Arraiza Juan

Aginako Naiara
Vicomtech-IK4, San Sebastián, Spain

Kioumourtzis Georgios

Leventakis George
Center for Security Studies, Athens, Greece

Stavropoulos Georgios

Tzovaras Dimitrios
CERTH-ITI, Thessaloniki, Greece

Zotos Nikolaos

Sideris Anargyros
Future Intelligence, London, UK

Charalambous Elisavet

Koutras Nikolaos
ADITESS, Nicosia, Cyprus

Fighting Volume Crime: an Intelligent, Scalable, and Low Cost Approach

Keywords

volume crime, petty crime, surveillance systems

Abstract

Volume Crimes (aka petty crimes) take place on a daily basis affecting citizens, local communities, as well as business and infrastructure owners. In this paper, we present a novel intelligent surveillance solution (P-REACT) that integrates video and audio analytics both on-site (using an embedded platform connected to local sensors) and centrally on a cloud service. This intelligent surveillance system has been conceived and designed to anticipate volume crimes in areas where video surveillance is allowed by current legislation and more specifically in shops and public transportation systems; intended as a modular and low cost solution. The capability of dynamically adapting the analytic algorithms that are performed on-site provides a more accurate detection of crime evidences.

1. Introduction

Volume crime incidents undermine the social fabric of the community as they are associated with elevated rates of fear of strangers and general alienation from participation in community life. There is a direct link between the fear of crime and social outcomes, by-passing behavioural responses to fear. The idea communicated by this link is that the fear of crime, if widespread, can feed directly into attitudes that have broad social consequences, regardless of the behavioural responses that people make to fear.

The economic impact is also very high when compared to the gravity of volume crime. In Italy, a study carried out in 2010 and 2011 by FORMAT Research [9] showed that in 2010 more than 15% of

businesses pointed out crimes as one of the main causes limiting the competitiveness of Small and Medium Enterprises (SMEs). In 2010, shops and businesses spent about 5.2 billion Euros to protect their premises from crimes; around 2.5 % of added value produced by this economic sector.

The goal of the P-REACT project is to design and develop a low cost surveillance platform that will ensure communication between key users with a focus on increasing the ability of the ground police and security personnel to respond. The solution will encompass low cost video and audio sensors for detecting volume crime incidents, a cloud based monitoring, intelligent alert detection and storage platform. When a suspicious event is detected by sensors a work flow will be initiated which, when verified as a potential crime by the cloud, will alert security personnel and/or police with the relevant video and information, ensuring an immediate response.

This paper provides an overview of the P-REACT platform and the research findings throughout the first twelve months of the project implementation. The rest of this paper is organised, as follows. In Section 2 current state-of-the art solutions are presented. An introduction of the P-REACT solution, highlighting its main contributions, is provided in Section 3. Section 4 presents a discussion on the obtained results as well as some conclusions on the implications that those results might have.

2. State of the Art

CCTV surveillance for crime prevention is certainly not a new research topic [5], [7] and it has been tackled from many different angles. Some researchers for example have focused in the legal, ethical and privacy aspects of the CCTV surveillance [13], [16]. Others are of interest in the identification of the visual cues that are used in CCTV images to accurately predict a criminal act. Grant and Williams [11] established a link between better crime prediction accuracy in CCTV images and a more sophisticated awareness of the social context present in the scene. More recently, Dadashi et al. [6] have shown that when an automated or semi-automated CCTV system provides reliable system confidence information to operators, workload significantly decreased and spare mental capacity significantly increased. In particular, the use of CCTV surveillance systems to prevent and fight volume crime is not a new area of research [1], [12] and [19] either. There has also been research in several scientific areas related to volume crime prevention, such as identifying infrequent events in surveillance video [15] or pre, post, and actual fight events detection [2].

The European Commission has been funding several research projects related to smart surveillance systems, being some of the most relevant the following ones: SAVASA, ADVISE, VIDEOSENSE, SMARTPREVENT or ADDPRIV [18]. The main goal of each one of them will be briefly described. The SAVASA project, similarly to what is done in the ADVISE project, focuses in video archive searches and analysis to allow authorized users to perform semantic queries over various remote and non-interoperable video archives. The VIDEOSENSE project, on the other hand, focuses in the domain of ethically-aware data and video analytics. Also with the aim of pre-serving privacy but with a different approach the ADDPRIV project focuses in enriching the current video surveillance systems through an automatic discrimination of relevant data recorded.

3. Introducing the P-REACT solution

3.1. Key characteristics

The main idea behind the P-REACT solution is to provide a system that will allow the installation of low-cost components in the premises of small businesses or transportation operators, networked with one or more cloud based services, monitoring and analysing their activity.

The end-users of the proposed system are small business owners and transport operators/agencies, in which premises a system consisting of a number of sensors will be installed. These sensors, depending on the end-user's needs, may include a combination of cameras, depth sensors and microphones. Although most of these types of sensors are being utilized in already existing surveillance systems, the proposed solution advances over them by employing state of the art methods for automatic volume crime detection by the system itself. To the best of our knowledge, existing surveillance systems only monitor the areas, and do not perform any kind of processing, expecting from a user to detect any events, or just store the data as evidence. This is one of the key characteristics of the proposed solution; the system is capable of performing the necessary analysis for detecting events that might indicate volume crime incidents in real time and alert the security personnel and/or the authorities. Despite its advanced characteristics over the existing systems, the end-user solution will be a low cost system, tailored to the needs of the specific user, and will be dynamically reconfigurable in order to be updated with new algorithms and software components. The low cost requirement is achieved by removing the

need for storing recorded data constantly, since the system will know when a relevant event has actually happened. Finally, the fact that the system can be dynamically reconfigurable ensures that the end-user will constantly have an up to date surveillance of their property.

A key element of the proposed system is that unlike most existing systems, in which the intelligence of the system is either embedded in the on-site system's sensors or available only in the central servers, the proposed solution will include intelligence at both sides. This way, when a suspicious event is detected by the local system, the cloud service is informed and receives data from the on-site system in order to perform more in depth analysis, which can allow determining whether there are profound indications of a volume crime event being detected. The latter is also facilitated by having the cloud service alerting the neighbouring sensors to the one that raised the alarm, thus providing more evidence about the suspect(s), i.e. a more clear view of them, or evidence about the escape route.

In summary, the proposed solution, while being innovative, is still very close to the market and mostly applies to related studies showing where the main problem with volume crimes exists.

3.2. Architecture and main components

The architecture design has two main components: (i) local embedded platform, (ii) cloud service. Moreover, the architecture includes an end-user interface and a mobile app for first responders. P-REACT architecture has been designed in a modular way to support, in the future, more powerful embedded systems and algorithms for video and audio analytics. In addition to this, the architecture allows using the appropriate video and audio analytics algorithm(s) for a specific case, which can be downloaded from the cloud to the embedded system improving the flexibility and scalability of the proposed platform. This is achieved by either having the new video or audio analytics algorithm automatically pushed to a neighbouring embedded system when an event trigger is sent to the Video Software as a Service (VSaaS) cloud or when a trigger is sent to the VSaaS cloud and needs further examination at the local point. *Figure 1* depicts the main components of the architecture.

The P-REACT architecture includes modules and components at two different layers, the local embedded platform and the cloud service. At the local embedded platform there are five different modules that are described below.

The Sensor Management module is responsible for management of various sensors connected at the embedded system. More specifically, it can manage

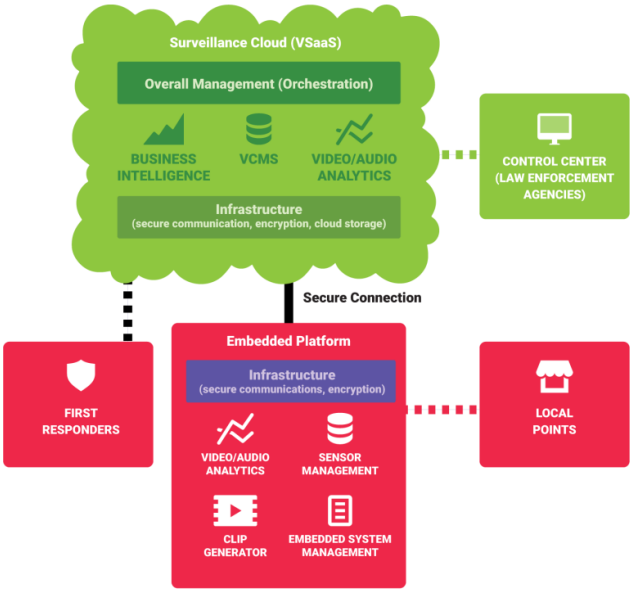

Figure 1. P-REACT High level Architecture

different video cameras (network, USB, etc.), different microphones (network, USB, etc.), depth sensors (Kinect, etc.), infrared and other sensors (smoke detection, etc.). Moreover it enables the connection of a panic button either legacy or software one (e.g. a specific word is specified).

The Video and Audio analytics module is responsible of performing the first step of sensor data analysis. The focus of this module is at the analysis of videos and audio flows coming from the sensors but also to combine the analysis with data from other sensors if they exist. The output of this module is to inform the embedded system manager that an event that might indicate a volume crime has been detected and therefore a trigger alarm can be created.

The Clip generator creates a P-REACT clip that consists of a JavaScript Object Notation (JSON) file, which includes all metadata related with the nature of the content, the analysis results, the detected evidence and the triggered alarm. The clip generator module can be configured or adapted to change some parameters of the clip (e.g. the length of the clip in MB depending on the network type used).

The Embedded System Manager module is the "brain" of the embedded system which retrieves all the information regarding the status of embedded system modules. The most important function of the embedded system management module is that it can modify/adapt automatically the operation of the embedded system module according to the circumstances (e.g. network conditions, another audio/video analysis algorithm needed to be loaded). It also, based on the events coming from analytics

modules, triggers alarms that are sent to the cloud system.

Finally, the Infrastructure and Secure Communications module is mainly responsible for establishing the communication to the cloud using a VPN channel. It is also responsible of encrypting the data that will be sent to the cloud. This module monitors the connection characteristics in order to inform the embedded system management module with the communication status each time.

On the other hand, there are five modules also at the cloud level that we describe below.

The Infrastructure and secure communications module is mainly responsible for establishing the communication to the embedded system using a secure VPN channel. It is also responsible of encrypting the data exchanged between the cloud and the embedded system. This module monitors the connection characteristics in order to inform the cloud overall management module with the communication status each time. Moreover this module is responsible for the Infrastructure as a Service (IaaS) (power, storage, network, memory) status of the cloud infrastructure and receiving the data from embedded systems' sensors creating the appropriate format of the data received for the VCMS.

The Video Content Management System (VCMS) is a Platform as a Service (PaaS), which is responsible for storing the clip objects with the appropriate format to the cloud storage service database in order to be searchable from the end-user or in case the system needs to have access to the content for further analysis.

The Business Logic module is the "brain of P-REACT" VSaaS Cloud, which is responsible for all the intelligent actions that have to be done in order to optimise the whole operation of sensor data analysis. Business logic module determines whether the quality of the results is enough to reach a safe outcome (without false alarms) for a specific event, or to request further analysis that may be needed, combining data from sensors (other embedded systems or additional sensors, except video and audio ones) or information from past events, even information/evidence that comes in real time from first responders. It is also responsible for the organisation of all the deployed embedded systems. After the analysis of an event, business logic module takes the decision of what are the appropriate neighbour embedded systems in order to enable their operation. Moreover, it is the module that decides what is the appropriate analysis algorithm that should be downloaded to the specified embedded systems in order to recognise better and faster an event, or to track the responsible for this event in case of a proven volume crime.

The Video/Audio analytic module contains the actual analysis modules, which are divided into two categories: (i) video analysis and (ii) audio analysis. Video and Audio analytics modules are responsible for performing the second step and in-depth analysis of video and audio data. The focus of this module is at the analysis of videos and audio flows coming from the embedded systems with appropriate algorithm/s. The module pick-ups the flows at the appropriate format as stored from the VCMS module and it applies complicated procedures for the analysis. This procedure includes face and gait recognition in one or multiple flows, object recognition, audio classification from complicated audio streams coming from multiple and different voice channel with high noise level. The output of each of these modules is fed into the business logic module, in order for the latter to make appropriate decisions regarding an event.

The Orchestrator module is responsible for orchestrating all the cloud modules achieving the smooth operation of P-REACT cloud platform. All the modules should inform the overall management system about their actions and they receive feedback from overall management module on what action/s should follow. Overall management module holds all the functions needed for the communication with end-user interface and first responders mobile app and forward the receiving information to the appropriate modules into P-REACT cloud.

Besides, the P-REACT solution includes an end-user interface which is an interface helping the administrator or a single operator of the P-REACT platform to visualise the information regarding the management of the platform or/and the operation performed by the platform. In general, it is the tool that enables the end-user to interact with platform according to the specific end-user's access control rights.

Finally, the P-REACT solution includes a first responder's mobile app which is the application that provides additional information/evidence coming from the first responder deployed on the field to P-REACT platform, helping towards a better tracking and identification of the suspect(s) in the crime scene. This application can also receive notifications from the P-REACT plat-form for additional actions from the first responder side, based on geo-location data.

3.3. Workflow description

There are two different workflows in P-REACT system: (i) dynamic configuration workflow and (ii)

content analysis workflow. System configuration not only encompasses the user assisted configuration to update the performed analysis but also the dynamic configuration of algorithms running in analytics modules at embedded sys-tem level. This dynamic configuration is led by the Business Logic module in the cloud; as the received clips analysis is performed, this module detects the needs of P-REACT system and sends new configuration parameters to the Embedded System Manager in order to adapt light-weight algorithms and their settings to the requirements.

Dynamic configuration workflow: (i) P-REACT user/administrator installs configuration parameters using P-REACT GUI. This configuration affects the Business Logic (BL) module at cloud level and the Embedded System Manager (ESM) at embedded level; (ii) The Business Logic module analyses the specified configuration rules and establishes if additional configuration parameters are required by the embedded system analytic modules; (iii) Once the preloaded configuration is done, the system starts analysing the scene and extracts knowledge from the captured content; (iv) Depending on the detected events/alarms, Business Logic module deduces, based on the analysis results, on one hand, the video and audio analytics algorithms that should be applied at the cloud level, and on the other hand, new algorithms that should be loaded to the ESM, in order to extract more useful information that can enrich P-REACT solution. The low processing capacity of embedded systems does not allow parallelisation of many analysis algorithms; therefore, it is much more valuable to have a dynamic configuration of the ESM for the inclusion of diverse algorithms while processing. This dynamic configuration driven by Business Logic module is one of the main advances of P-REACT.

Figure 2 and *Figure 3* depict the P-REACT content analysis workflow at embedded and cloud levels respectively.

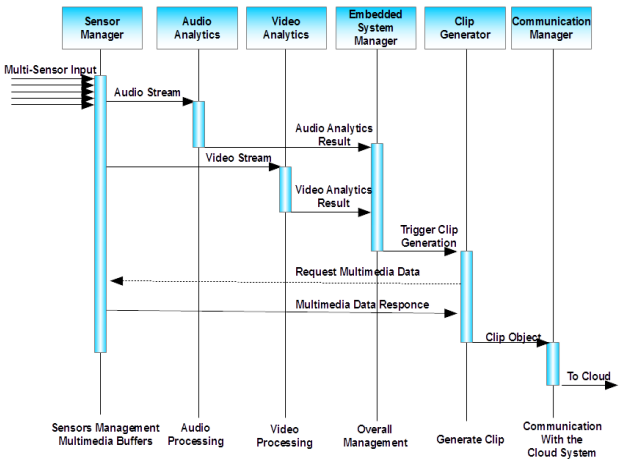

Figure 2. Content analysis workflow overview (embedded level)

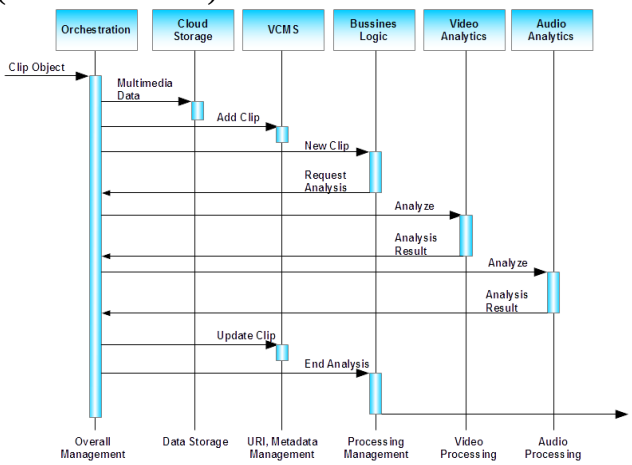

Figure 3. Content analysis workflow overview (cloud level)

4. Preliminary analytics results

As aforementioned, P-REACT platform includes also video analytics solutions (including depth cameras) that are computed both at embedded level and cloud level. Due to the lower processing power of the embedded systems, light algorithms are implemented at embedded side and algorithms needing more computing power at cloud level. Moreover audio analytics will be performed at the embedded system level enhancing video analytics results (audio covers a 360-degree area, enabling a video surveillance system to extend its coverage beyond a camera's field of view). The continuous developed modules are designed in order to fit in most of the embedded systems available in the market. The paragraphs below presents preliminary results on the various analytics algorithms deployed in P-REACT solution for volume crimes detection. The results are based on adaptations on currently available algorithms, which have been modified for volume crimes events detection and can be utilised from an embedded system with limited computing resources. Further analysis may be needed to perform at cloud level, resulting in a distributed computing platform with increased efficiency.

The embedded system used for the preliminary results has the following main technical characteristics:

i. Allwinner A20 dual core Cortex-A7 processor, each core typically running at 1GHz and dual-core Mali 400 GPU
ii. 1GB DDR3 RAM memory

4.1. Video Analytics

Within P-REACT video analytics is used to detect abnormal behaviors such as fighting, chasing, and running. The detection of these events is divided into two detection phases. First, movement is detected based on background detection methods such as VIBE [20]. When motion objects are present in the videos, the system generates evidence which will incur in an alarm only if the other analysis parameters are also considered above a predefined threshold levels. In the following table there are some testing results regarding the embedded platform taking into account the inclusion of the Sensor Manager deployed with Gstreamer. For the comparison between webcam captured files and already saved videos, BEHAVE [3] dataset have been used always using 25 fps videos.

Table 1. Testing results

Considering Sensor Manager		**Reading directly with OpenCV**	
Capturing from Web cam (1600x1200)	Reading from file (640x480)	Capturing from Web cam (1600x1200)	Reading from file (640x480)
CPU: Gstreamer = 48% VIBE=47%	**CPU:** Gstreamer = 52% VIBE=45%	**CPU:** VIBE = 82%	**CPU:** VIBE = 94%
Memory usage 5,8%	**Mem.** 5,7%	**Mem. usage** 5,7%	**Mem.** 5,8%

The video analytics module residing on the cloud platform will receive its input from the various Embedded Systems hooked on the P-REACT platform. This module will rely on depth and shape information to perform gait and face recognition on the corresponding components. At Cloud Level there is a Rule Manager for Event Detection which is in charge of determining between the abnormal behaviours that have been considered in P-REACT. BEHAVE has been used the training and testing of this Rule Manager as it suits with the needs of P-REACT.

Figure 4. Running and chasing detection in BEHAVE dataset videos

The proposed P-REACT platform also utilizes depth sensors in order to be able to detect abnormal events in indoor environments. Depth sensors advance over the regular color or IR cameras since they provide 3D information on the scene and are immune to changes in lighting conditions while maintaining a low cost profile. On the other hand they require more processing power for the image processing, but the market provides low-cost solutions that are suitable for depth image processing.

For the needs of the P-REACT's solution, the depth sensors are being used to enhance the accuracy of abnormal events detection (such as fighting or chasing) in indoor areas, like a small shop or the waiting area of a bus/train station. The current approach utilizes the HON4D descriptor proposed [17]. The proposed descriptor calculates a histogram, capturing the distribution of the surface normal orientation in the 4D spatiotemporal space, thus capturing not only the changes in shape and motion but their correlations as well. This way the descriptor is very robust against noise and occlusions, something very common in real life environments. The authors of [20] propose to use Support Vector Machines (SVM) in order to classify the descriptors. During experiments on a proprietary database collected specifically for the needs of P-REACT project, it was determined that SVM classifiers did not work very efficiently for the type of events the system needed to detect, so Random Forest Trees [4] were utilized. The latter achieved an accuracy of > **85%** in a proprietary database. The aforementioned database was specifically recorded for the needs of P-REACT project and includes ~50 sequences of depth images, depicting fighting, chasing and neutral events.

4.2. Audio Analytics

Audio analytics is performed though the means of classification which aims at identifying to which of a set of categories, a new uncategorised artefact belongs; on the basis of a training set of artefacts whose class/type is known. Within the implemented system, the classification has to be executed on a low-cost embedded system with limited processing and storage power. Therefore only methods where the classification of an event takes minimum time and space could be considered; decision tree classification methods service these requirements very well.

The implemented audio analytics module is capable of identifying key alarming events like screaming, glass breaking and gunshots. The detection of events is accomplished with the C5.0 classification algorithm [14] the selection of this method was the

result of a concrete experiment where several classification methods were called to categorise recorded audio as screams or false alarms. During this experiment bootstrapping with replacement [8] was used along with the 5x2 cross validation f-test technique for minimizing the bias over any sample and producing statistically valid results. The C5.0 reported the best results with respect to classification AUC rates and computational time.

For the classification, the captured audio is transferred to the connected embedded system where the continuous stream is split into blocks. Each block is then processed through the already trained classification tree to be assigned in one of the following classes: scream, gunshot, glass breaking and false (no alarm). The categorization of each block is performed based on a number of extracted features; these are essentially a number of MFCC and filter bank coefficients [10]. The trade-off between processing time and classification error (measured as the misclassification rate) was considered critical and as a result, the parametrisation of the algorithm was determined based on the results (see *Table 2*) of a comprehensive experiment; involved the following parameters: audio sampling frequency, block size (expressed in ms), nfft frame, number of filter bank and MFCC coefficients, misclassification error and average (block) processing time.

Table 2. The results of a comprehensive experiment

Parameter	***Configuration***		
	#1	#2	#3
F_S (kHz)	8	8	8
Block (ms)	140	100	140
FBank	22	22	22
MFCC	13	13	10
Classification Error	1.40%	1.60%	1.60%
Time (ms)	85.74	89.06	65.13

5. Discussion and Conclusion

In this paper we have presented the low cost video surveillance P-REACT solution and the preliminary results of the video, depth and audio analytics that are being developed to detect volume crimes. These preliminary results suggest that the overall design of the solution will be proved valid in order to meet its main goals: (i) to effectively detect volume crime incidents; (ii) to remain a low cost solution; and (iii) to be dynamically re-configurable.

Work remains in evaluating newly available low cost embedded platforms with heterogeneous multiprocessing and separate GPU. These more powerful embedded platforms will open new possibilities for running more resources consuming and video, audio, and depth analytics, even on parallel processing. There is also one very important aspect however that will require further work in the following months, which is to find the right combination and distribution of video, depth, and audio analytics per each of the volume crime use cases between the embedded system and the cloud service in order to minimize the number of false alarms whilst maximizing the detection rate. In the second half of the P-REACT project the overall solution will be fully developed for at least one volume crime use case, anti-social behaviour.

Acknowledgements

The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement number 607881, project titled P-REACT.

References

[1] ACPO, (2011). *Practice Advice on the use of CCTV in criminarl investigations*, London.

[2] Blunsden S.J. & Fisher, R.B. (2005). Pre-fight detection: Classification of fighting situations using heirachical AdaBoost, VISAPP 2009 - *Proc. Fourth Int. Conf. Comput. Vis. Theory Appl.*, vol. 2, pp. 3003–308.

[3] Blunsden, S. & Fisher, R.B. (2010). The BEHAVE video dataset: Ground truthed video for multi-person behavior classification, BMVA, No. 4, 1-12.

[4] Breiman, L., Random, F., Machine Learning, Kluwer Academic Publishers, 2001, doi=10.1023/A:1010933404324,

[5] Burrows, J.N. (1979). *The impact of closed circuit television on crime in the London Underground*, London.

[6] Dadashi, N., Stedmon, W. & Pridmore, T.P. (2013). Semi-automated CCTV surveillance: the effects of system confidence, system accuracy and task complexity on operator vigilance, reliance and workload. *Appl. Ergon.*, vol. 44, no. 5, 730–8.

[7] Ditton, E. & Short, J., (1998). Evaluating Scotland's first town centre CCTV scheme. In *Surveillance, Closed Circuit Television and Social Control*, 55–73.

[8] Efron, B., & Tibshirani, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Statistical science, 54-75.

[9] FORMAT, "FORMAT research.". Available: https://www.formatresearch.com/eng/home/.
[10] Ganchev, T., Fakotakis, N., & Kokkinakis, G. (2005). Comparative evaluation of various MFCC implementations on the speaker verification task. *Proc. of the SPECOM*, Vol. 1, 191-194.
[11] Grant, D. & Williams, D. (2004). The importance of perceiving social contexts when predicting crime and antisocial behaviour in CCTV images. *J. Pers. Soc. Psychol*. vol. 16, no. 2, 307–322.
[12] HomeOffice, (2004). Defining and measuring anti-social behaviour, London.
[13] Krishan, S.K. & Yoshiura, N. (2010). Restrained surveillance towards community benefit. *Procedia - Soc. Behav. Sci*., vol. 2, no. 1, 28–35.
[14] Kuhn, M., & Johnson, K. (2013). Applied predictive modeling, New York: Springer.
[15] Little S., Connor, N.E.O., Smeaton, A.F., Clawson, K., Wang, H. & Nieto, M. (2013). An Information Retrieval Approach to Identifying Infrequent Events in Surveillance Video. *ACM International Conference on Multimedia Retrieval*, 16–19.
[16] Macnish, K., (2012). Unblinking eyes: the ethics of automating surveillance. *Ethics Inf. Technol*., vol. 14, no. 2, 151–167.
[17] Oreifej, O. & Liu, Z. (2013). HON4D: Histogram of Oriented 4D Normals for Activity Recognition from Depth Sequences, CVPR 2013
[18] P-REACT related projects, PREACT, 2014. [Online]. Available: http://P-REACT.eu/links/. [Accessed: 14-Nov-2014].
[19] Rossy, Q., Ioset, S. Dessimoz, D. & Ribaux, O. (2013). Integrating forensic information in a crime intelligence database. *Forensic Sci. Int*., vol. 230, no. 1–3, 137–46.
[20] Van Droogenbroeck, M. & Barnich, O. (2014). ViBe: A Disruptive Method for Background Subtraction. In T. Bouwmans, F. Porikli, B. Hoferlin, and A. Vacavant, editors, Background Modeling and Foreground Detection for Video Surveillance, chapter 7, pages 7.1-7.23. Chapman and Hall/CRC.