-1538-1) proposes a method of geolocation of city landmarks in photographs. The geolocation is achieved using techniques from the field of content-based image retrieval. The content of the query image taken at a given location in a city is compared against a geolocated reference database of street view images. The visual similarity between

the query image and the reference image is evaluated according to selected MPEG-7 descriptors, previously computed for the content under analysis.

The second paper, titled “Classification of video sequences into chosen generalized use classes of target size and lighting level” (10.1007/s11042-013-1546-1), refers to recommendations of the Video Quality in Public Safety Working Group (VQIPS). VQIPS is an initiative formed by the US Department of Homeland Security to improve the way in which video technologies serve the public safety community. The algorithms presented in this paper are capable of automatically classifying input video sequences into generalized use classes (GUCs). The definitions of GUCs refer to the scene content and the use parameters that have an impact on the recognition task. The classification accuracy of the framework is high; for example, in the case of lighting level (scene content), it is approx. 93 %.

3.4 Digital watermarking

The next category of papers refers to new digital watermarking algorithms as different aspects of their application. In “Dual watermarking algorithm based on the fractional Fourier transform” (10.1007/s11042-013-1531-8), a novel approach which embeds two different watermarks into each image is presented. The first is the robust watermark, embedded to protect the copyright of the image, while the second is the fragile watermark, applied as a means of detecting tampering.

Both parts of the approach are based on the FRFT transformation and the grey relational analysis. As reported, the scheme is robust to various types of shearing, compression and Gaussian noise attacks. In addition, its fragile component shows good tampering detection capabilities.

The next watermarking-related paper, titled “Analysis of the impact of audio modifications on the robustness of watermarks for non-blind architecture” (10.1007/s11042-013-1636-0), presents a new digital watermarking scheme for audio content. The embedding part of the proposed non-blind watermarking system operates in the DWT domain. The scheme is examined in relation to selected possible attacks, including lossy compression and low-pass filtration. In addition, subjective and objective analyses are performed and their results are compared against those given by the audio watermarking tools (AWT) encoder.

The final paper in this section, “Real data performance evaluation of the CAISS watermarking scheme” (10.1007/s11042-013-1544-3), presents an analysis of the performance parameters of the recently proposed digital watermarking scheme known as correlation-and-bit-aware improved spread spectrum. The results of this analysis, performed in the domain of the discrete cosine transform, show that CAISS significantly outperforms traditional spread spectrum techniques in many aspects, including robustness against JPEG compression, additive Gaussian noise and image scaling. In terms of its bit error rate, it is also better than the improved spread spectrum scheme.

3.5 Integrated security architectures

The final category includes a single paper, titled “Integrated security infrastructures for law enforcement agencies” (10.1007/s11042-013-1532-7). An overview of the security architecture designed as one of the results of the INDECT project is provided in the paper. The security infrastructures that have been deployed so far during this project are also presented. The paper introduces and discusses a range of new ideas and innovative approaches for user management, communications security, new cryptographic algorithms, and a public key infrastructure developed as a result of the INDECT project.

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.



Andrzej Dziech holds the position of a full professor at the Department of Telecommunications of the AGH University of Science and Technology in Kraków. He was awarded his M.Sc. and Ph.D. by the Institute of Electrical Engineering in Saint Petersburg in 1970 and 1973, respectively, and the D.Sc. by the Technical University of Poznań in 1978. He is the author and co-author of six books and approx. 230 publications. He has supervised 18 Ph.D. students. His fields of interest are related to digital communication, image and signal processing, data analysis and compression, information and coding theory, random signals, computer communications networks, object recognition, and watermarking technology. He has received four awards for research achievements by the Ministry of Education of Poland. Prof. Dziech participates in numerous international research projects, such as Tempus, Knixmas and Calibrate. In addition, he has been the coordinator of several large projects as Imcop, Insigna and Tapas. He is currently coordinating the EU FP7 integrated project INDECT.



Remigiusz Baran was awarded his the M.Sc. in Electrical Engineering from the Faculty of Electrical and Control Engineering, Kielce University of Technology in 1993, and the Ph.D. in Telecommunications from the Faculty of Electrical, Control, Electronic and Computer Engineering, AGH University of Science and Technology in Kraków in 2004. He is currently working as an Assistant Professor at the Kielce University of Technology. He is the author or co-author of over 40 publications in the field of digital signal processing, focusing on image compression and feature extraction. The main areas of his academic interest are feature- and appearance-based

object detection and recognition techniques, and microprocessor technology and embedded systems. He also serves as a reviewer of international journals and conferences. Dr. Baran has been working on numerous international and national research projects including OASIS Archive, Calibrate, INWAS, INSIGMA, TAPAS and INDECT. At present he is the Project Manager of the international project IMCOP “Intelligent Multimedia System for Web and IPTV Archiving. Digital Analysis and Documentation of Multimedia Content”, which is carried out in with an Israeli partner.



Mikolaj Leszczuk started his professional career in 1996 at COMARCH SA as manager of the Multimedia Technology Department, and then at COMARCH Multimedia as the CEO. Since 1999 has been employed at the Department of Telecommunications at the AGH University of Science and Technology in Kraków. In 2000 he moved to Spain for a 4-month scholarship at the Universidad Carlos III de Madrid. After returning to Poland, he was employed at the Department of Telecommunications as a research and teaching assistant, and in 2006, he successfully defended his doctoral dissertation as an Assistant Professor. His current research interests are focused on multimedia data analysis and processing systems, with particular emphasis on Quality of Experience. He has authored over 100 scientific publications. He has participated more than 20 major research projects, including FP4, FP5, FP6, FP7, Horizon 2020, OPIE, Culture 2000, PHARE, eContent+, and Eureka!. Between 2009 and 2014, he was the administrator of the major international INDECT research project, dealing with solutions for intelligent surveillance and automatic detection of suspicious behavior and violence in urban environments. He is a member of VQEG (Video Quality Experts Group, board member), IEEE (Institute of Electrical and Electronics Engineers), and GAMA (Gateway to Archives of Media Art). The latter organization collaborates with the VQiPS (Video Quality in Public Safety) working group.