Detection and Classification of a Moving Object in a Video Stream

Asim R. Aldhaheri and Eran A. Edirisinghe

Abstract— In this paper we present a new method for detecting and classifying moving objects into humans and vehicles from a video surveillance scene. In this approach, the moving objects are firstly detected from the background using a background subtraction technique. Background subtraction algorithms are implemented in a MATLAB environment. A comparison for all the algorithms was made to determine which background subtraction algorithm performs better with the proposed classifying algorithm. The algorithms were then tested using more than a video and many frames of the video was tested. Secondly, edge detection of moving objects was performed using Canny or Prewitt operations, while bounding boxes were implemented over the moving objects. Edge detection and bounding boxes were used before the classification step to simplify the classification.

Finally, classifying the moving object into humans and vehicles was accomplished by finding the height-width ratio of the bounding box around the moving object in each frame and estimating the speed of the moving object from two consecutive frames in the video stream. The results found were fairly good and the wrong classification was due to the bounding boxes not correctly covering the moving objects. The background subtraction step and the classification step were tested using different video sequences. A comparison was made between the combination of each background subtraction technique and the classification algorithm; the false classification results are given. Although the proposed classification algorithm in this paper is simple and easy to be implemented, the results obtained are very satisfactory and the accuracy is very good

Keywords—background subtraction, classification, moving object detection.

Asim R. Aldhaheri
Department of Electrical & Computer Engineering
King Abdulaziz University
P.O. Box 80204
Jeddah 21589, Saudi Arabia

Eran A. Edirisinghe
Department of Computer Science
Loughborough University
Loughborough, UK.

I. Introduction

The classification and detection of an event in a video camera scene is a relatively new research area in computer science and, because of its broad applicability in real life this has been growing more and more. The CCTV is one of the main reasons for the growing interest and use of video in security systems [1], [2]. However, the majority of the CCTV systems that are currently available in the market have a limited functionality which comprises only the capture, storage and visualisation of a video that is gathered from one or more cameras. However, some CCTV systems already include motion detection algorithms and are able to constrain the video recording only when a variation in the scene's foreground is detected. The main utility of the system is recording conventional and non-conventional events and detecting the objects and non-conventional events that appeared in the scene. However, the majority of these systems do not have any embedded intelligence which is able to provide a classification of the events. For example, there is no mechanism to warn operators when non-conventional events occur. Such an attribute would be very helpful in preventing and detecting, in an active fashion, the occurrence of a non-conventional event.

Although detection and classification moving objects is widely used in security systems, it is also used in other applications such as video compression and robot technology. Detecting and classifying an object from a video sequence is a challenging task because of the requirement for real time computation and the large amount of data that are used.

A. *Moving Object Detection*

Moving object detection in a video stream is an essential step in video surveillance applications [1], [3-5]. Moreover, it is also an important research topic in the field of computer vision. The main difficulty with moving object detection is caused by changes in the scenes. In some algorithms, the moving objects may become part of the scene when they come to a stop. Also the scene maybe affected by changes in the light, leaves swaying, cameras shaking, etc.

Many algorithms for moving object detection have been proposed in recent years. These involve background subtraction, optical flow, temporal difference and many other algorithms for detecting moving objects. From these, the most widely used algorithm is background subtraction which has many algorithms such as frame difference, approximate median, Gaussian mixture, Running Gaussian and Kernel density background.

B. *Object Classification*

After detecting an object, the final step is to analyze the objects in order to identify their behavior. In this step, classification is used to identify whether the detected object is a vehicle or a human [6]. There are many algorithms and approaches that can be used for this, depending on various measurements to differentiate between humans and vehicles. The approach used in this study is mainly dependent on finding the height-width ratio of the bounding box around the moving object and on the moving speed of the object, provided that the human and the vehicle are in motion.

C. *Programming Steps for the Proposed Detection and Classification Algorithm*

The first step is to read the video stream. The video can take any format such as: AVI, MPEG, etc. Each frame in the video can include one or more than one moving object. Sometimes, the movie frames need to be resized to 120 by 160 pixels to reduce the computational complexity. Moreover, if the video is large, the number of frames is reduced to approximately 300 to 400 frames. The next step is to apply one of the back ground subtraction techniques to detect the moving object. The different background subtraction algorithms are discussed in detail in Section II.

The Region of Interest (ROI) is determined and labelled by drawing a bounding box around the moving object that is of interest.

The final step is to apply the classification algorithm to the object to classify whether it is a human or a vehicle. The classification algorithm is discussed in Section III and the results from this work are presented in this section.

II.   Background Subtraction Techniques

Background subtraction is a process commonly used for detecting moving objects in a video scene that uses a static camera [1-16]. The difference between the current frame and the reference frame is the approach that is used to detect the moving object. The background frame is often called the background image. The background image in the majority of background subtraction methods uses an image of the scene with no moving object. Moreover, the background image must be updated so it can adapt to the conditions in terms of changes of light, wind movement or the intrusion of any unwanted moving object. Both newcomers and experts can be confused about the benefits and limitations of all these methods which vary in terms of space, speed, memory requirement and accuracy. The methods that have been reviewed, implemented and tested here are: Frame Difference Method (FD), Approximate Median (AM), Modified Approximate Median (MAM), Running Gaussian Average (RG) and Mixture of Gaussian (MOG).

A. *Frame Difference Method*

The frame difference method is the simplest form of background subtraction [1], [3-7], [9-16]. In this method, the current frame is simply subtracted from the background frame. If the difference in pixel values for every pixel is greater than the threshold T_S, then the pixel is considered part of the foreground.

The Frame Difference method for background subtraction has been implemented and tested using MATLAB. The input in the Frame Difference Function takes the video in a gray or coloured scale and outputs the background frame, the video in gray scale and the video after removing the background. The Frame Difference Function in MATLAB takes the first frame as a background. This video contains a human as the moving object. Shots for the video showing the background frame, a frame in the video with the moving object and the same frame in gray scale after removing the background are presented Fig.1.

Fig.1 Background Subtraction using Frame Difference

B. *Approximate Median Method*

In this method, the median filtering buffers the previous N frames of the video. Then, the background frame is calculated from the median of the buffered frame and the background is subtracted from the current frame to give the foreground pixel. The Approximate Median method checks whether the pixel in the current frame has a value that is larger than the corresponding background pixel. If that is the case, the background pixel is incremented by one. However, if the pixel in the current frame has a value that is smaller than the corresponding background pixel, the background pixel is decremented by one.

The Approximate Median function that is implemented in MATLAB carries out a simple subtraction between the median frame and the test frame in order to obtain the foreground frame; this is done for every frame in the video. After the subtraction, the difference frame is compared with the threshold to obtain an accurate foreground.

The Approximate Median function was tested on a VisTest_01.mpeg video, using 6 Median frames and a threshold of 20 as shown in Fig.2.

International Journal of Advances in Computer Science and Its Applications– IJCSIA
Volume 4: Issue 2   *[ISSN: 2250-3765]*

Publication Date : 25 June 2014

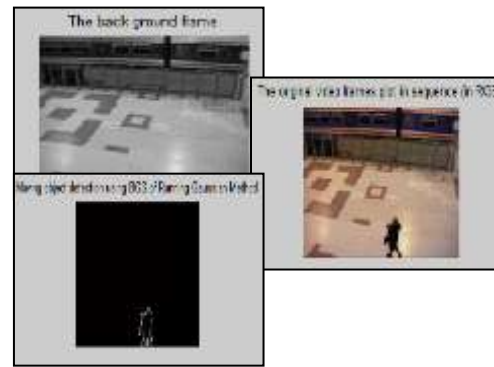Fig.2 Background Subtraction using Approximate Median

Fig.3 Background Subtraction using Running Gaussian

C. *Modified Approximate Median Method*

The difference between the Modified Approximate Median and the ordinary Approximate Median is as follows. After the subtraction between the test frame and the median frame has been implemented in order to obtain the frame without the background and with only the moving object in view, the background frame is increased by one if the test frame is larger than the background frame and decreased by one if the test frame is smaller than the background frame. Also, a two dimensional median filter is applied to the foreground frame to reduce the noise.

D. *Running Gaussian Average Method*

The running Gaussian Average algorithm is based on fitting a Gaussian probability density function (pdf) to the last n pixel's values [4-6], [12], [14]. A running average method is computed in order to avoid fitting the pdf from scratch at the time of each new frame. The running average method is:

$$\mu_t = \alpha I_t + (1 - \alpha)\mu_{t-1}$$

where μ_t is the previous average, α is the empirical weight and I_t is the pixel's current value. The standard deviation, δ_t is the other parameter of the Gaussian pdf that needs to be computed. After the two parameters (μ_t, σ_t) are calculated, the I_t pixel's value can then be classified as a foreground pixel at each t frame if the following inequality holds:

$$I_t - \mu_t > k\sigma_t$$

Otherwise, I_t will be classified as a background pixel.

The Running Gaussian function was implemented on the video hu.avi, using 10 median frames and an alpha equalling 0.02. The background and frame 70 before and after the background subtraction is shown in Fig.3.

E. *Mixture of Gaussian Method*

The Mixture of Gaussian (MoG) is a method that can handle multi-modal distribution [3], [4], [6], [10], [11], [15]. For example, in a scene with a leaf waving against a blue sky, there are two modes: the leaf and the sky. In the MoG method, both objects can be filtered out and each pixel location is represented by a mixture of Gaussian functions that come together to form a probability distribution function F:

$$F(i_t - \mu) = \sum_{i-1}^{k} w_{i,t}.\eta(\mu, \sigma)$$

The Mixture of Gaussian function was tested on the MOV04164.MPG movie and the results of Frame number 100 before and after applying the function is shown below.

Fig.4 Background Subtraction using Mixture of Gaussian

The background subtraction worked quite well using the Mixture of Gaussian algorithm, but there was noise, and the detection of the moving object using the background subtraction algorithm did not appear to be very accurate in this test.

Some of the most relevant background subtraction methods have been described above. Comparison can be made regarding the methods' complexity in terms of memory, speed, requirement and accuracy, which can act as an effective guide when selecting the best method for a specific application.

III. Comparing the Background Subtraction Algorithms

All the algorithms that were implemented in this study were tested on more than one video. Also, any algorithm that included a constant was tested with various constant values in the range of that constant. A comparative performance analysis of the background subtraction algorithms was carried out based on speed, memory requirement, background image accuracy and the accuracy of the background subtraction.

To compare all the background subtraction algorithms would require many tests to be carried out on various kinds of video, such as videos with one object, such as a human or a vehicle, and others with more than one object of the same kind, or a mixture of vehicles and humans. A test for all the background subtraction algorithms on video: 'hu.avi' and frame number 100 is shown below.
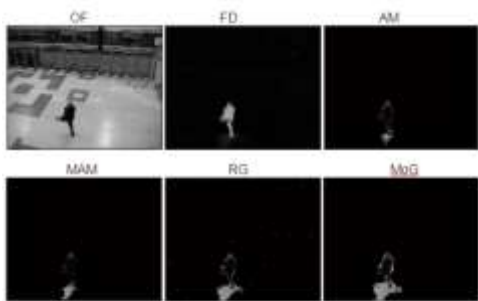

Fig.5 Test for using various Background Subtraction algorithms

This is an example of a test that was performed; many other tests were carried out on many videos taking many frames. After conducting this large number of tests, a conclusion concerning the differences in performance was reached in terms of speed, memory requirement and background subtraction accuracy.

IV. Human and Vehicle Classification

To keep track of the moving objects, bounding boxes and edge detection are employed to classify whether the moving object is a human being or a vehicle. The bounding boxes are drawn around the moving object. These bounding boxes with the edge detection simplify the tracking by refereeing to the boxes whenever necessary rather than referring to object itself.

A. Bounding Boxes

The bounding box for an object is a rectangular box with sides that are parallel to the coordinate planes that contain the object. In MATLAB a bounding box is implemented around the moving object in every frame of the video so every moving object will have a bounding box around it. This process is illustrated in Fig.6.

B. Edge Detection

Edge detection is a fundamental tool used in most image processing applications to obtain information from the frames as a step to feature extraction and for object segmentation. In the process of detecting the object, it detects the outlines of an object and boundaries between objects and the background.

Prewitt and Canny edge detection algorithms were implemented and tested on many frames for more than one video; the accuracy of the two algorithms varied from one video to another. A frame from a video with a moving object with the bounding box and another with Prewitt edge detection is shown in Fig.6.
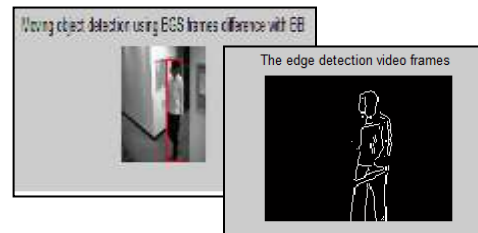

Fig.6 Bounding box and Prewitt edge detection

V. Classification function

The classification function is employed after the background subtraction function is implemented on the video in order to classify the moving object. After this has been accomplished, the bounding box function is implemented on the moving object. In the classification function in this study, the important fields were extracted from the moving object to distinguish between a human and vehicle. The basic factors extracted from the moving object are the ratio and the speed of the moving object. The ratio is calculated via the height and the width of every object detected in the bounding box. The speed is calculated via the moving distance of the object and the time of the movement.

The main advantage of the classification function is that it is implemented after any background subtraction algorithm: i.e. Frame Difference, Approximate Median, Modified Approximate Median, Running Gaussian or Mixture of Gaussian. After the background subtraction function has been implemented and the moving object detected, the classification function is then implemented to classify the object, depending on the ratio and the speed.

A. Classification for one moving human

The procedure outlined in Sections 4 and 5 was tested on the video 'hu.avi', the frames that were tested are frames 50 to 215 which contain one moving human, Fig. 7 shows how the moving object is tracked by drawing the bounding box around it after removing the background.

Fig.7 Test1, classification frames 130 and 131

The classification accuracy is very high in this test and it gave more than 97%. In other frames when the detection is not succeeded or when bounding box is not covering the whole object, the accuracy drop down, see Table 1 for some of the tests that we run.

B. *Classification for moving vehicles*

This test was carried out on video 'traffic.avi' that contained more than one moving vehicle. The classification was performed on about 90 frames. Fig. 8 shows frames 70 to 95. Notice that the bounding boxes do not contain all moving objects and sometimes do not cover the boundaries of the objects.
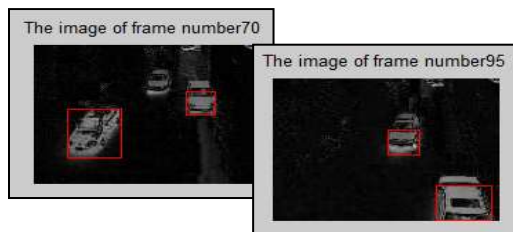

Fig.8 Classification frame 70-95, multi vehicles

The proposed classification function gave an accurate result in most of the frames but, in some frames, a moving object was not detected and consequently, the classification process failed. In some few cases, the failure results from the classification pre-process such as edge detection and drawing the bounding box, and hence the height-width ratio and the speed can be calculated.

Many videos were tested and a large number of frames were taken from every video to detect and classify the moving object. Some of the test results presented involved a variety of moving objects: one moving human, one moving vehicle, many moving humans, many moving vehicles, and moving humans with vehicles. A high proportion of accurate results were achieved from the classification function but there were still some inaccurate results given by the function for the different types of moving object. The classification function is one factor but there are many more. The factors that changed the accuracy of the results included: The background subtraction algorithm, the bounding box, the quality of the video and the angle of the camera.

One of the main classification factors is the background subtraction; this results in the accurate placing of the bounding box. Using the same frame from two background subtraction algorithms, the bounding box was drawn and then viewed to see the difference. Frames from the same video applying the

Frame Difference and Modified Approximate Median algorithms are shown below.
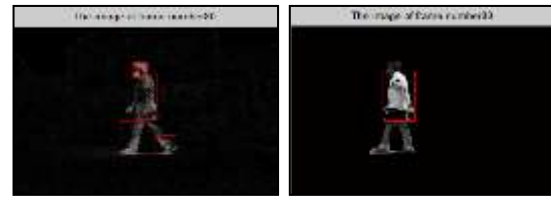

Fig.9 Classification Difference

Table 1 illustrates the classification results from different videos. Please notice that not all the frames contain moving objects and the accurate and the false results represent the frames with moving objects.

Video	B-S	Frames	Type of moving object	Acc. results	False results	Aver. accuracy
hu.avi	FD	50-215	H	162	4	97.59%
hu.avi	FD	220-400	H	160	20	88.89%
MOV04164.MPG	FD	50-100	H	42	6	87.5%
MOV04164.MPG	MAM	50-100	H	48	2	96%
MOV04206.MPG	FD	40-130	H	76	14	84.44%
MOV04206.MPG	MAM	40-130	H	90	0	100%
MOV04196.MPG	FD	60-160	V	98	2	98%
Traffic.avi	FD	10-100	V	74	16	82.22%

Table.1 Classification results

Classification of multi moving humans and/or multi vehicles were also carried out and the results are still promising, but accuracy is not that high and further research are still carried on to enhance the proposed algorithm.

VI. **Conclusions**

In this paper, a novel and simple approach to classify moving objects from a video stream into vehicles and human beings was proposed. The approach is based on two main steps: background subtraction and classification algorithms.

The classification is accomplished by finding the height-width ratio of the bounding box around the moving object in each frame and estimating the speed of the moving object from two consecutive frames in the video stream. The results of this work were fairly good and improvement to this problem can be extended.

References

[1] S. V. Mashak, B. Hosseini, and M. Mokji, "Background Subtraction for Object Detection Under Varying Environments," in *International Conference of Soft Computing and Pattern Recognition*, pp. 123-126, Dec. 2010.

[2] B. P. L. Lo and S. A. Velastin, "Automatic congestion detection system for underground platforms," *Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing. ISIMP 2001*, pp. 158-161, 2001.

[3] R. Cucchiara, C. Grana, M. Piccardi, and a Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337-1342, Oct. 2003.

[4] B. Benfold and I. Reid, "Stable multi-target tracking in real-time surveillance video*", IEEE Conference on Computer Vision and Pattern Recognition,* pp. 3457-3464, June 2011.

[5] M. Narayana, "Automatic Tracking of Moving Objects in Video for Surveillance Applications," MSc thesis, University of Kansas, 2007.

[6] O. Javed and M. Shah, "Tracking and Object Classification for Automated Surveillance," in *Proceedings of the 7th European Conference on Computer Vision-Part IV*, pp. 343-357, 2002.

[7] S. S. Mohamed, N. Tahir, R. Adnan, and U. T. Mara, "Background Modelling and Background Subtraction Performance for Object Detection," in *6th International Colloquium on Signal Processing & Its Applications (CSPA)*, pp. 1-6, 2010.

[8] S. Saravanakumar, A. Vadivel, and C. G. Saneem Ahmed, "Multiple Human Object Tracking using Background Subtraction and Shadow Removal Techniques," in *International Conference on Signal and Image Processing (ICSIP)*, pp. 79-84, 2010.

[9] Y. Long, X. Xiao, X. Shu, and S. Chen, "Vehicle Tracking Method Using Background Subtraction and Mean Shift Algorithm," *International Conference on E-Product E-Service and E-Entertainment (ICEEE)*, no. 9, pp. 1-4, 2010.

[10] M. Piccardi, "Background subtraction techniques: a review," *international Conference on Systems, Man and Cybernetics,* pp. 3099-3104, Oct. 2004.

[11] Y. Benezeth, P. M. Jodoin, B. Emile, H. Laurent, and C. Rosenberger, "Review and evaluation of commonly-implemented background subtraction algorithms," in *19th International Conference on Pattern Recognition*, pp. 1-4, 2008.

[12] S.-T. Su and Y.-Y. Chen, "Moving Object Segmentation Using Improved Running Gaussian Average Background Model," *Digital Image Computing: Techniques and Applications*, pp. 24-31, 2008.

[13] a Elgammal, R. Duraiswami, D. Harwood, and L. S. Davis, "Background and foreground modeling using nonparametric kernel density estimation for visual surveillance," *Proceedings of the IEEE*, vol. 90, no. 7, pp. 1151-1163, Jul. 2002.

[14] Z. Yi and F. Liangzhong, "Moving Object Detection Based on Running Average Background and Temporal Difference," in *International Conference on Intelligent Systems and Knowledge (ISKE)*, pp. 270-272, 2010.

[15] S. S. Ali and M. F. Zafar, "A robust adaptive method for detection and tracking of moving objects," in *International Conference on Emerging Technologies*, pp. 262-266, 2009.

[16] S. S. Ali, H. M. F. Zafar, and F. Khan, "Proposing a new Feature set for Recognition of Moving Humans in Video Frames," in *International Conference on Information and Emerging Technologies (ICIET-2010)*, pp. 1-6, 2010.