

Efficient Human Motion Detection with Adaptive Background for Vision-Based Security System

Fadhlan Hafizhelmi Kamaru Zaman[#], Md. Hazrat Ali^{*}, Amir Akramin Shafie^{*}, Zairi Ismael Rizman^{*}

[#]*Department of Computer Engineering, Universiti Teknologi MARA, Shah Alam, Selangor, Malaysia
E-mail: fadhlan@salam.uitm.edu.my*

^{*}*School of Engineering, Nazarbayev University, Astana, Khazakstan
E-mail: md.ali@nu.edu.kz*

^{*}*Kulliyyah of Engineering, International Islamic University Malaysia, Gombak, Selangor, Malaysia
E-mail: aashafie@iiu.edu.my*

^{*}*Faculty of Electrical Engineering, Universiti Teknologi MARA, Dungun, Terengganu, Malaysia
E-mail: zairi576@tganu.uitm.edu.my*

Abstract— Motion detection is very important in video surveillance system especially for video compression, human detection, and behaviour analysis. Various approaches have been used for detecting motion in a continuous video stream but for real-time video surveillance system; we need a motion detection that can provide accurate detection even in non-static background regardless of surroundings (outdoor or indoor), object speed and size, robust to camera noisy pixels or sudden change in light intensity. This is very important to ensure that the security of a monitored parameter or area is not compromised. In this paper, we propose a method for human motion detection that employs adaptive background subtraction, camera noise reduction and white pixel count threshold for real-time video streams.

Keywords— adaptive background; motion detection; video surveillance

I. INTRODUCTION

One of the most attractive research areas in machine vision and image processing is motion detection analysis. Motion analysis is the basis of several advanced image processing and machine learning tasks which includes (1) human and object detection and tracking, (2) human behaviour interpretation as well (3) advanced behaviour analysis [1], [17].

Generally, motion detection is essential in security system because it provides advanced security measure and can be used to trigger other types of detection and enables video compression. Consider secure premises like a secure vault or money vault in banks, where we do not want any unauthorized person to enter or get near that secured area. Motion detection can be implemented in this case, to detect if anyone passed or appear in the secured parameter and trigger alarm when it is necessary. It can deliver accurate result with simpler and low-cost implementation. It also can overcome human sight limitations, while not at all compromising the security.

Other alternatives to visual motion detection, especially in video surveillance system such as using detection sensor can be implemented by using passive or active infrared sensors (PIR and AIR sensor) or using ultrasonic sensors. Motion detection sensors can also provide an additional layer of security and can be paired with security or surveillance cameras, where these detectors can be connected to a monitoring system. Even though these devices can capture movements of objects which move slowly and can be used at short or long ranges, filling gaps, gates or doors, there are several limitations of these sensors when compared against visual motion detection. This includes the fact that:

- These devices work only directionally. Thus it might not be able to cover the whole targeted secured area
- Measurements quite sensitive to temperature and to the angle of the target
- Works less effectively in rain, snow or if the emitter's photocell is in direct sunlight
- Ultrasonic detectors are able to detect only sharp movements

- PIR and AIR are prone to false positives resulting from warm air flow from radiators, air conditioners, and etc
- Installation and placement of a detection sensor system is very critical in avoiding easy bypass and defeat

Therefore, due to aforementioned limitations of motion detection sensors, there are circumstances where these motion sensors are not fully capable of carrying security tasks. Thus, we believe that it is important to improve existing method of visual motion detection such that it can augment the use of motion sensors in video surveillance system. In this paper, we propose a visual motion detection technique which employs adaptive background subtraction with white pixel counts as motion level identifier. The paper is arranged as follows. In next section, we list some related work. In section II, we discuss in details our proposed method. In section III, we present some experiments results. Finally, we conclude the paper in section IV.

A. Related Works

In the case of continuous video stream such as in video surveillance system, several methods exist for detecting motion and object in the specific area. However, most of the approaches employ a similar strategy that is by comparing the frames from current video stream against frames previously received or simply comparing it against the background frame. For object detection and recognition from continuous video streams, quite a number of methods have been proposed in the past years [2], [3]. This includes background subtraction or frame differencing method, which is often the preliminary step of detecting motion or object of interest [4]-[6]. Other motion detection method that relies on image processing and those that do not rely on image processing usually employ on-body monitoring sensors [7]-[10].

One example of background subtraction is where light-independent background subtraction is proposed. Disparity verification is used in this technique where this approach is invariant to rapid changes in illumination, especially during the run-time. However, the biggest flaw of this technique happens when the object is moving slowly and smoothly where the relatively small changes in motion due to the frame differencing will only contribute to significantly minute changes between frames. So, it is difficult to get the whole moving object. There is also an instance when the object is moving too slowly up to a point that the background subtraction algorithms will not give any frame difference at all [11].

Another popular approach in acquiring a reliable frame as the background is similar to finding the median of several images, where images are averaged over a period of time thus acquiring a static scene except where motion does occur. This could be very effective situations where the background is observable over a substantial period of time. However, this approach is not robust against scenes that involve fast paced moving objects. This approach also depends on a predetermined threshold for the whole scene and cannot handle bimodal backgrounds. It also recovers slowly when the background is not affected by moving objects.

Alternatively, background subtraction can be carried out by comparing the current frame against the first frame from the continuous video sequence. Given that there is no object in this initial frame, the problem mentioned above could be avoided. We can now acquire the whole moving object regardless of its moving speed. However, the biggest flaw of this approach could render the whole approach useless. Consider a situation where there is a vehicle in the first initial frame, and then it is gone. It will cause the algorithm to always detect motion at the place where the vehicle initially appears. This flaw can be reduced by continuously renewing the initial frame after a certain interval of time, but still, there is no guarantee that the newly obtained initial frame only contains a static background.

As mentioned earlier, motion detection usually serves as a basis for more advanced image processing task. Thus, normally the results from the background subtraction are transmitted to higher level processing such as object detection and object tracking. The information acquired during object detection could, in fact, be used to improve background subtraction. In this case, pixel-based background subtraction decides whether the pixel belongs to the background (BG) or foreground object (FG) which is discussed using Bayesian [12], [13]. Assuming that the value of a pixel at time t in RGB or some other colour space is denoted by $\vec{x}(t)$, the Bayesian decision R can be made by as follows

$$\frac{p(\text{BG}|\vec{x}(t))}{p(\text{FG}|\vec{x}(t))} = \frac{p(\vec{x}(t)|\text{BG})p(\text{BG})}{p(\vec{x}(t)|\text{FG})p(\text{FG})} \quad (1)$$

Additionally, adaptive background proposed is carried out by analyzing the pixel-level approach. Using Gaussian mixture probability density, an efficient adaptive algorithm is proposed. The parameters are updated using recursive equations and also to at the same time choose the proper number of components for each pixel. Similarly, we use the Gaussian mixture probability density to analyze the motion at pixel level [14].

II. MATERIAL AND METHOD

To overcome the aforementioned problems, we propose this efficient algorithm which aims to model the background of the scene and to compare each foreground frames with the background frame. In this section, we elaborate further on the steps involved in the proposed motion detection method. There are several steps involved in this motion detection, where each step are important to ensure the right features are used in the subsequent steps. They are video streams processing, image acquisition, image pre-processing and process of determining whether the motion is detected based on motion threshold. These general steps are graphically described by the following flow chart in Fig. 1, and detailed steps are illustrated by image snapshots in Fig. 2.

A. Video Processing

The video stream is first forwarded to video processing, where the video stream will be locked and broken into multiple image frames to be processed independently. Every image frame is converted to bitmap image with uniform size. This is to ensure that the pixel scanning process is carried out uniformly.

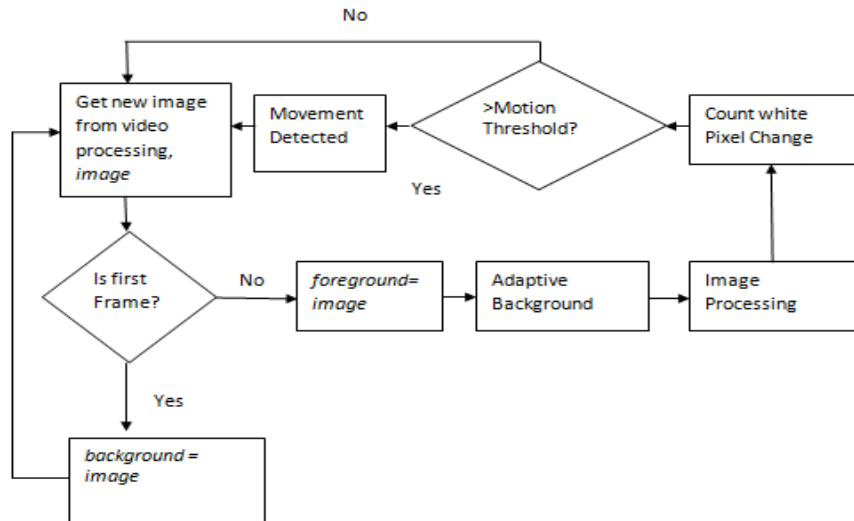


Fig. 1 General motion detection process flow used in this technique



Fig. 2 Image processing in proposed technique: (a) Grayscale (b) Background subtraction (c) Thresholding (d) Opening filter (e) Edge filter (f) Final image with motion region boundaries

B. Image Acquisition

Bitmap images are used as the standard format for the image as a result of preliminary video processing that produces individual image frames. The images are normalized to be 640 x 480 pixels, and the background of the scene is acquired from the first frame acquired from the video stream.

C. Image Pre-Processing and Noise Reduction

For straightforward processing, we use grayscale images. Initially, the background and the images will be converted to

a grayscale image. To detect any changes in the subsequent frames, the image will be compared to the background. Then, thresholding is applied which will transform the grayscale image to binary image (black and white image) [16]. Low colour with 0 intensity is assigned to pixels with intensities below the pre-defined threshold. On the other hand, the maximum intensity of 1 is assigned to pixels having intensities higher than said threshold. This will produce black and white image necessary for further processing.

The pre-defined threshold must be able to conserve the silhouette of human body shape. A highly-deformed human

body silhouette, especially in the region between head and chest, could be obtained as a result of thresholding, which is due to the difference in light intensity due to the shadow which we consider as noise in motion detection. Thus, the threshold value must be carefully selected such that it is able to compensate between the object shadow and light illumination. The effect of the higher threshold is it will increase the prominence of shadow in the image and increase the pixel noise, while lower threshold will deform the human shape silhouette. Hence, the selected threshold value should minimize both effects. We choose threshold equal to 25 based on various experiments that we have conducted.

Subsequently, the opening filter is applied to remove the camera pixel noise and noises due to illumination changes. In this process, small white pixels will be removed. Finally, edge filter is used to detect the edge of the image.

D. Motion Threshold

Then, the resulting image from the opening filter will have its pixels counted. The approach used to count the number of white pixels in the image is by using histogram technique, which following this histogram as follows

$$h'_f(z) = \frac{1}{2K+1} + \sum_{j=-K}^K h_f(z+j) \quad (2)$$

where K is a constant representing the size of the neighbourhood used for smoothing.

This value white pixel count is later denoted as motion level, which indicates the magnitude of the motion observed in the foreground. A Higher magnitude of motion would cause the greater difference between foreground and background, thus indicating higher motion level. Then, the motion level obtained is compared with a specified motion threshold. Similarly, the value for motion threshold is designated so that it can reduce the effect of ‘false motion’ that is actually caused by light illumination change, light flickers, and camera pixels’ noise. This implementation will help reduce false detection motion. Even though most of the noises from those sources were already reduced in image pre-processing, there is still likelihood that such noise might still exist in the later stage of motion detection. Based on various experiments carried out, the motion threshold of 10 is chosen. In another word, if the motion level is more than 10, motion detection is triggered whereas if motion level is less than 10, no motion is detected. This motion threshold value is an optimal value which gives the best trade-off between false acceptance and false rejection.

E. Adaptive Background and Gaussian Mixture Probability Density

In the implementation of adaptive background, we move the background frame slightly in the direction of the current frame. The pixel intensities in the background frame are changed by one level per frame. Employing this procedure, the background is slowly adapted to the current image pixels by pixels (1 level per frame) which will ensure that the system will be able to detect even minute change with respect to the background in the image.

In order to adapt new scene to the background, the new object needs to be static in the foreground long enough. The discussion on the period of time which can be considered

‘long enough’ is given [14]. Consider some additional clusters with small weights π_m represent the intruding foreground objects in the new scene. Thus, the background model by the first B largest clusters can be approximated as shown as follows

$$p(\vec{x}|X_T, BG) \sim \sum_{m=1}^B \hat{\pi}_m N(\vec{x}; \hat{\mu}_m, \sigma_m^2 I) \quad (3)$$

Assuming c_f is a measure of the maximum portion of the data that can belong to foreground objects without influencing the background model. If the components are sorted to have descending weights π_m , we then have Equation (4):

$$B = \arg \min b \left(\sum_{m=1}^b \pi_m > (1 - c_f) \right) \quad (4)$$

For example, it will probably generate stable cluster if a new object comes into a scene and remains static for some time. Since the old background is occluded, the weight of π_{B+1} and the new cluster will constantly be increasing. The weight of an object becomes larger than c_f if the object remains static long enough, and it can later be considered to be part of the background. The mixing weights denoted by $\hat{\pi}_m$ are non-negative and add up to one. Given a new data sample $\vec{x}(t)$ at time t , the recursive update equations are given in [15]:

$$\hat{\pi}_m \leftarrow \hat{\pi}_m + \alpha(o_m^{(t)} - \hat{\pi}_m) \quad (5)$$

$$\hat{\mu}_m \leftarrow \hat{\mu}_m + o_m^{(t)}(\alpha/\hat{\pi}_m) \bar{\delta}_m \quad (6)$$

$$\hat{\sigma}_m^2 \leftarrow \hat{\sigma}_m^2 + o_m^{(t)}(\alpha/\hat{\pi}_m) (\bar{\delta}_m^T \bar{\delta}_m - \hat{\sigma}_m^2) \quad (7)$$

Looking at Equation (5), we can conclude that the object should remain static in the foreground for approximate $\log(1 - c_f) = \log(1 - \alpha)$ frames. For example, for $c_f = 0.1$ and $\alpha = 0.001$, the object needs to remain static in the foreground for about 105 frames. From the experiments, we get the average time required for the object to remain static in the scene before it is totally adapted to the background is 190 frames which are equivalent to 10 seconds at 19.0 frames per second.

F. Motion Boundaries

Finally, we merge the image acquired from edge filter with the original foreground image thus producing coloured edges indicating motion area. The different colour channel can be used to emphasize this motion area which specially used to indicate the region, where the motion is detected. This process will yield a final image with a coloured indicator for motion area.

III. RESULTS AND DISCUSSION

A. Indoor and Outdoor Test

The test was carried out both indoor and outdoor using high-resolution DSP colour camera, 1/3" colour CCD camera at 30 frames per second. This algorithm was implemented in our own-developed software and used to carry out the test.

Both indoor and outdoor test produced a good result and the boundaries produced to indicate the motion region fits the whole moving object in most of the test subject especially human. The effect of the shadow does not reduce the accuracy of motion detection, but it reduces the accuracy of motion region detected. The presence of shadow in the scene produced by the moving object made the boundaries indicated seems to be much larger than the actual moving object and occlude other objects. However, the camera pixel noise and sudden light intensity change can be overcome by this algorithm's threshold value. Therefore, it does not provide a high percentage of false alarm.

In the outdoor test, the shadow effect is less than the indoor, and the boundaries produced include the shadow of the moving object. But, it does not affect the motion detection accuracy in terms of false detection or false rejection. The sample result from outdoor in indoor highlighting the motion region boundaries is illustrated in Fig. 3. The motion detection in a video sequence is illustrated in Fig. 4. As illustrated in the figures, we can see that the algorithm only highlights the moving parts of the human body if the movement is partial.



Fig. 3 Top row: Outdoor test result, Bottom row: Indoor test result



Fig. 4 Motion detection in video sequences for frame 1, 6, 11 and 16

B. Comparison Against Existing Techniques

We compared the result of our proposed technique with other techniques such as simple background subtraction that use only two frames difference to find the motion and the low precision background modelling, which is the preliminary result before we found our proposed technique. Low precision background modelling also uses adaptive background, but it is more prone to camera noise and light

changes. The sample result of the comparison is illustrated in Fig. 5. The overall result from 50 scenes from outdoor and indoor is summarized. For the purpose of performance evaluation, three well-known measurements are used namely False Rejection Rate (FRR), False Acceptance Rate (FAR) and Accuracy. Result presented in Table 1 gives the comparison of this method, which also includes the computational efficiency measured from average CPU usage.



Fig. 5 Motion detector comparison: (Clockwise) original image, simple background subtraction (two frame differences), proposed technique and low precision background subtraction

TABLE I
PERFORMANCE COMPARISON

Motion Detection Techniques	Accuracy (%)	False Acceptance Rate (%)	False Rejection Rate (%)	Average CPU Usage (%)
Simple background subtraction (2-frame difference)	92.0	2.0	6.0	9.0
Low precision background modelling	95.0	5.0	0	10.0
Proposed method	99.1	0.9	0	11.0

IV. CONCLUSION

This proposed algorithm performs fast and can detect human motion even only tiny movement involves in the image while producing good motion boundaries around the moving human. It actually calculates the white pixel count corresponding to the change in the image, which is called motion level. This parameter motion level will determine how much change actually occurs in the scene/image. The alarm will be raised when the motion level passed motion threshold value. The threshold value is set to compensate for uncontrolled background changes like changes due to wind, light illumination change, and camera noisy pixels.

ACKNOWLEDGMENT

The work presented here is sponsored by Universiti Teknologi MARA Malaysia under grant 600-IRMI/DANA 5/3/SINERGI (0001/2016).

REFERENCES

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: A review," *Computer Vision and Image Understanding*, vol. 73, pp. 420-440, Jun. 1999.

[2] M. H. Yang, D. Kriegman and N. Ahuja, "Detecting face in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 34-58, Jan. 2002.

[3] E. Hjelmås, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, pp. 236-274, Sep. 2001.

[4] I. Haritaoglu, D. Harwood and L. Davis, "W4: A real-time system for detection and tracking of people and monitoring their activities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 809-830, 2000.

[5] J. Bernd, *Digital Image Processing: Concepts, Algorithms and Scientific Applications*, 3rd ed., Berlin, Germany: Springer-Verlag, 1992.

[6] J. Zhang, J. Gao and W. Liu, "Image sequence segmentation using 3-D structure tensor and curve evolution," *IEEE Transactions on*

Circuits and Systems for Video Technology, vol. 11, pp. 629-641, 2001.

[7] J. Bang D. Kim and H. Eom, "Motion object and regional detection method using block-based background difference video frames," in *Proc. IEEE ICERTCSA'12*, 2012, p. 350.

[8] A. N. Ansari, M. Sedky, N. Sharma and A. Tyagi, "An Internet of things approach for motion detection using Raspberry Pi," in *Proc. IEEE ICICIT'15*, 2015, p. 131.

[9] Y. Geng, J. Chen and K. Pahlavan, "Motion detection using RF signals for the first responder in emergency operations: A PHASER project," in *Proc. IEEE AISPIIMRC'13*, 2013, p. 358.

[10] Y. L. Hsu, Y. T. Chen, P. H. Chou, Y. C. Kou, Y. C. Chen and H. Y. Su, "Golf swing motion detection using an inertial-sensor-based portable instrument," in *Proc. 2016 IEEE ICCE'16*, 2016, p. 1.

[11] Y. Ivanov, A. Bobick and J. Liu, "Fast lighting independent background subtraction," *International Journal of Computer Vision*, vol. 37, pp. 199-207, Jun. 2000.

[12] M. Harville, "A framework for high-level feedback to adaptive, per-pixel, mixture-of-Gaussian background models," in *Proc. ECCV'02*, 2002, p. 543.

[13] P. Withagen, K. Schutte and F. Groen, "Likelihood-based object detection and object tracking using color histograms and EM," in *Proc. ICIP'02*, 2002, p. 589.

[14] K. Wang, Y. Liang, X. Xing and R. Zhang, "Target detection algorithm based on Gaussian mixture background subtraction model," in *Proc. CIAC'15*, 2015, p. 439.

[15] A. Khoshrou and A. P. Aguiar, "Unsupervised learning of Gaussian mixture models in the presence of dynamic environments," in *Proc. CONTROLO'14*, 2014, p. 387.

[16] I. M. Yassin, A. Zabidi, M. S. A. M. Ali, N. M. Tahir, H. A. Hassan, H. Z. Abidin, and Z. I. Rizman, "Binary particle swarm optimization structure selection of nonlinear autoregressive moving average with exogenous inputs (NARMAX) model of a flexible robot arm," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 630-637, Oct. 2016.

[17] M. N. M. Nor, R. Jailani, N. M. Tahir, I. M. Yassin, Z. I. Rizman, and R. Hidayat, "EMG signals analysis of BF and RF muscles in autism spectrum disorder (ASD) during walking," *International Journal on Advanced Science, Engineering and Information Technology*, vol. 6, pp. 793-798, Oct. 2016.