

## Video Surveillance for Real Time Objects

N.R. Raajan, G. Shiva, P. Mithun and G.N. Jayabhavani  
Department of ECE, School of Electrical and Electronics Engineering,  
SASTRA University, Thanjavur, Tamil Nadu, India

**Abstract:** In this study, we describe an integrated solution for video surveillance in a fortified environment. The focus of this study is on identification of real time objects on different environments. The system is composed of robust object detection module, which normally detects the presence of abandoned objects, concealed objects hidden inside the human clothing, objects in dark environment and performs image segmentation with the intention of facilitating human operator's task of retrieving the cause of a buzzer. The abandoned objects are detected by image segmentation based on temporal rank order filtering. Image fusion technique which fuses a color visual image and a corresponding IR image for concealed objects in guarded environment and in some cases like dark environment heat signature can be used for detecting real time objects etc. In the clips of interest, the key frame is the one depicting a person leaving a dangerous object and is determined on the basis of a feature indicating the movement around the dangerous region.

**Keywords:** Content based retrieval, image fusion, surveillance systems

### INTRODUCTION

Video Surveillance is the monitoring of the activities usually for the purpose of influencing or protecting. It is occasionally done in a secret manner for the observation of highly constrained area by any organizations. It is generally known that, if a person is exposed to this type of job for several hours, his focus decreases, thus the likelihood of missing risky situations increases.

After acquisition, Region of focus (objects and people) must be then coordinated in the GUI and events deemed of special significance must be tracked throughout the scene. Wherever possible, tracked events should be classified and their dynamics analyzed to alert an authority of a potential danger. In the progress of highly developed visual-based surveillance systems, a number of key issues critical to successful operation must be addressed. The requirement of working with difficult scenes considered by high variability requires the use of sophisticated algorithms for stuff identification in different environments. We propose a novel approach for a video-based surveillance system for the automatic detection of real time objects in fortified environments. This surveillance system integrates a sophisticated real-time detection method with video indexing capabilities in order to establish a logical relationship between a real time object, the person who left it and even in unlike environment like dark environment using thermal camera by allowing the human operator to easily retrieve the

image or the clip of interest while scanning a large video library. In future study in addition to camera, Kinect sensor device can be used to track, detect real time objects

The application presented in this study executes semantic video-shot detection and indexing in the way analogous but uncomplicated than the one discussed in Courtney (1997), where a surveillance system sense various real time objects (e.g., abandoned object detection, Budgeting object detection, etc.). The system is based on the movement analysis performed by deriving a temporal graph of the objects present in a scene. A portion of this graph represents the clip related to an event and the indexing is performed by using features describing a sub-graph. On implementing this to the application discussed in this study we found several downsides. Primarily stagnant objects are detected if they hang about two successive frames and then the difference between the seated person and lost object is not taken into consideration.

The presented system in this study is designed in such a manner that it is capable of detecting real time objects using temporal rank order filtering (Regazzoni *et al.*, 1997; Murino *et al.*, 1994) and the classification of extracted static region. This system detects video shots by following a hybrid approach based on low-and semantic-level features. At the low level the system is able to detect permanent changes in a scene by means of color-based features. At the higher level, a neural network classifies detected regions on the basis of different features (color,

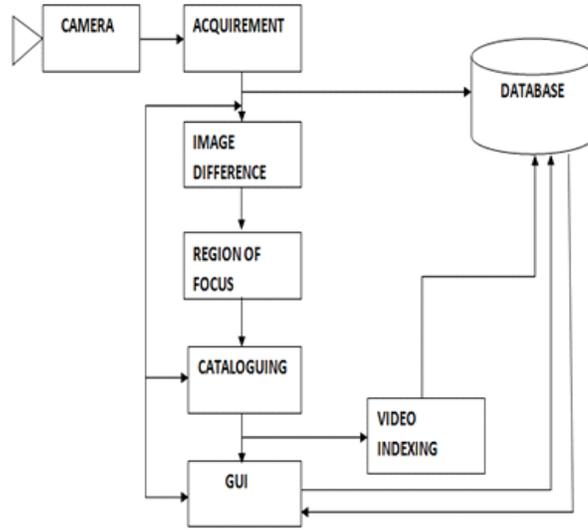


Fig. 1: Structural design of the proposed surveillance system

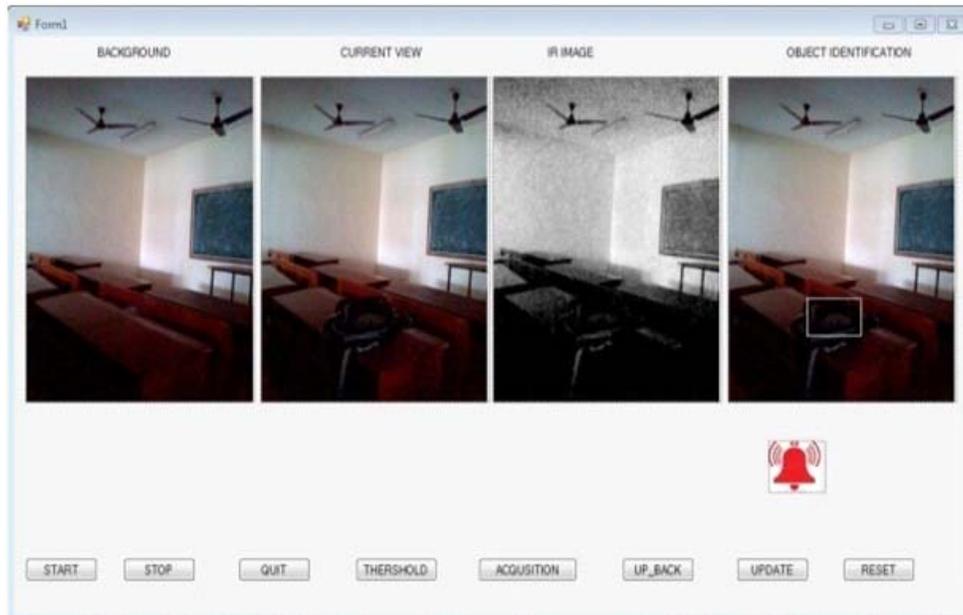


Fig. 2: Graphical user interface. The human operator is alerted when real time objects are detected

shape, movement, etc.). After the detection of real time objects the procedures of video-shot detection and indexing start.

On the other hand for identifying the concealed objects hidden in the person clothing the visual and IR images have been processed and the color visual image is fused with corresponding IR image through image registration by which the object hidden inside a person's clothing will appear darker than the human body due to the temperature difference.

## METHODOLOGY

**Structural design of proposed surveillance system:** The general architecture of the proposed system is shown in Fig. 1. The system aspires to provide information to human operator about the presence of real time objects (it may be even concealed objects inside the human body) in the scene. An alarm signal in the situation of detecting a hazardous object and possibility to regain the sequence in which a person leaves an object which causes the alarm

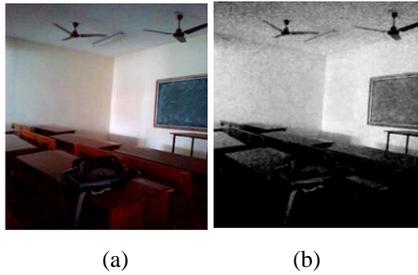


Fig. 3: Concealed object identification

signal. The Graphical User Interface (GUI) provides information to the human operator regarding background, current view and object identification. The information flow starts with the acquirement of data from the camera and terminates with an alarm signal to the human operator.

**A video-based surveillance system for detecting the presence of abandoned objects:** This system aims at providing the human operator with:

- Information about the presence of stationary objects in a Scene.
- An alarm signal in the case of the presence of real time objects.
- Information about the concealed object in the scene or thermal image if the environment is dark
- Indexing the frames by the operator can find which person has left the object that caused the alarm signal.

This information is presented to the human operator by means of a GUI (Graphical User Interface) (Fig. 2) One obtains the architecture shown in Fig. 1, in which the information flow starts with the data acquired with a sensor (a TV camera) and ends with the alarm signaling to the human operator. In this section, we analyze in detail the main algorithms used by image-processing tools for detecting the presence of real time objects in a guarded environment at different environments. Two images were used for experimental test in color image fusion method. These test images were selected to test the fusion algorithm. In Fig. 3 (a) shows the original image and (b) shows the IR image. If there is any concealed object is detected in IR image and it will be shown in dark intensity due to its low temperature compared to the object in which it is hidden.

## RESULT AND DISCUSSION

**Region of focus:** This section aims at marking the areas that contain different real time objects when there is any

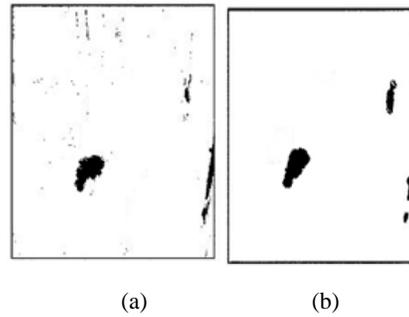


Fig. 4: (a) Input and (b) output of region of focus module

change in the scene furthermore, the section provides the human operator with a record of significant regions where the real time objects are causing the significant changes in the scene. The segmentation method should as simple as possible so that the system should guarantee real time performance even it should meet the robustness requirements like overlapped objects, object split by noise.

The algorithm developed for the proposed application involves the following operations:

- Statistical erosion (Yuille *et al.*, 1992) of the binary image ensuing from change detection in order to separate the regions. The size of the structuring element should be selected in an adaptive way even for threshold selection should be done in the same way like of structuring element
- Labelling of the regions sealed by the preceding filtering
- Splitting and merging of the labelled regions
- Dilation of the regions
- Splitting and merging of the dilated regions

After the segmentation step, blobs (objects) are highlighted by the rectangles around the obtained regions. As an example, the input and output of the segmentation procedure are shown in Fig. 4.

**Change-detection module:** An array of transform revealing methods have been developed for predetermined cameras. Here, the foremost methods used in video-based supervision systems are detailed. Threshold selection is a decisive task and the methods proposed by Kapur *et al.* (1985), Otsu (1979), Ridler and Calvard (1978), Rosin, (2002) and Snidaro and Foresti (2003) had shown the range of approaches that have been engaged.

The simple detection method is the effective one, but it is very susceptible to noise and illumination changes that openly affects the gray level recorded in the scenario, which could be improperly interpreted as structural



147	147	146	145	144	143	142	142	156
124	124	126	128	129	131	131	133	115
091	093	096	100	104	108	111	112	108
077	078	080	083	086	088	091	092	095
091	091	091	090	090	090	090	090	076
107	106	105	104	102	101	100	099	103
095	095	095	096	097	098	098	099	121
069	071	073	077	081	084	087	088	085
088	087	086	084	082	080	079	078	072
123	123	121	120	118	117	115	115	113
184	184	183	182	181	180	179	178	180
055	054	054	054	053	053	053	053	053
071	072	072	072	073	073	073	073	071

Fig. 5: Image and its matrix values without an object



148	146	144	141	140	139	138	137
129	129	130	130	131	131	134	135
104	107	110	113	115	116	107	108
092	093	095	096	097	098	092	093
091	089	088	086	085	085	105	104
092	090	087	084	082	081	101	099
089	088	088	087	087	086	087	086
083	085	088	090	092	093	093	093
067	067	067	067	067	067	071	071
115	115	115	115	115	115	117	117
191	191	191	191	191	191	189	189
068	068	068	068	068	068	064	064
082	082	082	082	082	082	078	078

Fig. 6: Image and its matrix values with an object detected

changes. To triumph over the obstacle of noise sensitivity, the derivative model method considers  $n \times n$  pixel regions in the two input images and computes a likelihood ratio  $P_{ij}$  by using the means and the variances of the dual regions  $R_i$  and  $R_j$ . The binary image output is obtained as:

$$B(x, y) = \begin{cases} 0 & \text{if } P_{ij} < PTh \\ 1 & \text{otherwise } \forall (x, y) \in R_i, R_j \end{cases} \quad (1)$$

where,  $PTh$  is the threshold.

The shading method models the intensity at a known point  $I_q(u, v)$  as the product of the illumination  $I_p(u, v)$

and a shading coefficient  $S_p$  which is projected for each point according to Phong's illumination model. To set up whether a change has taken place in a known region  $R_i$  over two successive frames,  $I_{t-1}(u, v)$  and  $I_t(u, v)$ , it is enough to approximate the variance  $\sigma_i$  of the intensity ratios  $I_t/I_{t-1}$  in that area. If  $\sigma_i$  is close up to zero, no changes will take place.

The noise level affecting the output image is lower in DM methods than that generated by the simple detection method; however, the precision of the detected objects, in terms of position, shape and size, is worse. In particular, object contours are significantly distorted and the original object shape is partly lost. The local intensity gradient method (Carpenter and Grossberg, 1991) is based on the postulation that pixels at locations having a towering gray-level gradient form a part of a blob and that almost all pixels with alike gray levels will also be a part of the same blob. The intensity gradient is computed as:

$$G(u, v) = \min \{I(u, v) - I(u \pm 1, v \pm 1)\} \quad (2)$$

Thereafter, the  $G(u, v)$  image is divided into  $m \times m$  sub images in order to bound the effects of illumination change on the computation of confined means and deviations. The regional means and deviations are foremost smoothed using the adjacent regions and then interpolated to refill a  $m \times m$  region. Finally, a threshold method is implemented to isolate object pixels from the backdrop. The local intensity gradient method gives pleasing results, even if is not satisfactory to completely distinguish the object from the backdrop. For example difference between the matrix values of image with an object and image without an object are shown in Fig. 5 and 6.

**Classification module:** The idea behind this part is to classify the objects within the analyzed blobs. This has been implemented by a multilayer perception (Parker, 1991). The detected blobs are classified as belonging to one of the Following classes:

- Abandoned objects.
- **Persons:** A person seated on a bench generates a stable change in a scene similar to the change generated by an abandoned object.
- **Lighting effects:** The class room has at least one door and often also windows. The opening of the door or windows and persons passing near the windows can generate some determined local changes in the lighting of fortified environments. In this case, not performing the classification step would generate a false alarm.
- Structural changes.

An alarm is generated only when an abandoned object is recognized.

Several features are used for the classification step; they are as follows:

- **Three-dimensional position of an object in a guarded environment:** This feature has been introduced because there are events that are located in the scene. As an example, Let us consider a structural change caused by an open door: this event only can happen in a specific region of the guarded environment. It is possible to notice that lighting effects and structural changes are located in fixed regions. Other features are introduced in order both to avoid that an object left near a region of possible structural change may be classified in a wrong way and to separate the clusters related to persons and abandoned objects.
- **Elongation features:** This uniqueness give an idea of the shape of an object inside a blob. A dangerous object may present similar elongation values with respect to  $x$  and  $y$  image axes, whereas a person exhibits different elongation values. These two features are computed as follows:

$$e_x = \max_{i \in A} \sqrt{(x(i) - x_0)^2}$$

$$e_y = \max_{i \in A} \sqrt{(y(i) - y_0)^2}$$

where

- A : Denotes the set of pixels belonging to an object inside the considered blob
- $(x_0, y_0)$  : Represents the center of gravity of the blob in the image plane
- $(x_i, y_i)$  : Represents the position of the pixel in the image plane

Experience suggests that persons show elongation features larger than other objects, especially in the vertical direction in the case of a standing person.

**Video-shot detection and indexing:** Video-shot detection is executed by the projected system on the basis of the linguistic content of a scene Let us consider, at the moment  $t_k$ , an alarm is detected by the system. Due to the structure of the change detector, it is feasible that an objective was left about 16 frames before, i.e., near the moment  $t_{k-15}$ . The video shot is defined as the position of frames related to the time from  $t_{n-k}$  to  $t_k$ , with in the proposed system. In the detected video-clip, the frame shows the authentication of the dangerous object; for this cause, we consider this frame as the major one. The detection of the key frame is made probable by a progress analysis performed in the sequential window shown in Fig. 4; this frame is theoretical to be the one with the maximum progress near the left object's position in a time close to the object deposition. The basic idea is that,

during the object deposition, a person departure the object is moving near the region in which the object is missed. The number of movement can be easily performed in terms of the progress value defined in the previous section. Indexing is performed based on index table in which each area finds out a detected clip. The index consists the time to time alarm detection features and the features of the discarded object extracted by the cataloguing module (colour, shape, etc.).

## RECOMMENDATIONS

Kinect sensor device can be used to identify the various real time objects. This gadget features an RGB camera, depth sensor which provides full-body 3D capture. The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures real time data in 3D. The sensing range of the depth sensor is adjustable and the Kinect sensor device is capable of automatically calibrating physical environment, accommodating for the presence of obstacles. The death map is visualized using color gradient from white (near) to blue (far).

## CONCLUSION

In this study, An integrated solution for video surveillance in a fortified environment for the detection of abandoned and hidden object has been proposed. The system is capable of detecting the abandoned object within sensible period and it is capable to make difference between the abandoned object and a static object and also it detects the object hidden inside a person's clothing. When a person leaves an object, the system detects causing an alarm. The sequence is categorized by a key frame represented by the image that shows the maximum movement around the object. In most cases, the key frame is sufficient to identify the person leaving the object else to find the cause of alarm the entire video shot can be recovered and visualized. In the case of detecting the hidden object the visual image is fused with IR image resulting in the detailed description of the people in the scene and any hidden object detected by the IR image. It is feasible to conclude that the proposed system can efficiently achieve indexing and retrieval task in wide range of application.

## REFERENCES

- Carpenter, G.A. and S. Grossberg, 1991. Pattern Recognition by Self-Organizing Neural Networks. MIT Press, Cambridge, MA.
- Courtney, J.D., 1997. Automatic video indexing via object motion analysis. Pattern Recogn., 30: 607-625.
- Kapur, J.N., P.K. Sahoo and A.K.C. Wong, 1985. A new method for gray-level picture thresholding using the entropy of the histogram. Comput. Vis. Graph., Image Proc., 29(3): 273-285.

- Murino, V., *et al.*, 1994. Interpretazione di Immagini e Segnali Multisensoriali per la Sorveglianza di Stazioni Ferroviarie non Presenziate. (in Italian), Univ. Genoa, Italy, Tech. Rep. DIBE/PFT2/93.019 11.PF74/06.94/VM, DIBE.
- Otsu, N., 1979. A threshold selection method from gray level histograms. *IEEE T. Syst. Man Cy.*, 9(1): 62-66.
- Parker, J.R., 1991. Gray-level thresholding in badly illuminated images. *IEEE T. Pattern Anal.*, 13(8): 813-819.
- Regazzoni, C.S., A. Teschioni and E. Stringa, 1997. A long term change detection method for surveillance applications. *Proceeding of 9th International Conference on Image Analysis and Processing*, Florence, Italy, pp: 485-492.
- Ridler, T.W. and S. Calvard, 1978. Picture thresholding using an iterative selection method. *IEEE T. Syst. Man Cy.*, 8(8): 630-632.
- Rosin, P.L., 2002. Thresholding for change detection. *Comput. Vis. Image Understand.*, 86(2): 79-95.
- Snidaro, L. and G.L. Foresti, 2003. Real-time thresholding with Euler numbers. *Pattern Recogn. Lett.*, 24(9-10): 1533-1544.
- Yuille, A., L. Vincent and D. Geiger, 1992. Statistical morphology and Bayesian reconstruction. *J. Math. Imag. Vis.*, 1(3): 223-238.