

# Guest Editorial: Intelligent Audio-Visual Observation Systems for Urban Environments

Andrzej Dziech<sup>1</sup> · Mikołaj Leszczuk<sup>1</sup> · Remigiusz Baran<sup>2</sup>

Published online: 10 August 2016

© The Author(s) 2016. This article is published with open access at Springerlink.com

## 1 Introduction

The security of citizens constitutes an issue of primary importance. Constant improvements in security should be one of the strategic activities of any government and its agencies such as police, city police, firemen, local administration, etc.

One of the tasks in ensuring security of citizens is to develop both fixed and a man portable, integrated, fast, wide-area intelligent audio-visual, behavioral observation systems for individuals, platforms, and goods in complex urban environments. It should meet surveillance and security tasks including compound security, trafficking of illegal goods, safety monitoring and evacuation on a 24-h/7-day basis. This includes the integration of sensor technologies, data fusion, and intelligent observation systems to enable stand-off detection and analysis through barriers, of substances and weapons, of carriers and people, as well as behavior analysis to separate potential perpetrators from crowds and neutralize the threat.

The aims of this special issue are threefold as follows: (1) introducing innovative research in intelligent audio-visual observation systems, (2) presenting new ways of applying them, and (3) discussing different aspects of security architectures for urban environments.

## 2 Review process

The special issue of *Intelligent Audio-Visual Observation Systems for Urban Environments* collects reports of scientific research conducted on a wide range of topics. In general, issues addressed by the submissions concern various fields of digital signal processing, ranging from audio to visual intelligent observation systems for urban environments. There are certain characteristics which are common to all the works reported; as stressed in [Section 1](#) of this

---

✉ Mikołaj Leszczuk  
leszczuk@agh.edu.pl

<sup>1</sup> AGH University of Science and Technology, Kraków, Poland

<sup>2</sup> Kielce University of Technology, Kielce, Poland

editorial, all the approaches aim to improve intelligent audio-visual observation systems for urban environments.

Each submission was reviewed by at least three experts, both during the first and the second round. Ultimately, a total of 19 papers were accepted for this special issue.

### 3 Guide to included papers

This special issue includes nineteen papers. They fall into the following six main categories:

- i. Intelligent audio observation systems for urban environments
- ii. Intelligent visual observation systems for urban environments
- iii. Spatial data analysis
- iv. Security management
- v. Video analysis including intelligent monitoring
- vi. Audio analysis/processing

The abstracts of these nineteen papers follow.

### 4 Intelligent audio observation systems for urban environments

This category includes two papers. In “Detection, classification and localization of acoustic events in the presence of background noise for acoustic surveillance of hazardous situations” ([10.1007/s11042-015-3105-4](https://doi.org/10.1007/s11042-015-3105-4)), evaluation of sound event detection, classification, and localization of hazardous acoustic events in the presence of background noise of different types and changing intensities is presented. The methods for discerning between the events being in focus and the acoustic background are introduced. The classifier, based on a support vector machine (SVM) algorithm, is described. The set of features and samples used for the training of the classifier are introduced. The sound source localization algorithm based on the analysis of multichannel signals from the acoustic vector sensor (AVS) is presented. The methods are evaluated in an experiment conducted in the anechoic chamber, in which the representative events are played together with noise of differing intensity. The results of detection, classification, and localization accuracy with respect to the signal to noise ratio (SNR) are discussed. The results show that the recognition and localization accuracy is strongly dependent on the acoustic conditions. The authors also find that the engineered algorithms provide sufficient robustness in moderate noise in order to be applied to practical audio-visual surveillance systems.

The second paper, titled “Efficient acoustic detector of gunshots and glass breaking” ([10.1007/s11042-015-2903-z](https://doi.org/10.1007/s11042-015-2903-z)), presents an efficient acoustic events detection system—EAR-TUKE. The system is capable of processing continuous input audio stream in order to detect potentially dangerous acoustic events, specifically gunshots or breaking glass. The system is programmed entirely in C++ language (core mathematical functions in C) and was designed to be self sufficient without requiring additional dependencies. In the design and development process, the main focus was put on easy support of new acoustic events detection, low memory profile, low computational requirements to operate on devices with low resources, and on long-term operation and continuous input stream monitoring without any maintenance. In order to satisfy these requirements on the system, EAR-TUKE is based on a custom approach

to detection and classification of acoustic events. The system is using acoustic models of events based on hidden Markov models (HMMs) and a modified Viterbi decoding process with an additional module to allow continuous monitoring. These features in combination with weighted finite-state transducers (WFSTs) for the search network representation fulfill the easy extensibility requirement. Extraction algorithms for mel-frequency cepstral coefficients (MFCC), frequency bank coefficients (FBANK), and mel-spectral coefficients (MELSPEC) are also included in the pre-processing part. The system contains cepstral mean normalization (CMN) and our proposed removal of basic coefficients from feature vectors to increase robustness. This paper also presents the development process and results evaluating the final design of the system.

## 5 Intelligent visual observation systems for urban environments

The second category includes a total of five papers; the first paper is titled “A smart camera for the surveillance of vehicles in intelligent transportation systems” ([10.1007/s11042-015-3151-y](#)). The paper presents a smart camera aimed at security and law enforcement applications for intelligent transportation systems. An extended background is presented first as a scholar literature review. The smart camera components and their capabilities for automatic detection and recognition of selected parameters of cars, as well as different aspects of the system efficiency, are described and discussed in detail in subsequent sections. Smart features of make and model recognition (MMR), license-plate recognition (LPR), and color recognition (CR) are highlighted as the main benefits of the system. Their implementations, flowcharts, and recognition rates are described, discussed and finally reported in detail. In addition to MMR, three different approaches, referred to as bag-of-features, scalable vocabulary tree, and pyramid match, are also considered. The conclusion includes a discussion of the smart camera system efficiency as a whole, with an insight into potential future improvements.

The next paper is titled “Complexity analysis of the Pawlak’s flow graph extension for re-identification in multi-camera surveillance system” ([10.1007/s11042-015-2652-z](#)). The idea of Pawlak’s flow graph turned out to be a useful and convenient container for knowledge of objects’ behavior and movements within the area observed with a multi-camera surveillance system. Utilization of the flow graph for modeling behavior admittedly requires certain extensions and enhancements, but it allows for combining many rules into a one data structure and for obtaining parameters describing how objects tend to move through the supervised area. The main aim of this article is presentation of the complexity analysis of proposed modification of flow graphs. This analysis contains considerations of issues such as memory efficiency and computational complexity of operations on the flow-graph. The measures related to space and time efficiency were also included.

The third paper in this category, “CCTV objects detection with fuzzy classification and image enhancement” ([10.1007/s11042-015-2697-z](#)), proposes a novel approach for pattern recognition problems with non-uniform classes of images. The main concept of this classification method is to describe classes of images with their fuzzy portraits. This approach is a good generalization of the algorithm. The fuzzy set is calculated as a preliminary result of the algorithm before the final decision or rejection that solves the problem of uncertainty at the boundaries of classes. The authors use the method to solve the problem of knife detection in still images. The main aim of this paper is to test fuzzy classification with feature vectors in a real environment. The authors used selected MPEG-7 descriptor schemes as feature vectors. The method was experimentally validated on a dataset of over 12,000 images. The article describes the results of six experiments

which confirm the accuracy of our method. In addition, the authors conducted a test with enhanced images, achieved with two state-of-the-art super-resolution algorithms.

The next paper, titled “INSIGMA: an intelligent transportation system for urban mobility enhancement” ([10.1007/s11042-016-3367-5](https://doi.org/10.1007/s11042-016-3367-5)), presents intelligent transportation systems (ITS) that aim to improve safety, mobility, and environmental performance of road transport. More particularly, the ITS proposed by the INSIGMA project provides a fresh look at the possible innovations in this field, by enhancing the functionality and accuracy of ITS in urban environments. This paper describes the architecture, sensors, processing algorithms, output modules, and advantages of the developed system. A comparison of existing ITS systems has been provided as background. Special attention has been given to performance and privacy issues, as the system includes social aspects such as location monitoring.

The final paper in this section, “Simple gait parameterization and 3D animation for anonymous visual monitoring based on augmented reality” ([10.1007/s11042-015-2874-0](https://doi.org/10.1007/s11042-015-2874-0)), presents a method for video anonymization and replacing real human silhouettes with virtual 3D figures rendered on a screen. Video stream is processed to detect and to track objects, whereas anonymization stage employs animating avatars accordingly to behavior of detected persons. Location, movement speed, direction, and person height are taken into account during animation and rendering phases. This approach requires a calibrated camera and utilizes results of visual object tracking. A procedure for transforming objects visual features and bounding boxes into gait parameters for animated figures is presented. Conclusions and future work perspectives are provided.

## 6 Spatial data analysis

This category includes five papers. In “Multi-focus image fusion method of Ripplet transform based on cycle spinning” ([10.1007/s11042-014-1942-1](https://doi.org/10.1007/s11042-014-1942-1)) the cycle spinning method is adopted to suppress the pseudo-Gibbs phenomena in the multi-focus image fusion. On the other hand, a modified sum-modified-Laplacian rule based on the threshold is proposed to make the decision map to select the Ripplet coefficient. Several experiments are executed to compare the presented approach with other methods based on the curvelet, sharp frequency localized-contourlet transform, and Shearlet transform. The experiments demonstrate that the presented fusion algorithm outperforms these image fusion works.

The second paper, titled “Mobile context-based framework for threat monitoring in urban environment with social threat monitor” ([10.1007/s11042-014-2060-9](https://doi.org/10.1007/s11042-014-2060-9)), proposed a lightweight context-aware framework for mobile devices that uses data gathered by mobile device sensors and perform online reasoning about possible threats, based on the information provided by the social threat monitor system developed in the INDECT project.

The third paper is titled “Two-stage neural network regression of eye location in face images” ([10.1007/s11042-014-2114-z](https://doi.org/10.1007/s11042-014-2114-z)). In this paper, a new solution to the problem of automatic eye localization is proposed. Eye localization is posed as a nonlinear regression problem solved by two feed-forward multilayer perceptrons (MLP) working in a cascade. The input feature vector of the first network is constructed from coefficients of a two dimensional discrete cosine transform (DCT) of a face image. The second network generates corrections based on small-image patches. Feature extraction and neural network prediction have known and efficient implementations, thus the entire procedure can be very fast. The paper hints at the neural network structure and the procedure for generating artificial training samples from a low number of face images. In terms of accuracy, the method is comparable to state-of-the-art techniques; however, it is based on

numerical procedures that could be highly optimized (fast Fourier transform and matrix multiplication).

The next paper is titled “A new secure and sensitive image encryption scheme based on new substitution with chaotic function” (10.1007/s11042-014-2115-y). In this paper, a new image encryption scheme is proposed with high sensitivity to the plain image. In proposed scheme, two chaotic functions and an XOR logical operator are used. Image encryption process includes substitution of pixels and permutation. Using the new method of substitution, algorithm sensitivity somewhat has elevated to changes in the plain image that by changing a single pixel of the plain image, amount of NPCR reaches 100 %. Results of tests show that the cipher image does not give any information of statistical such as entropy, histogram, and correlation of adjacent pixels to attackers. Also, the proposed scheme has the wide key space and is so safe to the noise ratio and compression.

The final paper in this section, “Piecewise-linear sub-band coding scheme for fast image decomposition” (10.1007/s11042-014-2173-1), introduces a novel two-channel piecewise-linear sub-band coding scheme (PL-SBC). Its most important advantage is exceptionally fast and easy for implementation computational algorithms combined with good reconstruction quality. The proposed PL-SBC system is based on the dual-filter bank consisting of a single filter at the decomposition stage and a single one at the reconstruction stage. For the presented PL-SBC scheme, the conditions of perfect reconstruction and the piecewise-linear approximation filter bank have been defined. The compression properties of a new sub-band coding scheme have been analyzed for different categories of images and in comparison with well-known signal transforms, such as cosine, piecewise-linear, piecewise-constant, and wavelet transforms. The research includes a comparison of the reconstruction error both in objective as well as subjective approach.

## 7 Security management

The second category includes a total of three papers; the first paper is titled “Risk assessment for a video surveillance system based on fuzzy cognitive maps” (10.1007/s11042-014-2047-6). The paper discusses an application of a new risk assessment method, in which risk calculation is based on fuzzy cognitive maps (FCMs) to a complex automated video surveillance system. FCMs are used to capture dependencies between assets and FCM-based reasoning is applied to aggregate risks assigned to lower-level assets (e.g., cameras, hardware, software modules, communications, people) to such high-level assets as services, maintained data, and processes. Lessons learned indicate that the proposed method is an efficient and low-cost approach giving instantaneous feedback and enabling reasoning on effectiveness of security system.

The next paper is titled “Authentication in virtual private networks based on quantum key distribution methods” (10.1007/s11042-014-2299-1). This article introduces a new concept: quantum distribution of pre-shared keys. This approach provides end users with very secure authentication, impossible to achieve using currently available techniques. Secure authentication is a key requirement in virtual private networks (VPN)—popular protection in computer networks. The authors simulated quantum-based distribution of a shared secret in a typical VPN connection. Using a dedicated simulator, all individual steps of the quantum key distribution process were presented. Based on the created secret, a secure IPsec tunnel in a Strong-Swan environment was established between AGH (Poland) and VSB (Czech Republic). It allows end users to communicate at very high security levels.

The final paper in this section, “Security architecture for law enforcement agencies” (10.1007/s11042-014-2386-3), presents an integrated security architecture for law enforcement

agencies (LEAs) that is able to provide common security services to novel and legacy information and communication technology (ICT) applications, while fulfilling the high security requirements of police forces. By reusing the security services provided by this architecture, new systems do not have to implement custom security mechanisms themselves and can be easily integrated into existing police ICT infrastructures. The proposed LEA security architecture features state-of-the-art technologies, such as encrypted communications at network and application levels, or multi-factor authentication based on certificates stored in smart cards.

## 8 Video analysis including intelligent monitoring

The third category includes a total of three papers as well; the first paper is titled “Video analytics-based algorithm for monitoring egress from buildings” ([10.1007/s11042-014-2143-7](https://doi.org/10.1007/s11042-014-2143-7)). The paper presents a concept and a practical implementation of the algorithm for detecting of potentially dangerous situations of crowding in passages. An example of such a situation is a crash which may be caused by obstructed pedestrian pathway. Surveillance video camera signal analysis performed online is employed in order to detect holdups near bottlenecks like doorways or staircases. The details of the implemented algorithm which uses optical flow method combined with fuzzy logic are explained. The experiments were carried out on a set of gathered video recordings from the surveillance camera installed in the campus of Gdansk University of Technology. The results of experiments performed on gathered video recordings show that efficiency of the algorithm is high.

The second paper is titled “Recent developments in visual quality monitoring by key performance indicators” ([10.1007/s11042-014-2229-2](https://doi.org/10.1007/s11042-014-2229-2)). The concept proposed there, monitoring of audio-visual quality by key performance indicators (MOAVI), is able to isolate and focus investigation, set-up algorithms, increase the monitoring period and guarantee better prediction of perceptual quality. MOAVI artifacts key performance indicators (KPI) are classified into four categories, based on their origin: capturing, processing, transmission, and display. In the paper, the authors present experiments carried out over several steps with four experimental set-ups for concept verification. The methodology takes into the account annoyance visibility threshold. The experimental methodology is adapted from the International Telecommunication Union—Telecommunication Standardization Sector (ITU-T) Recommendations: P.800, P.910, and P.930. The authors also present the results of KPI verification tests. Finally, the authors also describe the first implementation of MOAVI KPI in a commercial product: the NET-MOZAIC probe. Net Research, LLC, currently offers the probe as a part of NET-xTVMS internet protocol television (IPTV) and cable television (CATV) monitoring system.

The final paper in this section is “Applications for a people detection and tracking algorithm using a time-of-flight camera” ([10.1007/s11042-014-2260-3](https://doi.org/10.1007/s11042-014-2260-3)). This paper outlines a method and applications for the detection and tracking of people in depth images, acquired with a low-resolution time-of-flight (ToF) camera. This depth sensor is placed perpendicular to the ground in order to provide distance information from a top-view position. With usage of intrinsic and extrinsic camera parameters, a ground plane is estimated and compared to the measured distances of the ToF sensor in every pixel. Differences to the expected ground plane define foreground information, combined to associated regions. These regions of interest (ROI) are analyzed to distinguish persons from other objects by using a matched filter on the height-segmented depth information of each of these regions. The proposed method separates crowds

into individuals and facilitates a multi-object tracking system based on a Kalman filter. Furthermore, the authors present several applications for the proposed method. Experiments with different crowding situations—from very low to very density—and different heights of camera placements have proven the applicability and practicability of the system.

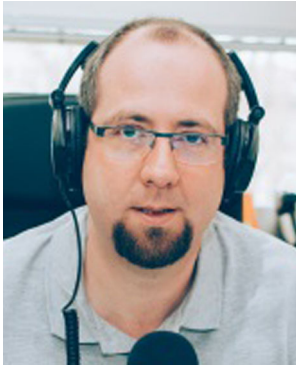
## 9 Audio analysis/processing

The last category includes only one paper; the paper is titled “Processing of acoustical data in a multimodal bank operating room surveillance system” (10.1007/s11042-014-2264-z). An automatic surveillance system capable of detecting, classifying, and localizing acoustic events in a bank operating room is presented. Algorithms for detection and classification of abnormal acoustic events, such as screams or gunshots are introduced. Two types of detectors are employed to detect impulsive sounds and vocal activity. A support vector machine (SVM) classifier is used to discern between the different classes of acoustic events. The methods for calculating the direction of coming sound employing an acoustic vector sensor are presented. The localization is achieved by calculating the direction of arrival (DOA) histogram. The evaluation of the system based on experiments conducted in a real bank operating room is presented. The results of sound event detection, classification, and localization are provided and discussed. The practical usability of the engineered methods is underlined by presenting the results of analyzing a staged robbery situation.

**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, Ph.D. Hab.** He received his M.Sc. and Ph.D. in telecommunications from the Electro-technical Institute in Leningrad, Faculty of Automation and Computer Science, in 1970 and 1973. He received a postdoctoral degree in engineering, Science Tech, from the Poznan University of Technology, Faculty of Electrical Engineering in 1978. He is a full professor from 1986. His fields of interest are related to digital communication, image and data processing, data compression, information and coding theory, random signals, computer communications networks, and signal processing. He has worked in number of foreign universities. Most recently in 2001–2003, he worked as a visiting professor at the University of Wuppertal in Germany. He co-authored 180 publications including five books. He promoted 18 Ph.D. students and approximately 100 M.Sc. He was awarded four times by the Ministry of Education of Poland for his research achievements. He serves as a coordinator of FP7 project INDECT.



**Mikolaj Leszczuk, Ph.D.** He 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, he has been employed at the AGH Department of Telecommunications. 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 (co-)authored approximately 130 scientific publications of which 23 are publications in journals of the JCR database. He has been teaching at the undergraduate and graduate levels. He has co-supervised one Ph.D. student and supervised (promoted) approximately 40 M.Sc. students of various nationalities. 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.



**Remigiusz Baran** He was awarded his M.Sc. in Electrical Engineering from the Faculty of Electrical and Control Engineering, Kielce University of Technology in 1993, and 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 with an Israeli partner.