



## Intelligent Surveillance System for Street Surveillance

**Y. M. Mustafah<sup>1\*</sup>, N. A. Zainuddin<sup>1</sup>, M. A. Rashidan<sup>1</sup>, N. N. A. Aziz<sup>1</sup> and M. I. Saripan<sup>2</sup>**

<sup>1</sup>*Mechatronics Engineering Department, Kulliyah of Engineering, International Islamic University Malaysia, 50728 IIUM, Kuala Lumpur, Malaysia*

<sup>2</sup>*Department of Computer & Communication Systems, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia*

### ABSTRACT

CCTV surveillance systems are widely used as a street monitoring tool in public and private areas. This paper presents a novel approach of an intelligent surveillance system that consists of adaptive background modelling, optimal trade-off features tracking and detected moving objects classification. The proposed system is designed to work in real-time. Experimental results show that the proposed background modelling algorithms are able to reconstruct the background correctly and handle illumination and adverse weather that modifies the background. For the tracking algorithm, the effectiveness between colour, edge and texture features for target and candidate blobs were analysed. Finally, it is also demonstrated that the proposed object classification algorithm performs well with different classes of moving objects such as, cars, motorcycles and pedestrians.

*Keywords:* Adaptive background modelling, distributed cameras tracking, object detection, object classification, intelligent surveillance

### INTRODUCTION

Surveillance system is becoming more common in Malaysia. Many areas such as banks, shop areas, pedestrian streets and many more areas are monitored by CCTVs. Moreover, the surveillance systems are getting cheaper nowadays, making its easily deployable (Shearing & Johnston, 2013). What is lacking is, an active monitoring of the surveillance video footage to detect, track and classify the moving object accurately. Having human operator to monitor the video feed is very costly and inefficient. Human tend to become bored due to the dull nature of the monitoring

---

#### Article history:

Received: 02 March 2016

Accepted: 14 December 2016

---

#### E-mail addresses:

yasir@iium.edu.my (Y. M. Mustafah),  
fiqahzainuddin@gmail.com (N. A. Zainuddin),  
ariff rashidan@gmail.com (M. A. Rashidan),  
normadirahaziz89@gmail.com (N. N. A. Aziz),  
iqbal.saripan@gmail.com (M. I. Saripan),  
\*Corresponding Author

activity. Thus, the development of intelligent surveillance technology becoming more critical as the number of CCTV is rising. Video processing of surveillance camera has many challenges such as variable illumination condition, inconsistent complex background, and differences of object appearances across camera views (Chen, Huang, & Tan, 2011). Therefore, an intelligent surveillance system that is able to solve background modelling problem, accurately track the objects and correctly classified the detected objects will produce a good prevention tool for security and good monitoring tool for surveillance purposes.

## RELATED WORKS

Intelligent surveillance system for street scene may involve the application of monitoring the traffic, and handle congestion problem. There are many proposed techniques with different focus are proposed to solve the related in the street scene. One of the common approaches to be used is computer vision as taking the advantage of excess supply of CCTV on the streets. Normally for most of the street surveillance system, there are three major stages involved which are moving object detection, tracking and classification.

Particularly for moving object detections, the most common method used is background subtraction. Background subtraction is widely used method and there are a lot of proposed works related to background subtraction (Ridder, O. Munkelt, & Kirchner, 1995) (Stauffer & Grimson, 2000) (Elgammal, Duraiswami, Harwood, & Davis, 2002) (Mukerjee & Das, 2013) (Asaidia, Aarabb, & Belloukic, 2014) (Asif, Javed, & Irfan, 2014) (Hung, Pan, & Hsieh, 2014) (Lee & Lee, 2014) (Nimse, Varma, & Patil, 2014). Basically, in any background subtraction approach, the interested foreground is detected by image differencing and represented in binary form of foreground mask. Information of the foreground mask is essential to the higher level processing, which are tracking and classification. The main difference between each of the proposed methods in background subtraction is the background modelling stage. Stauffer et al. (2000), Mukerjee et al. (2013) and Nimse (2014) used Gaussian mixture model to model the background. Meanwhile, Elgammal et al. (2002) claimed that GMM is ideal for indoor scenes only. Thus, they introduced the use of kernel density estimation (KDE) for background reconstruction. Gao et al. (2009) and Lee et al. (2014) also reconstructed the background using KDE approach.

Meanwhile, for tracking stage, the appearance model gives a priori information about the interested object and can be updated for each new frame. The appearance can be modelled using shape, templates, histograms or parametric representations of distributions, extracted from the object detection stage. The second step in visual tracking is how to use the model to find the location of the object in the next frame. One of the simplest ways is by correlating the new frame with the appearance model and finding the maximum response. Recently, most of the surveillance systems rely on multiple features to model the object's appearance. A multi-feature fusion scheme has achieved high boosting performance or robustness in computer vision field (Deori & Thounaojam, 2014). The commonly used features for tracking are colour, texture, centroid, height, and width.

Similarly, in classification of moving objects, the features extracted from the object detection stage can be used to cluster the interested objects to different classes. There are several

remarkable methods that are proposed to classify the moving objects, which are, neuro-fuzzy classifier, support vector machine (SVM) classifier, K-nearest neighbourhood (kNN) and boosted classifier. Basically, the classification algorithm involves learning and testing phase. In learning phase, features extracted are learned by the classifier to recognize the distinctive attributes of the moving objects.

## PROPOSED SYSTEM

Our proposed system consists of three real-time modules: (1) adaptive background modelling to effectively model the background of the scene for more accurate foreground object detection, (2) optimal trade-off features tracking to track moving objects across distributed overlapping and non-overlapping camera views and (3) detected objects classification to classify moving objects in the surveillance scene.

### Object Detection

Object of interest in street surveillance are normally moving object which can be detected using background subtraction algorithm. We introduce a background subtraction algorithm which consist of Maximum Occurrence Patch based Background Modelling (MOP-BM). As pixel by pixel analysis is not preferable for real-time analysis, image frames are divided into several patches and the analysis of the patches will be done in order to reconstruct the background model.

MOP-BM utilizes the assumption that background pixels are the most frequently observed pixels in the entire video sequence. After the frames have been segmented into  $m \times n$  patches, the first stage of MOP-BM is the maximum occurrence calculation of  $M$  patches. The process is repeated until  $f_i = f_N$ . Based on the maximum occurrence patches, the reconstruction of the background model is built. The patches with foreground objects will be omitted. Figure 1 shows the overall process of MOP-BM.

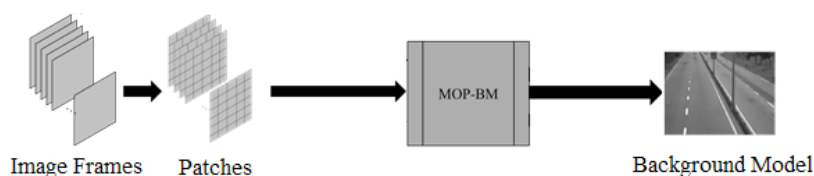


Figure 1. Details of MOP-BM algorithm

Let  $f_1, f_2, f_3, \dots, f_n$  represent the frames from the same video sequence. First, each of the frames must be segmented into  $m \times n$  patches, where  $m$  and  $n$  can be the same number. Assume  $N$  patches are obtained, where we let each patch be marked as in the notation of  $p_1, p_2, p_3, \dots, p_N$ . Experiments were conducted to find the optimum size of the patches. The result of the experiments shows that the size of the patches is not as sensitive as the sampling frame value. It does effect the result, but in a small percentage error. Then, at each frame,  $f_1, f_2, f_3, \dots, f_n$

the intensity value of pixel  $(x,y)$  in the area of specified patch,  $p_i$  is marked as  $i_1, i_2, i_3, \dots$ , in where  $i = 1,2,3, \dots,n$ . Then, mean  $(m_{p,f})$  at each frame for the specified patch is calculated as in the formula below:

$$m_{p,f} = \frac{\sum_{j=1}^q i_j(x, y)}{m \times n} \tag{1}$$

In order to calculate the MOP, the calculated mean is grouped into matrices called *mean\_matrix<sub>p,f</sub>* according to the specified sampling time. Assume that the sampling time is denoted as  $T$ . Thus, the mode at that time interval for patch 1 is:

$$n_{mode} = \max(m_{1,1}, m_{1,2}, \dots, m_{1,T-1}) \tag{2}$$

### Multi-camera Tracking

Tracking the same object within different cameras' view is essential in many surveillance applications. In our tracking algorithm, colour, edges and local binary pattern (LBP) are chosen as the tracking features. For the colour feature, HSV colour space is chosen. The first step is to extract the HSV colour space from the intensity image. The second step is to compute the mean value for each HSV channel. The similarity of each target and candidate blob is computed based on Equation [3]. The object position is computed using Euclidean distance based on Equation [4].

$$dist(HSV) = \frac{\sqrt{(P(H_1) - P(H_2))^2 + (P(S_1) - P(S_2))^2 + (P(V_1) - P(V_2))^2}}{\sqrt{P(H_1)^2 + P(S_1)^2 + P(V_1)^2} + \sqrt{P(H_2)^2 + P(S_2)^2 + P(V_2)^2}} \tag{3}$$

$$d_{Euclidean} = (p_1 - q_1)^2 + (p_2 - q_2)^2 \tag{4}$$

where  $p = (p_1, p_2)$  and  $q = (q_1, q_2)$ .

In order to make the system to be more robust, edge is used as an additional feature since it is insensitive to illumination. Object appearance may be differentiated in terms of its clothing such as checked shirt or plain shirt. It can be differentiated by using the edge feature of the clothing. The edge feature is obtained by using Canny Edge Detector. The frequency value of the edge is then used to calculate the similarity of tracked object. The similarity of each target blob and candidate blob are computed based on Equation [5].

$$dist(P(p_1) - P(p_2)) = \frac{\sqrt{(P(p_1) - P(p_2))^2}}{\sqrt{P(p_1)^2} + \sqrt{P(p_2)^2}} \tag{5}$$

The computational of LBP feature for each patch is represented by  $mn$  matrix, which is number of rows and columns respectively. The patch matrices are concatenated to obtain a set of decimal values. Then, the decimal values are divided into 5 groups of bins. The similarity between target and candidate blob are computed based on Equation [6].

$$dist(P(p_1) - P(p_2)) = \frac{\sqrt{(P(p_1) - P(p_2))^2}}{\sqrt{P(p_1)^2 + \sqrt{P(p_2)^2}}} \quad [6]$$

The features are combined based on Equation [7], where  $N$  is the received candidates. Each candidate, has  $K$  different features, namely colour, texture and edge features. For each feature, similarity score is calculated and given a weight. The similarity scores are referred to Equation [3], [5], and [6].

$$O = \arg \max_{i \in N} \sum_{j \in K} (w_j s^i_j) \quad [7]$$

### Classification Model

Our classification of moving objects is based on Adaptive Neuro-Fuzzy Inference System (ANFIS). ANFIS is a general artificial intelligence connectionist model of inference system, and its structure is shown in Figure 2. In this model, the reasoning system is based on Takagi-Sugeno-Kang (TSK) rules which constitutes of five layers, and the function of each layer is interdependent on each other.

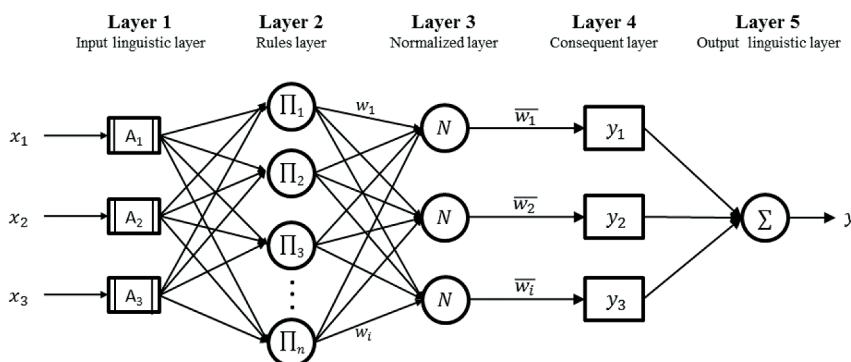


Figure 2. Structure of the ANFIS

In the training stage, three FIS were developed, in which each system is specifically for each output class. Each network has one output node, which will eventually be adjoined to recognized three different classes of moving objects including pedestrian, motorcycle, and car. Initially, the ANFIS training was done using our own datasets. All moving objects in the dataset are divided into three categories: pedestrian, motorcycle, and car. 200 positive samples, 1434 negative samples, and 200 test samples were randomly selected for each class in the ANFIS parameter training. The output of the network is in terms of probability value in the range of  $[0, 1] \in \mathbb{Z}^+$ . As the desired outputs were to be in finite value, a simple threshold is applied. Therefore, the outputs were (1, 0, 0), (0, 1, 0), and (0, 0, 1) for the pedestrian, motorcycle, and car respectively. During the testing stage, the feature vector of each region-of-interest that

detected by segmentation process will be extracted. This discriminant vector will be given to the network as an input. The output of linguistic layer will determine the belonging class of the moving object, which would be in the sequence of  $(y_1, y_2, y_3)$ .


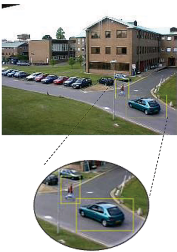
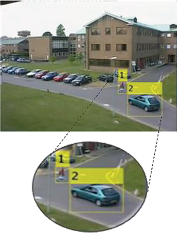
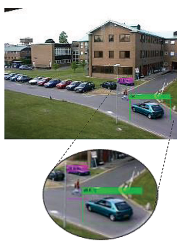





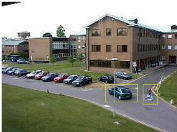

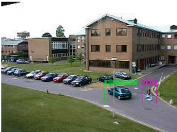
## RESULTS AND DISCUSSION

PETS 2001 dataset is used to evaluate the performance of the proposed system. The dataset consists of 2688 frames with interested class of pedestrians and car. Table 1 shows the parameter setup for moving object detection, tracking and classification stage of the proposed system. Table 2 shows the result of proposed system for three random frames, 535, 578 and 596 respectively. From the table, we can deduce that, in the object detection stage, the moving objects are successfully detected and negative effect from the surrounding such as illumination variation are suppressed. The IDs for tracking the moving objects also managed to be assigned correctly. As for the moving classification, the car and pedestrian are annotated by the correct pre-defined blob colour for each classes.

Table 1  
*Initial System Parameters*

Stage	Parameter	Value
Pre processing	Frame spatial resolution	240×320 pixels
	Median filtering structuring element	5×5
Moving object detection	Sampling Frame	$F_s = 80$
	Patch Size	$P_N = 10 \times 10$
Moving object tracking	Threshold for Hue Colour	$T_{HUE} = 0.27$
	Threshold for Edge & Texture	$T_T T_E = 0.07$
	Threshold for Top Weighted Ratio	$T_{Cl\_top} = 0.04$
	Threshold for Bottom Weighted Ratio	$T_{Cl\_bottom} = 0.04$

Table 2  
*Result of Proposed System for PETS 2001 Dataset*

Image Frame	Detected Foreground	Object Detection	Object Tracking	Object Classification
#0535				
#0578				
#0596				

Graph in Figure 3 shows the overall performance for detection, tracking and classification for PETS 2001 dataset. From the quantitative result, we can see that the proposed system achieved more than 90% in accuracy, precision and recall value for detection stage. At same time, the proposed system also get high value of accuracy, precision and recall for tracking stage with 0.8974, 0.9412, and 0.9488 respectively. For the classification, the high value for accuracy, precision and recall indicates that the proposed system able to differentiate the moving objects accurately.



Figure 3. Overall Performance of the Proposed System with PETS 2001 Dataset

## CONCLUSION

This paper presents a solution for background modelling, object tracking and object classification for intelligent surveillance system. From the result of background modelling, it shows that the proposed system able to reconstruct the background successfully and there is no need to have the non-moving object in the video sequence to reconstruct the background. For tracking moving objects under distributed cameras' view, hybrid of features, namely position, edge, colour and texture are combined. From the experimental results, it shows that the tracking performance using multiple features benefit the system particularly on the accuracy value. For classification of the objects, the classifier was developed using Neuro-Fuzzy approach. From the experimental results, it shows satisfactory results in detecting different classes of objects which are the pedestrian and car under complex street scene condition. The proposed system would be very helpful for in increasing the efficiency of CCTV system.

## REFERENCES

- Asaidia, H., Aarabb, A., & Belloukic, M. (2014). Shadow elimination and vehicles classification approaches in traffic video surveillance context. *Journal of Visual Languages & Computing*, 25(4), 333–345.
- Asif, S., Javed, A., & Irfan, M. (2014). Human Identification On the basis of Gaits Using Time Efficient Feature Extraction and Temporal Median Background Subtraction. *International Journal Image, Graphics and Signal Processing*, 3(2), 35-42.
- Beale, M. H., Hagan, M. T., & Demuth, H. B. (2010). *Neural Network Toolbox 7. User's Guide*. MathWorks.
- Breitenstein, M. D., Reichlin, F., & Leibi, B. (2011). Online Multiperson Tracking by detection from a Single, Uncalibrated Camera. *Pattern Analysis and Machine Intelligence*, 33(9), 1820-1833.



- Chen, X., Huang, K., & Tan, T. (2011). Direction-based stochastic matching for pedestrian recognition in non-overlapping cameras. In *IEEE 18th International Conference Image Processing* (pp. 2065-2068). IEEE.
- Deori, B., & Thounaojam, D. M. (2014). A Survey on Moving Object Tracking in Video. *International Journal on Information Theory (IJIT)*, 3(3), 31-46.
- Elgammal, A., Duraiswami, R., Harwood, D., & Davis, L. (2002). Background and Foreground Modeling Using Nonparametric Kernel Density Estimation for Visual Surveillance. In *Proceedings of the IEEE*, 90 (pp. 1151-1163). IEEE.
- Gao, T., Zhang, J., Gao, W., & Liu, Z. (2009). A Robust Technique for Background Subtraction in Traffic Video. In *15th International Conference on Neural Information Processing* (pp. 736-744). Springer Berlin Heidelberg, Auckland, New Zealand.
- Hung, M. H., Pan, J. S., & Hsieh, H. C. (2014). A Fast Algorithm of Temporal Median Filter for Background Subtraction. *Journal of Information Hiding and Multimedia Signal Processing*, 5(1), 33-41.
- Khosravi, A., Nahavandi, S., Creighton, D., & Atiya, A. F. (2011). Comprehensive review of neural network-based prediction intervals and new advances. *IEEE Transactions on Neural Networks*, 22(9), 1341-1356.
- Kim, K., Chalidabhongse, T. H., Harwood, D., & Davis, L. (2005). Real-Time Foreground-Background Segmentation using Codebook Model. *Real-Time Imaging*, 11(3), 172-185.
- Lee, S., & Lee, C. (2014). Low-complexity background subtraction based on spatial similarity. *EURASIP Journal on Image and Video Processing*, 2014(30), 1-16.
- Mukerjee, S., & Das, K. (2013). An Adaptive GMM Approach to Background Subtraction for Application in Real Time Surveillance. *International Journal of Research in Engineering and Technology*, 2(1), 25-29.
- Nimse, M., Varma, S., & Patil, S. (2014). Shadow Removal Using Background Subtraction and Reconstruction. *International Journal of Emerging Technology and Advanced Engineering*, 4(4), 324-327.
- Ridder, C., Munkelt, O., & Kirchner, H. (1995). Adaptive Background Estimation and Foreground Detection Using Kalman Filter. In *Proceedings of International Conference on Recent Advances in Mechatronics* (pp. 193-199). Istanbul, Turkey.
- Shearing, C. D., & Johnston, L. (2013). Public and Private Interest Overlap. In *Governing Security: Explorations of Policing and Justice* (p. 121). Routledge.
- Sonka, M., Hlavac, V., & Boyle, R. (2008). Image Processing, Analysis, and Machine Vision. In *Mathematical morphology* (pp. 559-597). Thomson.
- Stauffer, C., & Grimson, W. E. (2000). Adaptive Background Mixture Models for Real-Time Tracking. In *IEEE Computer Society Conference* (Vol. 2). IEEE. Colorado.
- Vilas, G. L., Spyrakos, E., & Palenzuela, T. (2011). Neural network estimation of chlorophyll from MERIS full resolution data for the coastal waters of Galician. *Remote Sensing of Environment*, 115(2), 524-535.
- Zheng, Y., Ma, X., Zhao, X., & Wang, X. (2011). Mean shift target tracking algorithm based on color and edge features. *Journal of Optoelectronics*, 8(26), 1231-1235.

