European Journal of Scientific Research
ISSN 1450-216X Vol.38 No.3 (2009), pp.441-453

http://www.eurojournals.com/ejsr.htm

Curve Analysis for Real-Time Crowd Estimation System

Lim Eng Aik
Institut Matematik Kejuruteraan, Universiti Malaysia Perlis, 02000 Perlis, Malaysia
E-mail: eng_aik@yahoo.com
Tel: +604-9855485

Zarita Zainuddin
School of Mathematical Science, Universiti Sains Malaysia, 11800 Penang, Malaysia
E-mail: zarita@cs.usm.my

Abstract

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. There are different kinds of crowd monitoring systems in the market with their principles of operation different from each other. Some may use sensors as the principle of the operation. In this paper, the principle of operation of the crowd monitoring system is based on image processing techniques where information extracted from the image is based on its color density. The information in the number represented by the image will be rearranged and changed in an appropriate way. Then, by analyzing the pattern and format of these numbers, a computer can obtain the information from the image like a human being. In addition, this paper presents a new people counting system to count the number of people in a certain image in real-time. The proposed system employs a Curve Analysis (CA) method for the crowd monitoring system where evaluation on its performance provides promising results.

Keywords: Cluster Segmentation, Curve Analysis, Edge Detection, Median Filtering

1. Introduction

A 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 situations which are able to provide a qualitative output corresponding to the occurrence of an alarm condition. Previously, real-time crowd estimation research has focused on using visual information [1], [2], [3], and [4]. Other researchers use a distributed extended Kalman filter system with both spatial and temporal information as a basis for model optimization [4], [5]. Kalman filter type algorithm is slow in 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 is by using template matching to track people. In [13], the direction of motion is applied to detect people motion using a CCTV from a certain height and to estimate the crowd density. But the approach in [13] involves a highly computational complex algorithm and showed poor performance when people walk close to the camera [13]. Recently, a few neural network based vision systems for estimating the crowd level of an underground station platform has been reported [6], [7], [8].

In this paper, we describe a visual based system designed for performing accurate people counting. The principle of the operation of the crowd monitoring system is based on image processing techniques. In an image, there is a lot of information that can be utilized. In human perception, image represents its information as color intensity. By observing the color intensity of the image, we can obtain certain information from the image. However, in computer vision, image represents its information as numbers. It is meaningless if this invaluable information is not being processed. Hence, by using image processing techniques, the numbers represented by the image will be rearranged and changed in an appropriate way. Then, by analyzing the format of these numbers, a computer can obtain the information from the image just like a human being. Therefore, the main issues discussed in this paper are how to extract the essential data from unconstrained images and perform accurate people counting by using a Curve Analysis method.

In general, automatic people counting systems [9], [10] can be divided into two stages: image processing and image understanding. Image processing will localize the objects of interest from the background and other objects. Meanwhile, feature extraction techniques are applied for image understanding. Template matching used in [9], [10] to count and track people will locate the objects in images and the number of people will be obtained. Clearly, there are two problems that occur. When a system is directly based on image information, the people counting system will have a high input dimension, hence, this type of counting system can only deal with a limited size images. The other problem is the sensitivity to noise. Thus, the overall capability of this type of system greatly depends on the performance of the image processing techniques.

This paper is organized as follows: the second part briefly explains the condition involved and the characteristic of the background for this people counting system; the third part deals with the image processing stage; section four discusses the method of Curve Analysis which is the image understanding stage while section five gives the performance and discussion on our proposed system.

2. Condition and Characteristic of Image Background

The image handled in the system is monochrome image. This means that the color of the image is in gray color ranging from black to white. There are 256 gray-levels altogether varying from black to white. The black color in the image is represented by 0 while the white color in the image is represented by 255. In a monochrome image, each pixel of the image has a certain value in the range from 0 to 255 to represent the level of gray color of the pixel. We say that it is the pixel value of the pixels. The gray-level color band is shown in Figure 1. In order to choose a suitable environment for people counting, two factors must be considered: one is the illumination condition of the environment and the other is the characteristic of the background.

Figure 1: The 256 gray-level colour band

The illumination condition of the environment has a crucial effect on the image captured. With different illumination levels, the pixel intensity of the image of a constant scene captured by the camera will be different. Consequently, a stable illumination level is very important. If the variation of the illumination level of the environment is large, the variation of the image captured by the camera will also be very large. Hence, it will be very difficult to process the image.

Generally, the illumination level is more stable in an indoor environment than the outdoor environment because the illumination source of an outdoor environment is mainly from the sunlight which cannot be controlled. Here, the environment for people counting in the image which is handled by the system is a subway at a railway station which is an indoor environment. Thus, it provides an excellent illumination condition to be processed.

Apart from the illumination condition of the environment, another factor that has a significant effect on the image processing is the characteristic of the background of the image. In order to limit the variation in the image captured, an image with a static background is preferred. Furthermore, the background of the image must have contrast with the object in the image. This is due to the fact that in order to separate the object from the background of the image the difference of pixel intensity between the object and the background is made used. 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.

3. Image Processing

In a crowd monitoring system, the people in the image needs to be counted. The first thing that ought to be done is to separate the object from the background. One of the approaches to separate the object from the background is to change the image into binary image. This entails representing the object-of-interest in the image by black pixels while the background by white pixels.

3.1. Image Pre-Processing

Before the image undergoes the process of segmentation, the image acquired from the camera has to be pre-processed. The resolution of the image processed by the system is 300 x 225 pixels. Regardless of the resolution of the image, 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 system is monochrome image. As a result, 8-bit bitmap format is used instead of 24-bit RGB image. The processing time will be faster as the computer handles fewer bits.

3.2. Cluster Segmentation

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

In the 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 containing some noise pixels in the output image.

In the segmentation process, these noise pixels can be eliminated when a sub-block of pixels is treated as a unit. Since there are 9 pixels in each sub-block, it is impossible that all pixels in a sub-block are noise pixels. As a result, the noise effect of the total pixels intensity of a sub-block is relatively much smaller than that of an individual pixel. If the sub-block of pixels is found that it is not occupied by the object, then the whole sub-block will be changed to white pixels including the noise pixels in the sub-block. Also, 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, the sub-block of the reference background image at the same position will also has noise pixels. The noise pixels will then compensate each other during comparing and subtraction of the pixel value of the two sub-blocks.

3.3. Sub-Block Processing

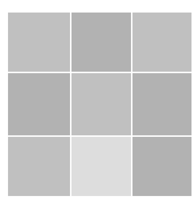
Initially, the counting zone of the image is divided into several sub-blocks. In the system, each sub-block consists of 3x3 pixels (see Figure 2). Let us focus on one sub-block in the counting zone. The 9 pixel values of the pixels of the sub-block will be added together and then the total pixel value of the sub-block can be obtained. 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 slightly darker than that of the background image due to the shadow of the people but it can be seen that if the sub-block of the input image belongs to the background, the difference in 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 belongs 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 not be very large.

As before, 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. The absolute difference of the total pixel value between the two sub-blocks will be very large. 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 to belong 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 a threshold value, the sub-block will be considered as belonging to the background and all the pixels of the sub-block will be changed to white 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 depends 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 be changed to a block of black pixels to represent the object. Otherwise the sub-block of the input image will be changed to a block of white pixels to represent the background.

Figure 2: Sub-block with 3 x 3 pixels



3.4. The Use of Sobel Filter

In order to solve the problem of shadow corruption, the Sobel filter edge detector has to be used [12]. The concept is that after applying the Sobel filter on the image, the outline (edge) of the object can be obtained. By analyzing the edge points of a certain area, it can be estimated whether the area belongs to the background or otherwise. To build a Sobel filter, two 3x3 convolution kernels are built as in Figure 3:

Figure 3: Sobel convolution kernel

-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)

Figure 3(a) is the 3x3 convolution kernel that use to calculate O1(i, j) as in (1). Figure 3(b) is the 3x3 convolution kernel that use to calculate O2(i, j). Figure 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 Figure 3(a) and Figure 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].

3.5. The Use of Median Filter

In order to eliminate the noise and preserve the edge points in the output image when the image is processed by the 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 a median filter. This is clearly illustrated in Figure 4(b) where it can be seen that the image has been smoothed.

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

(a) (b)

Examining Figure 4(b) closely, it can be seen that unwanted detail has been removed. Although the image is slightly blurred, it does not have any significant effect on further processing. The median filter algorithm 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 the median pixel value [12]. As the dark small point on the floor has a relatively low gray level, when the median filter sorts the pixels in a 5x5 sub-block, the small dark pixels will be filtered out. Then, the Sobel filter will be applied to the filtered image to obtain the edge points of the object.

The part in the counting zone of the image that is corrupted by shadow has a relatively dark gray level. Hence, the segmentation process misclassifies this part as an object rather than the background. However, an edge detector is used to detect the sharp variation of pixel intensity between pixels rather than detect the gray level of the pixel. As a result, edge points are resistant to shadow. Edge points are marked only if there is sharp change of pixels intensity between pixels.

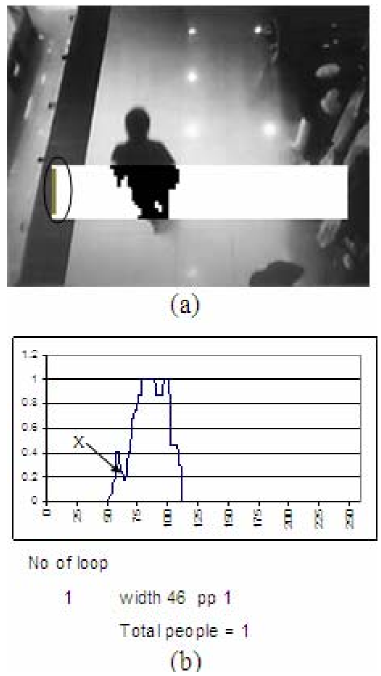
According to this characteristic, if a large area is found to have very few white pixels, this area will be considered as the background and it will be changed to white pixels to represent the background. Now, the edge white pixels are valuable information; it can be used to estimate whether the area belongs to the background in the counting zone of the image or belongs to the image.

In order to check the number of edge white pixels in certain areas, a template has to be designed to scan through the counting zone of the image. The operation of the template is that it will 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 found to be very small and lower than a threshold value, the area that is covered by the template will be considered as background and the area will be changed to white pixels. Figure 4 and Figure 11 each show the result of median filtering and Sobel filtering.

4. Curve Analysis

The method of Curve Analysis is developed in this paper to count the objects in the image after the input image is processed by segmentation. First, a template is produced to scan the counting zone of the segmented image. This is done on the first scanning in order to determine the location of the object which is the first column of black pixels found.

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



The template is indicated at the left hand most side of the counting zone as shown Figure 5(a). The template scans the counting zone of the image from left to right horizontally. At the left hand most side, the position of the template is 1 and the position will be increased when the template moves to the right. At the right hand most side, the position of the template is 287. When the template scans from left to right, it will gather the information of the percentage of black pixels covering a particular single column in the template area. There are a total of 287 columns for the template to go through. If the column of the area where the template scans contains only white pixels, the value of 0 will be assigned to the y-axis. On the contrary, if the template scans an area made up of only black pixels, a value of 1 will be assigned to the y-axis (representing 100%).

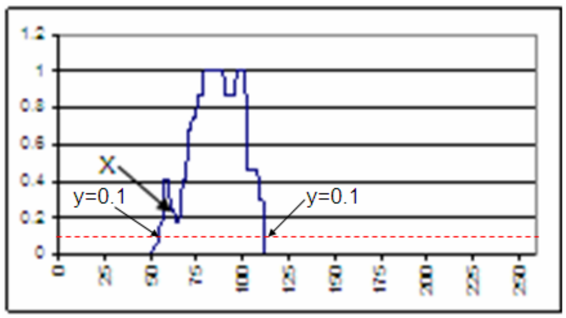
Now, for the second time of scanning, the template will move from position 1 to position 287 again in the counting zone of the image, but this time the template is not used to find the object. When the template moves from position 1 to position 287, the percentage of the area of the template that is covered by black pixels is recorded and will be represented as the y value of the curve and the position of the template will be represented as the x value of the curve. As a result, when the template scans through the counting zone of the image, a curve is plotted where the x-axis represents the position of the template and the y-axis represents the percentage of the area of the template that is covered by black pixels.

It can be seen in Figure 5(b) that when the template is at position 57 where the object is detected, a certain percentage of the area of the template is covered by the black pixels of the object. As a result, in Figure 5(b), it can be seen that when x =57, the y-axis will attain a certain non –zero value indicating the percentage of the area of the template which is covered by black pixels.

Obviously, when an object is in the counting zone, a curve that increases and then decreases will be plotted. This means that a loop will be plotted and the object in the image is represented by the plotted loop. Consequently, by counting the number of the loops and analyzing the width of the loop, it teaches the computer to recognize and count the target object in the image. In this paper, the width of the loop at y-axis = 0.1 is set based on the height of the camera for the current experiment. Different heights indicate different zooming for the image captured which leads to different sizes of the object captured by the webcam. This fact explains the different loop width obtained in the counting zone for the same object. For example, upon capturing an image of a static single object using the webcam placed at 5 meter high, produces a loop with a width of 65 in the counting zone, but the image captured for that same object from 10 meter high produces a loop with only a width of 45 in the counting zone.

In Figure 5(a), we observed a single person. After processing, we obtained the single loop plotted out is 0.1 at the y-axis for detecting a single person. A "loop" stands for a complete curve crossing up and down from y-axis at 0.1; this can be observed in Figure 6. This value gives a good reference point. After processing more sample images, it can be seen that the best y-axis of the loop that represent a typical person is about 0.1 to 0.2. To have some margin, the threshold value of the y-axis of one person is set at 0.1 in the system, which is the minimum value, to enable the system to detect people more accurately.

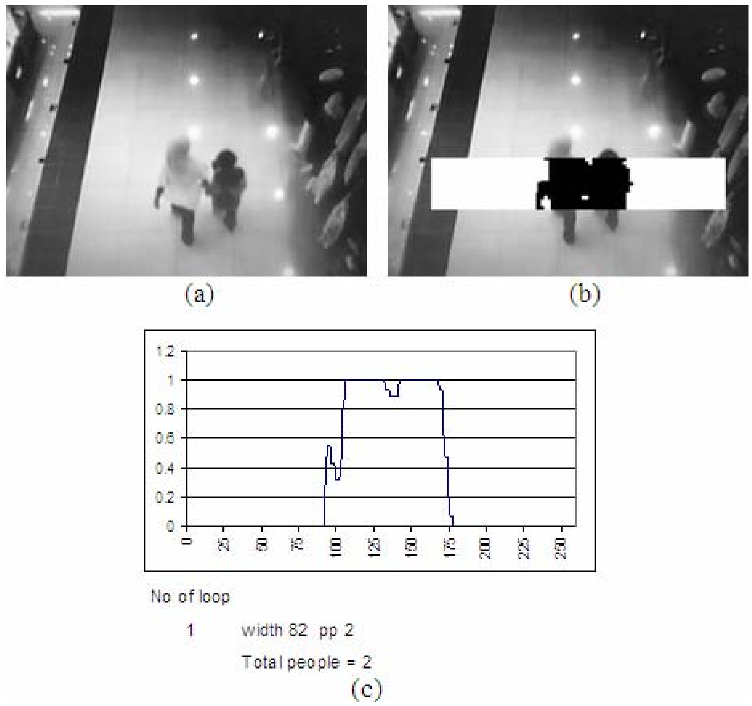
Figure 6: Plot of percentage of black pixels percentage from Figure 5(b).



For example, in Figure 5(b), the width of the loop is 44 which means that the width of the loop of the curve at the position y-axis = 0.1 is 44. If there are two people clamped together as in Figure 7(a), after processing the image (refer Figure 7(b)) and plotting the curve as in Figure 7(c), 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 because the width of the loop in Figure 7(c) is wider than that of the loop that represents one people. This entails having to set a suitable threshold value for the width of one person to detect and recognize the number of people represented by the loop.

Figure 7(a) shows two people whose pixel values are very close to the background. As a result, in Figure 7(b), it can be seen that there are some black pixels which should represent the object being missed even though a low threshold value is used for segmentation. However, the proposed method can recognize the people and correctly count the people. It can be observed from the loop plotted out in Figure 7(c) that the shape of the curve does not change substantially even though some black pixels that represent the object in the segmented image are missed. This demonstrates that the adaptation of the proposed counting method to the problem where the pixel value of the people is so close to the background is very high.

Figure 7: Two people are clamped together. (b) The counting zone of the input image. (c) The plot of percentage of black pixels of two persons walking in the counting zone. Result = 2.



5. Experiments

The digital images for this experiment come from a video at a shopping Mall consisting of 50 selected images. In these images, there are people in various conditions: some people are standing, some are walking, some are clamping together, some are in white or black clothing, etc. We have taken all the 50 images for testing. The system has been evaluated on its performance, and the average people counting accuracy is 0.86. Figure 8, Figure 9, and Figure 10 gives some images for testing and the results of our system. This demonstrates that the Curve Analysis method can overcome the noise from the shadow and similar pixel intensity problems.

Figure 8: Three people are clamped together. Result CA = 3.

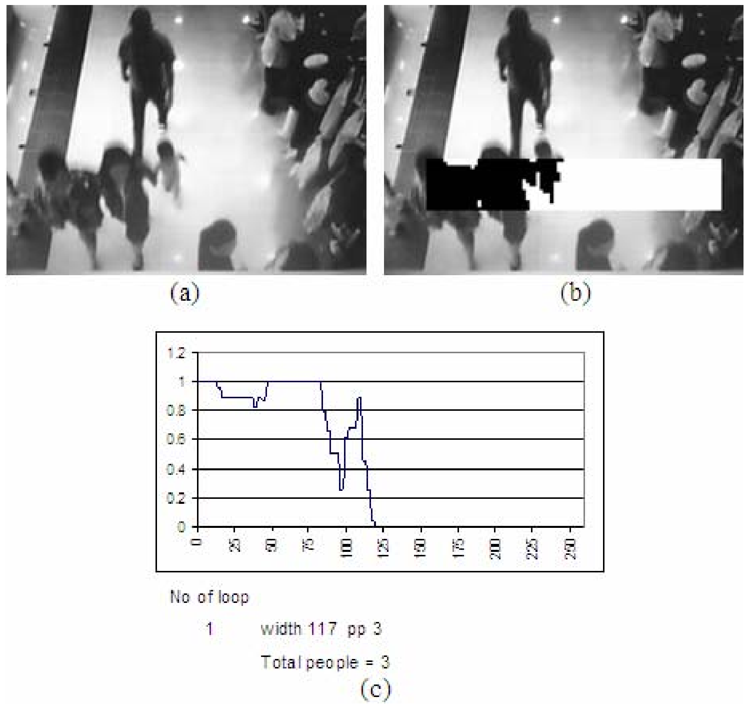


Figure 9: Two people case. Result CA = 2.

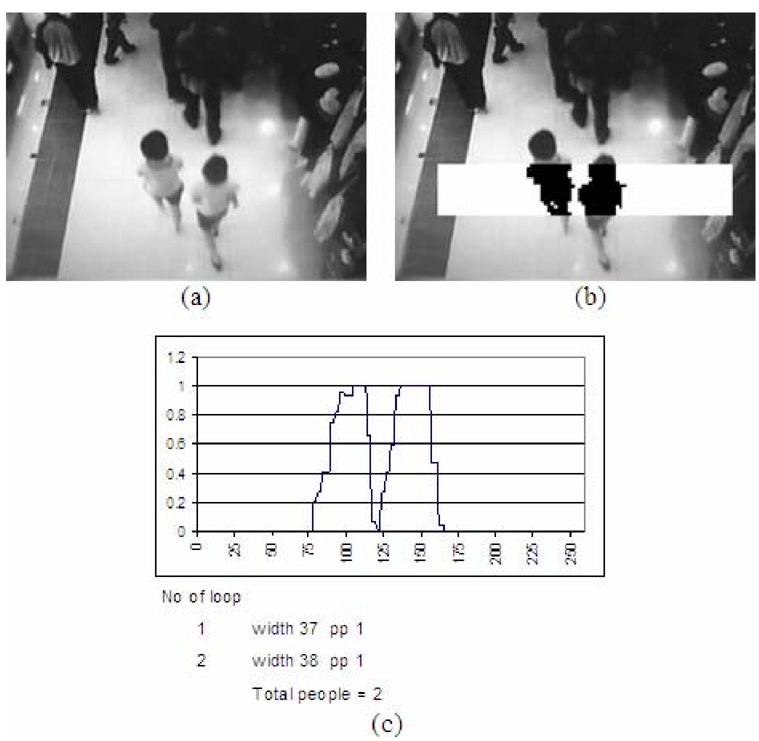


Figure 10: Five People case. Result CA = 5.

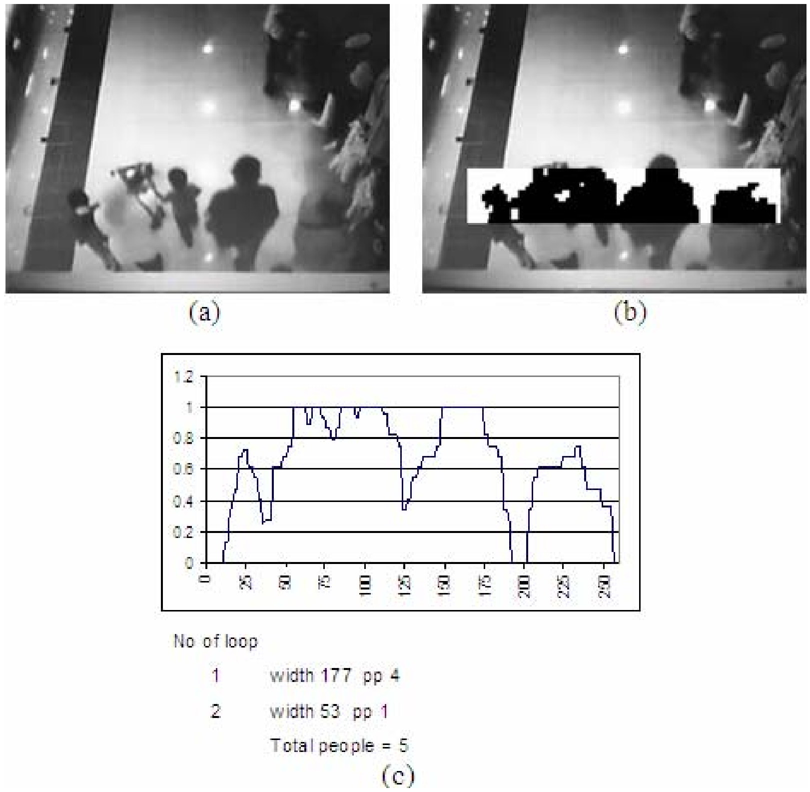


Figure 11: (a) Image after median filtering. (b) Result of image (a) when apply Sobel filtering.

6. Discussion and Evaluation

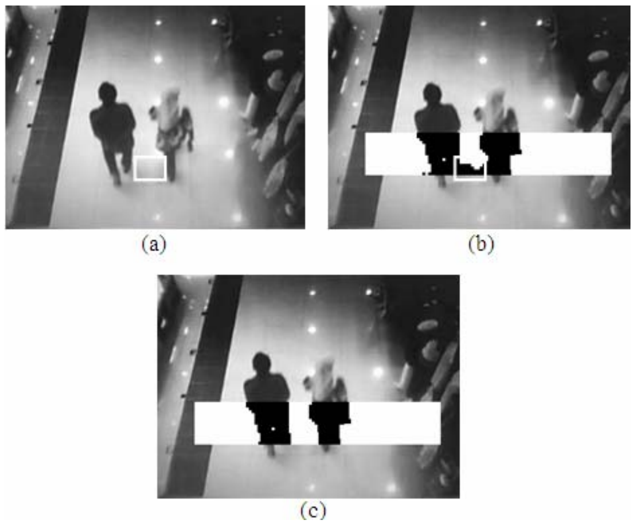
In this paper, cluster segmentation is used to extract the object from the background and the Curve Analysis method has been developed to count the extracted object. In the ensuing discussion, performance and the adaptability to the problems of these image processing techniques will be discussed.

6.1. 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 occur 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 found and is represented as white pixels. By analyzing the edge white pixels, the area corrupted by shadow can be found and correction can be made accordingly. 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 Figure 12.

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



6.2. 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. Thus, the method of Curve Analysis was able to count the object in the image more accurately and produces correct results. This is due to the output of Curve Analysis on the image where the object is in and being represented by a loop of certain width.

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

7. Conclusion

A crowd monitoring system that can be used for people counting by the techniques of image processing has been presented. The image processing techniques, cluster segmentation from Section III, is used to extract the object from the background. The advantage of this method is that it can 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 just employing the 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, known as the Curve Analysis method, 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 both a medium density and low density crowd is good. If there is no high density crowd, the accuracy rate is very high and exceeds 85%. But when there is a high density crowd present in the counting zone, occlusion of people occurred due to the angle of view of the side-mounted camera. This will decrease the accuracy of people counting. Nevertheless, it is possible to have better performance in a high density crowd when the mounting position of the camera has been changed to top mounting.

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, 1999. Digital Image Processing, John Wiley & Son, Inc, New York.

[11] J. Teuber, 2001. Digital Image Processing, Prentice Hall, New York, 2001.

[12] I. Pitas, 2004. Digital Image Processing Algorithms, Prentice Hall, New York, 2004.

[13] B.A. Boghossian, S.A. Velastin, "Evaluation of motion-based algorithms for automated crowd management", *International Conference on Vision Systems*, 1999, pp. 80-96.