International Journal of Computer and Electrical Engineering, Vol. 1, No. 1, April 2009
1793-8198

Real-Time People Counting System using Curve Analysis Method

Lim Eng Aik, and Zarita Zainuddin

***Abstract*— A new people-counting system is described in this paper. The proposed system showed its effectiveness and flexibility to count the number of people in a certain image and performing real-time people counting. A Curve Analysis method is developed for the crowd monitoring system to count the objects. Promising results were obtained and the evaluation on their performance indicates the proposed Curve Analysis method counting method provides good results.**

***Keywords*—Cluster segmentation, curve analysis, edge detection, median filter.**

I. Introduction

The crowd monitoring system has been widely used in many areas such as transportation control and safety management. The aim of a crowd monitoring system is to count the number of people in a confined area. Real time monitoring, estimation and control of crowd appear to be essential for overcrowded situation which able to provide a qualitative output corresponding to the occurrence of an alarm condition. Previously, a few real-time crowd estimation researches have focus on using visual information as in [1], [2], [3], and [4]. Other researcher also using a distributed extended Kalman filter system with both spatial and temporal information as based for model optimization [4], [5]. Kalman filter type algorithm is of slow convergence due to its complexity in computation. Another research approach [9], [10] in counting the number of people flowing in and out of the station by using template matching to track people. Currently, a neural network based vision system for estimating the crowd level of an underground station platform was reported [6], [7], [8]. Applying neural network in the vision system will increased computational time and system complexity.

In this paper, we are describing a visual based system designed for performing an accurate people counting. The principle of the operation of the crowd monitoring system is based on image processing technique. In computer vision, an image has a lot of information such as color intensity, and color intensity is represents by number. By using image processing technique, these number represented by the image will be rearranged and changed in an appropriate way. Then by analyzing the format of these numbers, a computer can get the information from the images. In this paper, the main issues discussed in this paper are how to extract the essential data from unconstrained images and perform accurate people counting by using Curve Analysis method.

In general, automatic people counting systems can be divided into two stages as reported in [9], [10]: image processing and image understanding. Image processing will localize the objects of interests from the background and other objects. Meanwhile, features extraction techniques are applied for image understanding. Template matching used in [9], [10] to count and track people will located the objects in images and the number of people was obtained. But, there are two problems occur in their solution, that are, the system is directly based on image information, which contributed to high input dimensions for the people counting systems, meaning, this type of counting systems can only deal with a given size of images. Another problem is the sensitivity to noises. So, overall results of these system are greatly depends to the performance of image processing technique.

This paper is organized as follow: Section 2 briefly explains about the condition involved and characteristic of background for the proposed system; Section 3 deals with the image processing stage; Section 4 discussed on Curve Analysis which is the image understanding stage; Section 5 gives the performances; Section 6 will gives the discussion and evaluation on our proposed system.

II. Condition and Characteristic of Image Background

The image handled in the system is monochrome image, meaning that the color of the image is in gray color from black and white. There are 256 gray-levels between black and white. The black color in the image is represented as 0 while the white color in the image is represented as 255. In monochrome image, each pixel value of the pixels in the image has a certain value from range of 0 to 255 to represent the level of gray color of the pixel. The gray-level color band is shown by Fig. 1. There are few factors must be considered in choosing a suitable environment for people counting, that are the illumination condition of the environment and the characteristic of the background.

Fig. 1 The 256 gray-level colour band

The illumination condition e.g. different illumination level of the environment has a serious effect on the image captured

Manuscript received January 6, 2009. This work was supported in part by the Universiti Malaysia Perlis under Grant 9001-00102.

Lim Eng Aik is with the Institute of Engineering Mathematic of Universiti Malaysia Perlis, Perlis, 02000 Malaysia. Phone: +6012-4686859; e-mail: eng_aik@yahoo.com.

Zarita Zainuddin is with the School of Mathematical Sciences, Universiti Sains Malaysia, Penang, 11800 Malaysia. e-mail: zarita@cs.usm.my.

such as the pixel intensity of the image of a constant scene captured by the camera will be different. Therefore, a stable illumination level is very important. It is very difficult to process the image with a large variation of the illumination level for the environment that will also making the variation of the image captured by the camera become very large.

Generally, the illumination level for indoor environment is more stable than the outdoor environment due to the illumination source of outdoor environment is mainly from sunlight that cannot be controlled. Now, the environment for people counting in the image handled by the system is the walkway of a shopping mall which is an indoor environment. Thus, it provides a good illumination condition to process.

Apart from the illumination condition of the environment, another factor that has serious effect on the image processing is the characteristic of the background. In order to limit the variation on the image captured, an image with a static background is preferred to be processed. Also, there must have contrast between the background of the image and the object in the image. This is because the principle to separate the object from the background in the image is made use of the fact that there is difference of pixel intensity between object and the background. If the pixel intensity of the background and the object is too similar, then it is very difficult to separate the object from the background.

III. Image Processing

In the crowd monitoring system, the people in the image are needed to be counted. The first thing that needed to do is to separate the object from the background and the best way to it is by changing the image into binary image. This means that the object-of-interest in the image is represented by black pixel while the background is represented by white pixel.

A. Image Pre-Processing

Images acquired from the camera needs to be preprocess before undergoing for the process of segmentation. The resolution of the image processed by the system is 318 x 238 in pixels. Either the resolution of the image is large or small, the two criteria must be considered;

1. If the resolution of the image is too large, noise become dominated and the image is blur and the processing time will be slower as more pixels has to be handled by the computer.
2. If the resolution of the image is too small, it is so difficult to see and represent the information.

The image handled by the proposed system is monochrome image. As a result, 8-bit bitmap format is used instead of 24-bit RGB image. Hence, the processing time will be faster as the computer handles fewer bits.

B. Cluster Segmentation

Segmentation is an image processing technique, which is used to extract the object out of the background in the image. Each object occupies certain pixels in the image. Generally, there is some difference of the pixel value and gray level of the occupied pixels of the object with reference to the pixels of the background image.

In the proposed system, the absolute difference of pixel values between the input image and a reference background image is used for segmentation. Also, the detection of the difference of pixel value between the input image and the reference background image is based on a sub-block of pixels rather than individual pixels. This is due to the texture of the floor of the counting zone contain some noise pixels in the output image.

These noises pixels are eliminated when sub-block of pixels is treated as a unit during segmentation process with each sub-block contains 9 pixels. Not all pixels in a sub-block are noise pixels. Hence, the noise effect of the total pixels intensity of a sub-block is relatively much small than that of an individual pixel. If the sub-block of pixels is not occupied by the object, then the whole sub-block will changed to white pixels including the noise pixels in the sub-block where, the total pixel value of the sub-block will be compared with the sub-block at the same position of the reference background image. If a sub-block of the input image has noise pixels, then the sub-block of the reference background image at the same position will also has noise pixels, which will then compensate each other during comparing and subtraction of the pixel value of the two sub-blocks.

C. Sub-Block Processing

Initially, the counting zone of the image is divided into several sub-blocks. In the proposed system, each sub-block consists of 3x3 pixels as shown in Fig. 2. The 9 pixel values of the pixels of the sub-block will be added together and then we obtained the total pixel value of the sub-block. The total pixel value of the sub-block in the image will then be compared with the sub-block of the background image at the same position.

Normally the gray-level of the sub-block of the processing image is a bit darker than the background image due to shadow of the people but it can be seen that, if the sub-block of the input image that belonged to the background, the difference of the gray level and the total pixel value of the two sub-blocks will not be very large due to similar pixel value and gray level. As a result, if the sub-block of the input image is belonged to the background, then the absolute difference of the total pixel value between the sub-block of the input image and the sub-block of the background image will be very small.

As mention above, the pixel value of the 9 pixels of the sub-block will be added together and the total pixel value of the sub-block is obtained and then the total pixel value will be compared with the ones of the sub-block of the background at the same position, which resulting a very large absolute difference of the total pixel value between the two sub-blocks. As a result, if the absolute difference of the total pixel value of the two sub-blocks is very large and greater than a threshold value, the sub-block will be considered as belonged to the object and all the pixels of the sub-block will be changed to black pixels. If the absolute difference of the total pixel value of the two sub-blocks is small and less than the threshold value, the sub-block will be considered as belonged to the background and all the pixels of the sub-block will be changed to white pixels.

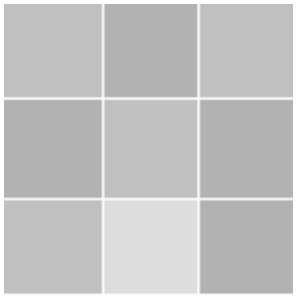

Fig. 2 Sub-block with 3 x 3 pixels

Whether the sub-block of the input image is changed to a block of white pixels to represent the background or changed to a block of black pixels to represent the object is depended on the threshold value. If the absolute difference of the total pixel value between the sub-block of the input image and the sub-block of the background image is large and greater than the threshold value, the sub-block of the input image will change to a block of black pixels to represent the object. Otherwise the sub-block of the input image will change to a block of white pixels to represent the background.

D. The use of Edge Detector "Sobel Filter"

To solve the problem of shadow corruption, the edge detector Sobel filter has to be used. The concept is that, after apply Sobel filter on the image, the edge of the object can be obtained. By analyzing the edge points in certain area, it can be estimated whether the area is belonged to the background or not. To build a Sobel filter, two 3x3 convolution kernel are built as in Fig. 3:

-1	-2	-1
0	0	0
1	2	1

(a)

-1	0	1
-2	0	2
-1	0	1

(b)

K_{11}	K_{12}	K_{13}
K_{21}	K_{22}	K_{23}
K_{31}	K_{32}	K_{33}

(c)

Fig. 3 Sobel convolution kernel

Fig. 3(a) is the 3x3 convolution kernel that use to calculate $O_1(i, j)$ as in (1). Fig. 3(b) is the 3x3 convolution kernel that use to calculate $O_2(i, j)$. Fig. 3(c) shows the coordinate system of the kernel. The formula of convolution is as below:

$$O(i, j) = \sum_{k=1}^{3}\sum_{l=1}^{3} I(i+k-2, j+l-2)K(k,l) \qquad (1)$$

where $I(i, j)$ is the pixel value of the pixel (i, j) of the input image. $K(k, l)$ is the value of the convolution kernel . $O(i, j)$ is the calculated pixel value of the pixel (i, j) of the processing image.

The 3x3 convolution kernels shown by Fig. 3(a) and Fig. 3(b) are applied to the pixels of the counting zone of the input image. Then $O_1(i, j)$ and $O_2(i, j)$ of each pixels of the counting zone can be found. After finding $O_1(i, j)$ and $O_2(i, j)$ of each pixel, then the gradient magnitude at each pixel $G(i, j)$ are needed to be found. The formula of magnitude of the gradients as below:

$$G(i, j) = \sqrt{O_1^2(i, j) + O_2^2(i, j)} \qquad (2)$$

From (2), the gradient magnitude $G(i, j)$ of each pixel in the counting zone has to be found. If $G(i, j)$ is large and greater than the threshold value τ, the pixel (i, j) is considered as edge and will be represented as white pixel [11].

E. The use of Median Filter

In order to eliminate the noises and preserve the edge points in the output image when the image is processed by Sobel filter, some pre-processing work has to be done on the input image. Preprocessing work involves smoothing the input image to remove noise and unwanted detail by median filter. As shown in Fig. 4(b), it can be seen that the image has been smoothed.

(a) (b)

Fig. 4 (a) is the original input image. (b) is the output image by median filtering on image (a).

Although the image is a little blurred after smoothed, it has no harm effect on further processing. The algorithm of median filter will sorts the pixels in a 5x5 sub-block in ascending order according to the pixel value of the pixels in the 5x5 sub-block and replaces the center pixel of the sub-block with the pixel which has median pixel value [12]. If the small dark point in the counting zone has relative low gray level, when the median filter sorts the pixels in a 5x5 sub-block, the small dark pixels will be filter out. After the pre-processing work, Sobel filter will then be applied to the filter image to get the edge points of the object.

The part in the counting zone of the image that corrupted by shadow has relative dark gray level which will cause the segmentation process to misclassified this part as object rather than background. Due to the capabilities of edge detector to detect the sharp variation of pixel intensity between pixels and pixels rather than detect the gray level of the pixel [12], as a result, the edge points are resistance to the shadow. Edge points are marked only if there is sharp change of pixels intensity between pixels and pixels.

According to this characteristic, if a large area is founded to have very little white pixels, this area will be estimated as background and will be changed to white pixels to represent background. The edge white pixels here are the valuable information which can be used to estimate whether the area is belonged to background in the counting zone of the image or belonged to object.

In order to check the number of edge white pixels in certain areas, template has to be designed to scan through the counting zone of the image. The operation of the template is to scan through the counting zone of the image from left to right and the number of white pixels inside the template will be counted. If the number of white edge pixels inside the template is less than a threshold value, the area that covered by the template will be considered as background and the area will be changed to white pixels. Fig. 11. showed the result of median filtering and Sobel filter.

IV. Curve Analysis

Curve Analysis is the method developed in this paper in order to count the objects in the image after the input image is processed by segmentation. A template is made to scan the counting zone of the segmented image.

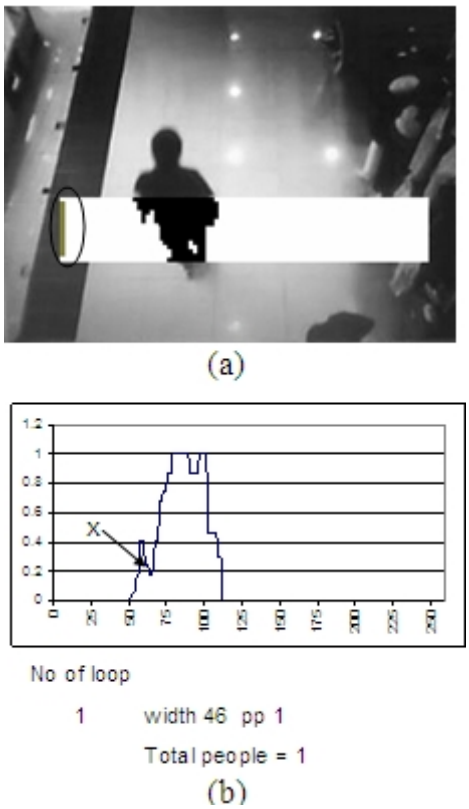

Fig. 5 Plot of black pixels percentage from a person walk in counting zone

The template is shown at the most left hand side of the counting zone in Fig. 5(a), where the oval is found. The template will scan the counting zone of the image from left to right. At the most left hand side, the position of the template is 1 and the position will be increased when the template move to right. At the most right hand side, the position of the template is 256.

This means that the template will move from position 1 to position 256 but this time the template is not used to find the matched object. When the template moves from position 1 to position 256, the percentage of the area of the template that covered by black pixels is recorded and will be represented as *y*-axis of a curve and the position of the template will be represented as *x*-axis of the curve. As a result, when the template scans through the counting zone of the image, a curve is plotted where *x*-axis shows the position of the template and y-axis shows the percentage of the areas of the template that covered by black pixels.

When the template is at position 57 where the object detected, it can be seen that a certain percentage of the area of the template is coved by the black pixels of the object. As a result, in Fig. 5(b), it can be seen that when $x = 57$, *y*-axis will marks at a certain value that indicate the percentage of the area of the template which is coved by black pixels.

Obviously, when an object is in the counting zone, a curve that rises and then drops will be plotted. This means that a loop will be plotted and the object in the image is represented by the loop. As a result, by counting the number the loops and analyzing the width of the loop, it can teach the computer to recognize and count the target object in the image. In this paper, the width of the loop at *y*-axis = 0.2 is calculated.

For example in Fig. 5(b), the width of the loop 46 means that the width of the loop of the curve at the position *y*- axis = 0.2 is 46. In Fig. 6(a), it can be seen that two people are clamped together. After processes the image and plots the curve, in Fig. 6(b), it can be seen that there is one loop rather than two loops. But the computer can recognize that there are two people rather than one people, it is because the width of the loop in Fig. 6(b) is wider than that of the loop that represents one people. The thing is that we need to set the suitable threshold value for the width to detect and recognize the number of people represented by the loop. After processes more samples images, we have seen that the width of the loop that represent a typical fat people is about 55 – 65. To have some margin, the threshold value of the width of one people is set at 46 in the system. The number of people will be count as the result from the width of loop divided by width of one people which is then round to the nearest integer. For example, for Fig. 6(b), the width of the loop is 82, the computer will calculate it as 82/46 = 1.78, and round it to the nearest integer which is 2, so the computer will recognize that there are two people. The same principle can be applied to the case that 3 or 4 people are clamped together as in Figure 7(b). In Figure 7(c), the width of the loop is very large, that is 117. Now, the computation is done by 117/46 = 2.54 and round it to the nearest integer 3, so the computer can successfully recognize that three people are representing by the loop.

Fig. 8(a) show a people who pixel values is so closed to the background. As a result, from Fig. 8(b), it can be seen that the some black pixels that should represent the object are missed even a low threshold value is used for segmentation. However, the computer can recognize and correctly count the people. Fig. 8(c) has shown the loop plotted out. We observe that the shape of the curve will not be changed a lot if some black pixels that represent the object in the segmented image are missed unless almost all black pixels are missed. But it is a rare case that almost all part of the body of the target person cannot be segmented out. It can be seen that the adaptation to the pixel value that similar to background problem for the counting method Curve Analysis is very high.

V. Experiments

The digital images in this experiment come from video in a walkway of shopping mall. We select 50 images. In these images, there are people in various conditions: some people

are standing, some are walking, some are clamping together, some are in color that similar to background, etc. We have taken all the 50 images for testing. The system has been evaluated on their performance, and the average people counting accuracy is 86%. Fig. 6, Fig. 7, Fig. 8, Fig. 9, and Fig. 10 gives some images for testing and the results of our system. They show that the Curve Analysis (CA) method can overcome the noise of shadow and similar pixel intensity problems.

VI. Discussion and Evaluation

In this paper, cluster segmentation is used to extract the object from the background and then the method has been developed to count the extracted object is Curve Analysis. Now the performance and the adaptability to the problems of these image processing techniques will be discussed.

A. Shadow

Dark shadow will be induced on some area if parts of light rays are blocked by people when people come together. Then there is a great change of pixels intensity with reference to the background image. When segmentation process detects such changes of pixels intensity, it will misclassify the area as objects and change it into black pixels. As a result, over-counting will occurred when Curve analysis is applied to count the object.

However, with the aid of median filter and Sobel edge detector, the edge and outline of the object can be founded and is represented as white pixels. By analyzing the edge white pixels, the area that corrupted by shadow can be founded out and correction can be made. With the help of Sobel filter, the segmentation can adapt the problem of shadow on most images and the accuracy is greatly increased. The result is shown in Fig. 12.

B. Similar Pixel Intensity of the People to the Background

Sometimes, the object in the image may have similar pixel intensity to the background. The object in the input image cannot provide a significant change of pixel intensity with reference to the background image if the object has similar pixel intensity to the background. Then the segmentation process may not successfully extract the object as some black pixels that represent the object are missed.

As the problem of shadow has to be solved with the aid of Sobel filter, a low threshold value can be used for segmentation. Then more black pixels can be preserved to represent the object and the result is improved. So the method Curve Analysis was able to count the object in the image more accurate and the result is correct. This is due to the output of Curve Analysis on image where the object is, were represented by the loop of a certain width.

Although some black pixels are missed for the image object, but the shape of the curve plotted will not be changed too much for such small loss of black pixel. The counting result is mainly depended on the shape of the curve. As a result, the method Curve Analysis shows a higher adaptation and flexibility for the images.

VII. Conclusion

A crowd monitoring system that use for people counting by the techniques of image processing has been presented. The image processing techniques, cluster segmentation is used to extract the object out from the background. The advantages of this method are it simplicity to performed people counting and ability to eliminate the noise without building up any filter. However, by the effect of shadow, some results are not very good. But we have shown that by using edge detector Sobel filter and median filter, the shadow induced in the segmentation process can be removed and then a low threshold value can be used for segmentation in order to obtain the black pixels of the object as many as possible. Also, a more reliable counting method Curve Analysis has been developed which has high adaptation to the problem of similar pixel intensity of the object to the background.

The performance of the crowd monitoring system in medium crowd rate and low crowd rate is good. If there is no high crowd rate, the accuracy rate is very high and exceeds 85%. But when high crowd rate presence in the counting zone, occlusion of people occurred due to the angle view of the side-mounted camera. This will decrease the accuracy of people counting. But it is possible to have better performance in high crowd rate when the mounting position of the camera has been changed to top mounting.

Appendix

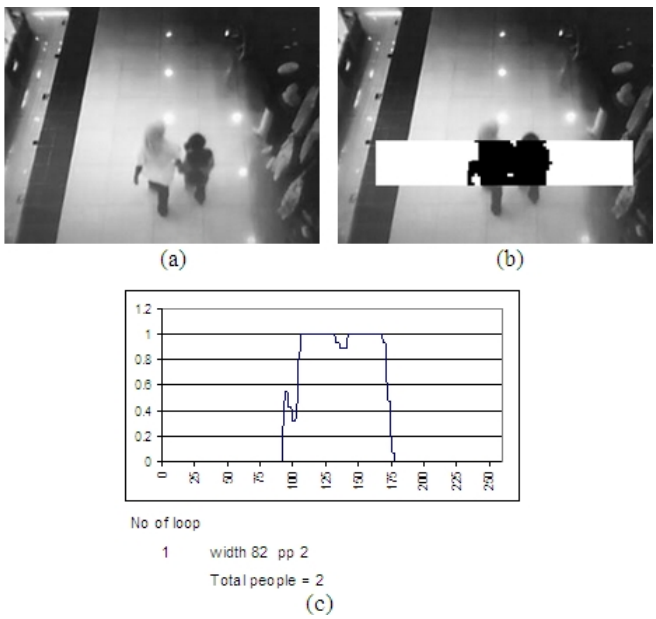

Fig. 6 Two people are clamped together. Result = 2

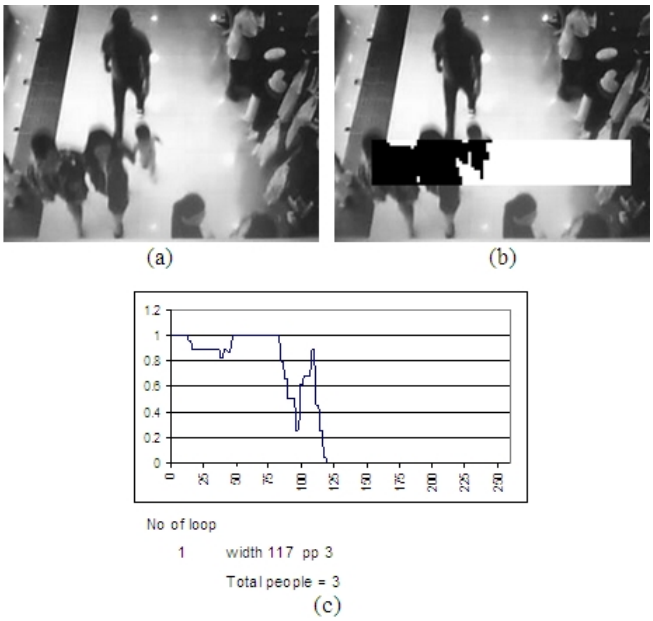

Fig. 7 Three people are clamped together. Result CA = 3.

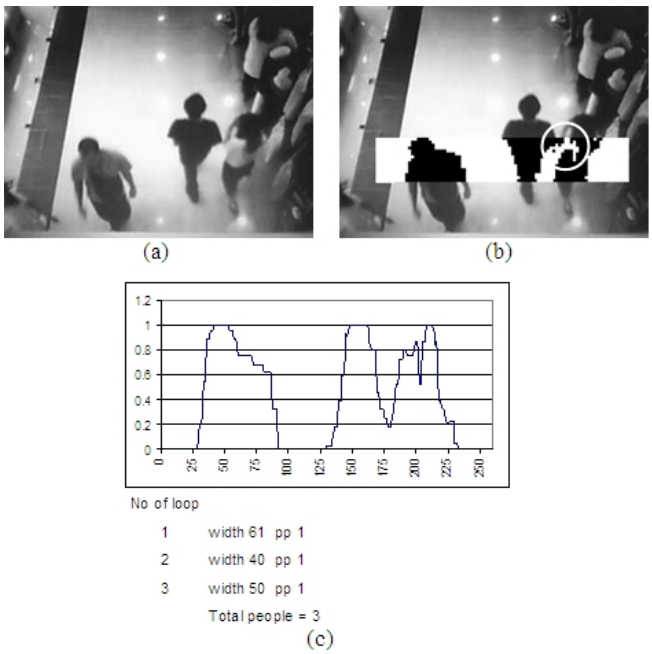

Fig. 8 Three people case with the pixel value of the people is so closed to the background (refer to the circle in (b)). Result CA = 3.

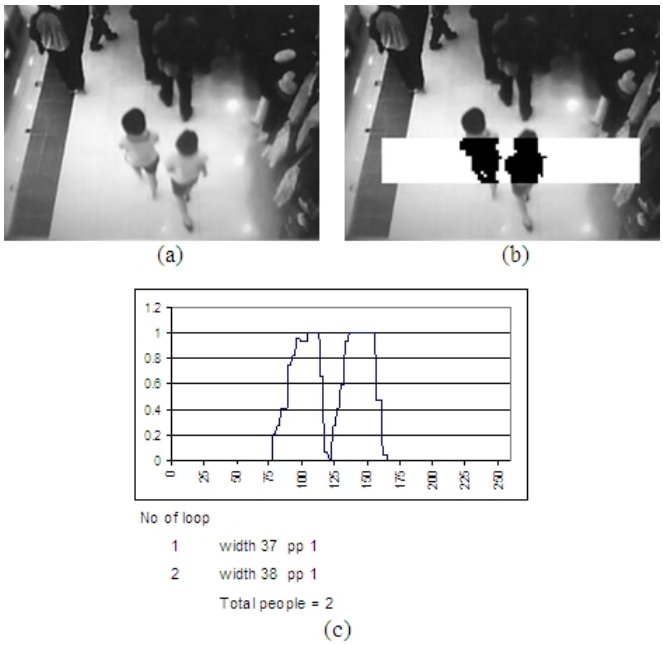

Fig. 9 Two people case. Result CA = 2.

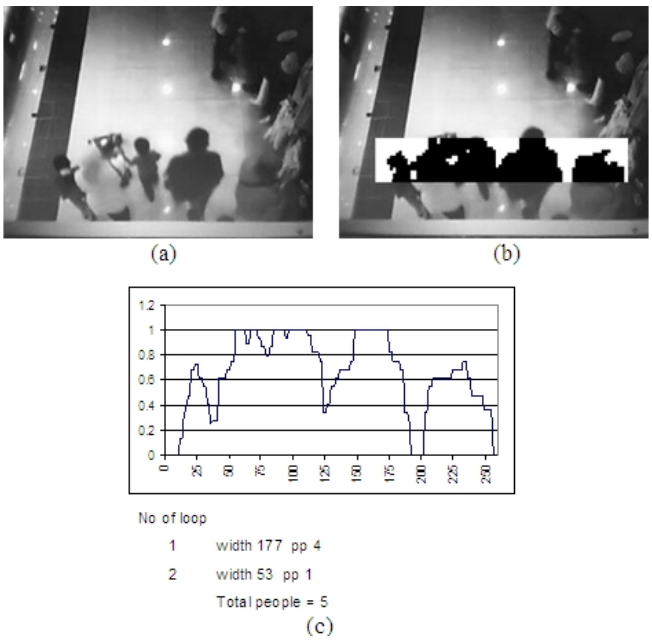

Fig. 10 Five People case. Result CA = 5.

Fig. 11 (a) Image after median filtering; (b) Result of image (a) when apply Sobel filtering

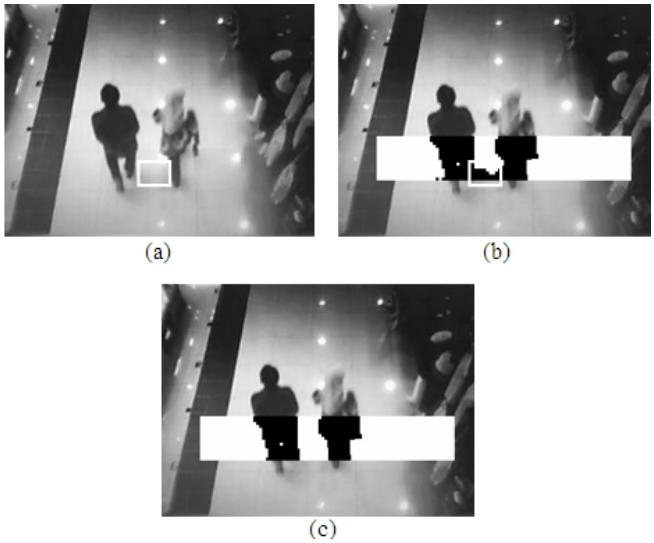

Fig. 12 (a) Original image; (b) Noise from shadow. (c) Result after median filter and Sobel filter

ACKNOWLEDGMENT

The authors would like to thanks Lim Lee Sian for providing the video for this paper.

REFERENCES

[1] A.C. Davies, J. H. Yin, and S. A. Velastin, "Crowd Monitoring Using Image Processing" *Electronic & Communication Engineering Journal*, Feb. 1995, pp. 37-47.
[2] C.S. Regazzoni, A. Tessei, and V. Murino, "A Real-Time Vision System for Crowd Monitoring", *Proc. Of IECON'93*, 1993, pp. 1860-1864.
[3] C.S. Regazzoni, and A. Tessei, "Distributed data fusion for real-time crowding estimation", *Signal Processing*, **Vol. 53**, 1996, pp. 47-63.
[4] F. Cravino, M. Dellucca, and A. Tessei, "DEKF system for crowding estimation by a multiple-model approach", *Electronics Letter*, **Vol. 30, No. 5**, 1994, pp. 390-391.
[5] Y.Cho, and T.W.S. Chow, "A fast neural learning vision system for crowd estimation at underground stations platform", *Neural processing letters*, **Vol. 10, no. 2**, 1999, pp. 111-120.
[6] Y.Cho, T.W.S. Chow, and C.T. Leung, "A neural based crowd estimation system by hybrid global learning algorithm", *IEEE Trans on Systems, Man, and Cybernetics, Part B*, **Vol. 29, no. 4**, 1999, pp. 535-541.
[7] D. Huang, T.W.S. Chow, and W.N. Chau, "Neural network based system for counting people", *IEEE Trans on Systems, Man, and Cybernetics, Part B*, **Vol. 31, no. 4**, 2002, pp. 2197-2200.
[8] C.K. Ng, "A system for counting people using image processing", Thesis for M. Phil, University of Hong Kong, 2001.
[9] M. Rossi, and A. Bozzoli, "Tracking and counting moving people", *Proc. of IEEE International Conference on Image Processing 1994*, **1994(3)**, pp. 212-216.
[10] W.K. Pratt, *Digital Image Processing, John Wiley & Son, Inc*, New York, 1999.
[11] J. Teuber, *Digital Image Processing, Prentice Hall*, New York, 2001.
[12] I. Pitas, *Digital Image Processing Algorithms, Prentice Hall*, New York, 2004.

Lim Eng Aik received the M.S. degree in Applied Mathematic and currently pursues his Ph.D. in Applied Mathematic at Universiti Sains Malaysia, Malaysia. His current research interests include neural networks, pattern recognition and image processing.

Zarita Zainuddin is an Associate Professor in the School of Mathematical Science at Universiti Sains Malaysia in Malaysia. Her research interests include pattern recognition, mathematical computing, neural network, signal processing and image processing. She received her Ph.D. in computer science from Universiti Sains Malaysia. She is a member of the Neural Network Society.