

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Human Silhouette Extraction Using Background Modeling and Subtraction Techniques

S. Sulaiman, A. Hussain, N. Tahir, S.A. Samad and M.M. Mustafa  
Department of Electrical, Electronics and Systems, Faculty of Engineering,  
Universiti Kebangsaan Malaysia, 43600 UKM Bangi, Selangor DE, Malaysia

**Abstract:** The main objective of this study is to develop an algorithm that is capable of detecting the presence of human based on motion and background subtraction technique along with post-morphology processes to extract the foreground pixels from its background thus separating the human silhouette. The extracted silhouette information can later be used for traffic monitoring and analysis, human tracking and monitoring, video surveillance since silhouette-based technique tend to offer speed and simplicity. Results obtained indicate that the developed algorithm achieves its objective and successfully extract human silhouette from the analyzed video scenes.

**Key words:** Silhouette extraction, background modeling and subtraction, scene analysis, human motion detection

### INTRODUCTION

A very fundamental and critical task in any human tracking and monitoring system or video surveillance system is to identify moving objects (Elbasi *et al.*, 2005; Bharu and Zou, 2004). A common approach to this critical task is to first perform background modeling to yield a reference model. This reference model is used in background subtraction in which each video sequence is compared against it to determine possible deviation. A deviation between current video scenes to that of the reference in terms of pixels signifies existence of moving objects. The deviation which also represents the foreground pixels are further processed for object localization and extraction. In this research project, a method for detecting human motion, performing background subtraction and subsequently extracting its silhouette information is presented. This measure is the first step in computer vision applications such as in traffic monitoring and analysis system. In such a system, human or pedestrians are the most vulnerable traffic participants (Curio *et al.*, 2000; Gavila 2000; Xiong and Jaynes, 2004).

The main objective of this study is to develop an algorithm that can detect human motion at certain distance. The idea is to place the system as depicts in Fig. 1. Then, video scene analysis of the recorded scene is performed to detect and extract moving objects specifically human. Such a detection system could be



Fig 1: On-board implementation of the pedestrian detection system (Picture source: Gavila, 2000)

useful in assisting and providing the driver with convenient information of the presence of the pedestrians (Gavila, 2000).

Basically, the task can be divided into 3 subtasks that is, motion detection, background modeling and subtraction and foreground object detection.

### MATERIALS AND METHODS

The motion detection and silhouette extraction algorithm consists of several sequential processes. Overall, the processes can be described using a flow chart (Fig. 2).

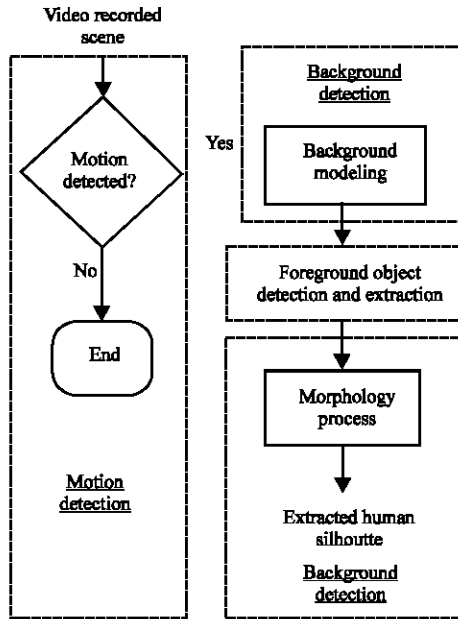


Fig. 2: Flow chart for the human silhouette extraction

**Motion detection using sum of absolute difference (SAD):**

Motion detection technique is used to detect any motion that exists in a scene. In this research, we use the Sum of Absolute Difference (SAD) algorithm that is based on the image differencing techniques. It is mathematically represented using the following equation:

$$D(t) = \frac{1}{N} \sum |I(t_i) - I(t_j)| \quad (1)$$

Where,

N = The number of pixels in the image and also used as the scaling factor

I(t<sub>i</sub>) and I(t<sub>j</sub>) = The images at time i and j, respectively while

D(t) = The normalized sum of absolute difference for that time instance. In an ideal case, when there is no motion, the following condition holds

$$I(t_i) = I(t_j) \text{ and } D(t) = 0 \quad (2)$$

Nevertheless, in most instances the images consist of noise and therefore, a better model for an image in the absence of motion will be:

$$I(t_i) = I(t_j) + n(p) \quad (3)$$

where, n(p) is noise signal.

**Background subtraction and background scene modeling:**

Background subtraction is a commonly used technique for segmenting out objects of interest in a scene. This technique involves subtracting an image that contains the objects, with the previous background image that has no foreground objects of interest. The areas of the image plane where there is a significant difference within these images indicates the pixel location of the moving objects (Haritaoglu *et al.*, 2000; McIvor, 2000). These objects, which are represented by groups of pixel, are then separated from the background image by using the threshold technique. In outdoor environment, several characteristics such as scene illumination and local surface properties and orientation may change over time. Thus, a simple background subtraction technique does not work well for these scenes, as it cannot handle the small motion and changes of the background pixels. In order to improve the technique, we use the background scene modeling technique, where sequences of several background images are observed during a training period (Haritaoglu *et al.*, 2000; McIvor, 2000; Monnet *et al.*, 2003). The background scene is modeled by representing each pixel by three values, which are:

- Minimum m(x), intensity values.
- Maximum n(x) intensity values.
- Maximum difference d(x) between consecutive frames.

**Foreground object detection:** The foreground object of interests, in this case are the pedestrians. The foreground objects are segmented from the background frame of the current video sequences by performing several processes such as thresholding, post-morphological image processing and object detection. To segment out the foreground pixel, each pixel of an observe image is classified by using the background model. By taking the minimum m(x) and maximum n(x) intensity values and the maximum difference d(x) between consecutive frames that represent the background scene model B(x), pixel x from image I is a foreground pixel if:

$$B(x) = \begin{cases} 0 & \text{background} & \begin{cases} (I(x) - m(x)) < kd(x) \\ (I(x) - n(x)) < kd(x) \end{cases} \\ 1 & \text{foreground} & \text{otherwise} \end{cases} \quad (4)$$

where, k is the Gaussian distributions constant, in this case we consider k is ranging from 2 to 5.

**RESULTS AND DISCUSSIONS**

The analyzed video involves scenes of both indoor and outdoor modes. However, in this research, the analysis mainly focuses on the outdoor scene analysis

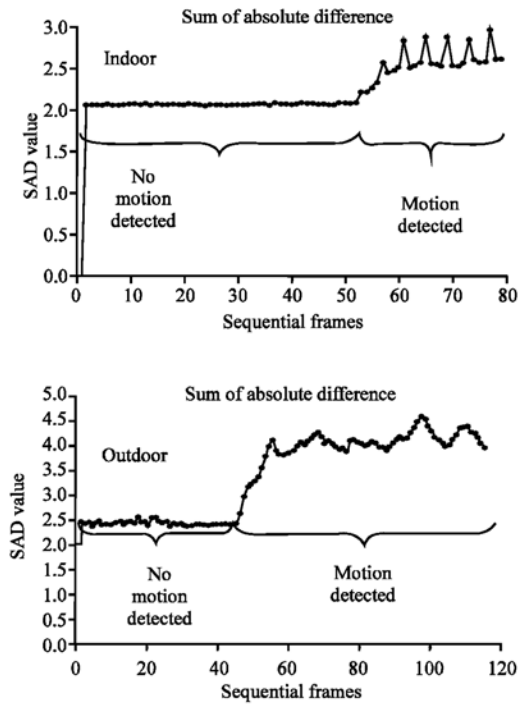


Fig. 3: SAD graphs for both indoor and outdoor scenes

since it will reflect the real situation in a human motion detection system for surveillance application. As mentioned earlier, outdoor scene analysis is much more challenging than the indoor scene. A monocular approach using a single video camera is used to capture the outdoor scene consisting of various human actions. Prior to that, the background scene without any foreground objects is recorded for 2-4 sec.

To show the object motion detection results, the Sum of Absolute Difference (SAD) values for each image frames is plotted on a graph. The results for both indoor and outdoor scene analysis results. It can be seen that the indoor object motion is easier to detect as compared to the outdoor scene (Fig. 3). SAD graph of the outdoor scene normally consists of noise due to variations in illumination, local surface and orientation which are beyond the operator's control and as such, makes it more difficult to analyze than the indoor scene. In this work, the main aim is to perform pedestrian detection and therefore, we have applied background scene modeling to achieve our goal. The background scene is modeled using a set of background image frames, which basically consists of 5-50 consecutive images.

Next, using Eq. 4, the object pixels are segmented out from its background followed by post-morphological processes such as erosion and dilation to eliminate noisy pixels (Gonzalez and Woods, 2000) thus producing better results. However, some limitations

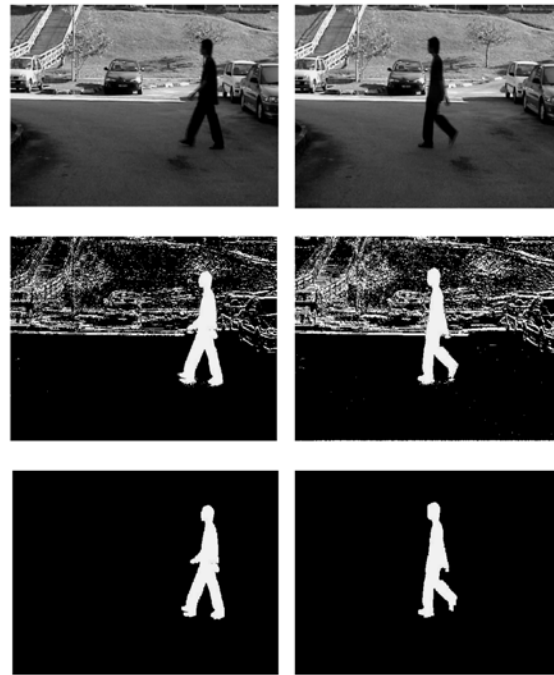


Fig. 4: Result for background subtraction of human; original image (top), extracted foreground (middle) and extracted silhouette (bottom)

have been imposed on the developed system. In this implementation, we avoid uncontrollable conditions such as illumination changes, movement of other objects such as swaying trees and moving vehicles and also varied weather condition. As such, in our scene analysis we have only considered a stable and fine weather scene and avoided scenes with swaying trees.

The developed algorithm is experimented to detect human motion in three main cases namely single human image, single human while carrying an object and finally multiple human motion. All the detections are performed in an outdoor environment. Figure 4-7 depict the results of the implementation for various human poses namely walking, walking while carrying backpack, walking with a carry-bag and walking while holding a file. In Fig. 8, results for the multiple human cases are presented.

Present study has successfully achieves its objective specifically to detect human motion and extract its silhouette information using the background subtraction algorithm developed. Motion detection is performed once the system detects the presence of motion in the video scene input. Results obtained are similar to findings by Stauffer *et al.* (1999).

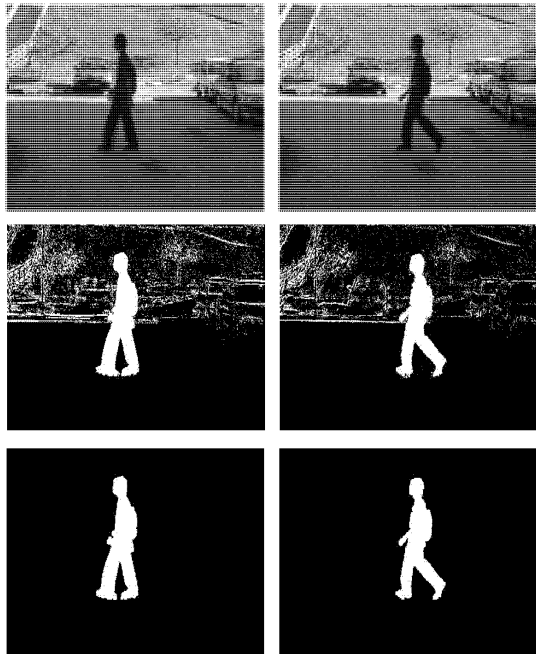


Fig. 5: Result of silhouette extracted for human carrying a backpack: original image (top), extracted foreground (middle) and extracted silhouette (bottom)

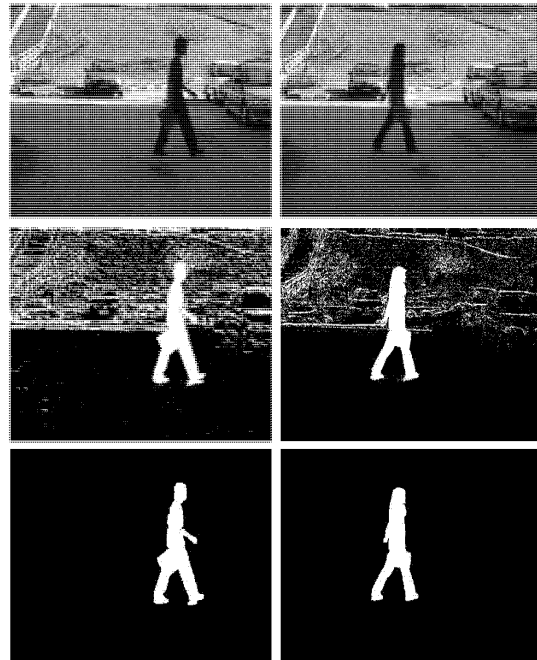


Fig. 7: Result of silhouette extracted for human carrying a file his side; original image (top), extracted foreground object (middle) and extracted silhouette (bottom)

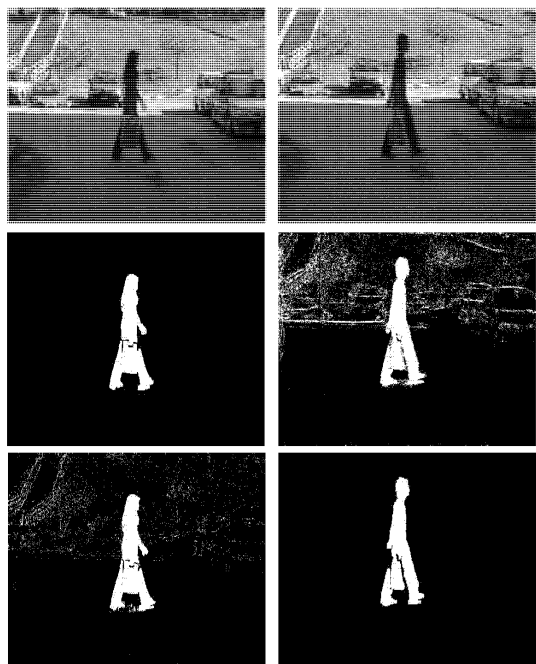


Fig. 6: Results for the case of human carrying a bag: original image (top), extracted foreground (middle) and extracted silhouette (bottom)

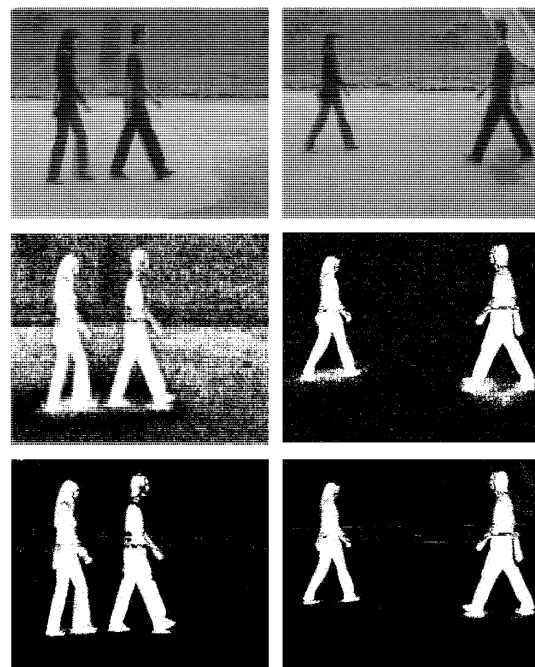


Fig. 8: Results for the case of multiple human detection: original image (top), extracted foreground object (middle) and extracted silhouette (bottom)

## CONCLUSIONS

In conclusion, an algorithm capable of detecting motion and extracting object information which involves human as object has been described. The algorithm involves modeling of the desired background as a reference model for later used in background subtraction to produce foreground pixels which is the deviation of the current frame from the reference frame. The deviation which represents the moving object within the analyzed frame is further process to localized and extract the information. Further work is being pursued to remove the constraints imposed in this implementation to make the algorithm more impermeable and relevant to the pedestrian detection system of an intelligent vehicle.

## REFERENCES

- Bhanu, B. and X. Zou, 2004. Moving human detection based on multi-modal sensor fusion. *Proceeding of IEEE Computer Vision and Pattern Recognition*.
- Curio, C., J. Edelbrunner, T. Kalinke, C. Tzomakas and W. Seelen, 2000. Walking pedestrian recognition. *IEEE Trans. Intell. Transport. System*, 1: 155-163.
- Elbasi, E., L. Zuo, K. Mehrotra, P. Mohan and P. Varshney, 2005. Control charts approach for scenario recognition in video sequences. *Turk. J. Elect. Eng.*, pp: 13.
- Gavrila, D.M., 2000. Pedestrian detection from a moving vehicle. In: *Proceeding of European Conference on Computer Vision*, pp: 37-49.
- Gonzalez, R. and R. Woods, 2000. *Digital Image Processing*. 2nd Edn., Prentice Hall, New Jersey, USA.
- Haritaoglu, I., D. Harwood and L.S. Davis, 2000. W4: Real-time surveillance of people and their activities. *IEEE Trans. Pattern Anal. Machine Intell.*, 22: 809-830.
- McIvor, A.M., 2000. Background subtraction techniques. In: *Proceeding of Image and Vision Computing*, 1: 155-163.
- Monnet, A., A.M.N. Paragios and V. Ramesh, 2003. Background Modeling and Subtraction of Dynamic Scene. In *Proceeding of Ninth IEEE International Conference on Computer Vision*, 2: 1305-1312.
- Stauffer, C. and W.E.L. Grimson, 1999. Adaptive background mixture models for real-time tracking. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2: 246-252.
- Xiong, Q. and C. Jaynes, 2004. Multi-resolution background modelling of dynamic scenes using weighted match filters. In: *Proceeding of the ACM 2nd International Workshop on Video Surveillance and Sensor Networks, IEEE Conference on Multimedia*, New York, NY, USA., pp: 88-96.