BOOK OF PROGRAM

The 6th International Conference on Engineering Technology and Industrial Application

ICETIA 2019

Intelligent Engineering in Industrial Revolution 4.0: Renewable Energy, Resource and Material Development

Surakarta, 11 - 12 December 2019

ASOSIASI PROFESI APTI TEKNIK INDONESIA

BOOK OF PROGRAM

THE 6th INTERNATIONAL CONFERENCE ON ENGINEERING, TECHNOLOGY, AND INDUSTRIAL APPLICATION 2019

Intelligent Engineering in Industrial Revolution 4.0: Renewable Energy, Resource and Material Development

Held by:

Supported by:

The Use of Gaussian Mixture Model for Counting Human Object on Video

Moch Arief Soeleman[1,*], Irfanul Haq, Muslih[1] , Karis W[1], R. Anggi P[1]

[1] Faculty of Computer Sciences, Dian Nuswantoro University, Semarang, Indonesia

Abstract

Video monitoring has been widely used in various places, for example, tourist attractions, stations, terminals, offices, supermarkets, minimarkets, and other places. The use of video monitoring has the aim to improve security aspects in a place that is considered quite helpful. The rapid development of technology, video monitoring has been implemented for purposes other than for security, such as people counting systems. People counting to find out the number of visitors in a place or building is a difficult job to do, requires a lot of time and often the data obtained is not appropriate.

Gaussian Mixture Model (GMM) is a background reduction method, which is used to identify the background and foreground. Background reduction is an approach that is widely used to detect moving objects on video from static cameras. While blob detection is detecting a group of pixels connected in an image that has a different colour (white or black). Blob is used for object classification, whether the object is a person or not.

From the results of system testing shows that counting people can be done well, with an average percentage of 95.83% recall, 94.5% precision and 90.88% accuracy of the entire video test data.

The use of GMM in counting people in this study can also count people who carry objects properly as long as they are not drawn as in the morning test video. In addition, this system can store the calculated data.

Keywords: Gaussian Mixture Model, counting, video, detect

1. Introduction

Real time report on CCTV cameras for monitoring has been widely used in various places, for example, tourist attractions, stations, terminals, offices, supermarkets, minimarkets, and other places. The use of video monitoring has the aim to improve security aspects in a place that is considered quite helpful. The rapid development of technology, video monitoring has been implemented for purposes other than security, such as a people counting (M. Sivabalakrishnana & K. Shanthib,, 2015) system to determine the estimated flow of tourists, pedestrian traffic management, counting the number of shopping centre visitors and so on. Person counter is a process that is used to count (S. Yoshinaga, A. Shimada , & R. Taniguchi, 2010) the number of people who walk through the entrance or exit by (Phatel & Dhage, 2017). With automatic calculation, the work is more efficient compared to manual calculation to get data for various reasons. First, manual counting by people tends to lose concentration when calculating and thus the number is prone to errors.

Counting people requires two steps namely person detection and tracking (M. A. Kandavalli, , 2018) to count people in a certain direction (G. S. Patel & J. S. Dhage, , 2017). Reducing background (S. Jeeva , & D. Sivabalakrishnan, 2015) and foreground segmentation (Rahman, Ahmed, & Hosian , 2017) are the first step in many computer vision applications. This approach considers the difference between the incoming image and the background to detect foreground objects. There are several methods for implementing people counting (Phatel & Dhage, 2017), for example using the Vertical Kinect depth sensor, the K-Means grouping approach, the Histogram of Oriented Gradients (HoG) feature, and the Kalman filter. The speed and accuracy of background reduction depend on the results of the background extraction algorithm.

In previous studies mentioned the use of background subtraction (Syed , Kalpana , & Jyoti, , 2017) methods and Blob detection (K. Srinivasan, K. Porkumaran, & G. Sainarayanan, 2009) to detect objects that move abnormally and classify them. From these studies there are similarities between the method with its use. According to the research on object tracking, the approach begins with translating the object obtained by reducing the background into an entity in a scene. Objects are tracked in 2-dimensional shapes and classified as living things (people) or inanimate objects. Then, the motion features are calculated and recorded in the form of historical records.

This approach ensures real-time performance, adaptability, noise resistance and non-linear cameras. Previous research also mentioned that the Histogram of Oriented Gradients (HoG) method and the Mean shift algorithm were used for counting people, this study detected heads for the counting process (G. S. Patel & J. S. Dhage, , 2017).

Whereas in this study using the Gaussian Mixture Model (GMM) (Basri & Andani , 2015) method and particle filter for counting people (Xiaofeng & Caidi, 2018). Detection is done by detecting people as a whole, tracking using particle filters, and counting based on the number of centroids.

[*] Corresponding author: author@email.org.me

2. Related Research

In research (K. Rahangdale & M. Kokate, , 2016), it discusses how the application of background subtraction and Blob detection methods to detect objects that move abnormally and classify them. The approach in this study to the problem of tracking people automatically and detecting unusual or suspicious movements in surveillance video. The results of the research ensure real-time performance, adaptability, noise resistance and non-linear cameras, and the elimination of training required by the learning method.

In research (G. S. Patel & J. S. Dhage, , 2017), discussing the Histogram of Oriented Gradients (HoG) method and the Mean shift algorithm that is used for counting people, to do the calculation based on the head detected. The results of this study can achieve high accuracy and can serve continuous tasks with accuracy ranging from 91% to 100%. The accuracy of the results depends on the number of people crossing the counting zone simultaneously, occlusion, intensity variations in the video sequence.

In (Marwa abd el Azeem Marzouk, , 2010) define approaches to the problem of tracking people automatically and detecting unusual or suspicious movements in surveillance videos. Ensuring real-time performance, adaptability, noise resistance and non-linear cameras, and eliminating training needed by learning methods.

The counting of people is based on the head detected on the surveillance video. The results of achieving the accuracy of the results depend on the number of people who cross the counting zone simultaneously, occlusion, variations in intensity in the video sequence (Ajna & Kaur, 2017).

3. Methodology

At this stage a method for calculating CCTV-based human object calculation is proposed as shown in the following diagram

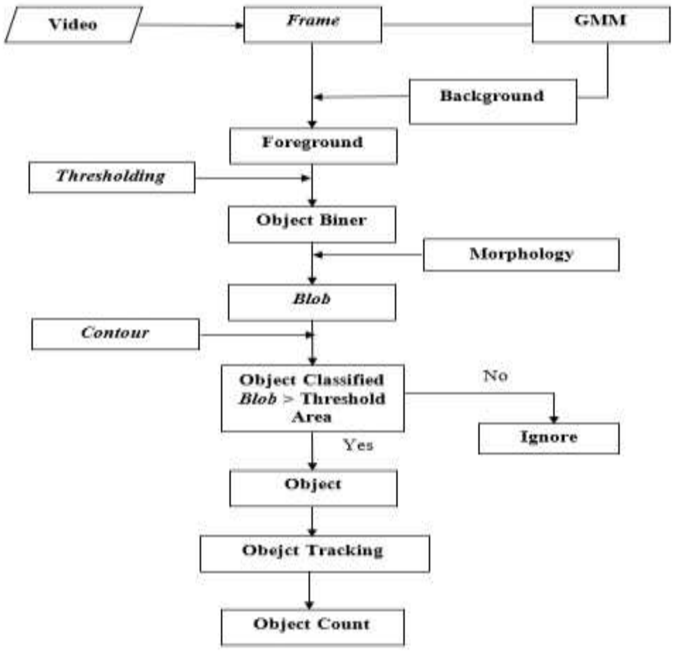

Figure 1. Block Diagram Method of Count Human Object

3.1 Gaussian Mixture Model Algorithm

This stage is to identify the background and foreground by using background reduction (C. Stauffer & W.E Grimson, 1999). Gaussian Mixture Model (GMM) (Xiaofeng Lu & Caidi Xu, 2018) will provide gaussian component functions for each pixel, with input being the colour of pixels where GMM models are formed based on time. The model defines an image that provides two descriptions of the background and foreground. Whereas for GMM measurements of each pixel, the pixel colour (RGB) can be modelled using a mixed distribution of K gaussian. In this study the gaussian n value used is 3. The GMM equation model is as follows:

$$P(X_t) = \sum_{i=1}^{K} \omega_{it} \eta(X_t, \mu_{i,t}, \Sigma_{i,t}) \qquad (1)$$

$$\eta(X_t, \mu_{i,t}, \Sigma_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}}\Sigma^{1/2}} exp^{\frac{1}{2}(x_t-\mu_t)^T}\Sigma^{-1(x_t-\mu_t)} \quad (2)$$

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha M_{(i,t)} \quad (3)$$

$$\mu_{i,t} = (1-\rho)\mu_{i,t-1} + \rho X_{i,t} \quad (4)$$

$$\sigma_{i,t}^2 = (1-\rho)\sigma_{i,t-1}^2 + \rho(X_t - \mu_{i,t})^2 \quad (5)$$

$$\rho = \alpha\eta(X_t, \mu_{i,t}, \sigma_{i,t}) \quad (6)$$

$$B = \arg min_b(\Sigma_{k=1}^b\, w_k > c_f) \quad (7)$$

3.2. Object Threshold

Threshold is a simple and effective technique for segmenting images (Z. Zeng,, J.Jia , & Y.Chen, 2016) by processing pixels on an image or eliminating some pixels and also maintaining a few pixels so as to produce a new image. With threshold computation, it can easily get the edge of an image. At this stage, the image of the Gaussian Mixture Model (GMM) (Yesong, XiaopingLi, Na Fu, & Qiongxin Liu, 2014) results is carried out thresholding (Z & F. Liangzhong, 2010) in order to get better results that distinguish between background and foreground. If the pixel value is greater than the threshold value, then a value of 1 (white) is set, otherwise a value of 0 (black) is specified.

$$for\ pixel\ I\ (w,t)\ scope\ the\ image\ I$$

$$\begin{aligned} &if\ I(w,t) > thresh \\ &\quad I(w,t) = 1 \\ &else \\ &\quad I(w,t) = 0 \\ &end \end{aligned}$$

3.3 Morphology Operation

This stage is useful for removing noise in the form of small pixels that are detected as the foreground and filling small holes in the object. Morphological operations used in this study are opening and closing. Opening morphological (A. Amer, 2002) operations are performed on images generated from the thresholding stage by using a 3x3 pixel kernel. Opening is an erosion process that is followed by dilation (C. Nagaraju, Nagamani, & G. Rakesh , 2012). Starting with erosion of the image then the results are carried out again.

Opening is used to remove small pixels that are detected as the foreground and can make the edges of the image smoother. Erosion aims to reduce or erode the edges of objects or by making pixels that are worth 1 (white) neighbourhood pixels that have value 0 (black) into pixels that have value 0 (black). Dilation aims to enlarge the object segment (binary image) by adding layers around the object or by making pixels that are 0 (black) neighbourhood pixels 1 (white) into pixels that are 1 (white).

3.4 Contour

This stage is to detect the edges of objects in the video frame. Contour here is useful for obtaining and drawing object shapes. The images used are binary (K. Mahesh & K. Kuppusamy,, 2012) images from the results of morphological operations. In this study, the method of taking contours only cares about external contours. The process of getting a starting point by scanning a pixel to find the boundary of an object. After the starting point is found, neighbouring pixels are used as a reference to find the next point that will become part of the contour

In this way, pixels that have been scanned or that are located (pixels that have a value of 1) may not be the next pixel that is part of the contour. Contour search ends after meeting again with the starting point. Contour will store the coordinates (x, y) of the boundary shape of the object. While the contour approach method in this study, by eliminating all redundant points and compacting the contours

4. Experimental Results

In this method experiment each frame taken from video data input, computation is implemented using the Python V2.7 programming language and OpenCV v3.4.4. The results of the person count will be stored in the MySQL database. Then the background reduction stage is carried out using the Gaussian Mixture Model (GMM) method. This stage is to

identify the background and foreground. Background is considered as something that is constant in a series of images or that remains static, whereas foreground is everything that changes (moves). Then thresholding is done to change to binary image and to get clearer results that distinguish background and foreground.
Furthermore, pixel cleaning is done using morphological operations by filling small holes from the thresholding results. Morphological operations are carried out with the aim of removing noise in the form of small pixels detected as the foreground and covering the foreground areas that are separated from each other, but close together. Furthermore, contour is used to detect edges on a BLOB. A blob which is a set of white pixels in a binary image that will be defined as an object. This contour will be the basis for the next step, object classification. Then the classification stage to separate whether the BLOB includes objects that are expected or not. By comparing the previously detected BLOB with a predetermined threshold value (BLOB> area threshold), the threshold area is a container that collects the minimum threshold value that has been determined to classify the detected BLOB is an object (person) or not. The area value is not fixed or depends on the video used. If the BLOB is less than the area then the BLOB is ignored and if the BLOB is more than the area then tracking is done. The tracking process by giving ID to the detected object and save its initial position, after the object crosses the specified limit have done.

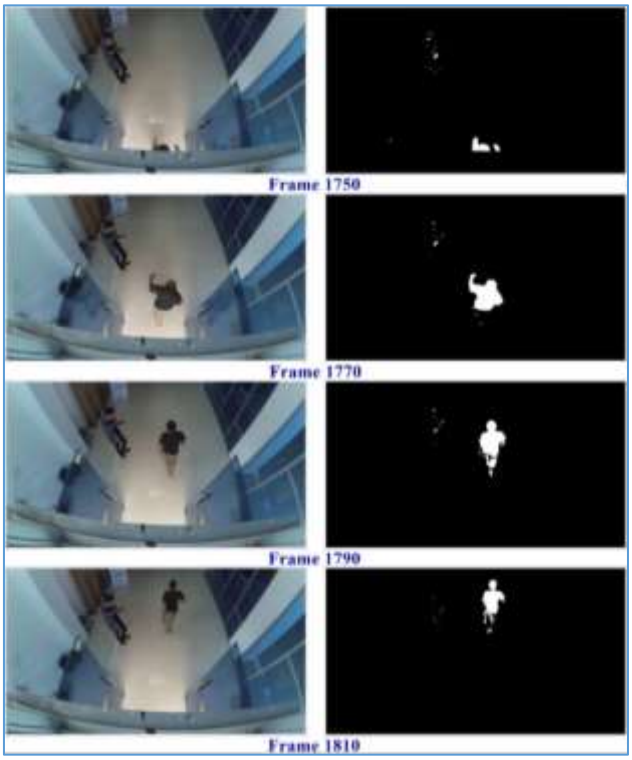

Figure 2. Result GMM from Video Frame

At this stage, the image of the Gaussian Mixture Model (GMM) results is carried out thresholding in order to get better results that distinguish between background and foreground. If the pixel value is greater than the threshold value, then a value of 1 (white) is set, otherwise a value of 0 (black) is specified. In this study the threshold value is set at 200, so that 1 if the image (w, t> 200 and 0 if the image (w, t) ≤200). The choice of the threshold value is due to the results of GMM sometimes there is still a little grey colour or that does not reflect the white colour caused by lighting or shading, so the value chosen is close to white

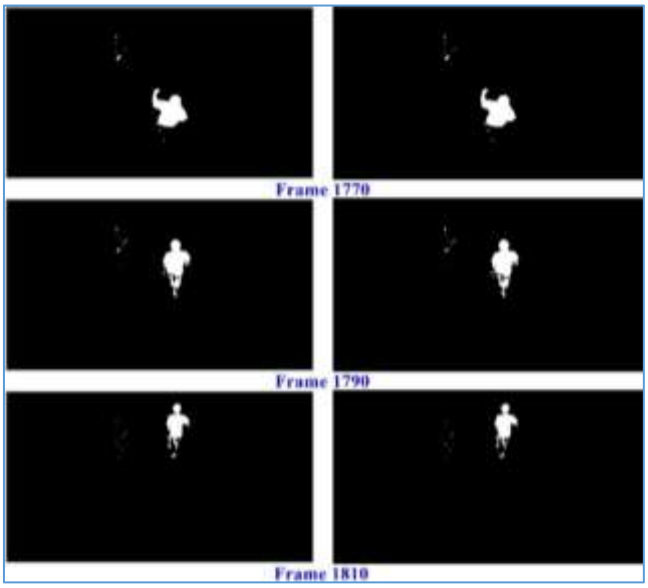

Figure 3. Result Thresholding from GMM

In the next stage, the closing morphology operation is carried out on the frame resulting from the opening morphology operation using a 11x11 pixel kernel. Closing is a process of dilation followed by erosion. Where the image is first carried out dilation which is then followed by erosion. Closing aims to fill small holes in objects and to join adjacent objects

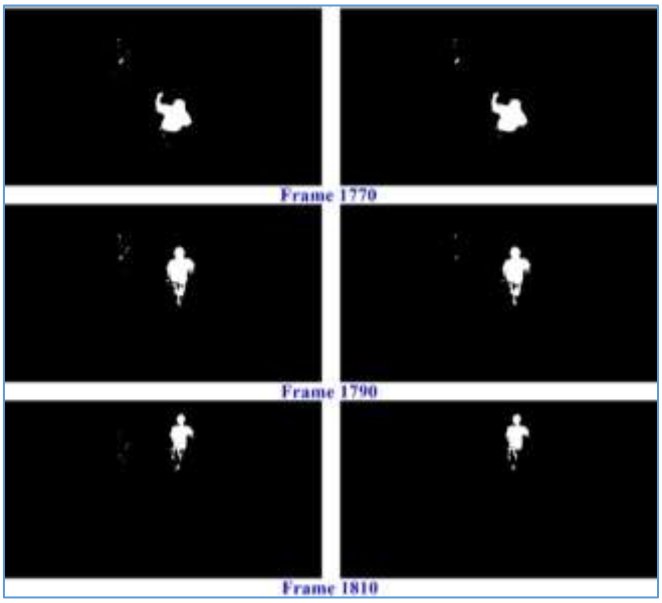

Figure 4. Result Opening Morphology from Thresholding frame

While closing morphology operations are performed on images resulting from opening morphological operations using a 11x11 pixel kernel. Closing is a process of dilation followed by erosion (A. Amer, 2002) (C. Nagaraju, Nagamani, & G. Rakesh , 2012). Where the image is first carried out dilation which is then followed by erosion. Closing aims to fill small holes in objects and to join adjacent objects.

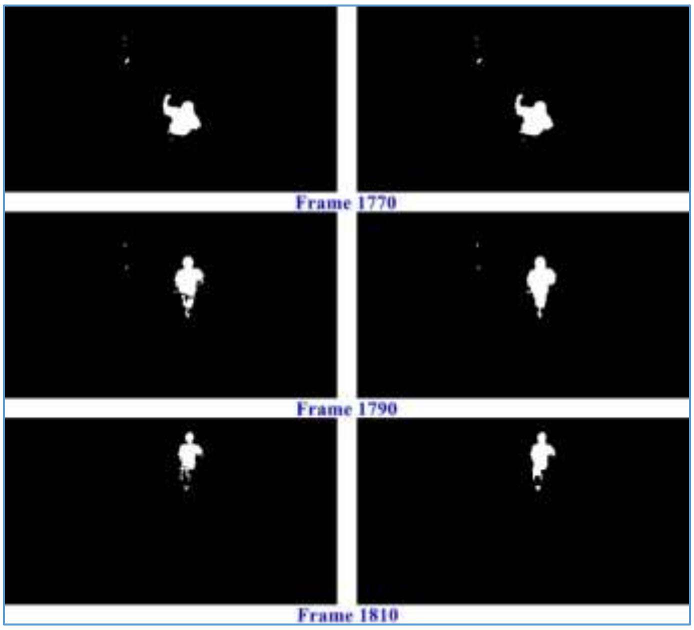

Figure 5. Frame Result of Closing from Opening Morphology

In the next step the determination of the contour is represented by a rectangle, so it only requires 4 points and saves memory which makes the system lighter because it does not store too many coordinates (x, y). At this stage a contour area (blob detected) will also be generated which will then be compared to the minimum threshold area value for classifying objects. Contour area value is the value obtained from the shape of the object by counting the number of pixels on the object. The number of pixels counted is the number of non-zero pixels in a binary image.

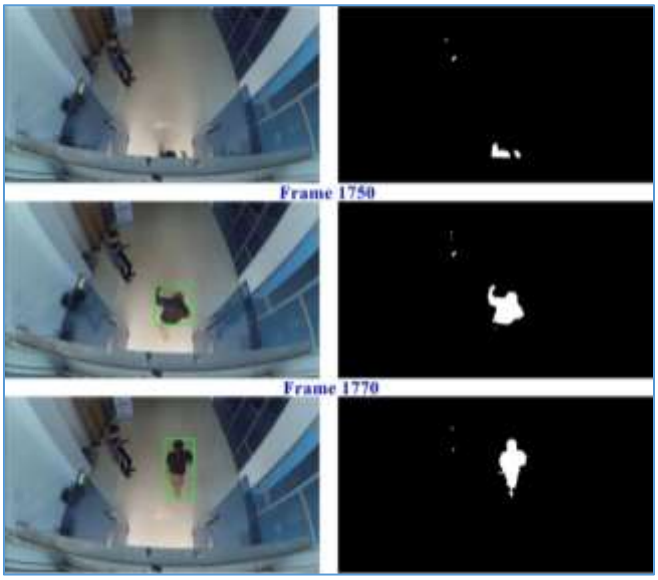

Figure 6. Contour Detection

At this stage to determine the detected BLOB which is included as a person object or not. The BLOB value is generated from the contour area value, which counts the number of non-zero pixels in a binary image. To classify objects, it is necessary to determine the minimum threshold area value, the area value is used to compare the detected BLOB values. If the BLOB> area then the BLOB is included BLOB from the person object and then tracking is done for the calculation. Determination of the value of these areas is very influential also on the accuracy of counting people. In this study to determine the value of the area by taking a video frame size, the size is 856x480 pixels.

In the next stage the object that has been classified and then carried out tracking the position of the object and then counted. To do tracking and counting it is necessary to determine the coordinates of the line position to start tracking the object, stop tracking the object and count the object. The counting line is a line where the system counts people entering or leaving, if the object of a person crosses the line. While the tracking line is a line where the system starts evaluating the direction of the object and tracking the position of the object and stop tracking the position of the object.

In this study, the calculation line is at 160 coordinates for counting people entering and 270 coordinates for counting people out. Whereas the tracking line in this study is at coordinates 75 to start evaluating the direction of the object, tracking objects from inside and stopping tracking objects after counting, while the other line at coordinates 350 is to start evaluating the direction of objects, tracking objects from outside and stopping tracking objects after exit count. These coordinates are the coordinates of the y or height of the video frame.

Figure 7. Tracking Object

5. Evaluation

Testing method for counting people is done by using 4 video files from a Closed-Circuit Television (CCTV) camera installed in the Building, the video used for testing is taken at different times namely morning, afternoon, evening and night. Furthermore, each video is counted by people in the video scene using the Gaussian Mixture Model (GMM) method and Blob detection that has been implemented in this study. The test by measuring the level of Precision and recall which are two calculations that are widely used to measure the performance of the system or method used. Precision is the level of accuracy between the information requested by the user and the answers provided by the system. While recall is the success rate of the system in finding back an information. Accuracy is defined as the level of closeness between the predicted value and the actual value. The formulas of precision, recall, and accuracy are:

$$Precision = TP/(TP + FP)$$

$$Recall = TP/(TP + FN)$$

$$Accuracy = (TP + TN)/(TP + TN + FP + FN)$$

The testing details of each video are as follows:

Table 1. Result Evaluation Performance

Dataset	Recall	Precision	Accuracy
Morning Video	100%	92.9%	92.9%
Afternoon Video	91.67%	100%	91.67%
Night Video	100%	85.7%	86.67%

In the night video scene, there are objects that are not people, namely cats, but are not detected and are not counted. In this video there are also people with scenes carrying trash cans but the person is counted correctly. Testing this video also occurs differences in the results of the calculation of the system with the actual number of people in the video scene. The difference in results is due to error detection in the 7th person, in the video scene that goes out, where sometimes the object is detected one sometimes two so that the object is counted twice, namely in and out. Subsequent counting errors in the 8th and 9th people in the adjoining video scene right at the door of the building make 2 of these people counted 3. The error is due to poor lighting and shadowy night videos that are more clearly visible than other videos. The calculation of the percentage of recall, precision and accuracy

5. Conclusion

Based on the results of testing and analysis conducted in this study, it can be concluded that in this study successfully implemented the Gaussian Mixture Model (GMM) method and BLOB detection for counting people using video data from CCTV installed in the room. The Gaussian Mixture Model (GMM) method is used in the system to separate the background from the foreground. In addition, GMM is quite tough with changes in light intensity, can adapt well, the resulting foreground is good as long as the background colour is not similar to the foreground colour. Although the shadows can be adapted well by the background, if the shadows look clearer than usual, change shape or change position quickly, the shadows disturb the resulting foreground. Detection results and counting people well as long as people in the video scene are not close together or walking side by side which makes two people in the video scene only count one. That's because the Bounding Box of the two objects together, so the system only counts one. The people counting system in this study can also count people who carry objects properly as long as they are not pulled as in the morning test video. From the results of testing the system shows that counting people can be done well, with an average percentage of 95.83% recall, precision 94.5% and 90.88% accuracy of the overall video test data used.

Acknowledgement

This research was funded by a PDUPT Research Grant number 060/A.38.04/UDN-09/V/2019, from the Ministry of Research, Technology, and Higher Education, Republic of Indonesia

References

A. Amer. (2002). New binary morphological operations for efective low-cost boundary detection. *International Journal of Pattern Recognition and Artitcial Intelligence, 17*(2).

Ajna, & Kaur. (2017). Review of Image Segmentation Technique. *International Journal of Advanced Research in Computer Science*, pp. 36-39,.

Basri , I., & Andani , A. (2015). Gaussian Mixture Models Optimization for Counting The Numbers of Vehicle by Adjusting The Region of Interest under Heavy Traffic Condition. *International Seminar on Intelligent Technology and Its Applications.* Makasar.

C. Nagaraju, S., Nagamani, G., & G. Rakesh , P. (2012). Morphological Edge Detection Algorithm Based on Multi-Structure Elements of Different Directions. *International Journal of Information and Communication Technology Research, 1*(1).

C. Stauffer , & W.E Grimson. (1999). Adaptive background mixture models for real-time tracking. *IEEE Computer Society Conference on Computer Vision and Patern Recognition.*

G. S. Patel , & J. S. Dhage, . (2017). Real-Time People Counting Method with Head Localization and Tracking System Consideration. *" International Journal of Innovative Research in Computer and Communication Engineering*, . 7974-7981.

K. Mahesh , & K. Kuppusamy,. (2012). A New Hybrid Video Segmentation Algorithm using Fuzzy C Means Clustering,. *International Journal of Computer Science,*, 1-9.

K. Rahangdale , & M. Kokate, . (2016). Event Detection Using Background Subtraction For Surveillance Systems. *International Research Journal of Engineering and Technology*, 1300-1308.

K. Srinivasan, K. Porkumaran, & G. Sainarayanan. (2009). Improved Background Subtraction Techniques for Security in Video Applications. *Proc. of 3rd International Conference on Anti-counterfeiting, Security, and Identification in Communication,.*

M. A. Kandavalli, . (2018). A novel approach on object detection and tracking using adaptive background subtraction method. *Second International Conference on Computing Methodologies and Communication.*

M. Sivabalakrishnana , & K. Shanthib,. (2015). Person Counting System Using EFV Segmentation and Fuzzy Logic. *2nd International Symposium on Big Data and Cloud Computing.* Chennai.

Marwa abd el Azeem Marzouk, . (2010). Modified background subtraction algorithm for motion detection in surveillance systems. *Journal of American Arabic Academy for Science and Technology, 1*(2), 112-123.

Phatel , P., & Dhage, J. (2017). A Survey On Real Time People Counting Method With Head Localization And Tracking System Consideration. *International Journal of Advance Engineering and Research Development*, 218-222.

Rahman, A., Ahmed, M., & Hosian . (2017). An adaptive background modeling based on modified running Gaussian average method. Bangladesh: International Conference on Electrical, Computer and Communication Engineering (ECCE).

S. Jeeva ,, & D. Sivabalakrishnan. (2015). Survey on Bankground Modeling and Foreground Detection for Real Time Video Surveillance. *2nd International Symposium on Big Data and Cloud Computing (ISBCC'15).* Chennai.

S. Yoshinaga, A. Shimada , & R. Taniguchi. (2010). Real-Time People Counting using Blob Descriptor p. 143–152, 10 December 2010. *Security Camera Network, Privacy Protection and Community Safety,*, (pp. 143-152).

Syed , T., Kalpana , G., & Jyoti, , S. (2017). Moving Object Detection Using Self Adaptive Gaussian Mixture Model for Real Time Applications. *Proceeding International conference on Recent Innovations is Signal Processing and Embedded Systems.*

Xiaofeng , L., & Caidi, X. (2018). Novel Gaussian mixture model background subtraction method for detecting moving object. *International Conference of Safety produce Information (IICSPI),.*

Xiaofeng Lu, & Caidi Xu. (2018). Novel Gaussian mixture model background subtraction method for detecting moving object. *International Conference of Safety produce Information (IICSPI).*

Yesong, XiaopingLi, Na Fu, & Qiongxin Liu. (2014). Fast Moving Object Detection Using Improved Gaussian Mixture Models. *ICALIP 2014.*

Z, Y., & F. Liangzhong. (2010). Moving object detection based on running average background and temporal difference. Hangzhou: IEEE International Conference on Intelligent Systems and Knowledge Engineering,.

Z. Zeng,, J.Jia , & Y.Chen. (2016). Pixel modeling using histograms based on fuzzy partitions for dynamic background subtraction. *IEEE Transaction In Fuzzy System,*, 1-10.