

Application of foreground object patterns analysis for event detection in an innovative video surveillance system

Dariusz Frejlichowski · Katarzyna Gościewska ·
Paweł Forczmański · Radosław Hofman

Received: 31 January 2014 / Accepted: 22 July 2014 / Published online: 15 August 2014
© The Author(s) 2014. This article is published with open access at Springerlink.com

Abstract SmartMonitor is an innovative surveillance system based on video content analysis. It is a modular solution that can work in several predefined scenarios mainly concerned with home/surrounding protection against unauthorized intrusion, supervision over ill person and crime detection. Each scenario is associated with several actions and conditions, which imply the utilization of algorithms with various input parameters. In this paper, focus is put on the analysis of foreground object patterns for the purposes of event recognition, as well as the experimental investigation of selected methods and algorithms which were developed and employed for the SmartMonitor system prototype. The prototype performs three main tasks: detection and localization of foreground regions using adaptive background modelling based on Gaussian Mixture Models, candidate objects extraction and classification using Haar and HOG descriptors, and tracking using Mean-Shift algorithm. The main goal of the work described here is to match system parameters with each scenario to provide the highest effectiveness and to decrease the number of false alarms.

Keywords SmartMonitor · Visual surveillance system · Video content analysis · Foreground object pattern · Pattern analysis · Event detection

D. Frejlichowski (✉) · K. Gościewska · P. Forczmański
Faculty of Computer Science and Information Technology, West
Pomeranian University of Technology, Żołnierska 52,
71-210 Szczecin, Poland
e-mail: dfrejlichowski@wi.zut.edu.pl

K. Gościewska · R. Hofman
Smart Monitor sp. z o.o., Niemierzyńska 17a, 71-441 Szczecin,
Poland

1 Introduction

Video surveillance systems have recently become more autonomic and functional. The advances in video content analysis (VCA) algorithms have undoubtedly contributed to the application of such systems in new areas and demanding locations. This has also resulted in lowering the demand for operators of monitoring systems, at the same time has facilitated the work of those who have to handle a large number of cameras and peripherals combined into a single system. Intelligent monitoring systems with VCA functionality are implemented mainly for monitoring wide areas and public buildings, and the infrastructure utilized for this purpose is specific and expensive. Despite this, there are still people who want to ensure their own safety, protect their houses or small businesses and surrounding areas. For this reason, the demand for systems that utilize common electronic devices, work without human control and are affordable for individuals arises. This causes that surveillance systems have to operate under different conditions; however, the concept of 'universality' cannot be applied here directly—there are no systems that work equally well in various environments and circumstances using the same parameters. Since the solution cannot be universal, it should offer a possibility to adjust it to enable customisation and better adaptation. In response to these needs, SmartMonitor was developed as a customizable visual surveillance system for personal use.

The SmartMonitor system is designed to work under several independent scenarios that provide homes and their surroundings protection against unauthorized intrusion, allow for supervision of people who are ill and detect suspicious behaviour. Each scenario is characterized by a group of performed actions and is activated when certain conditions are fulfilled, for instance a

movement is detected in a protected area or there is no movement for a specific period of time. In these cases, it is crucial to properly configure associated thresholds to avoid multiple false alarms. The most important parameters are associated with the duration of the actions performed by the object and an object's physical (two-dimensional) features. In the paper, they are investigated to find the most appropriate parameter values for the event detection in each scenario.

SmartMonitor is an innovative surveillance system that combines the advantages of closed-circuit television systems (CCTV) and visual content analysis algorithms. It gives the possibility to set individual safety rules and adjust the system's sensitivity degree to the actual requirements. This allows users to decide how the system should respond and what types of objects or events should it detect. The application of VCA algorithms and feature-based methods allows for eliminating a large part of human involvement which is needed only during initial calibration. The visual analysis algorithms are integrated into six main modules which are responsible for background modelling, object tracking, artefacts removal, object classification, event detection and system response. Background modelling is performed using Gaussian Mixture Models and two background models based on the intensity component Y (YIQ colour model) and hue component H (HSV colour model). The use of different colour information helps to detect and remove artefacts at a later stage. For object tracking, the Mean-Shift algorithm is used. Objects are classified using the Haar and HOG descriptors. Event detection involves the analysis of changes in an object during a period of time. The system's response is subject to logical rules and determined by selected threshold values. The simplified diagram of system modules is provided in Fig. 1. All of these algorithms together with some additional operations resulted in forming a new approach applied in the SmartMonitor system, that is experimentally tested in the context of the analysis of foreground object patterns. The analysis ultimately aims at finding the best system working parameters for each scenario.

Due to the fact that the SmartMonitor system is a combination of security and surveillance solutions with different degrees of advancement, it can be compared to alarm systems based on sensors, small CCTV, home automation, video surveillance and advanced systems based on video content analysis algorithms. To provide a background of existing solutions, we will focus on some examples and indicate the differences. The solutions like ADT Pulse [1] or vivint [2] generate alerts based on various sensors, in turn ZoneMinder [3] is intended for video surveillance with motion detection. The key features of these systems include: a use of wireless camera, remote access and some home automation functionalities.

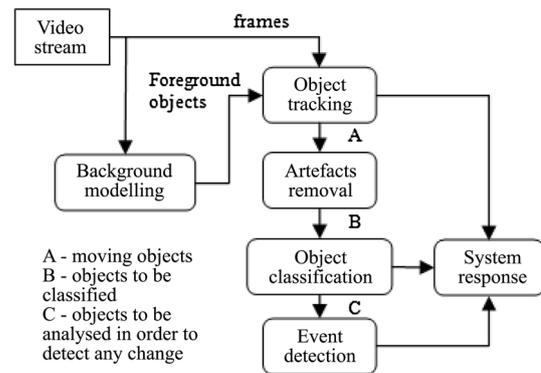


Fig. 1 Simplified diagram of the system modules

They have features similar to our system, but require human intervention and simultaneously do not provide an automatic differentiation of dangerous situations as well as an automatic response. AgentVi [4] provides solution for video analysis in large installations that is based on open architecture approach. The software is distributed between an edge device and a server. Another VCA-based industry solution is IVA 5.60 Intelligent Video Analysis by Bosch [5]. It is a guard assistant system based on intelligent video analysis which detects, tracks and analyses moving objects. The analytics is built into cameras and encoders, which increases the cost of installation. Both AgentVi and IVA 5.60 offer advanced video analysis, but are not intended for home use. They differ in respect of architecture, where SmartMonitor is a centralized solution and does not process any data on edge devices. Moreover, the mentioned systems do not enable the use of controlled devices. The more advanced solutions are also present on the market, such as AISight by BRS Labs [6] for behaviour recognition. This system is able to autonomously build a base of knowledge and generate real-time alerts to support the security team in various industry areas. It analyses traffic, detects perimeter intrusion, secures facilities, generates transit and manufacturing alerts, and identifies various events and activities with respect to an usual time of a day. Compared to AISight, the SmartMonitor system is not capable of self-learning. However, for the purposes of SmartMonitor, such functionality is redundant and might increase the final price. Moreover, this solution does not enable the use of controlled devices.

The rest of the paper is organized as follows. The second section provides a description of system working scenarios. The third section describes the algorithms that were applied in building the system prototype. The fourth section includes experimental conditions, and the fifth one discusses the results of the experiments and their explanation. The last section concludes the paper.

Fig. 2 Two sample frames presenting running objects in the garden—basic human/not-human classification excludes the alarm activation when dogs are detected (scenario A) [7]



Fig. 3 Sample frames presenting the simulation of an ill person fainting (scenario B) [7]



2 Brief description of the system scenarios

In the SmartMonitor system, a scenario is a combination of intended or predicted situations which are associated with various monitored scenes. This in turn implies different groups of performed actions among which the most important are movement detection, object tracking, object classification, object size limitation and event detection. All these actions concern activity and appearance of objects, especially human silhouettes, and for this reason an analysis of object patterns is a crucial task in a video surveillance system. Other actions relate to environmental conditions, for example weather conditions or the size of the monitored region. The SmartMonitor system is able to operate under three basic scenarios—home/surrounding protection against unauthorized intrusion (scenario A), supervision over ill persons (scenario B) and crime detection (scenario C). Scenario A focuses on movement detection. However, in contrast to the traditional monitoring systems, moving objects are classified before the alarm is triggered. As a result, only particular objects are taken into account, for instance the objects larger than a specified size or characterized by certain features. System working in scenario A can operate outside and inside the buildings, and should also be active at night. Due to various environments, weather conditions and the influence of light have to be considered. Sudden changes in lighting could affect image colours and cause the appearance of artefacts on foreground images. Fig. 2 contains sample frames typical for scenario A.

Scenario B focuses on incident detection, especially the recognition of faints and falls. These events result in changes in the object's shape and trajectory, such as changes in proportions or the lack of movement that lasts

longer than a pre-assumed time. It is important that the object is correctly localized and extracted without artefacts. Figure 3 contains sample frames showing a person who appears in the scene, falls after some time and remains lying on the ground.

The system working in scenario C operates inside the building and is intended mainly for offices, shops and other small enterprises. The system aims to detect events and for instance it reacts to raising one's hands or to unusual trajectories. It would prove helpful when a threatened person could not trigger the alarm, as the system detects suspicious behaviour and is able to send message and images to appropriate services. Figure 4 contains two sample frames with a person with hands raised up. Different lighting conditions can cause problems during background modelling on the basis of various colour components.

3 Methods and algorithms developed and employed for the system prototype

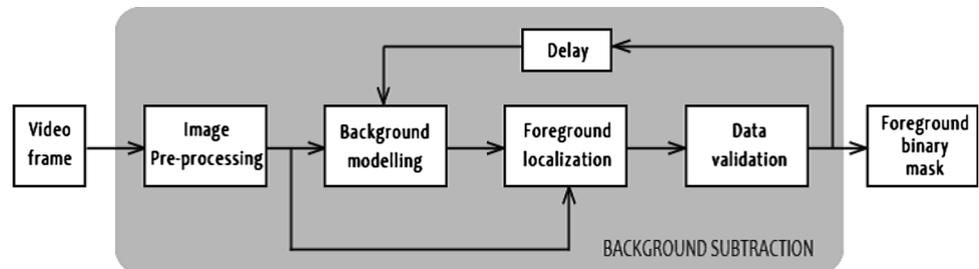
Several existing approaches have been investigated and the most effective ones were modified and adopted to create the best solution for the system. Particular attention was paid to real-time methods with respect to their computation time. Moreover, certain assumptions have been made to ensure proper operation of the surveillance system and are given below:

- A camera has to be placed in a fixed location and observe the same area in a continuous manner;
- Exposure parameters have to remain unchanged for a long period of time;

Fig. 4 Sample frames presenting the simulation of a crime scene (scenario C) [7]



Fig. 5 General scheme of applied background subtraction process [13]



- Frame resolution has to enable the extraction of single objects;
- Camera noise and weather conditions may not cause problems during the extraction of foreground objects;
- Individual frames of the video stream are processed;
- Information about future frames is not included.

A system prototype performs three main tasks: detection and localization of foreground regions using adaptive background modelling based on Gaussian Mixture Models (GMM), candidate objects extraction and classification using Haar and HOG descriptors, and tracking using Mean-Shift algorithm. Some additional algorithms were applied to connect all modules and give better results.

3.1 Background modelling and foreground extraction

The first and most important task was to build a reliable background model, which had to be robust to any environmental changes in the scene and, at the same time, sensitive enough to detect all objects of interest (OOI) for the system, where each OOI is a coherent region larger than a specified size. Background modelling is one of the stages of background subtraction process (see Fig. 5). Due to the fact that real scenes are prone to variability in time, the GMM was chosen to identify moving objects. It is an adaptive background modelling technique that can cope with long-term changes and the influence of light. GMM models each pixel as a mixture of multiple Gaussian distributions and is able to use various colour information as input. The model is parametric and can be adaptively updated with every consecutive frame without the necessity to keep a large video frames buffer [8, 9]. Other approaches to background modelling were investigated, i.e.

background models that are static [10] or averaged in time [11]. These approaches are relatively simpler than GMM, however they do not provide an accurate update of the background image when scene content changes over the time and may cause the appearance of artefacts. The problem of foreground extraction was also investigated in [12].

Because GMM can operate on various colour information [14], three different colour models were investigated in [7] to find the most appropriate one—the intensity model (Y component of the YIQ colour scheme), chrominance model (hue component of the HSV colour scheme) and RGB model. It was expected that the model would enable accurate detection of moving objects and that the extracted regions would not be affected by false detections. However, if any artefacts appeared, it was desired that they would be easy to distinguish from the actual OOIs in the data validation step. Some foreground images extracted by subtracting background model from the currently processed frames were visually evaluated, but unfortunately it was not easy to determine which background model is most effective. The RGB model introduced the lowest number of false detections, but the shadow was detected. In turn, in the chrominance model the shadow areas were not marked as a foreground region; however, some parts of the OOI's region were not detected appropriately (see Fig. 6). These problems stem from the fact that the GMM algorithm does not distinguish between moving objects and moving shadows. To solve this problem, the authors of [15] proposed an improved GMM that reinvestigates update equation. Another solution in this case is the utilization of the colour space that can separate chromatic and intensity components. The authors of [16] investigated the usage of

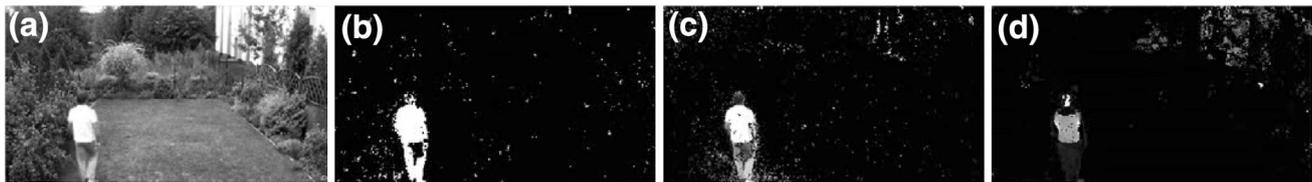


Fig. 6 Foreground extraction: a sample input image (a) and the results of foreground extraction using three models—RGB (b), intensity (c) and chrominance (d) [13]

H component of HSV colour scheme, which corresponds more closely to human perception. It was experimentally found that shadow darkens the region while the hue does not vary too much [16]. The experiments described in [7] confirmed this conclusion. Considering that localized foregrounds have to be further validated, the algorithm for artefacts removal has to enable utilization of the data obtained in the foreground localization step. Since there are various types of false detections, the usage of only one foreground image decreases the possibility of distinguishing false objects from actual OOIs. Therefore, as already mentioned, the combination of two background models, namely intensity and chrominance (hue) models, was selected for the false detection removal process and investigated in [13].

3.2 Types of false detections and the artefacts removal process

Depending on the environment and the utilized colour model, false detections can take various forms—from large coherent regions to single isolated pixels. Artefacts occur for several reasons, e.g. sudden illumination changes, shadows of moving objects, background movement and background initialization in the presence of moving objects [17]. Short and sudden changes in illumination may appear due to turning the light on and off or the sun shining through the clouds. It causes the background colours to change, leading to an increase in the difference between the model and the current frame. Moving shadows are usually detected when the intensity model is used and very often their regions are connected with the actual OOI's region. Hence, a shadow may be mistakenly classified as a foreground region. Background movement can be defined as relocation of part of the background caused, for example, by movements of the grass and leaves in the wind, and resulting in a high level of noise in the foreground areas. As building the background model usually includes the use of the first captured frame, the selected image cannot contain any moving objects—if it does, they are incorrectly incorporated into the background image and partially occlude it.

According to the previous paragraph, three types of false detections have to be eliminated, i.e. shadow areas, noisy

regions and background occlusion caused by objects present during initialization. The last one is solved by using an image with random values for algorithm initialization and adapting it with several first frames. Noisy areas are removed by means of morphological erosion and dilation in each of the two foreground images previously subjected to thresholding. In turn, shadow elimination requires the use of both foregrounds that are multiplied using an entry-wise product. As a result, the foreground binary mask is obtained and only objects larger than specified are further considered. The process of false detection removal is depicted in Fig. 7.

3.3 Object classification

During the classification stage each object is labelled as either human or not. The Haar and Histogram of Oriented Gradients (HOG) descriptors are used for this purpose. The reason for applying such a classification results from the necessity to discard all moving objects not present in the area of interest of the system, and simultaneously to accelerate further calculations. To obtain the HOG representation, firstly the gamma and colour of an input image are normalized. Next, oriented gradients are calculated using various directional filters. In the next step, the image is divided into cells and frequencies of oriented gradients are calculated for each cell. The frequencies are presented on histograms. Subsequently, cells are grouped into larger overlapping blocks which can be square or rectangular (called R-HOG), or located in the polar-logarithmic coordinate system (C-HOG). The final representation is obtained through concatenation of the oriented histograms in particular cells [18]. Figure 8 contains exemplary results of the experiment utilizing the HOG descriptor—a sample frame with a chosen template (left column), two frames (middle column) from the same video sequence which were scanned across to find matching regions and depth maps (right column), where the darker the colour the greater is the similarity of the region to the template.

The second classifier is based on Haar-like features [20], which are simple features combined into a cascade. The AdaBoost machine learning technique is used to select the most appropriate Haar features and set correct threshold values. During classification using the Haar-like features

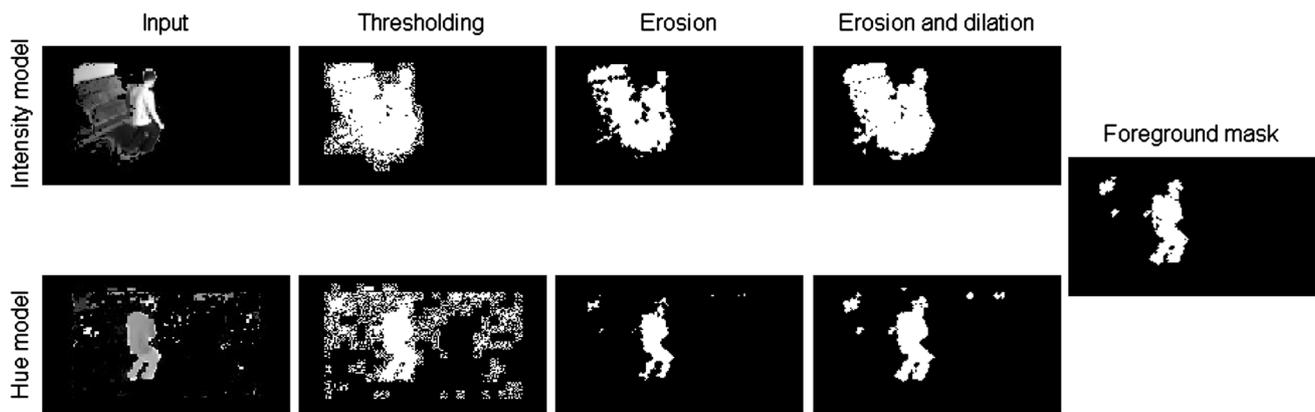


Fig. 7 The artefact removal process—exemplary images obtained at each stage



Fig. 8 Results of the experiment utilizing the HOG descriptor with a fixed template size [19]

cascade, subsequent object's features are calculated only when the answer of the previous feature is consistent with the learned value. Otherwise the object is rejected. The cascade is designed such as to reject the negative objects at the earliest stage of recognition [21].

3.4 Object tracking

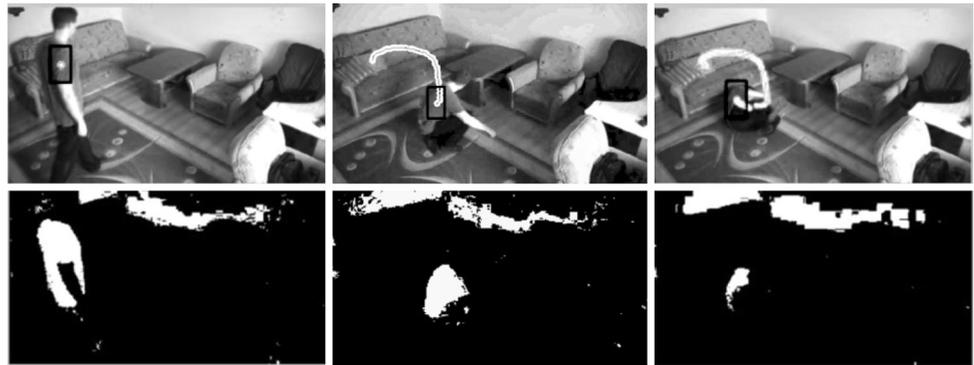
Since classification is not always performed, tracking module becomes active each time a moving object is detected. In automated surveillance, tracking aims to estimate moving object's trajectory and detect suspicious activities or unlikely events. In the process of tracking, a tracker assigns labels to the tracked object in every consecutive video frame [22]. Various tracking techniques have been tested, e.g. the Kalman filter [23], but ultimately the Mean-Shift algorithm was chosen, because it increases the continuity of tracking. The Mean-Shift is an iterative, simple and appearance-based method using features

defined by histograms, such as colour or detected edges. In the first step, the region including a tracked object (or a part of it, further called a template) is selected and the object's features are calculated. Then, for each processed frame, a region that is most similar to the template is found [24, 25]. Figure 9 presents three sample frames from the tracking process using a fixed template size (first row) and their corresponding binary masks in the HSV colour scheme (second row). The white masked regions indicate those regions that are similar to the template, a dark rectangle determines the template and the light points within the rectangle form the object's trajectory.

4 The experimental conditions

In the previous section, the main algorithms and methods developed and employed for the system prototype have been briefly presented, namely background modelling,

Fig. 9 Results of the experiment utilizing the Mean-Shift algorithm [19]



artefacts removal process, tracking method and human silhouettes classifiers. The reasons for selecting the described solutions were also provided along with examples of other approaches that were initially taken into account. Some experimental results of employing the individual solutions have been provided. The experiments proved the accuracy and efficiency of single approaches, however their fusion into prototype software has to be investigated. This section contains the explanation of conditions of the experiments investigating combined algorithmic approaches in the context of object pattern analysis and event recognition for the determination of appropriate system working parameters associated with each scenario.

The main goal of the experiments was to explore the algorithms selected to create a prototype software in real surveillance conditions. The experiments concerned three situations: surveillance of a protected area (protection against unauthorized intrusion; scenario A), faint/fall detection during supervision of ill person (scenario B) and attack detection (scenario C). The occurrence of a particular event activates the alarm when certain conditions are met, i.e. when threshold values of object-related parameters are exceeded. Thresholds determine the sensitivity level of the system and are directly associated with objects' spatial and temporal features. Three parameters were initially proposed to be verified during the experiments:

- P , defines the proportion of an object's bounding box;
- K , defines the maximum number of frames, that is the time in which a person stays in the protected area or does not move;
- T , defines the number of frames, that is the time in which the change in proportion, if other conditions are met, causes the activation of the alarm.

A set of test video sequences that simulate events corresponding to the system scenarios was prepared. The scenes were recorded inside as well as outside the buildings and contained various types of moving objects. The video sequences have been compressed using different

techniques and have various spatial resolutions. During the experiments, input files were retrieved on a sequential read, i.e. at a particular moment only an individual, currently processed frame was available. Each experiment consisted of many steps, but all of them were related to the detection and analysis of human silhouettes. The main and the most important steps are presented in a simplified form below:

- Step 1. Set initial parameters and threshold values.
- Step 2. Open input file (video sequence).
- Step 3. Build a background model.
- Step 4. Retrieve current video frame.
- Step 5. Localize foreground areas in each processed frame.
- Step 6. Perform Haar and HOG classification for each detected object.

Step 7. If the classification step gives a positive result, go to Step 8. Otherwise perform Mean-Shift tracking for the recently detected object.

Step 8. Check thresholds predefined for each scenario:

Step 8.1. For scenario A, check object's location—if an object remains in the protected area longer than K frames, then start the alarm;

Step 8.2. For scenario B, check object's position—if object's position does not change for more than K frames and object's proportions do not change to larger than P over T frames, then start the alarm;

Step 8.3. For scenario C, check object's location and proportions—if object's proportions changed and exceed P , then start the alarm.

Step 9. If it is not the end of the video sequence, go to step 4. Otherwise terminate the processing.

To properly simulate the operation of a real surveillance system, some initial parameters (Table 1) and threshold values were established. Assuming that the camera captures 15 images per second, as a compromise between the speed of frame acquisition and processing, $K = 15$ in scenario A enables that an object which accidentally appears in the protected area for less than a second will not trigger the alarm. In scenario B, $K = 50$ enables to trigger the alarm when a person does not move for more than three

Table 1 Initial system parameters.

Parameter	Value
Frame spatial resolution	640 × 360 pixels or less
Scaling factor	0.5–320 × 180 pixel frames are processed
Learning mode for background model	40 frames without any movements are used for learning (the model calibrates within 2–3 s)
Morphological operations	Median filter and double erosion using foreground binary mask
The size of the area subjected to further classification	Minimum 60 pixels
Proportions of object’s bounding box that is classified as a human	2:1 (height to width)
Minimum object’s size for the Haar classifier	24 × 12 pixels
Minimum object’s size for the HOG classifier	96 × 48 pixels
The area on which the tracked object is searched for using the Mean-Shift algorithm	Half of the minimum object’s size (width, height)

seconds. In scenario C, $K = 15$ refers to the time in which an observed person is stationary at least for a second. Parameter K can be adjusted to the specific needs of a user

and system working conditions. Parameter P , which refers to a ratio of an objects bounding box, has also been established during observations. In scenario C, it is equal to 0.6. In scenario B, $P = 0.7$ and is related to parameter $T = 15$, and the both enable the detection of a lying person whose bounding box remains unchanged for at least a second.

5 Practical verification of the developed approach

In this section, the experimental results of application of the SmartMonitor system prototype are presented. Several experiments were carried out to investigate the effectiveness of the combination of algorithmic approaches and the accuracy of object pattern analysis for the needs of event recognition related to system scenarios and alarm activation conditions. Experimental results are provided in two ways: as figures and descriptions of the object pattern analysis, object-related parameters and their thresholds. Each figure contains: a sample frame before an alarm activation, a sample frame after the alarm activation and three graphs—object trajectory as XY position, aspect ratio and area of an object’s bounding box.

In case of scenario A, the most important element is the position of the centroid of the detected object, as well as its

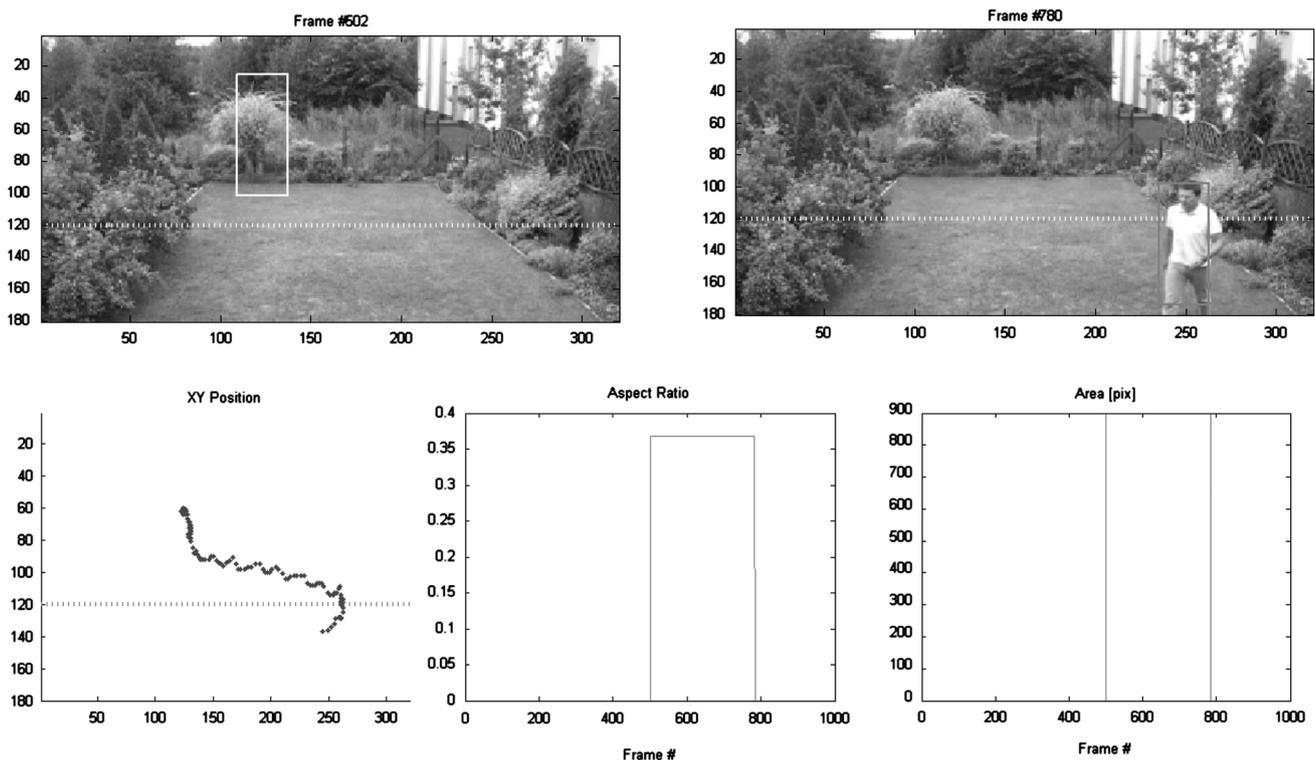


Fig. 10 Sample frames presenting the outdoor scene (scenario A). The two upper frames are (respectively): movement detected without triggering the alarm, and tracked object crossing a virtual line (alarm

was triggered). The lower plots present: tracked object position in the image plane, aspect ratio of the bounding box and the area in pixels

dimensions. Small objects are rejected at the stage of verification; hence the objects with assumed size are taken into consideration. The position of an object (probable human being) is tracked. The virtual line in the image plane defined an area that is protected. Crossing the line triggers an alarm. To eliminate false alarms, the centroid of object was tracked. As can be seen in Fig. 10, the tracked person crosses the line in frame no. 780, which triggers alarm. The next example for scenario A is shown in Fig. 11. As can be seen, the tracked person crosses the line in frame no. 2689, which triggers the alarm.

In case of scenario B, the most important element is the position of the centroid of the detected object, as well as its aspect ratio. Again, the position of an object (probable human being) is tracked. It was assumed that a standing persons' bounding box's aspect ratio is close to 0.5 (the ratio of width to height). If a person falls down, the aspect ratio of a bounding box changes rapidly and exceeds 0.5. It is calculated over several frames (depending on the frame rate, it may represent 1–2 s). Another rule is a change of the centroid position. If it does not change significantly over the same period, then the alarm can be triggered. As can be seen in Fig. 12, the tracked person falls down around frame no. 500, which triggers the alarm.

Scenario C is very similar to the previous one, however, the changes in position of the centroid and the changes in aspect ratio of the bounding box can occur less rapidly.

Again, the position of an object (probable human being) is tracked. We assume that a standing persons' bounding box's aspect ratio is close to 0.5 (width to height values). If a person raises his/her hands, the aspect ratio of a bounding box changes and it exceeds 0.5. We calculate it over several frames (depending on the frame rate, it may represent 2–4 seconds). Another rule is the change of the centroid position. If it does not change significantly over the same period, then we can trigger the alarm. As can be seen in Fig. 13, the tracked person raises his hands around frame no. 550, which triggers the alarm.

6 Summary and conclusions

The main goal of the paper was to provide experimental results of the algorithms prepared for the prototype SmartMonitor software. SmartMonitor is an innovative surveillance system based on image analysis that was created to ensure protection of individual users and their properties in small areas. The system enables the user to set individual safety rules, which in turn determine the degree of system's sensitivity. Human interaction is only required during calibration. The system is now prepared to be placed on the market.

To sum up, in the previous section some experimental results of investigating algorithmic approaches developed

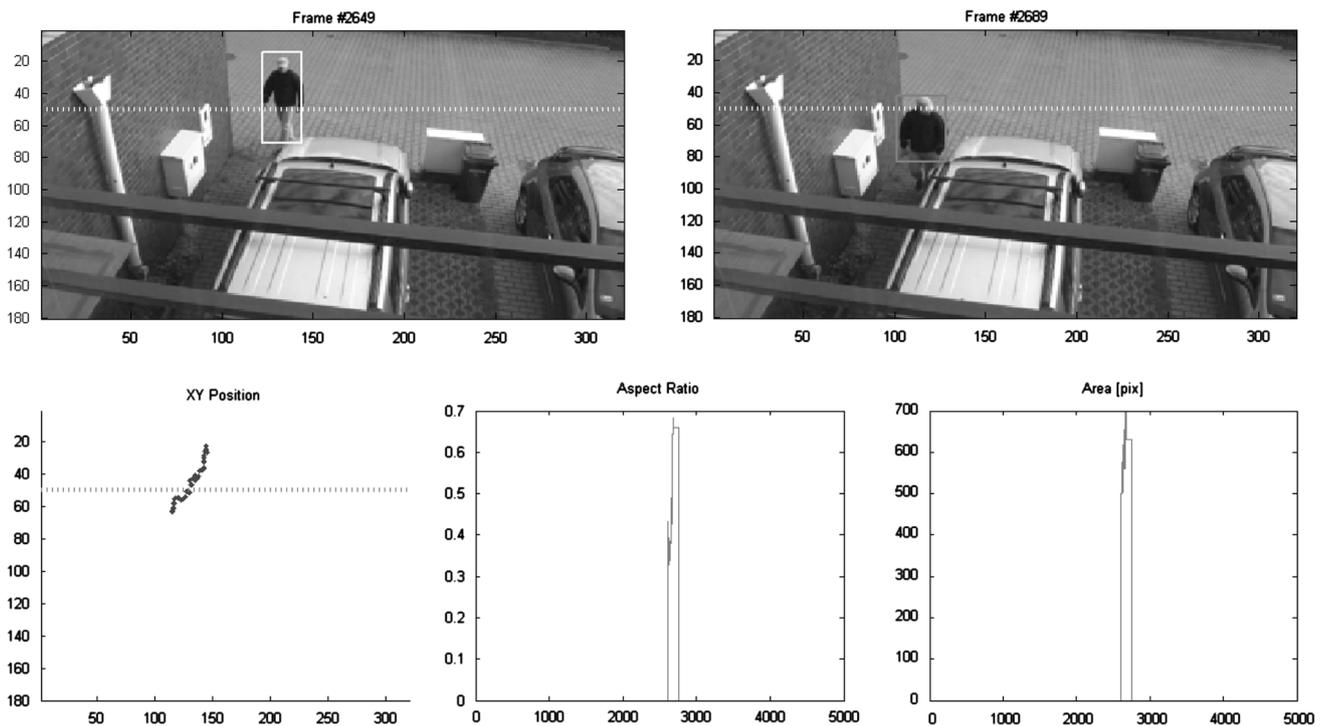


Fig. 11 Sample frames presenting the outdoor scene (scenario A). The two upper frames are (respectively): movement detected without triggering alarm, and tracked object crossing a virtual line (alarm was

triggered). The lower plots present: tracked object position in the image plane, aspect ratio of the bounding box and the area in pixels

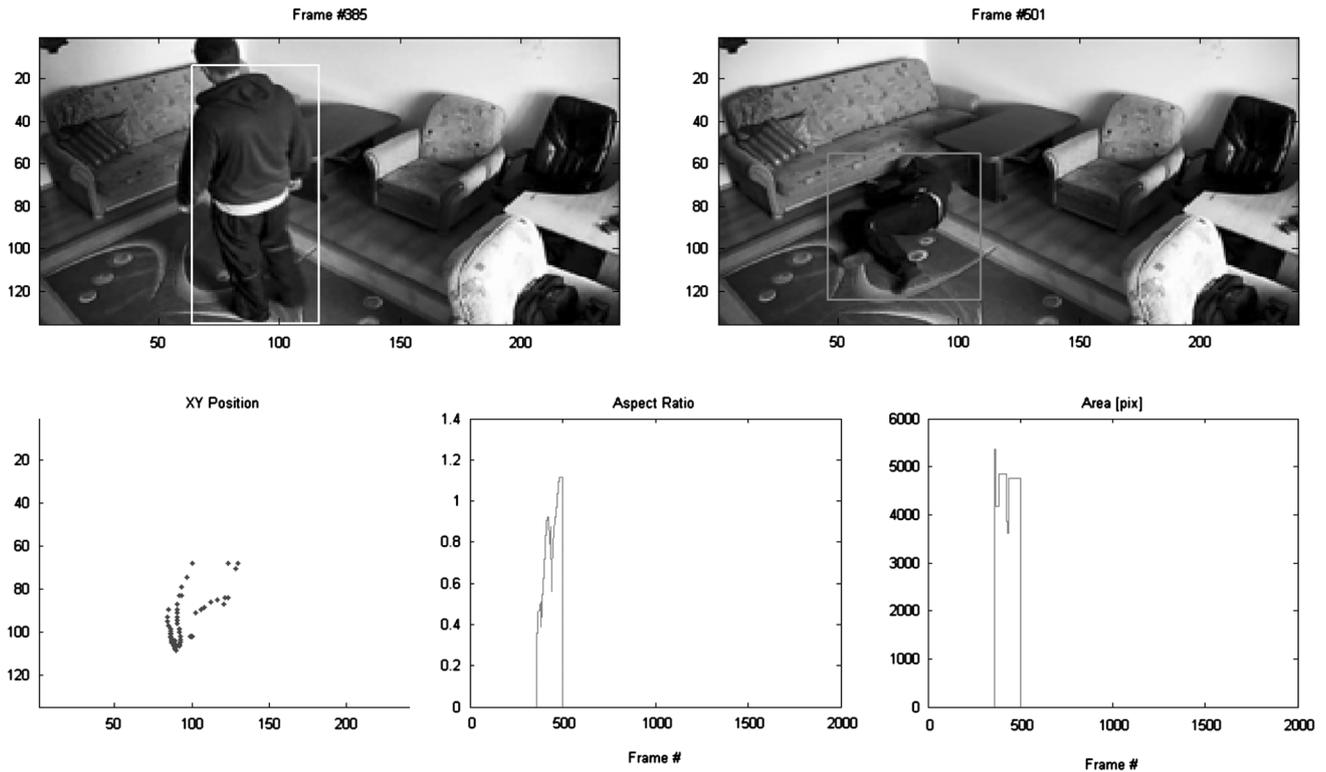


Fig. 12 Sample frames presenting the indoor scene (scenario B, supervision over ill person). The *two upper frames* are (respectively): movement detected without triggered alarm, and tracked object

changing its proportions (alarm was triggered as an effect of falling down). The *lower plots* present: tracked object position in the image plane, aspect ratio of the bounding box and the area in pixels

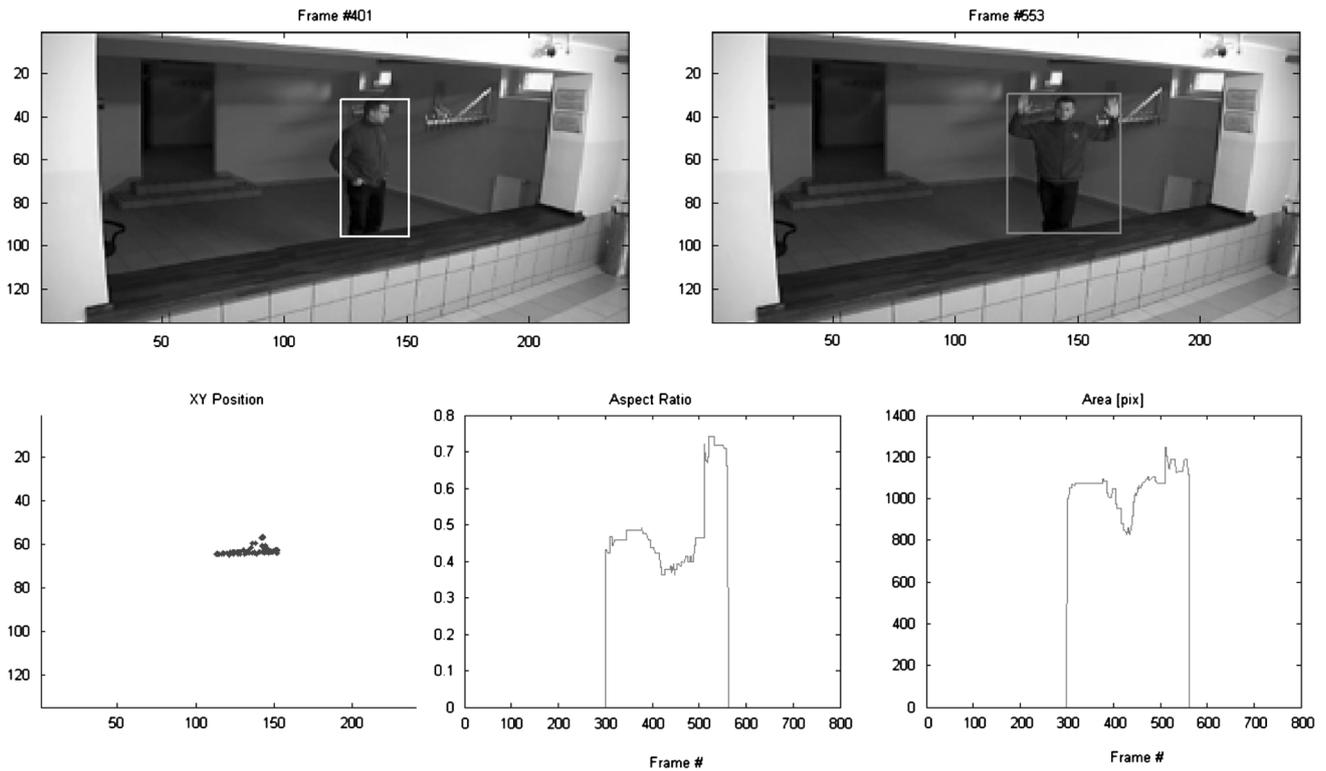


Fig. 13 Sample frames presenting an indoor scene (Scenario C, crime protection). The *two upper frames* are (respectively): movement detected without triggered alarm, and tracked object changing its

proportions (alarm was triggered as an effect of raised hands). The *lower plots* present: tracked object position in the image plane, aspect ratio of the bounding box and the area in pixels

Table 2 Threshold values of parameters determining alarm activation in each scenario.

Scenario	Parameter	Threshold values
A	K	Minimum 15 and maximum 45 frames, which represents 1–2 s
B	K	Minimum 50 and maximum 150 frames, which represents 2–10 s for 15 fps
B	P	Minimum 0.7 and maximum 1.2
B	T	The parameter turned out to be unnecessary
C	K	Minimum 30 and maximum 60 (in order to eliminate false alarms)
C	P	Minimum 0.5

for the prototype SmartMonitor system were presented. The solution constitutes a combination of individual methods and algorithms, namely background modelling based on adaptive Gaussian Mixture Models, morphological operators for artefacts removal, HOG and Haar human silhouettes classifiers and Mean-Shift tracking. Object position and change in proportion were the main features taken into account in the object analysis. The most important conclusions are associated with the pre-assumed object-related parameters and threshold values that were verified during the experiments and are provided in Table 2. Again, it should be emphasized that the effectiveness of the algorithms and the accuracy of object patterns analysis influence threshold values of parameters which determine the time when an alarm has to be triggered in each system working scenario.

System prototype consists of three key modules which are background modelling using adaptive Gaussian Mixture Models, object classification using the Haar and HOG classifiers, and tracking using Mean-Shift algorithm. The proposed combination of algorithms proved to be effective and appropriate for the system. The experiments helped to determine suitable threshold values of the parameters responsible for triggering the alarms in three various situations corresponding to system working scenarios. The most important task was to analyse the patterns of moving objects, especially human silhouettes, and their features. It turned out that the ratio of an object's bounding box and the time in which a person remains in the protected area or does not move constitute crucial parameters for the recognition of specific events.

Acknowledgments The project 'Innovative security system based on image analysis —“SmartMonitor” prototype construction' (original title: Budowa prototypu innowacyjnego systemu bezpieczeństwa opartego o analize obrazu—“SmartMonitor”) is the project co-founded by European Union (project number PL: UDA-POIG.01.04.00-32-008/10-02, Value: 9.996.604 PLN, EU contribution: 5.848.560 PLN, realization period: 07.2011-04.2013). European

Funds—for the development of innovative economy (Fundusze Europejskie—dla rozwoju innowacyjnej gospodarki).

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- ADT (2014) Webpage <http://www.adt.com/video-surveillance> Accessed 07 May 2014
- Vivint (2014) Webpage <http://www.vivint.com/en/solutions/> Accessed 07 May 2014
- ZoneMinder (2014) Online Documentation <http://www.zoneminder.com/documentation> Accessed 07 May 2014
- AgentVi (2014) Webpage http://www.agentvi.com/61-Products-62-Vi_System Accessed 07 May 2014
- Bosch (2014) Webpage http://us.boschsecurity.com/us_product/03_solutions_2/solutions Accessed 07 May 2014
- BRS Labs (2014) Webpage <http://www.brslabs.com/> Accessed 07 May 2014
- Frejlichowski D, Gościewska K, Forczmański P, Nowosielski A, Hofman R (2013) 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
- Stauffer C, Grimson WEL (1999) Adaptive background mixture models for real-time tracking. IEEE Computer Society Conference on computer vision and pattern recognition 2
- Zivkovic Z (2004) Improved adaptive Gaussian mixture model for background subtraction. In: Proceedings of the 17th International Conference on pattern recognition, vol 2, pp 28–31
- Gurwicz Y, Yehezkel R, Lachover B (2011) Multiclass object classification for realtime video surveillance systems. Pattern Recognit Lett 32:805–815
- Frejlichowski D (2008) Automatic localisation of moving vehicles in image sequences using morphological operations. In: Proceedings of the 1st IEEE International Conference on Information Technology, Gdansk 2008, pp 439–442
- Frejlichowski D, Forczmański P, Nowosielski A, Gościewska K, Hofman R (2012) SmartMonitor: an approach to simple, intelligent and affordable visual surveillance system. In: Bolc L et al. (eds) ICCVG 2012, Lect Notes Comput Sci, vol 7594, pp 726–734
- Frejlichowski D, Gościewska K, Forczmański P, Nowosielski A, Hofman R (2013) The removal of false detections from foreground regions extracted using adaptive background modelling for a visual surveillance system, In: Saeed K et al. (eds) CISIM 2013, Lect Notes Comput Sci, vol 8104, pp 253–264
- Sen-Ching SCS, Kamath C (2004) Robust techniques for background subtraction in urban traffic video. In: Bhaskaran V, Panchanathan S (eds) Visual communications and image processing, vol 5308, pp 881–892
- Kaewtrakulpong P, Bowden R (2001) An improved adaptive background mixture model for realtime tracking with shadow detection. In: Video based surveillance systems: computer vision and distributed processing, Kluwer Academic Publishers
- Forczmański P, Seweryn M (2010) Surveillance video stream analysis using adaptive background model and object recognition. Lect Notes Comput Sci Comput Vis Graph 6374:114–121

17. Javed O, Shafique K, Shah M (2002) A hierarchical approach to robust background subtraction using color and gradient information. *Workshop on Motion and Video Computing*, pp 22–27
18. Dalal N, Triggs B (2005) Histograms of oriented gradients for human detection. *IEEE Comput Soc Conf Comput Vis Pattern Recognit* 1:886–893
19. Frejlichowski D, Gościewska K, Forczmański P, Nowosielski A, Hofman R (2012) SmartMonitor: recent progress in the development of an innovative visual surveillance system. *J Theor Appl Comput Sci* 6(3):28–35
20. Viola P, Jones M (2001) Rapid object detection using a boosted cascade of simple features. *IEEE Comput Soc Conf Comput Vis Pattern Recognit* 1:511–518
21. Avidan S (2005) Ensemble tracking. In: 2005 IEEE Computer Society Conference on computer vision and pattern recognition, San Diego
22. Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. *ACM Comput Surv* 38(4):13
23. Welch B (2006) An Introduction to the Kalman filter. UNC-Chapel Hill, TR 95–041
24. Cheng Y (1995) Mean shift, mode seeking, and clustering. *IEEE Trans Pattern Anal* 17(8):790–799
25. Comaniciu D, Meer P (2002) Mean shift: a robust approach toward feature space analysis. *IEEE Trans Pattern Anal* 24(5):603–619