SmartMonitor: recent progress in the development of an innovative visual surveillance system

Dariusz Frejlichowski¹, Katarzyna Gościewska¹,², Paweł Forczmański¹, Adam Nowosielski¹, Radosław Hofman²

¹ Faculty of Computer Science and Information Technology, West Pomeranian University of Technology, Szczecin, Poland
² Smart Monitor sp. z o.o., Szczecin, Poland

{dfrejlichowski,pforczmanski,anowosielski}@wi.zut.edu.pl,
katarzyna.gosciewska,radekh}@smartmonitor.pl

Abstract: This paper describes recent improvements in developing SmartMonitor — an innovative security system based on existing traditional surveillance systems and video content analysis algorithms. The system is being developed to ensure the safety of people and assets within small areas. It is intended to work without the need for user supervision and to be widely customizable to meet an individual’s requirements. In this paper, the fundamental characteristics of the system are presented including a simplified representation of its modules. Methods and algorithms that have been investigated so far alongside those that could be employed in the future are described. In order to show the effectiveness of the methods and algorithms described, some experimental results are provided together with a concise explanation.

Keywords: SmartMonitor, visual surveillance system, video content analysis

1. Introduction

Existing monitoring systems usually require supervision by responsible person whose role it is to observe multiple monitors and report any suspicious behaviour. The existing intelligent surveillance systems that have been built to perform additional video content analysis tend to be very specific, narrowly targeted and expensive. For example, the Bosch IVA 4.0 [1], an advanced surveillance system with VCA functionality, is designed to help operators of CCTV monitoring and is applied primarily for the monitoring of public buildings or larger areas, hence making it unaffordable for personal use. In turn, SmartMonitor is being designed for individual customers and home use, and user interaction will only be necessary during system calibration. SmartMonitor’s aim is to satisfy the needs of a large number of people who want to ensure the safety of both themselves and their possessions. It will allow for the monitoring of buildings (e.g. houses, apartments, small enterprises, etc.) and their surroundings (e.g. yards, gardens, etc.), where only a small number of objects need to be tracked. Moreover, it will utilize only commonly available and inexpensive hardware such as a personal computer and digital cameras. Another intelligent monitoring system, described in [2], analyses human location, motion trajectory and velocity in an attempt to classify the type of behaviour. It requires both the participation of a qualified employee and the preparation of a large database during the learning process. These steps are unnecessary with the SmartMonitor system due to a simple calibration mechanism and feature-based methods. Moreover, a precise calibra-
tion can improve a system’s effectiveness and allow the system’s sensitivity to be adjusted to situations that do not require any system reaction. The customization ability offered by SmartMonitor is very advantageous. In [3], the problem of automatic monitoring systems with object classification was described. It was assumed that the background model used for foreground subtraction does not change with time. This is a crucial limitation caused by the background variability of real videos. Therefore, and due to planned system scenarios, the model that best adapts to changes in the scene will be utilized.

SmartMonitor will be able to operate in four independent modes (scenarios) that will provide home/surroundings protection against unauthorized intrusion, allow for supervision of people who are ill, detect suspicious behaviours and sudden changes in object trajectory and shape, and detect smoke or fire. Each scenario is characterized by a group of performed actions and conditions, such as movement detection, object tracking, object classification, region limitation, object size limitation, object feature change, weather conditions and work time (with artificial lighting required at night). A more detailed explanation of system scenarios and parameters is provided in [4].

The rest of the paper is organised as follows: Section 2 contains the description of the main system modules; algorithms and methods that are utilised in each module are briefly described in Section 3; Section 4 contains selected experimental results; and Section 5 concludes the paper.

2. System Modules

SmartMonitor will be composed of six main modules: background modelling, object tracking, artefacts removal, object classification, event detection and system response. Some of these are common to the intelligent surveillance systems that were reviewed in [5]. A simplified representation of these system modules is displayed in Fig. 1.

![Figure 1. Simplified representation of system modules](image)

Background modelling detects movement through use of background subtraction methods. Foreground objects that are larger than a specified size and coherent are extracted as objects of interest (OOI). The second module, object tracking, tracks object locations across consecutive video frames. When multiple objects are tracked, each object is labelled accordingly. Every object moves along a specified path called a trajectory. Trajectories can be compared and analysed in order to detect suspicious behaviours. The third module, artefacts removal, is an important step preceding classification and should be performed correctly. In this, all
artefacts, such as shadows, reflections or false detection results, enlarge the foreground region and usually move with the actual OOI. The fourth module, object classification, will allow for simple classification using object parameters and object templates. The template base will be customizable so that new objects can be added. A more detailed classification will also be possible using more sophisticated methods. The key issue of the fifth, i.e. the event detection module, is to detect changes in object features. The system will react to both sudden changes (mainly in shape) and a lack of movement. The final module defines how the system responds to detected events. By eliminating the human factor it is important to determine which situations should set off alarms or cause information to be sent to the appropriate services.

3. Employed Methods and Algorithm

For each module we investigated the existing approaches, and modified them to apply the best solution for the system. Below we present a brief description and explanation of this.

Background modelling includes models that utilize static background images [3], background images averaged in time [6] and background images built adaptively, e.g. using Gaussian Mixture Models (GMM) [7, 8]. Since the backgrounds of real videos tend to be extremely variable in time, we decided to use a model based on GMM. This builds per-pixel background image that is updated with every frame, and is also sensitive to sudden changes in lighting which can cause false detections, mainly by shadows. It was stated in [9] that shadows only affects the image brightness and not the hue. By comparing foreground images constructed using both the Y component of the YIQ colour scheme and the H component of the HSV colour scheme, it is possible to exclude false detections that are caused by shadows. Following this, morphological operations are applied to the resulting binary mask. Erosion allows for the elimination of small objects composed of one or few pixels (such as noise) and the reduction of the region. Later the dilation process fills in the gaps.

For the object tracking stage we investigated three possible implementations, namely the Kalman filter [10], Mean Shift and Camshift [11, 12] algorithms. The Mean Shift algorithm is simple and appearance-based. It requires one or more feature, such as colour or edge data to be selected for tracking purposes. This can cause several problems with object localization when particular features change. The Camshift algorithm is simply a version of the Mean Shift algorithm that continuously adapts to the variable size of tracked objects. Unfortunately, the described solution is not optimal since it increases the number of computations. Moreover, both methods are effective only when certain assumptions are met, such as that tracked objects will differ from the background (e.g. through variations in colour). The Kalman filter algorithm was therefore selected to overcome these drawbacks. This constitutes a set of mathematical equations that define a predictor-corrector type estimator. The main task was to estimate future values in two steps: prediction based on known values, and correction based on new measurements. It is assumed that objects can move uniformly and in any direction but will not change direction suddenly and unpredictably.

After tracking the objects are classified (labelled) as either human or not human. A boosted cascade of Haar-like features [13] connected using the AdaBoost algorithm [14] can be utilized. However, at this stage, we replaced the AdaBoost classification with a simpler one. Objects can now be classified using their binary masks and the threshold values of two of their properties: area size and minimum bounding rectangle aspect ratio.

A specific and detailed classification can be performed using a Histogram of Oriented Gradients (HOG) [15]. A HOG descriptor localises and extracts objects from static scenes
through use of specified patterns. Despite its high computational complexity, the HOG algorithm can be applied to a system under several conditions such as those with limited regions or time intervals.

4. Experimental Conditions and Results

In this section we present some experimental results from employing the algorithms for object localization, extraction and tracking that have given the best results so far. In order to ensure the experiments were performed under realistic conditions, a set of test video sequences corresponding to certain system scenarios was prepared. These include scenes recorded both inside and outside the buildings, with different types of moving objects. A database also had to be created due to the lack of free, universal video databases that matched the planned scenarios.

The results of employing both the GMM algorithm and the methods for removing false objects are presented in Fig. 2. The first row contains the sample frame and background images for the Y and H components. The second row shows the respective foreground images for the Y and H components alongside the foreground object’s binary mask after false objects removal. It is noticeable that the foregrounds constructed using the different colour components strongly differ and that, by subtracting one image from another, we can eliminate false detections.

Specific objects can be localised and extracted using the HOG descriptor. This detects objects using a predefined patterns and extracted feature vectors. Below we present the results of the experiments utilizing HOG descriptor. The first experiment was performed using a fixed template size and two sample frames, the second one utilized various template sizes and one sample frame.

The results of the first experiment are pictured in Fig. 3. The figure contains: a sample frame with a chosen template (left column) and two frames (middle column) from the same video sequence which were scanned horizontally in an attempt to identify the matching regions. The depth maps (right column) show the results of the HOG algorithm — the darker the colour the more similar the region is. Black regions indicate a Euclidean distance between two feature vectors of zero.
In the next experiment, devoted to an investigation of the HOG descriptor, various template sizes were tested. The left column of Fig. 4 presents a frame with a chosen template marked by a white rectangle, the central column contains a frame that was scanned horizontally using two different template sizes (dark rectangles in the top left corners define the size of the rescaled template) and the right column provides the respective results of the HOG algorithm. Clearly, the closer the template size is to object size, the more accurate the depth map is.

As mentioned in the previous section, we investigated three tracking methods. The first one, the Mean Shift algorithm, uses part of an image to create a fixed template model. In this case we converted images to the HSV colour scheme. Fig. 5 presents three sample frames from the tracking process (first row) and their corresponding binary masks (second row). The white masked regions indicate those regions that are similar to the template, the dark rectangle determines the template and the light points within the rectangle create the object’s trajectory.

Camshift was the second tracking method investigated. This uses the HSV colour scheme and a variable template model. The first row in Fig. 6 presents sample frames from the tracking process: the starting frame with the chosen template, the central frame with an enlarged template and the finishing frame where the moving object leaves the scene. The second row in Fig. 6 shows corresponding binary masks for each frame. Both tracking methods, thanks
Figure 5. Results of the experiment utilizing the Mean Shift algorithm to their local application, were effective despite of the presence of many similar regions to the template.

Figure 6. Results of the experiment utilizing the Camshift algorithm

Fig. 7 shows a result of employing the third algorithm, the Kalman filter, to track a person walking in a garden. Light asterisks are obtained for object positions that were estimated using a moving object detection algorithm and dark circles are positions predicted by the Kalman filter.

5. Summary and Conclusions

In this paper, recently achieved results from the SmartMonitor system during the development process were described. We provided basic information about system characteristics and properties, and system modules. Investigated methods and algorithms were briefly described. Selected experimental results on utilizing various solutions were presented.

SmartMonitor will be an innovative surveillance system based on video content analysis and targeted at individual customers. It will operate in four independent modes which are fully customizable (and will also be combinable to make custom modes). This allows for individual safety rules to be set based on different system sensitivity degrees. Moreover, SmartMonitor will utilize only commonly available hardware. It will almost eliminate human involvement,
being only required for the calibration process. Our system will analyse a small number of moving objects over limited region which could additionally improve its effectiveness.

Currently, there are no similar systems on the market. Modern surveillance systems are usually expensive, specific and need to be operated by a qualified employee. SmartMonitor will eliminate these factors by offering less expensive software, making it more affordable for personal use and requiring less effort to use.

Acknowledgements

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 the European Union (project number PL: UDA-POIG.01.04.00-32-008/10-01, Value: 9.996.604 PLN, EU contribution: 5.848.800 PLN, realization period: 07.2011-04.2013). European Funds — for the development of innovative economy (Fundusze Europejskie — dla rozwoju innowacyjnej gospodarki).

References


