Crowd counting based on Difference Images

Jinyan Chen[1]
[1]School of Computer Software, Tianjin University,
Tianjin 300072, P.R. China, phone: 86-22-87201819
chenjinyan@tju.edu.cn

***Abstract*—Many crowd counting methods were proposed in recently years. Most of these methods were implemented by extracting the human silhouette from the background image. But under some conditions it is difficult get a clear background image. In this paper a crowd counting method based on the images difference is proposed, instead of extract silhouette from background, the surveillance was divided into frames. Difference image of two frames is calculated by images subtraction. Then image features were extracted based on difference image and the crowed count is calculated based on these features. Experiment result show that this method is feasible.**

***Index Terms*—Image recognition, crowd, silhouette extraction, image segmentation.**

I. INTRODUCTION

It is necessary to control the crowd size in public place in order to avoid overcrowding or for other security reason. In the past years researchers are try to find ways to estimate the crowd from the surveillance equipment automatically. The difficulty of estimating the crowd size comes from three ways:

1) There is often overlapping among the pedestrians. That mean it is difficult to separate one silhouette from others;
2) Under some conditions the crow is very crowded this made it is difficult to get the true background and we can't segment the silhouette from the background.
3) The estimation algorithm speed should be efficient for real-time computing and surveillance.

Several methods have been developed to estimate the size of crowd in the past years. Ryan [1] use foreground pixels and other local features to estimate the crowd size; Chan [2] counting the Pedestrians by segment the crowd into components of homogeneous motion; Kong [3] using background subtraction and edge detection to each frame and extracting edge orientation and blob size histograms as features and then using these features to estimate the crowd size. Gray [4] used mixture Gaussian model to get the background of the Video surveillance. Huang [5] detect heads from the stereo image by scale-adaptive filtering and then calculate the crowd size.

II. OUR APPROACH

In this paper a new method of estimate the crowd size was proposed. Instead of directly extract features from the frames in the video, we try to get features from the difference of the frames. This method is based on the following assumption: Though the background is changing as time goes by, but in a very short period (1-2 seconds) the background is approximately unchanged. We can compare two adjacent frames (or two frames whose shoot time is very near) and find the difference then try to estimate the crowd size using features extract features from the difference image.

The test database was gotten from [2]. The Region of Interested in was called ROI and we only estimate the crowd size appeared in ROI as showed in Fig. 1.

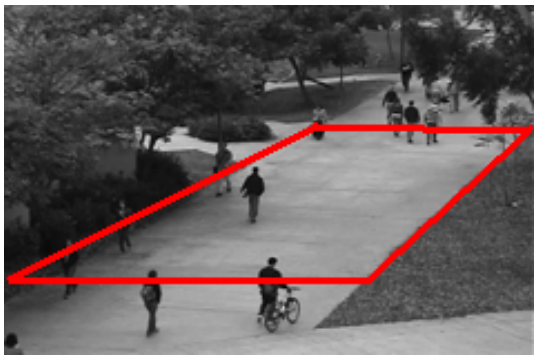


Fig. 1. The region of interest in the test database.

A. About the difference image

In surveillance video, given frames at time j, j+t, t means the frames interval between two frames, x, y means the coordinate in the difference images, we define the difference image as follows:

$$diff\_img_{j,t}(x,y)=\begin{cases}0 & if\ (img_j(x,y)=img_{j+t}(x,y),\\ 1 & if\ (img_j(x,y)\neq img_{j+t}(x,y).\end{cases} \quad (1)$$

a) b)

Fig. 2. Two frames from the test database (frames 1200,1201).

For example given frames 1200, 1201 from the test base as Fig. 2, we can get the difference image as Fig. 3.

Manuscript received March 19, 2012; accepted May 29, 2012.
Supported by Ph.D. Programs Foundation of Ministry of Education of China.No.20100032120011.

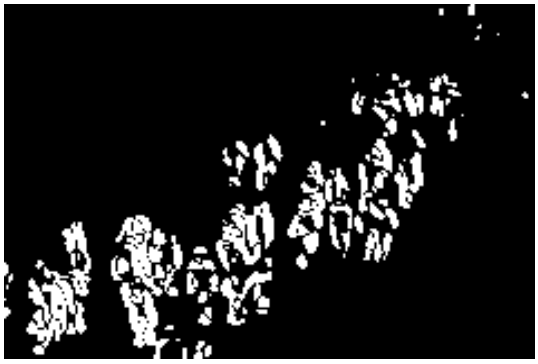


Fig. 3. The difference image ($img\_diff_{1200,1}$).

It is clear that the difference image between one image and the background image is the human silhouette image. The silhouette image is very useful for crowd size estimation. But under most condition it is difficult to get the extract background image so it is difficult to get the exact silhouette image.

Commonly speaking as t increase, most area of the difference image will be 1. If the video is shoot 10 frames per second and t=1 means the time interval between the two frames is 1/10 second.

B. Feature extraction from the difference image

To estimate the crowd size several features are extracted from the difference image.

1) Area of the difference images

$$Area_{j,t} = \sum (x, y). \tag{2}$$

That is mean the total area of "bright" points. But we need considerate the perspective effect and we will talk about it later.

2) Perimeter of the difference images contour

That is the total white pixels count in the edge detection map. The calculation of perimeter also should considerate the perspective effect.

We can get edge detection map from the difference image by using canny algorithm. Fig. 4 is the edge detection map of Fig. 3.

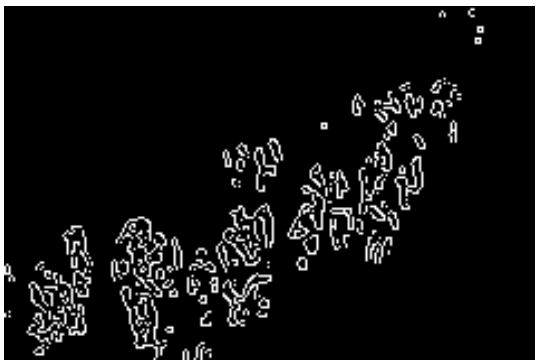


Fig. 4. Edge detection map of $img\_diff_{1200,1}$.

C. Perspective normalization

The total pixel count for the blob segment and each pixel is weighted by its value in the density map. Taking into account the perspective effect then the area (perimeter) can be express as

$$Area_{j,k} = \sum w(x, y)\ (x,y). \tag{3}$$

Because of the effect of perspective, the object closer to the camera will appear larger. It is important to normalize the feature before extract the feature from the crowd. In this paper the method mention in [2] is used to normalize the feature. Every point is assigned different weight according to it distance from the camera. The further the point, the more weight the point is assigned. In the following of this paper all the feather extracted from the image is multiplied by the weight w(x,y) of this point. The weight distribution of w(x,y) can be expressed as Fig. 5.

Fig. 5. The weight distribution of the ROI.

D. Selection of difference time

From above introduction we can see that the difference image is not only affected by the crowd size and the perspective, but is also affected by the time interval between two frames. The longer time interval means the more "bright area" in the difference image.

In this paper in order to find out the relationship between the difference image and the crowd size, we select the 1,2,3,4 as the frames interval to calculate the difference image. That is mean the time interval between two frames used to create the difference images is 0.1, 0.2, 0.3, 0.4 second.

III. EXPERIMENTS AND DISCUSSION

First of all we analysed the relationship between the crowd size and the features of the difference image.

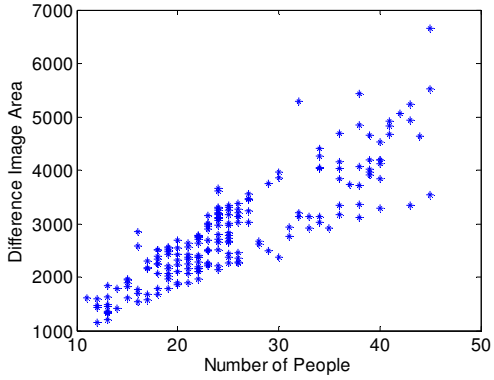


Fig. 6. Relationship between the crowd size and the difference image area. t=1(0.1 second).

From Fig. 6 and Fig. 7 we can see that the pedestrian

count and the features size or image perimeters does not have a simple linear relationship. We also set t=2, 3, 4 and get the similarly result.

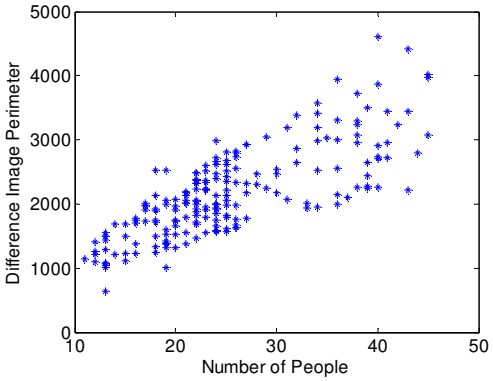


Fig. 7. Relationship between the crowd size and the difference image perimeter (t=1 (0.1 second)).

The correlation coefficient between the area and the perimeter shows in Table I.

TABLE I. CORRELATION COEFFICIENT BETWEEN CROWD SIZE AND AREA (PERIMETER).

Features	Gap	Correlation Coefficient
The area of difference	1	0.839
	2	0.844
	3	0.832
	4	0.542
The counter perimeter	1	0.831
	2	0.844
	3	0.847
	4	0.654

The ground truth of the test dataset is supplied by [2]. We totally get 2000 frames from the dataset. The crowd size of every frame is showed as Fig. 8.

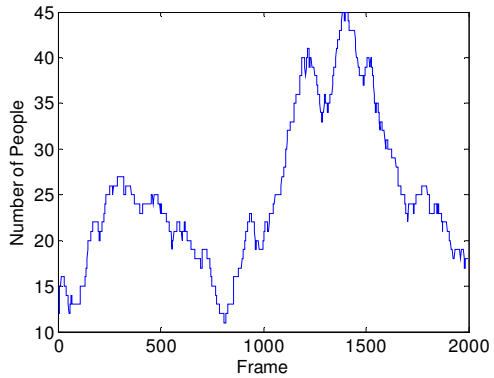


Fig. 8. The ground truth of the test data set.

For single camera it is a challenge problem to accurately counting the crowd size. Under such conditions, the pedestrian count and the features size does not have a simple linear relationship. This is mainly because of the occlusion and the moving of the pedestrian. We use neural network to find out the nonlinear mapping between the features and the crowd counting .The neural network has one hidden layer and one output. The output is the crowd size in the Region Of Interested. The input of the neural network is the futures we extracted from the input data: the area of the difference images, Perimeter of the difference images contour. We train this network using standard back propagation (BP) algorithm. Frames 600-1400 in dataset were used as the training data and frames 0-599 and frames 1401-2000 were used as test data..

The performance of the proposed system is assessed using two criteria:

1) *Error.* The mean value of the absolute difference between the crowd estimate and the ground truth.

2) *MSE.* The mean value of the error squared.

In this paper several features set combination were used to estimate the crowd size:

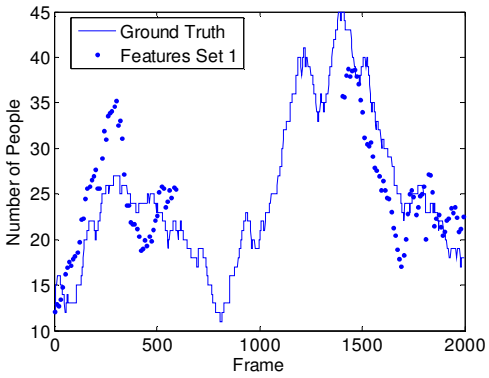
1) *Features set 1.* The difference image interval=1 that mean 0.1 second, using area and perimeter.

2) *Features set 2.* The difference image interval=1,2 that mean 0.1,0.2 second, using area and perimeter.

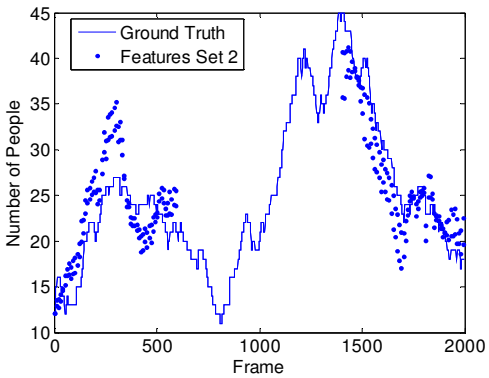
3) *Features set 3.* The difference image interval=1,2,3 that mean 0.1,0.2,0.3 second, using area and perimeter.

4) *Features set 4.* The difference image interval=1,2,3,4 that mean 0.1,0.2,0.3,0.4 second, using area and perimeter.

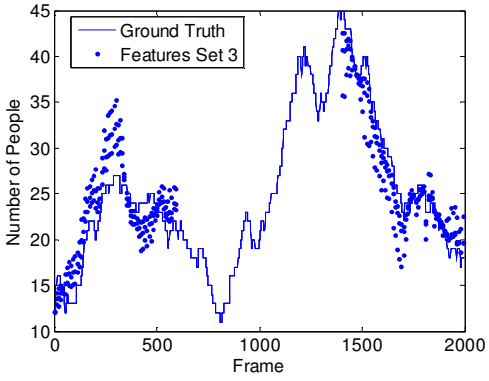
The experiment result was showed in Fig. 9, (a-c).



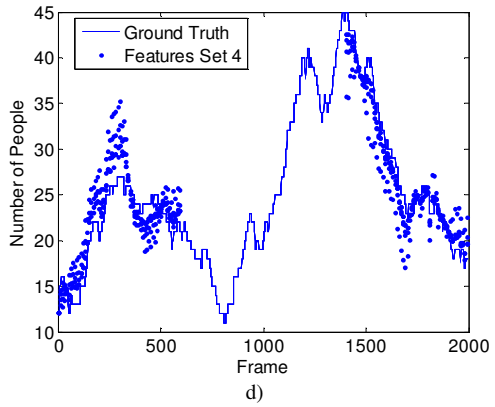
a)



b)



c)



d)

Fig. 9. The experiment result and the ground truth (a) features set 1, (b) features set 2, (c) features set 3, (d) features set 4.

From the experiment we can see that the error and MSE decline with increase of the features selected and will reach the optimization at features set 3.

The comparison of our method and other methods is showed in Table II:

TABLE II. ESTIMATE RESULT COMPARING WITH OTHER METHODS.

Method	Raw Estimate	
	Error	MSE
Proposed in this paper Features set 3	3.367	4.913
David Ryan[1]	1.306	2.684
Kong[3]	1.710	4.642
Holistic[1]	4.462	31.24

Comparing to other methods [1], [3], [6], the method proposed in this paper is not the best method. But the method proposed in this paper has the following advantage:

1) This method need depend on the background segmentation;

2) This algorithm is simple and can be implemented real time.

IV. CONCLUSIONS

In this paper, a new method of estimate the crowd size was proposed. The main idea of this method is to calculate the difference between two images, the time interval between the two images can be 0.1-0.5 second. Then the area and the perimeter were used to estimate the crowd size. Instead of segment the silhouette from the background, this method did not need to calculate the background. It is suit for the circumstance that the background is changing rapidly.

Because all the experiment were base on the dataset which was shoot 10 frames per second, in this paper we only get difference image at interval 0.1, 0.2, 0.3, 0.4 second. In the future a fine time interval should be used (e.g. 0.05 second and so on) to calculate the difference image.

REFERENCES

[1] V. M. D. Neil, *Tracking weakness links in cold chain.* BerkeleyCA : University of California Press, 2006, p. 342.

[2] B. Marija, B. Ludvik, V. Robert, "Stability of perishable goods in cold logistic chains", *International Journal of Production Economics,* vol. 93–94, no. 8, pp. 345–356, 2005.

[3] D. Ryan, S. Denman, C. Fookes, S. Sridharan, "Crowd counting using multiple local features", *Digital Image Computing: Techniques and Applications,* Melbourne, Australia, pp. 81–89, 2009.

[4] A. B. Chan, Z. S. J. Liang, N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking", in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition,* Anchorage, 2008, pp. 1–7.

[5] D. Kong, D. Gray, H. Tao, "Counting pedestrians in crowds using viewpoint invariant training", in *Conf. Machine Vision,* 2005.

[6] D. Kong, D. Gray, H. Tao, "A viewpoint invariant approach for crowd counting", in *Proc. of the 18th International Conference on Pattern Recognition,* Hong Kong, 2006,pp. 1187–1191.