Intelligent video surveillance systems for public spaces – a survey

Michał Zabłocki[1, 2], Katarzyna Gościewska[1, 2], Dariusz Frejlichowski[1], Radosław Hofman[2]

[1] West Pomeranian University of Technology, Faculty of Computer Science and Information Technology, Żołnierska 52, 71-210, Szczecin, Poland
[2] Smart Monitor sp. z o. o., Cyfrowa 6, 71-441, Szczecin, Poland

`{michal.zablocki,katarzyna.gosciewska,radekh}@smartmonitor.pl,`
`dfrejlichowski@wi.zut.edu.pl`

Abstract: *In recent years, a large number of cameras have been installed in public spaces as a part of intelligent video surveillance systems. Such systems are being continuously developed due to the advancements in the Video Content Analysis algorithms. In this paper, some of the latest state-of-the-art intelligent video surveillance systems will be presented in the context of their most desirable characteristics and features. Due to the variety of the solutions the following categories have been taken into consideration: systems based on object detection, tracking and movement analysis, systems able to warn against, detect and identify abnormal and alarming situations, systems based on vehicle detection and traffic or parking lots analysis, object counting systems, systems based on multiple integrated camera views, privacy preserving systems and systems based on cloud environment. The paper describes several solutions for each category and underlines main functionalities of the current intelligent surveillance systems.*

Keywords: *intelligent video surveillance system, public spaces*

1. Introduction

This paper presents an overview of contemporary issues of the intelligent video surveillance systems for public spaces described in the world's scientific publications. This survey indicates the capabilities of modern software in the context of functionalities planned to be implemented in the "SM4Public" system prototype. The related project is aimed at the construction and implementation of the innovative system prototype based on the Video Content Analysis algorithms that will ensure the safety of various public spaces using real-time solutions and typical computer components. The idea of the project was raised during the development of the previous system entitled "SmartMonitor" [1-4].

The analysis of alternative system applications has shown that there is a need to build the solution for public space video surveillance to effectively detect events threatening public safety, especially in places characterized by simultaneous movement of large number of people. In the comparison with existing systems, the main planned advantages of the "SM4Public" system prototype are high customizability and adaptability of system operation. According to that, "SM4Public" system will be able to work under scenarios specific for public spaces, such as scenarios associated with vehicle traffic (e.g. failing to stop at the red light, accident detection), infrastructure protection (e.g. devastation or theft detection),

breaking the law (e.g. drinking alcohol in public spaces, prohibited in many countries) or threats to life or health (e.g. a fall).

The rest of the paper is divided into three sections. The second section addresses main issues concerning intelligent surveillance systems implemented to monitor public spaces. In the third section, a group of selected intelligent video surveillance systems is presented. The described solutions have been classified according to planned functionalities of the "SM4Public" system and include: systems based on object detection, tracking and movement analysis, systems able to warn against, detect and identify abnormal and alarming situations, systems based on vehicle detection and traffic or parking lots analysis, object counting systems, systems based on multiple integrated camera views, privacy preserving systems and systems based on cloud environment. Each subsection of the third section contains the description of several systems relevant to the specific category. The fourth section contains brief summary and conclusions.

2. Main issues related to the intelligent monitoring systems used in public spaces

As a public space we understand an accessible to people and generally open areas such as roads, parks, pavements, parking, squares, etc. Some parts of buildings and other urban infrastructure can also be considered as public space if they are open to the public, e.g. public transport (railway platforms, halls and waiting rooms), shopping centre (hallways), government buildings (e.g. library or halls were the clients are served by the officials). Nowadays, video surveillance systems are rapidly being placed in public spaces to strengthen public safety. A large number of surveillance cameras have been installed due to their decreasing costs. The law enforcement community is increasingly relying on video surveillance for crime prevention and community safety. Video footage captured through surveillance cameras is routinely used to identify suspects and as evidence in the courts.

Originally, the video surveillance systems were designed for human operators in order to remotely observe protected space or concurrently to record video data as archive for further analysis. Watching surveillance video is a labour-intensive task when a large number of cameras need to be controlled. Furthermore, it is also a tedious task and human observers can easily loose attention. Automation can help to overcome both cost and performance issues. Moreover, it can free security personnel from routine tasks and let them focus on higher-level cognitive work that better utilize their abilities. Intelligent video surveillance [5] is of great interest in industry applications due to the increased demand for the reduction of the manpower analysing large-scale video data.

Regarding the terminology, Elliott has recently defined an intelligent video system (IVS) as "any video surveillance solution that utilizes technology to automatically, without human intervention, process, manipulate and/or perform actions to or because of either the live or stored video images" [6]. The industry and academics, being motivated by the availability of powerful and low cost computing hardware, have developed the key technologies for intelligent surveillance, such as object tracking [7, 8], pedestrian detection [9], gait analysis [10], vehicle recognition [11], privacy protection [12], face and iris recognition [13], video summarization [14] and crowd counting [15]. Video surveillance systems have been used for various applications, such as traffic monitoring, security, post-incident analysis, etc. IVS embeds computer vision technologies into video devices such as cameras, encoders, routers, digital video recorders, network video recorders, and other video management and storage devices [16].

Recent advancements in image processing, camera hardware, and wireless communication have led to an increase in the rate of installing new video surveillance systems in private and public places, such as airports, railway stations, banks, restaurants, schools, hospital, houses, etc. Video data collected by cameras present in the scene have many applications outside of the surveillance domain. The most noteworthy among them are smart environments where video data can be used to identify occupants and analyse activities taking place in these environments. The ability to "see" enables smart solutions to respond to the needs of their users. Existing intelligent video surveillance systems offer many functionalities. The following list presents selected and most significant features, however it does not exhaust all capabilities of modern software. Simultaneously, it is adequately express capabilities of the state-of-the-art's techniques in the field of Visual Content Analysis: moving object detection, particularly human silhouettes, namely classification of detected object as people [17-20], intrusion detection/perimeter protection [21], crowd behaviour analysis [22, 23], activity recognition and behaviour understanding [24-26], abnormal/irregular behaviours detection, such as fight [27], faint, fall [29-30], sudden events [31], counting the number of moving objects [32, 33], object tracking and trajectory analysis [34, 35], vehicle tracking and traffic analysis [36, 37], smoke/fire detection [38, 39], camera tampering detection [40], abandoned/taken object detection [41, 42].

The video surveillance systems used in public spaces have to deal with large traffic density, as well as heterogeneity, variability and occlusions of appearing objects. Such systems often have to detect sudden and complex events. However, there is another set of problematic aspects associated with privacy, i.e. privacy preservation of monitored area in public spaces. The widespread use of the video surveillance has raised concerns about the privacy and impairs people "right to anonymity". Video surveillance combined with biometric technology (e.g. face recognition) raises even more of these concerns [43]. Video cameras are constantly watching pedestrians' movements on the streets, at the stations, in shops, workplaces and so on.

People generally dislike when their activities are being recorded and watched by the others. With the increased automation and growing networking of computers and various information systems, the potential for linking information has driven increased concerns about the adverse effects of video surveillance on our privacy. The main challenge here is to understand and analyse various inference channels that can result in a breach of privacy. While such inference channels are well studied in the context of traditional data sharing applications (e.g. a hospital releasing patient records and GPS-based location aware services), it is challenging to understand inference channels embedded in semantically rich video [44]. For example, by incorporating an automatic face detection mechanism into the surveillance video system and linking it with the database or networked systems, significant information can be mined out. The surveillance systems are owned and operated by both public and private organizations which can misuse the surveillance data for their advantage. The surveillance data can be used to identify people, their day to day activities, collecting demographic information and so on. Thus mechanisms that can satisfy the objectives of surveillance while ensuring privacy to people are necessary [45].

3. A review of the selected intelligent video surveillance systems for public places

This section contains a description of selected intelligent video surveillance systems. The review includes both the solutions implemented in public places and those that may be applied there. Many of these solutions provide valuable knowledge that will be useful in the

development of the "SM4Public" system. To facilitate the understanding of the content and systematize collected information, the description has been divided into several subsections.

3.1. Systems based on object detection, tracking and movement analysis

Automatic detection and tracking of multiple objects in the scene is a challenging problem. There are multiple solutions for human detection, ranging from generic algorithms and ending on architectures specially developed to deal with human silhouette. The most common challenges involve collision and occlusion detection. Moreover, the algorithm has to be able to re-identify target in other camera views, or when it is temporarily totally or partially occluded by any element of the scene. This task is especially difficult when two or more people (tracked targets) occlude each other in front of the camera, what can cause an identification problem. The main idea of multi-object tracking systems is to include a priori information about the object of interest. In case of dealing with human object, information about its shape is used.

In [46], a strategy for human tracking under unpredictable trajectories is presented. The strategy is based on the omega-shaped descriptor. The tracing procedure uses particle filter system in combination with linear filter to predict the next position of a tracked person. The error produced by the particle filter along the successive frames is handled by the improved Viola-Jones and HOG feature-based SVM detector. When either a collision or occlusion event is detected, the particle filter (used to track every person in the scene) is disabled. The areas in which the lost target could occur are defined by elliptic blobs. Newly appeared target is compared against the lost target using colour-histogram representation.

The method proposed in [47] deals with detection and tracking. The detection includes face and eye detection. After the face is detected in the frame, the coordinates of eye region are calculated and tracked. Face, as the most individual part of the human body, is an important element of human identity. It varies from person to person but can be located with the use of certain context information. The complexity of face causes most difficulties in the development of the face model. Both internal (facial expression, beard, moustache, glasses) and external (scale, lightening conditions and orientation) factors have influence on human face. Face detection based on skin colour segmentation is the simplest technique and requires the least computations. The authors of [47] have decided to use YC_bC_r colour model to detect skin regions due to the fact that this model represents intensity and colour information separately. Additionally, the authors use morphological operations as a tool for extracting image components that are useful in the representation and description of the shape region. Two most important operations are erosion and dilation. Eye region is detected by the use of projection function and pixel counting methods. Finally, tracking is performed using the mean shift algorithm and Kalman filter algorithm.

In [48] authors propose another technique for the detection and tracking of multiple human objects in the video. The solution includes the classifier based on Haar-like features for object detection and particle filter for tracking. According to the experimental results, proposed technique has good performance in the presence of poor lighting conditions, variation of human objects in poses, shape, size, clothing etc. It handles varying in time number of people.

In [49] the problem of anomalous motion patterns in urban video surveillance is investigated. Anomalous motion patterns are typically caused by people merging into dense groups and initiating disturbances or threatening situations within the group. The authors try to automatically detect anomalous motion patterns in crowds. They use the unsupervised K-means clustering and the semi-supervised hidden Markov model (HMM) as a tool for crowd

analysis. By changing K-means values, and examining cluster density, cluster quality and changes in cluster shape, the results show some motion patterns that well correspond with the events on the video sequences. This brings the conclusion that the process of analysis can be automated. The results also illustrate that very accurate detections of people in the dense group would not be necessary. An uncertainty in the detections does not have essential influence on the analysis process.

3.2. Systems able to warn against, detect and identify abnormal and alarming situations

Video surveillance systems are able to manage hundreds or even thousands of cameras, e.g. to cover a large shopping mall building or a busy airport. The large number of cameras may cause communication problems. Therefore, by compressing the captured video in a local camera processor or by the nearby video server, it is possible to avoid communication bottlenecks. The compressed video is broadcasted to a central facility for storage and display. The key to effective and economical video surveillance is the real-time automatic abnormal motion detection. The automatically detected abnormal motion can trigger video transmission and recording, and can be used to attract the attention of the human observer to a particular video channel or even to alarm proper security services. Three related challenges characterize this problem. The first challenge requires persistence of irregular events detection and low false-alarm rate at the same time. The second one requires effective characterization of normal and abnormal motion, enabling their discrimination. The third challenge requires to perform detection in the limited time and using the limited computational power.

In reference to the above-mentioned challenges, [50] presents a novel real-time abnormal motion detection algorithm. The authors use macroblock motion vectors generated as a part of the standard video compression process. The algorithm is based on normal activity description, which is characterized by the joint statistical distribution of the motion features. This distribution is estimated during the training phase at the examined scene. Abnormal motion is indicated by the unlikely motion feature values. The approach based on motion vectors reduces the input data rate by about two orders of magnitude in comparison with the approach based on pixel data. Furthermore, it enables real-time operation under limited computational resources.

The authors of [51] present an intelligent video surveillance system that by analysing object movement is able to detect and identify abnormal and alarming situations. The main advantage is that by minimizing video processing and transmission the system enables a large number of cameras to be deployed. This advantage makes the system suitable for usage as an integrated safety and security solution in Smart Cities. The system detects abnormal and alarming events on the basis of parameters of moving objects and their trajectories. The system also employs a high-level conceptual language easy to understand for human operators using semantic reasoning and ontologies. Thus it is capable of raising enriched alarms with description of the situations in the video. Moreover, the system can automatically react to these alarms by alerting the appropriate emergency services via the Smart City safety network.

Another paper [52] concerns 3D spatial- and temporal-volume based event descriptors for video content analysis and event detection. The authors proposed a 3D shape-matching method to facilitate definition and recognition of human actions recorded in the videos. The method is based on Spatio-Temporal Volume (STV) data structure and region intersection (RI). The main idea of this approach is to compare the STV "shapes" derived from the video stream with the 3D event templates. In the event detection pipeline the STV model is con-

structed by an extended 3D hierarchical Pair-wise Region Comparison (PWRC) segmentation algorithm. The proposed approach is distinguished by the performance gain that stems from the application of the coefficient factor-boosted 3D region intersection and matching mechanism. The paper also reports the comparative study on techniques for the efficient STV data filtering able to reduce the amount of voxels (volumetric-pixels) that need to be processed in each operational cycle in the proposed system.

The authors of [53] present an intelligent framework for the detection of multiple events in surveillance videos. The basic framework principle is the compositionality. The authors of the system modularize the surveillance issues into a set of variables: regions-of-interest, classes, attributes and a set of event notions. Reasoning process results enable broader and integrated understanding of complex activity patterns in the scene. In general, according to the proposed method, the process of detecting multiple events is categorized into three broad levels. First level is associated with data acquisition process. In the second level, knowledge of the environment is constructed, followed by the analysis and reasoning based on principle of compositionality. In the third level any abnormal event triggers an alarm.

In [54], the abnormal human activities detection system working in an outdoor surveillance environment is described. In this system the human tracks are provided in real time by the baseline video surveillance subsystem. This subsystem has capability of multi-camera human tracking. It delivers humans' (actors in the scene) trajectories as the list of tracks (location information at each frame – x-y positions). Information about trajectory is analysed by the event analysis module in order to detect a presence of suspicious activity in the scene. Due to the real-time processing constrains, the further more detailed analysis is performed in an off-line manner. This prevents false alarm generation resulting from the video noise or non-human objects. The proposed hierarchical abnormal event detection methodology is able to perform as multi-tasking approach in the real and semi-real time, e.g. low level tasks, like human detection and tracking, basic abnormal event detection and identification are performed in real time. Low level real-time process utilizes actor's trajectory as an input, verify if the received information is associated with the already existing event objects, verify if multiple event objects are related and then merge them to remove redundancy. Subsequently, detectors browse events for specific events such as illegal entry, a fall or line formation. As an output of this low level process the VOD (Video-On-Demand) data request for the video of before-and-after scenes of the detected suspicious event is send. This process is relatively computationally inexpensive and it can be performed in real-time. VOD data is used as an input for the high level non-real-time process where it is verified whether event objects (actors) are humans. This process employs very computationally expensive algorithms, therefore it cannot be performed in real-time.

In [55] the authors propose a large-scale crowd behaviour perception method that is based on novel concept of spatio-temporal viscous fluid field models. The analysis of interaction forces in crowd is based on viscous fluid theory and represents the crowd as a fluid abstraction. The methodology explores both appearance of crowd behaviour and interaction among pedestrians in order to model crowd motion patterns. The local fluctuation of the video signals in both spatial and temporal domains is represented in a spatio-temporal variation matrix. To extract the principal fluctuations resulting in an abstract fluid field, an eigenvalue analysis is applied on the matrix. To characterize motion properties of a crowd, an interaction forces and fluctuations are explored based on shear force in viscous fluid. Dirichlet allocation model enables to recognize the crowd behaviours. The authors create a codebook by clustering neighbouring pixels with similar spatio-temporal features. The proposed method obtains high-quality results for large-scale crowd behaviour perception in terms of both robustness and effectiveness.

The [23] focuses on the video anomaly detection which plays a critical role in the intelligent video surveillance. The authors consider both spatial and temporal contexts of abnormal event detection in the system. The spatio-temporal video segmentation is performed to characterize the video. Then, using a novel region-based descriptor, motion and appearance information of the spatio-temporal segment is described. The authors formulate the abnormal event detection as a matching problem. They explain this approach pointing to the greater robustness of this solution in comparison to static model-based methods, especially when the training dataset has limited size. The process of abnormal event detection is based on searching for each tested spatio-temporal segment's best match in the training dataset, and is aimed to determine how normal it is using a dynamic threshold. Compact random projections are used to accelerate this process. The performed experiments and comparisons with the state-of-the-art methods confirm superiority of this approach.

The [56] presents a spatial-temporal hierarchical topic model. The model is used to represent a crowded traffic scene in the process of irregular behaviour. The authors have proposed the approach for pattern learning of behaviours occurring in crowded traffic scenes based on the non-object statistical algorithms which give an ability to detect locally and globally occurring behaviours over time and space. The spatial-temporal behaviour patterns of moving objects in the crowded traffic scenes are described during a spatial-temporal hierarchical probabilistic latent semantic analysis (ST-HpLSA) with the use of the bag-of-words representations of visual features. Complex video scenes usually are composed of multiple behaviours which are divided into local and global behaviours. Local behaviours, such as vehicle or pedestrian moving across the road, occur in small spatial and temporal region of the consecutive video frames. Distributions of local behaviours over space and time in the video represent the global behaviours. The general behaviour pattern consists of each event object from the video scene that is intrinsically affected by the other objects. To accurately describe these patterns, the hierarchical behaviour modelling based on correlations between local and global behaviours is required. By determining the log-likelihood of each behaviour distribution it is possible to detect irregular behaviour. The probability of occurrence of irregular behaviours is high when the log-likelihood is low. The experiments show favourable results of using the ST-HpLSA for the detection of irregular behaviours.

The framework described in [57] is able to detect complex events in surveillance videos. The system consists of three components. The first one detects moving objects in the foreground. The second component tracks and labels each detected object, and handles occlusion situations. The last component creates and trains rule-based event models using Markov Logic Networks (MLNs) in a manner that each rule is given a weight. In the training process, events are inferred using MLNs. The weight values assigned in this process are used to determine the occurrence of the event. The proposed system is able to handle many complex events simultaneously.

3.3. Systems based on vehicle detection and traffic or parking lots analysis

In the traffic surveillance, the segmentation of road information brings special benefits, e.g. it enables to automatically wrap regions of traffic analysis. Consequently, it speeds up flow analysis in the videos and helps with the detection of driving violations. Moreover, road segmentation improves contextual information in the videos of traffic. In turn in urban scenes it is related to the following challenges: the urban scenes exhibit more cluttered sceneries; these sceneries consist of many vehicles stopped on the roads; the roads in these scenes are not only parallel lines, they also veer and intersect under various angles; and fi-

nally, pedestrians can complicate the challenge since they can cause errors during road boundaries analysis.

The authors of [58] are interested in segmenting road regions from the remaining parts of an image. They aim to support traffic flow analysis tasks, where the road segmentation relies on the superpixel detection based on the novel edge density estimation method. The proposed approach firstly builds a background model with the use of the fast technique based on a median filter. The resulting background model is segmented by the edge-density-based superpixel detection process. All features are computed on the basis of grey level amount, texture homogeneity, horizon line and motion. Ultimately, it is determined whether a particular superpixel belongs to a road or not, and the comparison is based on the Support Vector Machine (SVM) supported by a vector composed of all features.

The [59] presents the video-based approach to traffic analysis and monitoring in the night light conditions. This task is especially challenging due to the fact that under this conditions the vehicle silhouettes are very low-contrasted. In this case many algorithms used in the day-time have decreased performance. The main features extracted from the image taken from an urban or inter-urban traffic camera are vehicle headlights. The proposed algorithm uses these features to obtain a number of vehicles per time unit and to extract other information, i.e. intensity, mean speed and traffic occupancy. The system was tested on the real-time video sequences from the Traffic Authority of the city of Valencia, Spain. Urban parking management is the topic that receives significant attention as well. In contrast to the systems based on many various non-video sensors, the video-based monitoring for on-street parking areas offers several advantages. The main advantage is that the cost of the solution is lower due to the use of inexpensive video cameras that can monitor and track several parking spots. Moreover, the maintenance of the video cameras is less inconvenient than the maintenance of the in-ground sensors. Video-based system can also deliver other tasks, like traffic surveillance, since cameras can capture a wider range of useful information including vehicle colour, license plate, vehicle type, speed, etc.

In [60], the video-based real-time on-street parking occupancy detection system is presented. It is composed of video cameras that monitor on-street parking areas. The cameras continuously capture video and transfer the image sequences to the central processing unit. Video data is processed in order to estimate the on-street parking occupancy. This information can be delivered to drivers via smart phone apps, radio, Internet, on-road signs, or global positioning system auxiliary signals. The proposed method is designed to handle several challenges associated with computer vision, such as shadows, reflections, illumination changes, occlusions, rain and other inclement weather conditions, as well as camera shake due to wind or vehicle-induced vibration. The method takes advantage of several components from the video processing and computer vision for background subtraction, and vehicle and motion detection.

The authors of [61] investigate several key challenges of automatic vehicle recognition in video surveillance systems composed of multiple cameras and propose new technologies for these systems. The authors focus on two main issues for vehicle recognition: automatic vehicle recognition (AVR) and license plate recognition (LPR). The automatic vehicle recognition tackles with such challenges as multiple-feature fusion, heterogeneous camera properties for region extraction and imperfect object detection/recognition. A novel idea for AVR multiple cameras for the video surveillance application presented in this paper is based on the proposed adaptive license plate detection technique, connected component analysis and level-based region comparison algorithm. Additionally, for improving LPR accuracy, the paper presents new LPR method based on self-organizing neural network to identify alphanumeric characters of license plates. Experimental results indicate simplicity and ef-

fectiveness of the proposed algorithms. The composed image has comparable quality to the results of the state-of-the-art methods.

3.4. Object counting systems

Estimating the number of objects in a video is constantly a challenging task in computer vision. From the video surveillance perspective, the most popular issues are related with people and vehicle counting in the sense of traffic dense estimation. Counting by detection is a common approach. However, it requires explicit object detection and modelling, and occlusion handling. The authors of [62] propose the counting method for arbitrary objects in an arbitrary scene. The method automatically determines the area of interest through the motion flow of objects in the scene. Additionally, a perspective effect is compensated when it is detected.

In [32] a novel approach to people counting and density estimation in crowd environment for human safety is presented. In order to avoid threatening situations in crowded public areas it is fundamental to estimate the number of people and react when the limit is exceeded. The proposed solution is based on density estimation of the crowd size and the number of people in the crowd. The main idea assumes that whereas crowd density increases, the occlusion between people also increases. To bypass such problem, the Improved Adaptive K-GMM Background subtraction method is used to extract foreground. The size of the crowd is estimated by the boundary detection algorithm. Canny edge detector, connected component labelling and bounding box with centroid algorithms are used to count the number of people in the crowd.

The other paper [63] presents an efficient self-learning people counting system. It can precisely count the number of people in a region of interest. To effectively detect the pedestrians a bag-of-features model is used. It provides acceptable distinction between static or slowly moving pedestrians, and the background. The proposed system has an ability of automatic pedestrian and non-pedestrian samples selection in order to update the classifier. It has ability to a real-time adaptation to the specific scene. Experimental results proved the robustness and high accuracy of the system.

3.5. Systems based on multiple integrated camera views

The monitored regions in the video surveillance are often wide. Due to the limited field of camera view and in order to cover the whole scene area, multiple cameras are required. In the traditional video surveillance system the monitored area is controlled by security guards. The guards can have difficulties with keeping track of targets since it is arduous to simultaneously look at many screens over a long period of time. Therefore, it is convenient to develop a surveillance system capable to integrate all the video streams captured by multiple cameras into a single comprehensive view.

In [64] a 3-D surveillance system based on multiple cameras integration is presented. The authors of the approach build the 3-D environment model using planar patches and dynamically maps texture on the 3-D model to visualize videos. By estimating homographic transformations for every pair of image regions in the video content and the corresponding areas in the 3-D model, the relationship between the camera content and the 3-D model is obtained in the form of lookup tables. These lookup tables accelerate the coordinate mapping in the video visualization processing. Complex environment requires smaller texture patches to be accurately rendered. On the other hand, it results in a higher computational cost. The patches of texture are automatically divided into appropriate size using mean-squared error (MSE) method to estimate the amount of distortion during the image patches

rendering. Finally, for better 3-D visual effects, moving objects that were segmented from the background are displayed via axis-aligned billboarding.

3.6. Privacy preserving systems

Video surveillance systems play a strategic role for a variety of critical tasks, such as personal safety, traffic control, resource planning and law enforcement. They are widely deployed in many strategic places such as airports, banks, public transportation or busy city centres. People usually appreciate the sense of increased security brought by these systems. However, some of them have introduced severe concerns about the privacy and trustiness of the captured data. In addition, wireless cameras represent another great risk factor for the security of the transmitted video streams. Illicit parties may intercept a visual data and use them for nefarious purposes such as obtaining sensitive data or maliciously manipulating to hide some evidence and/or introduce some false ones. Therefore, there is a need to introduce a reliable solution to protect the obtained images and videos, their source devices and transmission media.

The [65] introduces a video surveillance system that provides distributed trust and content confidentiality by using a hybrid cryptosystem based on a threshold multi-party key-sharing scheme. The approach delivers a flexible basic video content access-control scheme that can handle the problem of users losing their private shares. It can dynamically add new users that can participate in the content reconstruction. The system can be practically realized on the low-cost special purpose devices.

In [43] a Privacy Aware Surveillance System (PASS) that has an ability to enforce user-specific privacy policies is presented. In comparison with existing privacy-aware video surveillance systems the PASS stands out because of user interaction support. It collects privacy preferences and gets user consent to video data collection. Users can indicate desired privacy settings to the system by gestures. The system locates person presence in the scene using visual analysis, interprets gestures and applies requested privacy using privacy preserving filters.

Authors of [45] propose a novel, selectively revocable privacy preserving mechanism on-demand. The surveillance video can preferentially, for any individual, turn or revoke a view with complete privacy and without any influence on the privacy of the other pedestrians in the scene. This is achieved by tracking the pedestrians using a novel Markov chain algorithm with two hidden states. The system detects the head contour of the tracked pedestrian and obscures his/her face using the encryption mechanism. The encryption mechanism uses unique key derived from a master key for the privacy preservation purpose.

A crucial requirement of privacy solutions is an understanding and analysis of the inference channels that can lead to a breach of privacy. Privacy estimation of the video data has been reduced to the distinct identifiers such as faces in the video. Other important inference channels such as location (Where), time (When) and activities (What) are generally neglected. In [44] the privacy deprivation model is presented. The model emphasizes and take into account identity leakage through multiple inference channels that exist in a video due to what, when and where information. The approach models identity leakage and incorporates sensitive information to determine privacy deprivation. The proposed model enables consolidation of the identity leakage through multiple events and multiple cameras.

3.6. Systems based on cloud environment

Internet protocol-based cameras are slowly replacing Closed-circuit television (CCTV) devices. This is the result of the considerable advancement of audio and video compression

algorithms. Surveillance industry is changing. Network video recorders – a main component for the surveillance environments – gradually replace existing digital video recorders.

In [66] a scalable, flexible and reliable cloud-based video recording system is introduced. The solution designed in the Infrastructure as a Service abstraction layer of cloud computing manages the platform by the CVR service provider. The system consists of three parts. The first part, the Device part, is responsible for video data acquisition. The second one, the Cloud part, is a main part deployed in the cloud environment. It is responsible for video input stream handling in real-time, massive video data storage in distributed manner using Hadoop file system, access controlling, intelligent video content analysis and providing content managing system. The third part, the User part, is responsible for visualization, alerting and video data delivering.

In [67], a semantic based cloud environment is presented. Taking advantage of semantic content description and cloud computing the proposed solution facilitates the process of analysing and searching video surveillance data. The cloud computing gives sole reasonable utility in case of the following aspects: the massive amount of video data generated in real-time, the need to handle the data suitably for all end-users according to the Quality of Service (QoS) (in terms of appropriate content availability, content search delay, and content analysis delay), as well as the fact that processing outcomes will be consumed and processed by many end-users. The authors of [67] believe that an architecture integrating ontology for building video semantic descriptions and semantic searching upon them will bridge the gap between low-level features and high-level semantics of the video contents.

4. Conclusions

This paper contains a description of the selected intelligent video surveillance systems. The review reflects current research interests in both the solutions implemented in public places and the approaches that may be applied there. Nowadays, video surveillance systems are becoming more intelligent, automated and autonomic, which corresponds to the descending human operator's involvement. Furthermore, the newly designed algorithms and more advanced hardware devices increase the performance of video surveillance systems. This brings new applications and boosts capabilities of these systems – they work faster, are more effective and are able to manage more devices.

Research studies associated with the intelligent video surveillance systems development involve various features and characteristics. The most current trends concern the application of cloud computations. On the other hand, the main problem seems to be associated with legal aspects of privacy protection and anonymity preservation of people present in the captured scenes. Furthermore, there are some technical problematic aspects that involve difficulties with object movement analysis for occlusion handling and synchronization of multiple camera views during real-time system operation. Moreover, there is a continued interest in the studies and development of already existing standard functionalities of video surveillance systems, such as detection, tracking and classification of moving objects, as well as recognition of activities, behaviours and abnormal events. In case of systems deployed in public spaces a particular attention should be paid to early detection of threats, abnormal activities or security breaches. These events can involve various objects, such as vehicles, human individuals, and small groups of people or even dense crowds. The heterogeneity of objects and unpredictability of their behaviours make it difficult to develop reliable algorithms. However, the use of the Video Content Analysis algorithms is unquestionably a much more effective solution than watching images from the cameras by security employees. It automates repetitive tasks and enables the notification of a larger number of events in

a shorter time. Constant development of new techniques and increasing number of scientific publications indicate on intensive researches and the presence of unsolved problems. Therefore, there still is a large area for new advancements and new applications of the intelligent solutions which explains the need for the construction of the "SM4Public" system that will be able to increase security of public spaces using some of the latest achievements of science and technology.

Acknowledgements

References

[1] Frejlichowski, D., Forczmański, P., Nowosielski, A., Gościewska, K., Hofman, R.: SmartMonitor: An Approach to Simple, Intelligent and Affordable Visual Surveillance System. In: Bolc, L. et al. (eds.) ICCVG 2012. LNCS, vol. 7594, pp. 726–734. Springer, Heidelberg, 2012.

[2] Frejlichowski, D., Gościewska, K., Forczmański, P., Nowosielski, A., Hofman, R.: Extraction of the Foreground Regions by Means of the Adaptive Background Modelling Based on Various Colour Components for a Visual Surveillance System. In: Burduk, R. et al. (eds.) CORES 2013. Advances in Intelligent Systems and Computing, vol. 226, pp. 351–360. Springer International Publishing, 2013.

[3] Frejlichowski, D., Gościewska, K., Forczmański, P., Hofman, R.: 'SmartMonitor' — An Intelligent Security System for the Protection of Individuals and Small Properties with the Possibility of Home Automation. Sensors 14, 9922–9948, 2014.

[4] Frejlichowski, D., Gościewska, K., Forczmański, P., Hofman, R.: Application of foreground object patterns analysis for event detection in an innovative video surveillance system. Pattern Anal. Appl., 1–12, 2014.

[5] Singh, V., Kankanhalli, M.: Adversary aware surveillance systems, IEEE Trans. Inf. Forensics Security, vol. 4, no. 3, pp. 552–563, Sep. 2009.

[6] Elliott, D.: Intelligent video solution: A definition, Security, pp. 46–48, 2010.

[7] Avidan, S.: Ensemble tracking, IEEE Trans. Pattern Anal. Mach. Intell., vol. 29, no. 2, pp. 261–271, Feb. 2007.

[8] Khan, Z., Gu, I.: Joint feature correspondences and appearance similarity for robust visual object tracking, IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 591–606, Sep. 2010.

[9] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection, in Proc. CVPR, 2005, pp. 886–893.

[10] Wang, L.: Abnormal walking gait analysis using silhouette-masked flow histograms, in Proc. ICPR, 2006, vol. 3, pp. 473–476.

[11] Wang, S., Lee, H.: A cascade framework for a real-time statistical plate recognition system, IEEE Trans. Inf. Forensics Security, vol. 2, no. 2, pp. 267–282, Jun. 2007.

[12] Yu, X., Chinomi, K., Koshimizu, T., Nitta, N., Ito, Y., Babaguchi, N.: Privacy protecting visual processing for secure video surveillance, in Proc. ICIP, 2008, pp. 1672–1675.

[13] Park, U., Jain, A.: Face matching and retrieval using soft biometrics, IEEE Trans. Inf. Forensics Security, vol. 5, no. 3, pp. 406–415, Sep. 2010.

[14] Cong, Y., Yuan, J., Luo, J.: Towards scalable summarization of consumer videos via sparse dictionary selection, IEEE Trans.Multimedia, vol. 14, no. 1, pp. 66–75, Feb. 2012.

[15] Cong, Y., Gong, H., Zhu, S., Tang, Y.: Flow mosaicking: Real-time pedestrian counting without scene-specific learning, in Proc. CVPR, 2009, pp. 1093–1100.

[16] Venetianer, P. L., Deng, H. L.: Performance evaluation of an intelligent video surveillance system—A case study, Comput. Vis. Image Understanding, vol. 114, no. 11, pp. 1292–1302, 2010.

[17] Paul, M., Haque, S., Chakraborty, S.: Human detection in surveillance videos and its applications – a review. EURASIP Journal on Advances in Signal Processing 176, 2013.

[18] Nguyen, T.-H.-B., Kim, H.: Novel and efficient pedestrian detection using bidirectional PCA. Pattern Recognition 46, pp. 2220-2227, 2013.

[19] Hu, M.-C., Cheng, W.-H., Hu, C.-S., Wu, J.-L., Li, J.-W.: Efficient human detection in crowded environment. Multimedia Systems., 2014.

[20] Hu, C.-S., Hu, M.-C., Cheng, W.-H., Wu, J.-L.: Efficient human detection in crowded environment based on motion and appearance information. 5th International Conference on Internet Multimedia Computing and Service, ICIMCS 2013, pp. 97-100. Huangshan, China., 2013.

[21] Park, J.-H., Shin, Y.-C., Jeong, J.-W., Lee, M.-J.: Detection and Tracking of Intruding Objects based on Spatial and Temporal Relationship of Objects. ASTL(21), pp. 271-274, 2013.

[22] Zhang, D., Peng, H., Haibin, Y., Lu, Y.: Crowd Abnormal Behaviour Detection Based on Machine Learning. Information Technology Journal, 12(6), pp. 1199-1205, 2013.

[23] Cong, Y., Yuan, J., Tang, Y.: Video Anomaly Search in Crowded Scenes via Spatio-temporal Motion Context. IEEE Transactions on Information Forensics and Security, 8(10), pp. 1590-1599, 2013.

[24] Vishwakarma, S., Agrawal, A.: A survey on activity recognition and behaviour understanding in video surveillance. The Visual Computer, 29(10), pp. 983-1009, 2013.

[25] Hu, Q., Qin, L., Huang, Q.-M.: A survey on visual human action recognition. Chinese Journal of Computers, 36, pp. 2512-2524, 2013.

[26] Borges, P., Conci, N., Cavallaro, A. Video-based human behavior understanding: A survey. IEEE Transactions on Circuits and Systems for Video Technology, 11, pp. 1993-2008, 2013.

[27] Blunsden, S., Fisher, R. Pre-fight detection: Classification of fighting situations using hierarchical AdaBoost. Fourth International Conference on Computer Vision Theory and Applications, 2, pp. 303-308, 2009.

[28] Rougier, C., Meunier, J., St-Arnaud, A., Rousseau, J.: Robust Video Surveillance for Fall Detection Based on Human Shape Deformation. IEEE Transactions on Circuits and Systems for Video Technology, 21(5), pp. 611-622, 2011.

[29] Ngo, Y., Nguyen, H., Pham, T.: Study on fall detection based on intelligent video analysis. The 2012 International Conference on Advanced Technologies for Communications, ATC 2012, 2012.

[30] Albusac, J., Vallejo, D., Jimenez-Linares, L., Castro-Schez, J., Rodriguez-Benitez, L.: Intelligent surveillance based on normality analysis to detect abnormal behaviors. International Journal of Pattern Recognition and Artificial Intelligence, 23(7), pp. 1223-1244, 2009.

[31] Suriani, N., Hussain, A., Zulkifley, M.: Sudden event recognition: a survey. Sensors, 13(8), pages 9966-9998, 2013.

[32] Karpagavalli, P., Ramprasad, A.: Estimating the density of the people and counting the number of people in a crowd environment for human safety. 2nd International Conference on Communication and Signal Processing, ICCSP 2013, pp. 663-667, 2013.

[33] Conte, D., Foggia, P., Percannella, G., Vento, M.: Counting moving persons in crowded scenes. Machine Vision and Applications, pp. 1029-1042, 2013.

[34] Li, X., Hu, W., Shen, C., Zhang, Z., Dick, A., Van Den Hengel, A.: A survey of appearance models in visual object tracking. ACM Transactions on Intelligent Systems and Technology, 4(4)., 2013.

[35] Morioka, K., Kovacs, S., Joo-Ho, L., Korondi, P.: A cooperative object tracking system with fuzzy-based adaptive camera selection. International Journal on Smart Sensing and Intelligent Systems, pages 338-358, 2010.

[36] Liu, Y., Lu, Y., Shi, Q., Ding, J.: Optical flow based urban road vehicle tracking. Ninth International Conference on Computational Intelligence and Security, pp. 391-395. Beijing, China., 2013.

[37] Jiang, W., Xiao, C., Jin, H., Zhu, S., Lu, Z.: Vehicle Tracking with Non-overlapping Views for Multi-camera Surveillance System. 2013 IEEE 10th International Conference on High Performance Computing and Communications & 2013 IEEE International Conference on Embedded and Ubiquitous Computing (HPCC_EUC), pp. 1213-1220, 2013.

[38] Çetin, A. E., Dimitropoulos, K., Gouverneur, B., Grammalidis, N., Günay, O., Habiboglu, Y. H., Verstockt, S.: Video fire detection – Review. Digital Signal Processing, 23, pp. 1827-1843, 2013.

[39] Jiang, B., Lu, Y., Li, X., Lin, L.: Towards a solid solution of realtime fire and flame detection. Multimedia Tools and Applications., 2014.

[40] Li, Z., Li, Q.: Protection of regional object and camera tampering. Proceedings of the IEEE International Conference on Software Engineering and Service Sciences, ICSESS, pp. 672-675. Beijing, China., 2013.

[41] Muchtar, K., Lin, C.-Y., Kang, L.-W., Yeh, C.-H.: Abandoned object detection in complicated environments. 2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, APSIPA 2013. Kaohsiung; Taiwan., 2013.

[42] Miyahara, A., Nagayama, I.: An intelligent security camera system for kidnapping detection. Journal of Advanced Computational Intelligence and Intelligent Informatics, 17, pp. 746-752, 2013.

[43] Barhm, M., Qwasmi, N., Qureshi, F., el Khatib, K.: Negotiating Privacy Preferences in Video Surveillance Systems. W K. Mehrotra, C. Mohan, J. Oh, P. Varshney, and M. Ali (Editors), Modern Approaches in Applied Intelligence, vol. 6704, pp. 511-521. Springer Berlin Heidelberg., 2011.

[44] Saini, M., Atrey, P., Mehrotra, S., Kankanhalli, M.: W3-privacy: understanding what, when, and where inference channels in multi-camera surveillance video. Multimedia Tools and Applications, 68(1), pp. 135-158, 2014.

[45] Zhang, P., Thomas, T., Emmanuel, S.: Privacy enabled video surveillance using a two state Markov tracking algorithm. Multimedia Systems, 18(2), 175-199, 2012.

[46] Cancela, B., Ortega, M., Penedo, M.: Multiple human tracking system for unpredictable trajectories. Machine Vision and Applications, 25(2), 511-527, 2014.

[47] Tathe, S., Narote, S.: Real-time human detection and tracking. 2013 Annual IEEE India Conference (INDICON), pp. 1-5, 2013.

[48] Kushwaha, A., Sharma, C., Khare, M., Srivastava, R., Khare, A.: Automatic multiple human detection and tracking for visual surveillance system. 2012 International Conference on Informatics, Electronics Vision (ICIEV), pp. 326-331, 2012.

[49] Andersson, M., Gustafsson, F., St-Laurent, L., Prevost, D.: Recognition of Anomalous Motion Patterns in Urban Surveillance. IEEE Journal of Selected Topics in Signal Processing, 7(1), 102-110, 2013.

[50] Kiryati, N., Raviv, T., Ivanchenko, Y., Rochel, S.: Real-time abnormal motion detection in surveillance video. 19th International Conference on Pattern Recognition, ICPR 2008, pp. 1-4, 2008.

[51] Calavia, L., Baladrón, C., Aguiar, J. M., Carro, B., Esguevillas, A. S.: A Semantic Autonomous Video Surveillance System for Dense Camera Networks in Smart Cities. Sensors, 10407-10429, 2012.

[52] Wang, J., Xu, Z. STV-based video feature processing for action recognition. Signal Processing, 93(8), 2151-2168, 2013.

[53] Lim, M. K., Tang, S., Chan, C. S. iSurveillance: Intelligent framework for multiple events detection in surveillance videos. Expert Systems with Applications, 41(10), 4704-4715, 2014.

[54] Lee, S., Nevatia, R.: Hierarchical abnormal event detection by real time and semi-real time multi-tasking video surveillance system. Machine Vision and Applications, 25(1), 133-143, 2014.

[55] Su, H., Yang, H., Zheng, S., Fan, Y., Wei, S.: The Large-Scale Crowd Behavior Perception Based on Spatio-Temporal Viscous Fluid Field. IEEE Transactions on Information Forensics and Security, 8(10), pp. 1575-1589, 2013.

[56] Park, S., Yoo, C.: Video scene analysis and irregular behavior detection for intelligent surveillance system. 2012 9th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI), pp. 577-581, 2012.

[57] Onal, I., Kardas, K., Rezaeitabar, Y., Bayram, U., Bal, M., Ulusoy, I., Cicekli, N.: A framework for detecting complex events in surveillance videos. 2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW), pp. 1-6, 2013.

[58] Santos, M., Linder, M., Schnitman, L., Nunes, U., Oliveira, L.: Learning to segment roads for traffic analysis in urban images. 2013 IEEE Intelligent Vehicles Symposium (IV), pp. 527-532, 2013.

[59] Mossi, J., Albiol, A., Albiol, A., Ornedo, V.: Real-time traffic analysis at night-time. 2011 18th IEEE International Conference on Image Processing (ICIP), pp. 2941-2944, 2011.

[60] Bulan, O., Loce, R. P., Wu, W., Wang, Y., Bernal, E. A., Fan, Z. Video-based real-time on-street parking occupancy detection system. Journal of Electronic Imaging, 22(4), 41109-41109, 2013.

[61] Rao, Y.: Automatic vehicle recognition in multiple cameras for video surveillance. The Visual Computer, 1-10, 2014.

[62] Zhou, Y., Luo, J.: A practical method for counting arbitrary target objects in arbitrary scenes. 2013 IEEE International Conference on Multimedia and Expo (ICME), pp. 1-6, 2013.

[63] Li, J., Huang, L., Liu, C.: An efficient self-learning people counting system. 2011 First Asian Conference on Pattern Recognition (ACPR), pp. 125-129, 2011.

[64] Chen, Y.-Y., Huang, Y.-H., Cheng, Y.-C., Chen, Y.-S.: A 3-D surveillance system using multiple integrated cameras. 2010 IEEE International Conference on Information and Automation (ICIA), pp. 1930-1935, 2010.

[65] Castiglione, A., Cepparulo, M., De Santis, A., Palmieri, F.: Towards a Lawfully Secure and Privacy Preserving Video Surveillance System. W F. Buccafurri, G. Semeraro (Editors), E-Commerce and Web Technologies, vol. 61, pp. 73-84. Springer Berlin Heidelberg., 2010.

[66] Lin, C.-F., Yuan, S.-M., Leu, M.-C., Tsai, C.-T.: A Framework for Scalable Cloud Video Recorder System in Surveillance Environment. 2012 9th International Conference on Ubiquitous Intelligence Computing and 9th International Conference on Autonomic Trusted Computing (UIC/ATC), pp. 655-660, 2012.

[67] Xu, Z., Mei, L., Liu, Y., Hu, C., Chen, L.: Semantic enhanced cloud environment for surveillance data management using video structural description. Computing, 1-20, 2014.