Detecting Critical Situation in Public Transport

R. M. Luque, F. L. Valverde, E. Dominguez, E. J. Palomo and J. Muñoz

Department of Computer Science, University of Málaga, Málaga, Spain

Abstract. This paper presents a system information applied to video surveillance to detect and identify aggressive behaviors of people in public transport. A competitive neural network is proposed to form a background model for detecting objects in motion in the sequence. After identifying the objects and extracting its features a set of rules are applied to decide if an anomalous behavior is or not considered aggressive. Our approach is integrated in a CCTV system and is a powerful support tool for security operators to manage to detect in real time critical situations in public transport.

1 Introduction

Research in video surveillance systems has been an area in which scientific community has made a lot of number of contributions for the last years [1–3]. Developing systems to detect the movement of the objects in a scene could be applied to a wide range of potential applications, such as a security guard for important buildings, traffic surveillance in cities and highways, or supervision of suspicious behaviors in supermarkets. However, the increase of the research in video surveillance has not generated a huge rise of intelligent surveillance systems in operation. At present, passive supervision systems keep being used in many areas, in which the number of cameras exceeds the capability of human operators to analyze and monitor them.

The aim of this work is to develop an intelligent video surveillance system to analyze people behavior in public transport [4]. This paper is really focused on detecting unlucky events like hostile behaviors in public transport, using the steady cameras situated in these areas. These behaviors are relatively simples and based on action - reaction principle, in which a person is attacked by another person who is the aggressor. In this case, an event is thrown by the system to security people for checking if the signal received is significant.

In this work an unsupervised competitive neural network [5] is proposed for objects segmentation in real time. The proposed approach is based on competitive neural network to achieve background subtraction. This neural network is designed to serve both as an adaptive model of the background in a video sequence and a classifier of pixels as background or foreground. Neural networks posse intrinsic parallelism which can be exploited in a suitable hardware implementation to achieve fast segmentation of foreground objects.

The rest of the paper is organized as follows; section 2 shows a system overview and the interactions between human and machine; section 3 and 4 present a new proposed segmentation method and describe the object tracking phase; section 5 describes how the critical events are detected. Results and conclusion follow in section 6 and 7, respectively.
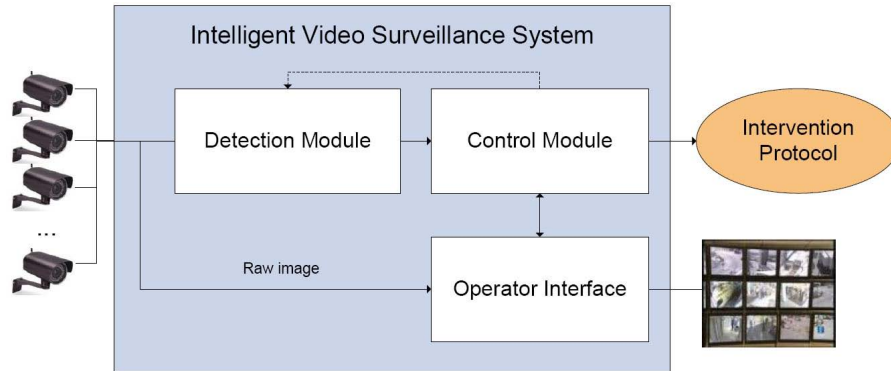
Fig. 1. Video surveillance system scheme.

2   A System Overview

Most video surveillance systems (SVS) require human interaction to detect critical situation. Usually, the number of cameras installed in SVS exceed the human capability and it is generally accepted that guards frequently miss incidents clearly visible on CCTV monitors. The vast majority of installed CCTV cameras remain unwatched and incidents are not likely to be detected while they are occurring.

In this work additional functionalities are added to the current surveillance systems, improving their effectiveness and becoming in a support tool for security enterprises. This information system analyzes an input sequence obtained through CCTV cameras, and informs about possible suspect behaviors in the passengers of the public transport.

An overall scheme of proposed system is shown in Fig. 1. A specific number of cameras can be connected to analyze independently each scene observed. A detection module manages to detect and extract features of the movement objects in the scene to, subsequently, inform about a suspect anomalous behavior. In this case, the operator validates or not the detected aggression through the user interface, by observing the sequence in which the action is performed.

The control module interacts with the other modules to keep synchronizing the system and executing an intervention protocol proposed by the company when an aggression is confirmed. This confirmation is sent towards to the detection module to learn what kinds of anomalous detected behaviors are right. Raw and processed images are shown to the security operators in the output monitors.

Certain information is required to decide what measures are taken in the intervention protocol when an aggression is detected. A short sequence in which the aggressive behavior is observed, the position of the vehicle inside its trajectory, the next station where passengers get off, and the time to arrive to that station are some parameters utilized to determine the immediate actions to perform.

Several steps are required by a typical video surveillance system to reach the objective. Detection module scheme of proposed system is observed in Fig. 2. At first, a video object segmentation method is required to obtain the objects in movement of the
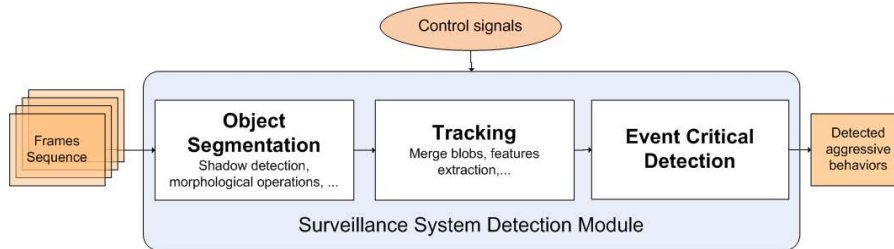
Fig. 2. Detection module architecture.

stream. A comparison among different techniques can be found in some works [6, 1, 7]. Next section describes in detail the proposed method to obtain the objects belong to the foreground.

In the second step and using the segmentation output, a tracking algorithm is applied to identify the objects in several frames of the sequence. A matching between each blob (or set of blobs) and an object previously recognized is done. In the literature, many works can be found about tracking people [2]. Finally, a set of rules have been employed to determinate possible dangerous movements. These events could be the reason of a critical situation in the analyzed stream.

3 Object Segmentation

This process aims at segmenting regions corresponding to moving objects from the rest of an image. Subsequent phase such as tracking and behavior recognition are greatly dependent on it. It usually involves environment modeling, segmentation motion, shadow detection and object classification.

Many works based on motion detection and, more concretely, on background subtraction using fixed cameras as the CCTV cameras installed in public transport can be found [8, 9, 4, 10]. All of these methods try to effectively estimate the background model from the temporal sequence of the frames and, subsequently get objects in motion of the scene.

The goal of the proposed video object segmentation algorithm is to be able to classify the pixel in a frame of the sequence as foreground or background, based on statistics learned by a neural network approach (NA). These statistics are learned from the frames observed in the video sequence. Each pixel is modeled by a competitive network and, by using its color components (RGB), the input pixels are classified as background or foreground at each frame.

Our model performs a clustering based on a crisp-fuzzy hybrid neighborhood. For each neuron, there is a value r_j, representing the crisp neighborhood of the corresponding synaptic weight w_j: $\mathcal{N}_j = \{x : \|x - w_j\| \leq r_j\}$. The fuzzy neighborhood of neuron j is given by a membership function μ_j defined over the entire input space, and taking values in the interval $(0, 1)$.

The use of this hybrid neighborhood allows us to better assign an input pattern to a class or category:

- If the input pattern x belongs to $\mathcal{N}_j$, the network assumes that the best matching category for x is the associated to w_j.
- If the input pattern x does not belong to any crisp neighborhood, its most likely category is represented by the neuron achieving the maximum value of the membership function for this particular input.

Usually, the membership function present in this model is of the form:

$$\mu_j(x) = e^{-k_j(\|x - w_j\| - r_j)} \tag{1}$$

The value of the parameter k_j is related to the slope of the function. The higher its value, the higher its slope is. For a great value of k_j, the fuzzy neighborhood of w_j will be more concentrated around $\mathcal{N}_j$.

The synaptic potential of the neuron j is defined as

$$h_j(x) = \begin{cases} 1, & \text{if } x \in \mathcal{N}_j \\ \mu_j(x), & \text{otherwise} \end{cases} \tag{2}$$

the winning neuron (with index denoted as $q(x)$) will be the one receiving the maximum synaptic potential:

$$h_{q(x)} = \max_j h_j(x) \tag{3}$$

The objective of this model is to study the color space associated to each pixel in the video sequence and to determine, at every step, the most likely category to which it belongs, according to the hybrid crisp-fuzzy neighborhood which measures the membership of every pattern to a given category.

The learning rule used by each network to model the input space is the standard competitive learning rule:

$$w_{q(x)}(t + 1) = w_{q(x)}(t) + \lambda \cdot (x - w_{q(x)}(t)) \tag{4}$$

where λ is the so-called learning rate parameter, usually decreasing to 0. It can be viewed as the stochastic gradient descent technique to minimize the squared error function (also named distortion function):

$$F(W) = \sum_i \|x_i - w_{q(x_i)}\|^2 \tag{5}$$

where W represents a matrix whose rows are the w_j.

When the segmentation results have been obtained, it will be necessary applying additional techniques to obtain clear foreground regions. We are usually analyzing indoor scenes (inside of a car train or in a building hall) so, it is probably objects in motion cast shadow on the background, confusing with foreground pixels and interfering in the correct detection of the scene objects. There are many shadow detection methods published [11] but the majority of them depend on scene features, types of shadow found { slighter, darker, sharper,... }, technique applied, etc. In [12], an overall review about them can be found. In our system, we develop the proposed technique cited in [13]. It is based on the RGB pixels vectors in shadow regions are in the same direction that the
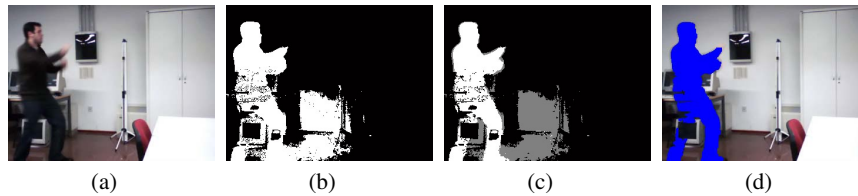
Fig. 3. An example of the objects detection algorithm is presented. (a) original frame in raw form. (b) segmentation result using the NA. (c) shadow detection is applied. (d) enhanced frame using morphological operations.

RGB vectors of the same pixels in the background, with a little variation and a smaller brightness between shadow pixels and background pixels. The results can be viewed in 3(c).

Additionally, morphological operations have to be applied to eliminate one-pixel thick noise and to fill those objects with pixels not correctly identified as foreground. As the final step of foreground region detection, a binary connected component analysis is applied to the foreground pixels to assign an unique label to each foreground object and grouped them into an initial blobs. We can observe the results of this mechanism in 3(d).

4 Tracking

In an object tracking phase, it is necessary to achieve a correspondence among obtained blobs and identified objects in motion. This tracking object must be sufficiently robust to overcome potential problems at the segmentation stage. It is possible to find incomplete objects in the scene, caused by an object becomes temporally occluded (by some fixed background object) or an object splits into pieces (due to a person depositing an object in the scene or a person is being occluded by small object).

To solve these situations, we have implemented a simple mechanism to merge obtained blobs based on their proximity. The system computes some features needed for each blob according to its size and aspect. With this information and the area and position of the objects previously identified, we can determinate which specific blobs belong to these tracked objects.

Our approach manages to solve person blobs merged, making a restriction in the scene based on the kind of events to detect. For our purpose of detecting aggressive actions, such as a fight, it is not essential to distinguish people who are within a blob. This situation would be starting in previous frames while the object features are being processed, and it would be definitely identified when the merging blobs occurs and the movement of the new merger object can be also considerate suspicious.

Typically, some aggressive behaviors involve the movement of some body parts, hence with just the features computed previously is not possible to recognize with what body part the aggression has been committed. Some works [14][15] try to obtain the position of these parts to decide more efficiently in critical situations, if an arm or a leg has caused damage to another person. However, this tracking body parts has cer-
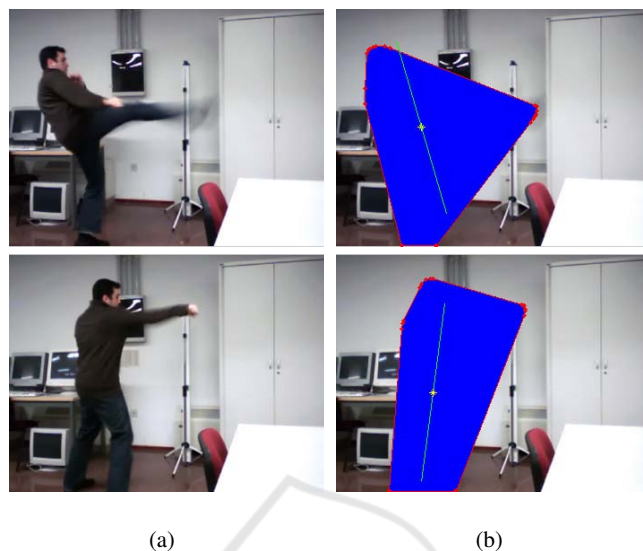
(a)   (b)

Fig. 4. The convex hull of the object has been calculated. (a) originals frame in raw form. (b) A polygonal figure has been calculated. In the upper-left image, we can deduce an arm in motion. In the lower-left image, a silhouette of a kick could be inferred.

tain complexity and could provoke the overall system does not carry out with temporal requirements.

An intermediate solution is to find the convex hull object, namely the minimum possible area where all the object points are contained [16]. Thus, in a state of aggression, the position of the points which have led to an increase in the object area are known (one of the vertices of the polygon), and we can estimate which kind of extremity it is, the speed increase from the previous frames, making the system more robust to possible errors or false positives. Figure 4 shows the convex hull of the object in a concrete frame.

Therefore, some features of the identified objects are extracted to achieve the aim of detecting aggressive behaviors in people of the sequence. Some of them are:

- The centroid of the object
- Its bounding box
- Object convex hull
- The number of the blobs to belong it and their positions
- Its area and the object orientation based on its major axes.

With these parameters, trajectories of single persons are easily obtained (Fig. 5). Our system is capable of dealing with possible object occlusions by computing a estimated trajectory using data of the previous frames. Object speed in these uncommon situations will be gained in terms of the median speed of the rest of the frames processed.

Fig. 5. Convex hull, centroid and principal axis of an person along several consecutive frames.

5 Detecting Critical Events

The information provided by the tracking task allows us to analyze the interactions between objects that appear on the scene, as well as detect events that might be considered unusual or suspicious. An automatic surveillance system is capable to identify some kind of diverse events like detection of abandoned luggage in buildings, unusual movements of passengers, people falling down, people walking in dangerous areas such as railway lines, aggressive behavior, etc.

In this work a set of rules are used for fights detection and dangerous behaviors. Detecting attacks on public transport is the focus at this work, using the existing CCTV in those places. The system informs to the personal security which must confirm whether the detected event is relevant or not.

An attack is composed by an active object and a passive object. Aggressive events are considered when an object is performing the action of assault, and another object reacting after that aggression. The system has three possible states (normal, suspect and aggression) about people behavior. The normal state is defined as no detection of uncommon behavior. The suspect state is defined by the detection of a person who is committing an aggression, whereas an aggression-state is defined by the detection of this aggressive person and an assaulted person.

A number of simplifications are considered to make this detection. For example, in public transport is not common to find people running or moving quickly through the wagon, hence the speed of the movement of objects is one of the main parameter for determining anomalous behavior.

In this sense, an aggressor individual is detected by analyzing several features: the irregular motion of the object centroid, a quick variation of the width and height of the object or the change in the area of the object convex hull. Moreover, a sudden change of distances of the vertexes of the object convex hull to its main axis, involves a sudden movement of the limbs of a person, either a possible kick or a possible punch.

The detection of an assaulted individual involves the previous detection of the aggressor person. The union and division of the objects recognized, the variation of the centroid object and the position of the alleged victim are significant signs to identify an aggression. A variation of the angle of the main axis of the object which envelop the victim provokes a fall or an attempt to dodge the blow.

6   Results

The proposed surveillance system has been applied to a set of proof sequences to show the validity of our method. These sequences have been grabbed in our laboratory and try to simulate some possible critical situations related with aggressive behaviors happened in public transport.
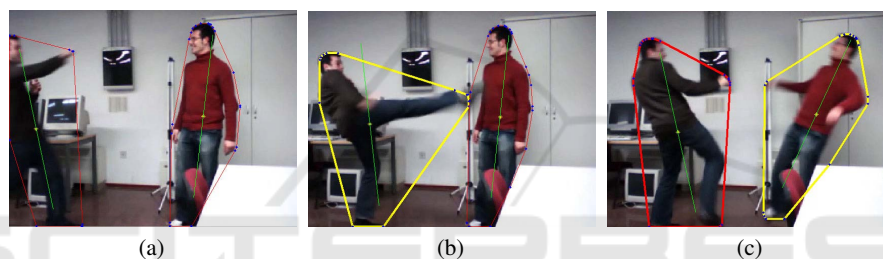

| (a) | (b) | (c) |

Fig. 6. Detecting of an aggressive behavior. In 6(b) an attack is detected. In 6(c) the reaction of the other person is observed.

They are middle-sized with 264x352 and they are captured through a web cam for domestic use. Our system has been proved in a standard PC with 3 GHz and 1 Gb of RAM memory. Despite using a low cost hardware, we have achieved very good results, like we can observe in Fig. 6. According to the previous section, initially, the system has detected a suspect behaviour in the left person (Fig. 6(b)) and, this aggressive behavior is confirmed when the reaction of another person is observed (Fig. 6(c)). To prove our detection technique of aggressive person a low people-density situation is assumed.

Figure 7 shows another kind of critical scene. We have represented a situation in which a person, who is sitting on a seat, is severely beaten by another person who is standing and moving along the public transport. We can observe the sudden movement of one person and the reaction of another one.

A similar scene happened in 2007 in a metro in Spain can be observed in Fig. 8 in which a man is hitting a girl with his leg. A sudden change in the area and the silhouette of the objects is detected, therefore a anomalous behaviors is identified. In this case, the reaction of the assaulted person is less obvious than in Fig. 8 and its movement has not been detected as an anomaly. The confirmation of security people would validity the suspicious behavior recognized.

We have to take into account the dependence of a good result in the segmentation phase. Without a great detection of the objects in motion is not possible develop a tracking system and, subsequently, detect and analyze simple people behaviors.
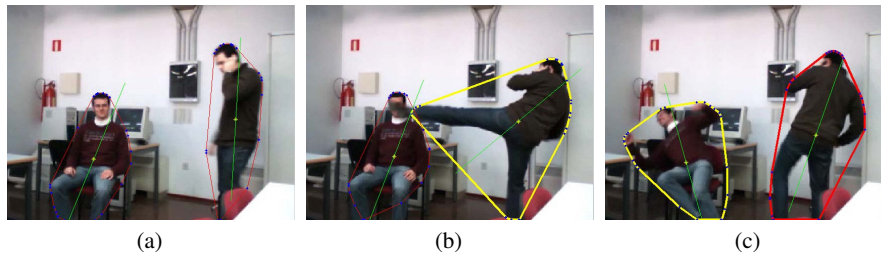
Fig. 7. Simulation of a public transport scene. In 7(a) two so-called passengers are observed. In 7(b) an aggression is being identified. The reaction of the assaulted person can be viewed in 7(c).

Fig. 8. Real scene in a public transport. In 8(a) two persons are detected. In 8(b) a sudden union between the two objects is found. A suspicious event keeps detecting in 8(c).

7 Conclusions

In this paper we have presented an information system for video surveillance in public transport which is very useful to detect aggressive behavior in real time. Human interaction is only necessary to confirm the alarm in case of aggressive behavior is detected. In this case, the information about the critical event (video of critical situation, real time location of the transport, next step, time to reach next step, video of actual situation,...) is provided by the system.

The proposed video surveillance system has been successfully applied to process real-time video sequences (more than 25 fps) using only a conventional PC and a low cost camera. Moreover, the speed of the proposed segmentation, can be improved in a parallel hardware implementation using field programmable gate arrays (FPGA). In addition, scalability is one of major advantages of neural networks, therefore we plan to design new neural models with new functionalities.

Acknowledgements

References

1. Hu, W., Tan, T., Wang, L., Maybank, S.: A survey on visual surveillance of object motion and behaviors. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews **34** (2004) 334–352

2. Yilmaz, A., Javed, O., Shah, M.: Object tracking: A survey. Acm Computing Surveys **38** (2006) 13–58

3. Fuentes, L., Velastin, S.: Tracking people for automatic surveillance applications. Pattern Recognition and Image Analysis, Proceedings **2652** (2003) 238–245

4. Haritaoglu, I., Harwood, D., Davis, L.: W4: real-time surveillance of people and their activities. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 809–830

5. Hertz, J.: Introduction to the theory of neural computation. Addison-Wesley (1995)

6. Piccardi, M.: Background subtraction techniques: a review. In: IEEE International Conference on Systems, Man and Cybernetics. (2004) 3099–3104

7. Radke, R., Andra, S., Al-Kofahi, O., Roysam, B.: Image change detection algorithms: a systematic survey. IEEE Transactions on Image Processing **14** (2005) 294–307

8. Lo, B., Velastin, S.: Automatic congestion detection system for underground platforms. In: Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, 2001. (2001) 158–161

9. Stauffer, C., Grimson, W.: Learning patterns of activity using real-time tracking. IEEE Transactions on Pattern Analysis and Machine Intelligence **22** (2000) 747–757

10. Elgammal, A., Duraiswami, R., Harwood, D., Davis, L.: Background and foreground modeling using nonparametric kernel density estimation for visual surveillance. In: Conference on Computer Vision and Pattern Recognition. (2002) 1151–1163

11. Mikic, I., Cosman, P., Kogut, G., Trivedi, M.: Moving shadow and object detection in traffic scenes. In: International Conference on Pattern Recognition. (2000)

12. Prati, A., Mikic, I., Trivedi, M., Cucchiara, R.: Detecting moving shadows: Algorithms and evaluation. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (2003) 918–923

13. Horprasert, T., Harwood, D., Davis, L.S.: A statistical approach for real-time robust background subtraction and shadow detection. In: Proceedings of IEEE ICCV'99 FRAME-RATE Workshop. (1999)

14. Park, S., Aggarwal, J.: Segmentation and tracking of interacting human body parts under occlusion and shadowing. In: Workshop on Motion and Video Computing. (2002) 105–111

15. Ju, S., Black, M., Yacoob, Y.: Cardboard people: a parameterized model of articulated image motion. In: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition. (1996) 38–44

16. Barber, C., Dobkin, D., Huhdanpaa, H.: The quickhull algorithm for convex hulls. ACM Transactions on Mathematical Software (TOMS) **22** (1996) 469–483