

Comparison of human detection using background subtraction and frame difference

Mohd Saifulnizam Zaharin, Norazlin Ibrahim, Tengku Mohd Azahar Tuan Dir

Department of Industrial Automation, Universiti Kuala Lumpur Malaysia France Institute, Malaysia

Article Info

Article history:

Received Dec 10, 2018

Revised Jan 7, 2019

Accepted Sep 26, 2019

Keywords:

Background subtraction

Frame difference

High-level feature extraction

Movement of human

Support vector machine

ABSTRACT

Image processing is mostly used for exploring image behaviour. There are several steps in image processing. Image acquisition, pre-processing, feature extraction, and classification are the processes used for the detection of human movement based on high-level feature extraction (HLFE), in which HLFE was used for feature extraction in this paper. This study proposed the use of background subtraction and frame difference. This research was conducted to analyse the difference of background subtraction and frame difference methods based on movement of human. Movement of human detected by using feature extraction were centroid image technique used. Furthermore, support vector machine (SVM) was used for classification.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Norazlin Ibrahim,

Department of Industrial Automation,

Universiti Kuala Lumpur Malaysia France Institute,

Seksyen 14, Jalan Teras Jernang, 43650 Bandar Baru Bangi, Selangor, Malaysia.

Email: norazlin@unikl.edu.my

1. INTRODUCTION

Nowadays, there are numbers of crime cases happened such as robbery, fighting and others. Therefore, CCTV is installed in high crime cases area and human is assigned to monitor the situation [1]. However, humans are prone to error, easily get tired and might missed out the crime event. This research will focus on detecting the human movement between walking and running by using background subtraction and frame difference. The problem in digital image processing is human required to monitor the behaviours of subjects in the usually complex scenes. Manual observation by human however is not appropriate, as it requires attentive and careful concentration over a long period of time [2]. Then the automated surveillance system is needed to detect human behaviour. Therefore, comparison of human detection is applied to acknowledge human movement in digital image processing. Research from S. A. Dhole have proposed techniques image acquisition, image pre-processing, image segmentation, feature extraction and classification but it's not suitable for this research.

The image acquisition in this research is a video of human walking and running. Pre-processing is the process of enhancing an image. It involves noise removal, segmentation, and other processes [3]. Feature extraction is the subsequent step after pre-processing that focuses on high-level feature extraction (HLFE). Two categories of feature extraction are HLFE and low-level feature extraction (LLFE). LLFE is based on motion directly while HLFE is based on shapes [4]. This research focused on the feature extraction used shape-based by applying background subtraction and frame difference. Kavitha and Tejaswini presented motion detection by overcoming the disadvantages of background subtraction algorithm [5]. The research has robust an efficiently computed background subtraction algorithm

which is able to cope with the problems of local illumination changes such as shadows and highlights as well as global illumination changes. The first step in background subtraction is to initialize image by having no motion or object in the frame. In this technique, it is enough to choose one frame image as a background subtraction image. The frame difference method is the subtraction of frames have motion as background subtraction. In order to ensure the feature used is suitable, a verification process through classification should be used. Support vector machine (SVM) technique was chosen in this study since it is good for validation feature.

2. RELATED WORK

Images that can be processed by a computer are called digital images, which can be obtained by image acquisition; it is a process that is performed to obtain a digital image [6]. In order to meet the demand of high speed, high efficiency, and high density, an image acquisition system was invented to realise automatic acquisition, which provides a variety of data sharing interface to realise direct docking with database [7]. Colour image can be represented by colour histograms that contain a one-column array that has 256 elements [8]. RGB to greyscale process is used to minimise the number of bits and pixels because RGB has 24 bits and the luminance of pixel of 0-255. This is a disadvantage for RGB; therefore, greyscale colour is used for better performance. The conversion of a colour image into a greyscale image also converts the RGB values of 24 bit into the greyscale values of 8 bit [9]. The greyscale colour is good for reducing processing time, which is achieved by reducing the storage from 24 storage to 8 storage.

Background subtraction is an image that contains no moving object in a video. Obtaining a background model can be done in two steps. The first model is background initialisation, where the background image is obtained from a specific time from the video sequence, and the second model is frame difference, where the background is updated due to the changes that may occur in the real scene [10]. Background subtraction and frame difference were used in this research. The basic way to separate a moving object from its background is to subtract the background from the image, leaving just the moving object or the foreground [11]. Background subtraction pixels from background subtraction can be extracted; thus, the moving object or human detection in this research can be defined by:

$$H(y) = \sum_{i=1}^m f(i, y) \quad (1)$$

$$H(x) = \sum_{j=1}^n f(x, j) \quad (2)$$

The process of background subtraction or frame difference consists of two steps. The first step is performing a thresholding on $H(x)$ and $H(y)$. Therefore, it will compute the first derivative, which can be defined by $\frac{d}{dn}H[n] = H[n] - H[n - 1]$, where the pixels change from 1 to 0 or vice versa in order to get the values [12]. Rakibe and Patil presented motion detection by developing a new algorithm based on the background subtraction algorithm [13]. Their first reliable background model was based on the statistical used, then subtraction between the current image and background image was done based on threshold. Detection of moving object is done after background image. After that, morphological filtering was initiated to remove the noise and solve the background interruption difficulty.

Since binary images are easy to operate, other storage format images are often converted into binary images when they are used for enhancement or edge detection [14]. However, each sample has one of the results that contains the desired object. In [13], they used filtering and binarising transformed greyscale image technique in order to remove holes, where the image was passed through the Gauss low-pass filter. I_{d1} was obtained by filtering the gray image. Then, I_{d1} image was binarised using binary threshold and I_{d2} binary image was produced.

$$I_{d2}(x, y) = I_{d1} \quad (3)$$

Where (x, y) represent pixel coordinates in the image. In image processing, noise commonly appears in an image. Noise exists in various types and characteristics [15]. Filtering is a basis function that is used to achieve many tasks such as noise reduction, interpolation, and resampling [16]. This research used Sobel

magnitude gradient for edge detection and gradient method to detect the edges by looking for the maximum and minimum in the first derivative of the image.

Gaussian smoothing filter, is considered a “perfect” blur for many applications, provided that kernel support is large enough to fit the essential part of the Gaussian [17]. When Gaussian filter is used for noise suppression, a large filter variance is effective in smoothing out noise, but at the same time it distorts those parts of the image where there are abrupt changes in pixel brightness [18]. This research used two-dimensional filter where the Gaussian distribution in one-dimensional form is:

$$G(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{x^2}{2\sigma^2}} \quad (4)$$

Where σ is the standard deviation of the distribution. Based on the one-dimensional Gaussian filter, two-dimensional digital Gaussian filter can be expressed as:

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \quad (5)$$

Where σ^2 is the variance of Gaussian filter, and the size of the filter kernel 1 (-1 5 x, y 5 1) is often determined by omitting values lower than five percent of the maximum value of the kernel.

This research used two-dimensional filter for reducing the noise in the images. Intelligent monitoring systems do not only detect motion but also separate moving objects in the foreground of the stationary objects in the background [19]. The goal of motion detection is to recognise the motion of objects found in two given images [13]. In this research, the detection of motion image depends on the process prior to background detection, which is Gaussian smoothing filter. This paper used binary large object (BLOB) to detect human or non-human by marking it BLOB, which is the technique used in image processing [20]. An accurate computation of BLOBs centre point is dependent on the accuracy of BLOB detection, particularly for BLOBs that are not perfectly circular or squarely-shaped [21].

HLFE was used in this research to identify the object, and then a description of the true object can be identified [20]. The moving objects are then tracked and finally a human modelling “Star Skeletonisation” is applied to detect human objects and the motion analysis used [10]. The centre of an image in the area where a human is moving is called centroid. The centroid is located at position (Cx, Cy), where Cx and Cy are respectively the average of the x and y coordinates for all white pixels with the boundary of the shape that consists of a series of boundary points [22]. Layer G and others have developed the framework for detecting the suspicious event from a video through three-steps LLFE, event classification, and event analysis with one assumption, which is unlabelled video sequences are known to contain only one event [23].

SVM had use in this research and it is trained using optimised centroidal profile of both categories to investigate their relative influence on human shape recognition [24]. The classification of pixel for background subtraction and frame difference is determined in a receiver operating characteristic (ROC) curve. SVM error can be used to differentiate data between background subtraction and frame difference. An ROC curve is a graphical plot that reveals the performance of a binary classifier. This curve is drawn by plotting the true positive rate against the false positive rate at various threshold settings [25].

3. METHODOLOGY

High level feature extraction starts with image acquisition for the input image of the study. Pre-processing is a critical process to determine a good image where there is no noise in the image. The feature extraction for this research used centroidal technique image, which is a step to obtain the characteristics of the image. It was used to identify the feature, which was the objective of the research. Classification of image is the last step in digital image and this research used SVM. Figure 1 shows the overall block diagrams of this research. All experiments were conducted using Matlab R2017a and graphics processing units (GPU) used for simulations were an Intel Core i7-6700K CPU with 4.00 GHz working frequency and 8 GB RAM.

3.1. Image acquisition

The image acquisition for this research was conducted using a video with the size 640×320 pixels. The video was converted to image sequences. The input image of this research was a video converted to images of human walking and running. This made it easier for the analysis of background subtraction and frame difference based on the frames of the images conducted.

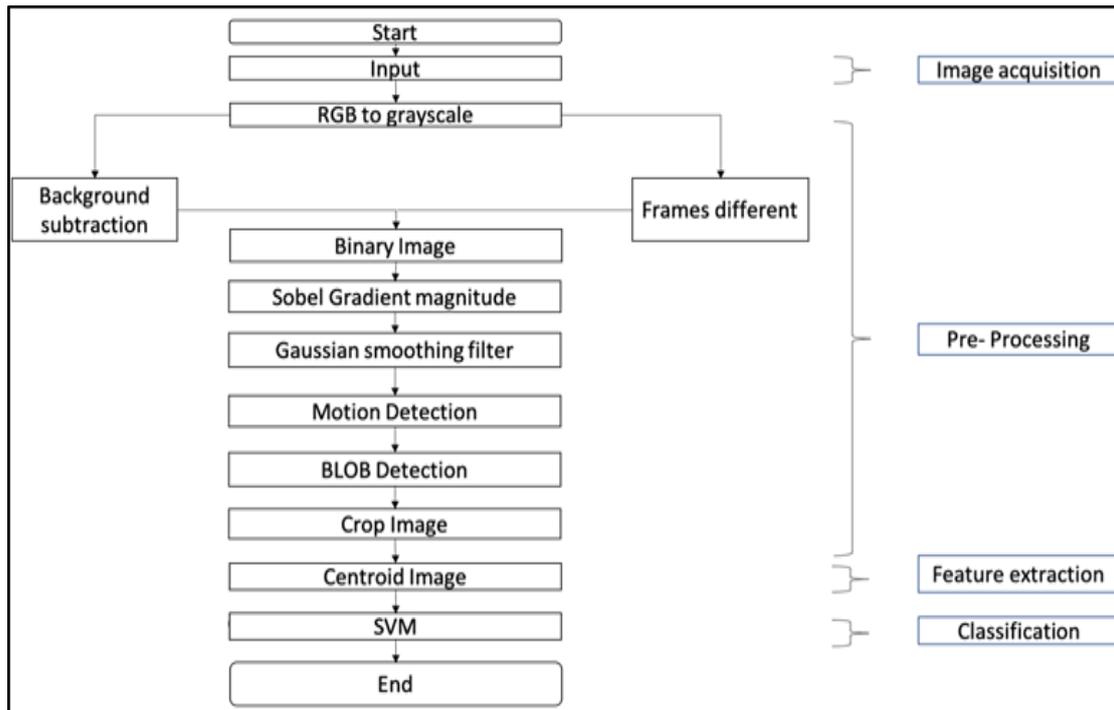


Figure 1. Summary for tracking individual activities to track suspicious activities

3.2. Pre-processing

A greyscale image was used to subtract an image matrix. It was easier using a greyscale image because the image consisted of a data matrix whose values represented intensities with some ranges from 0 and 1 or 0 to 255 for uint8. This research proposed the use of background subtraction. Two categories of subtractions, which were background subtraction by the first frame of the image and frame difference of the image were used. The video was recorded starting with no object so that several frames would be selected as background subtraction frames. The (8) shows the background subtraction equation:

$$BS(x, y, t) = I(x, y, t) - B(x, y, t) \quad (6)$$

Where t is the estimated time, BS is the background subtraction, and I is the image at time t , then B is the background image input at time t that is set up to the number of frame=1. Background subtraction of frame sequence is the subtraction of frame by frame, which is estimated by the previous frame. Morphological technique was involved to perform the segmentation of images after the background subtraction process, in which binary images were converted.

3.3. Feature extraction

Feature extraction is the enhancement of an image from the pre-processing of an image. Feature extraction is also known as shape-based or foreground detection. The image input will be different to the output after feature extraction. It will be a large data from the image that will be described by feature extraction and it will be reduced to achieve sufficient accuracy of data by the combination of various techniques. In this research, HLFEE was proposed to detect human movement, either walking or running. This is to determine the normal and abnormal movement of human. This paper used a centroid image as the feature extraction where the centroid image. The centroid image can determine if the movement is by a human or not. It was proven by the centroidal profile of 100 human and non-human images at 10° interval as show in Figure 2, which were extracted and yielded 36 centroidal feature profiles [24]. Centroid image was used to determine the movement of human based on the centroidal angle [24] show in Figure 3.

The features are the 80° , 90° , 100° , 110° profiles which are located in region 1, followed by four more features of region 2 located at the 180° , 190° , 200° and 210° of angle intervals. Region 3 has 7 features and they are located at the 250° , 260° , 270° , 280° , 290° , 300° , 310° of angle intervals and finally four features are positioned at the angle intervals of 350° , 360° , 10° and 20° in Region 4 as shown in Figure 3.

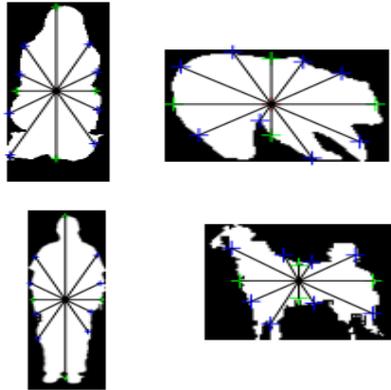


Figure 2. Example of the generalized centroidal profile for shape modelling [24]

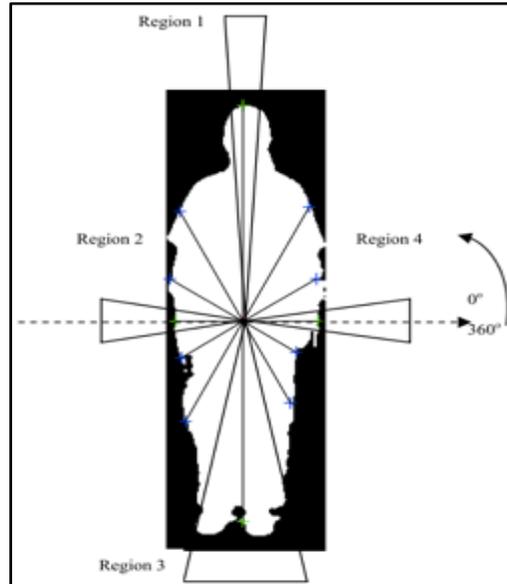


Figure 3. The generalized centroidal gait profile representing the four regions [24]

3.4. Classification

SVM was used to classify the results of abnormal or normal movement of human based on the results of HLFE. The pixels of each background subtraction and frame difference were tested and analysed. The prediction of HLFE image in background subtraction and frame difference sequences was computed for evaluation purpose. Classification of normal and abnormal movement was determined using SVM error. 10 classifications were tested for each background subtraction and frame difference for the prediction.

4. RESULTS AND DISCUSSION

The results for this research were divided into three sections: pre-processing, feature extraction, and classification.

4.1. Pre-processing

The pre-processing for this research covered the converted images of a video that were extracted by using multiple techniques as stated previously in the methodology. Images were extracted to obtain the movement of a human in the frame of images based on background subtraction and frame difference sequences. The results for the pre-processing techniques are shown in Table 1. The result of Table 1 shown background subtraction is better than frame different for pre-processing. The human movement can be detected efficiently in background subtraction compared frame difference.

4.2. Feature extraction

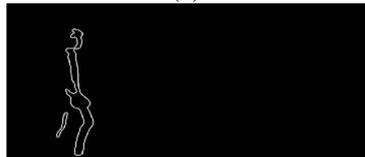
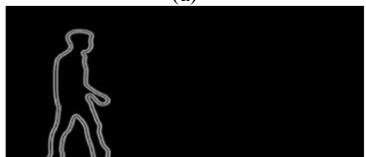
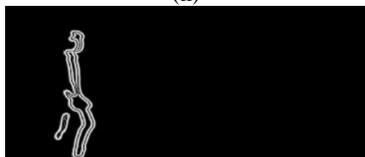
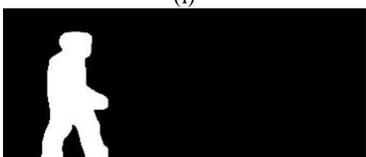
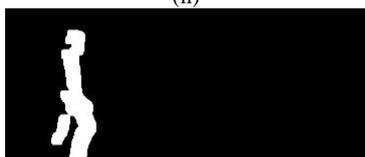
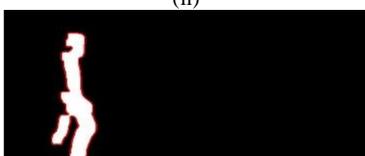
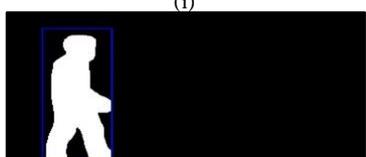
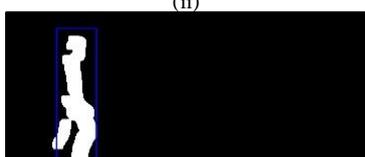
Feature extraction is the extraction of an image from pre-processing. This research used centroid image technique for feature extraction and confirmation of human body [26]. The centroid images can be illuminated as Figure 4. The centroid is the centre of images based on cropped images. The centroid image was divided into angles in 360° with 10° of each degree. The angle of degree in the centroid image was calculated and then classified to the image to determine either it is abnormal or normal from the movement of angle. An abnormal image is triggered when the number of change in angle has high degree of movement from frame to frame and if it is normal, there will be low degree of movement from frame to frame.

4.3. Classification

The process of classification of an image is important to ensure the data are correct. The movement of a human in the frames were calculated by the pixels. The pixels for background subtraction and frame difference were calculated and presented in Figure 5. Based on Figure 5, the background subtraction technique is stable than frame difference because the subtraction of frame difference is low compared

to background subtraction. There was zero reading in frame 0 to 40 because no movement was detected in the frame. Both techniques are good to detect movement in image processing but for this research, background subtraction technique is better than frame difference because movement of a human in background subtraction is more stable than the movement of a human in frame difference.

Table 1. Results for pre-processing techniques

Categories	Background Subtraction	Frame Difference
Greyscale image (a)	 (i)	 (ii)
Background subtraction/frame difference (b)	 (i)	 (ii)
Binary image (c)	 (i)	 (ii)
Sobel gradient magnitude (d)	 (a)	 (ii)
Smoothed gradient magnitude (e)	 (i)	 (ii)
Motion detection (f)	 (i)	 (ii)
BLOB detection (g)	 (i)	 (ii)
Cropped image (h)	 (i)	 (ii)

Based on Figure 6, the number error of background subtraction is lower than frame difference, where the average of error of background subtraction is 0.05 and frame difference is 0.24. The higher number

of error for background subtraction is 0.20 or 20% while frame difference is 0.45 or 45% of the testing. This shows that the data from background subtraction are good and the data from frame difference are not stable. The area under the curve for SVM is produced from the results of ROC curve.



Figure 4. Centroid images for, (a) Background subtraction, (b) Frame difference sequences

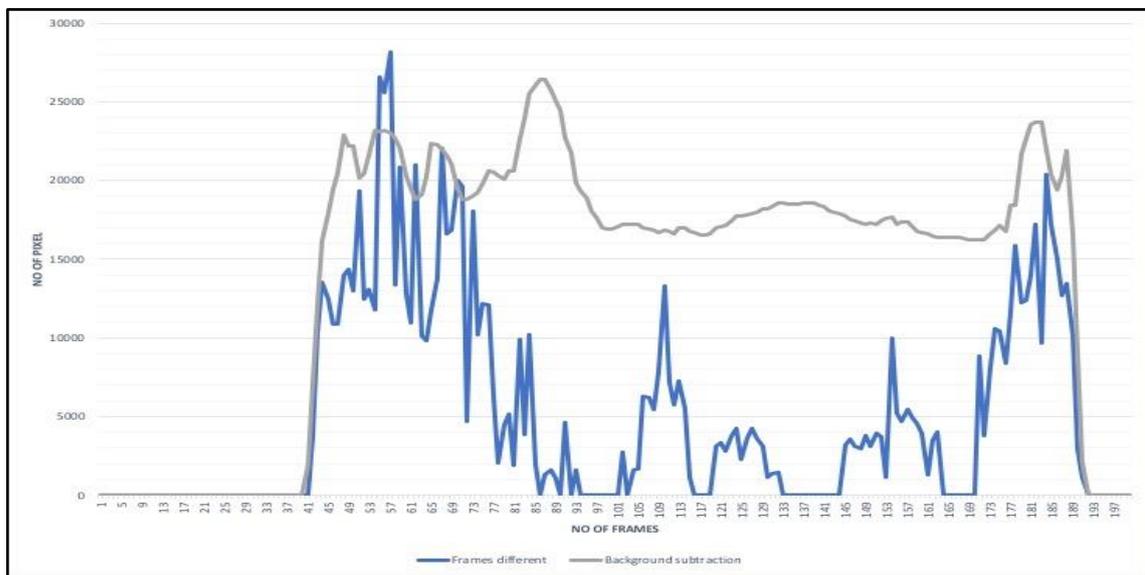


Figure 5. Graph of background subtraction vs. frame difference sequences

Based on Figure 7, background subtraction has good and accurate data because the data for area under the curve are approaching 1 or 100%, while the lowest is 0.80 and it is acceptable. The data for frame difference are not good based on the area under the curve. This is due to low stability of frame difference as shown in Figure 7.

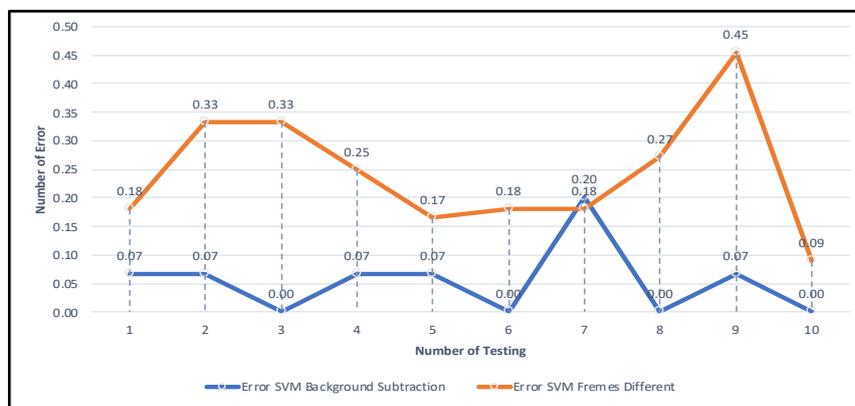


Figure 6. SVM error for background subtraction vs. frame difference sequences

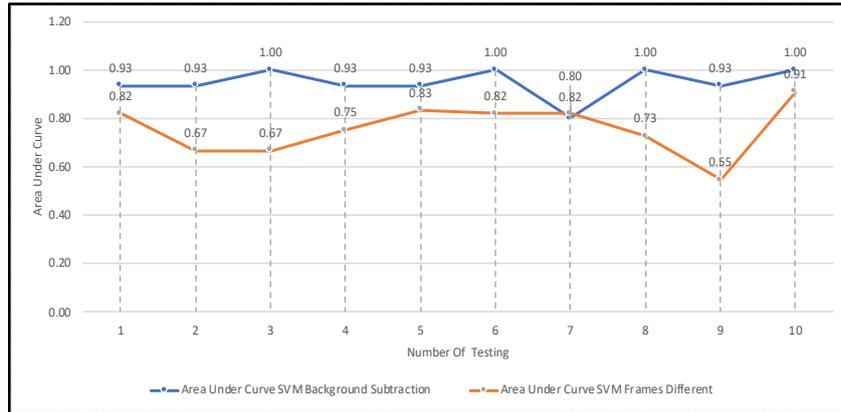


Figure 7. Area under the curve for SVM of background subtraction vs. area under the curve for SVM frame difference

Figure 8 shows the results of ROC curve by SVM based on background subtraction in the prediction of number six, where the number of error is 0 for test number six. The area under the curve for ROC curve based on Figure 7 is 1 and it has similar number of data as in Figure 8. Figure 9 shows the results of ROC curve by SVM based on frame difference in the prediction of number three where the number of error is 0.33 for test number three. The area under the curve for ROC curve based on Figure 9 is 0.67.

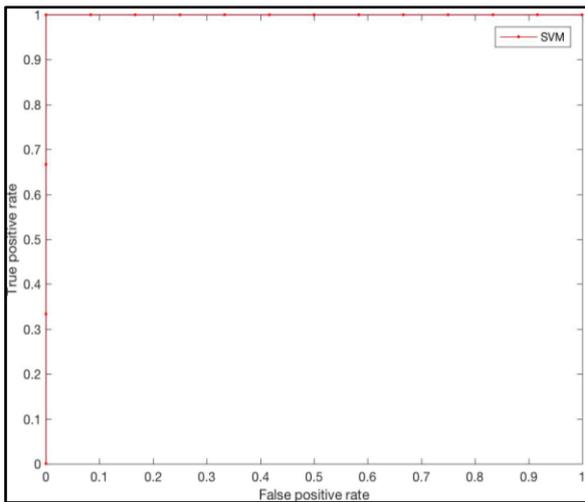


Figure 8. ROC curve by SVM based on background subtraction

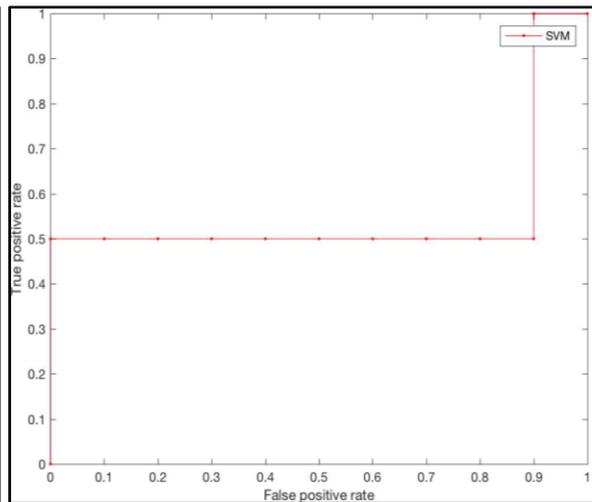


Figure 9. ROC curve by SVM based on frame difference

5. CONCLUSIONS

Based on the results of HLFE of background subtraction and frame difference, the technique that is better for HLFE is background subtraction. Normal and abnormal human movement have been classified from the results of feature extraction where SVM error of background subtraction is better than frame difference. This was based on the accuracy of background subtraction was 95%, which makes it more accurate than frame difference sequences that have 55% accuracy. Therefore, background subtraction should be implement in the automated human detection and the result can conclude the background subtraction better than frame different for human detection. The future work for the research is more analysis on classification technique are used for the research and combines HLFE as detect human normal and abnormal with other technique classification methods and make it a complete system such as a surveillance application system.

REFERENCES

- [1] H. H. Park, G. S. Oh, and S. Y. Paek, "Measuring the crime displacement and diffusion of benefit effects of open-street CCTV in South Korea," *Int. J. Law, Crime Justice*, vol. 40, no. 3, pp. 179-191, 2012.
- [2] S. L. Tang, Z. Kadim, K. M. Liang, and M. K. Lima, "Hybrid blob and particle filter tracking approach for robust object tracking," *Procedia Comput. Sci.*, vol. 1, no. 1, pp. 2549-2557, 2010.
- [3] N. Upadhyaya, M. Dixit, and M. Pradesh, "A review: relating low level features to high level semantics in CBIR," *Int. J. Signal Process. Image Process. Pattern Recognit.*, vol. 9, no. 3, pp. 433-444, 2016.
- [4] N. Ibrahim, S. Mokri, L. Siong, and M. Mustafa, "Snatch theft detection using low level features," *Proceedings of the World Congress on Engineering 2010 Vol II*, WCE 2010, London, U.K. vol. II, June 30-July 2, 2010.
- [5] K. Kavitha and A. Tejaswini, "VIBE: background detection and subtraction for image sequences in video," (*IJCSIT*) *International Journal of Computer Science and Information Technologies*, vol. 3, no. 5, pp. 5223-5226, 2012.
- [6] A. Yudhana, Sunardi, and S. Saifullah, "Segmentation comparing eggs watermarking image and original image," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 6, no. 1, pp. 47-53, 2017.
- [7] Min Luo, Yadong Luo, Hui Li, Xia Zhang and Yongkui Yang, "Design and implementation of high resolution face image acquisition system under low illumination based on the open source computer vision library," *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, Chengdu, pp. 680-683, 2017.
- [8] A. Al-rawashdeh and Z. Al-qadi, "Creating color image features using local contrast method," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 7, no. 3, pp. 367-376, 2018.
- [9] C. Saravanan, "Color image to grayscale image conversion," *2010 2nd Int. Conf. Comput. Eng. Appl. ICCEA 2010*, vol. 2, pp. 196-199, 2010.
- [10] N. Zarka, Z. Alhalah, and R. Deeb, "Real-time human motion detection and tracking," *2008 3rd Int. Conf. Inf. Commun. Technol. From Theory to Appl.*, pp. 1-6, 2008.
- [11] M. Nixon and A. S. Aguado, Chapter 9. Moving object detection and description, Third Edit. Elsevier Ltd, 2012.
- [12] N. Munera, C. Alvarez, S. Sastoque, and M. Iregui, "Human features extraction by using anatomical and low level image descriptors from whole body images," *2016 XXI Symp. Signal Process. Images Artif. Vis.*, pp. 1-6, 2016.
- [13] N. Singla, "Motion Detection Based on Frame Difference Method," *Int. J. Inf. Comput. Technol.*, vol. 4, no. 15, pp. 1559-1565, 2014.
- [14] T. Kumar and K. Verma, "A theory based on conversion of RGB image to Gray image," *Int. J. Comput. Appl.*, vol. 7, no. 2, pp. 5-12, 2010.
- [15] M. Syafiq, A. Alias, N. Ibrahim, and Z. M. Zin, "Multi-type Noise removal in lead frame image using enhanced hybrid median filter," *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 7, no. 5, pp. 1653-1659, 2017.
- [16] P. Kamboj and V. Rani, "A brief study of various noise model and filtering techniques," *J. Glob. Res. Comput. Sci.*, vol. 4, no. 4, pp. 166-171, 2013.
- [17] R. Chandel and G. Gupta, "Image filtering algorithms and techniques: A review," *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 10, pp. 2277-128, 2013.
- [18] G. Deng and L. W. Cahill, "An adaptive Gaussian filter for noise reduction and edge detection," *1993 IEEE Conference Record Nuclear Science Symposium and Medical Imaging Conference*, San Francisco, CA, USA, vol.3, pp. 1615-1619, 1993.
- [19] B. Nikolov and N. Kostov, "Motion detection using adaptive temporal averaging method," *Radioengineering*, no. 3, pp. 652-658, 2014.
- [20] M. S. Zaharin, N. Ibrahim, and T. A. T. Dir, "Motion detector using high level feature extraction," *Int. J. Appl. Eng. Res.*, vol. 12, no. 24, pp. 14580-14584, 2017.
- [21] A. Bochem, R. Herpers, and K. B. Kent, "Hardware acceleration of BLOB detection for image processing," *2010 Third Int. Conf. Adv. Circuits, Electron. Micro-electronics*, no. July, pp. 28-33, 2010.
- [22] H. H. Nooritawati Md Tahir, Aini Hussain, Salina Abdul Samad, "Human shape similarity and reconstruction using centroidal profile," in *MUSIC*, pp. 2-5, 2005.
- [23] G. Lavee, L. Khan, and B. Thuraisingham, "A framework for a video analysis tool for suspicious event detection," *Multimed. Tools Appl.*, vol. 35, no. 1, pp. 109-123, 2007.
- [24] N. Tahir, A. Hussain, and S. A. Samad, "Support vector machine for human shape recognition using optimized feature set," *J. Teknol.*, vol. 1, pp. 1-6, 2008.
- [25] R. K. Tripathi, A. S. Jalal, and S. C. Agrawal, "Suspicious human activity recognition: a review," *Artif. Intell. Rev.*, vol. 50, 283-339, 2017.
- [26] N. M. Tahir, A. Hussain, S. L. A. Samad, and H. Husain, "The denunciation of human beings is based on the centroid profile method and the artificial neural circuit welders (in Indonesia; Pengecaman insan berasaskan kaedah profil sentroid dan pengelas rangkaian neural buatan)," *J. Teknol.*, vol. 53, pp. 69-79, 2010.