TELKOMNIKA, Vol.12, No.4, December 2014, pp. 865~874
ISSN: 1693-6930, accredited A by DIKTI, Decree No: 58/DIKTI/Kep/2013
DOI: 10.12928/TELKOMNIKA.v12i4.362

Large Crowd Count Based on Improved SURF Algorithm

Haining Zhang*[1], Huanbo Gao[2]
School of Electronic Information Engineering, Xi'an Technology University
2 Xuefuzhonglu Avenue,Weiyang District, Xi'an 710021, China
*Corresponding author, e-mail: zhn1964@163.com[1], 1198309686@qq.com[2]

Abstract

This paper uses an analysis of Speeded up Robust Feature (SURF), based on the method of Linear Interpolation for camera distortion calibration, for high-density crowd counting. The eigenvalues are built on the Gray Level Co-occurrence Matrix (GLCM) features and the SURF features. Through the method of linear interpolation, weight values are interpolated to reduce the error, which is caused by camera distortion calibration. The optimized crowd's feature vector can be got then. Through the method of support vector regression, the crowd's number can be forecast by training model. The experiment result shows that the method of this paper has a higher accuracy than the previous methods.

Keywords*: crowd count, SURF, GLCM, perspective-correct, SVR*

1. Introduction

With the increase of the world population, unfortunate accidents in public places caused by high-density crowd occur frequently in recent years. At the same time, the video surveillance systems are ubiquitous [1]. If we make use of the existing resources, these intelligent systems can help us effectively forewarn and avoid disaster events. Compared with the traditional approach, the intelligent system of counting and density estimation can also improve the utilization rate of public facilities, and arrange the allocation of manpower and material resources effectively.

The algorithm of crowd counting can be divided into two categories: direct and indirect means. The direct way utilizes people's characteristics directly, such as color, shape, etc, to get the crowd's number. The people's head and face and some other characteristics can be selected as the statistical feature vector. This method is usually very complex, and is more suitable for monitoring low density populations. The major research method of counting high-density crowd in the world is the indirect way [2]. With this method, the number can be obtained by the method of regression through extracting the whole crowd's features [3]. But the statistical precision of this method is currently not accurate enough. The method still needs further in-depth research.

This paper uses a statistical regression method. First, the crowd's foreground image is extracted from the input image. Then, the GLCM features and SURF feature of crowd's foreground image are extracted [4]-[7]. Though the method of linear interpolation, weight values that caused by camera distortion calibration, are interpolated to reduce the error. Through the method of support vector regression, the crowd's number can be finally forecast by training model. Figure 1 shows the block diagram of whole algorithm.

2. SURF Feature Extraction

The research object of this paper is high-density crowd. SURF algorithm is used to describe the characteristics of population.

In 2006, Herbert Bay proposed a more practical feature detection algorithm of SURF. SURF is a local feature point detector with high robustness, and the operating speed of this algorithm is higher. Because of its good invariance of scale transformation and perspective transformation, it has become an important feature extraction algorithm in many ways.

Received September 1, 2014; Revised November 5, 2014; Accepted November 20, 2014

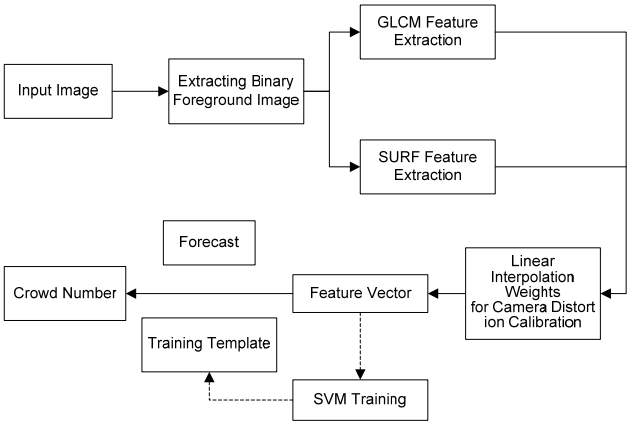


Figure 1. The block diagram of whole algorithm

The processing of crowd's SURF feature extraction is as follows:

(1) Running the global scanning of the original image, and obtaining the integral image [8].

In Figure 2, to any point (i, j) in the image, its value of $ii(i, j)$ in the integral image is the sum of all points' gray value on the diagonal. And the diagonal is from the point (i, j) to the top left vertex of original image. The value $ii(i, j)$ is as follows:

$$ii(i, j) = \sum_{i \le i, j \le j} p(i, j) \tag{1}$$

The sum of all pixels' gray value in any window W can be obtained by the value of the four points (i_1, j_1), (i_2, j_2), (i_3, j_3), (i_4, j_4) in the integral image.

$$I_W = ii(i_4, j_4) - ii(i_3, j_3) - ii(i_2, j_2) + ii(i_1, j_1) \tag{2}$$

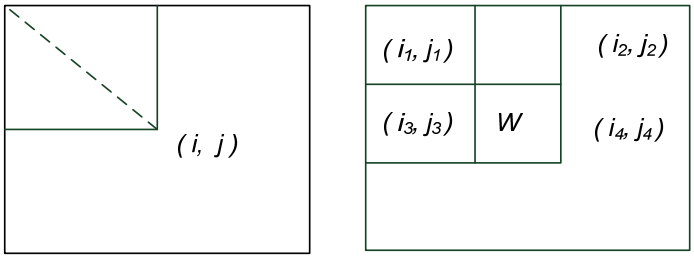


Figure 2. The calculation of integral image and the sum of all pixels' gray value in window W

When all the points' gray value in original image is 1, the value of $ii(i, j)$ in integral image represents the rectangular area from the point (i, j) to the top left vertex. I_w means the area of window W.

(2) The extreme points of scale-space can be got through Hessian matrix approximation. These extreme points are the feature points of what we need.

The Hessian matrix $H(x, \sigma)$ of *point* (x, y) in the image I is defined as follows:

$$H(x,\sigma)=\begin{bmatrix} L_{xx}(x,\sigma) & L_{xy}(x,\sigma) \\ L_{xy}(x,\sigma) & L_{yy}(x,\sigma) \end{bmatrix} \quad (3)$$

$$\det(H)=L_{xx}(x,\sigma)L_{yy}(x,\sigma)-\left(L_{xy}(x,\sigma)\right)^2 \quad (4)$$

σ means scale. $L_{xx}(x,\sigma)$ is the Laplacian of Gaussian of the image. It is the convolution of the Gaussian second order derivative $\partial^2 g(\sigma)/\partial x^2$ with the image. $L_{xx}(x,\sigma)$ and $L_{yy}(x,\sigma)$ have the same meaning. When the value of $\det(H)$ is greater than zero, if $L_{xx}(x,\sigma)$ is greater than zero, the point x is the local minimum point; if $L_{xx}(x,\sigma)$ is less than zero, the point x is the local minimum point. The feature point can be judged through computing the determinant of each pixel in the image [9].

For the convenience of applications, Herbert Bay proposed approximating second order derivatives with box filters. Using the box filters operation to replace the convolution operation L. The different scales of scale space are formed by expanding the size of the box.

Using D_{xx}, D_{yy} and D_{xy} to replace L, then the determinant is simplified as:

$$\det(H)=D_{xx}D_{yy}-\left(wD_{xy}\right)^2 \quad (5)$$

The weight w changes with the change of scale.

The feature points should be further confirmed after preliminary testing. In order to verify the extreme points in the scale-space, each sampling point should be compared with all its adjacent points. In other words, each point is compared with 26 points, which means those 18 points in the adjacent scale-space and 8 points in the same image. If the point is greater or less than these 26 points, it is the final feature point.

(3) The principal orientation of each feature point is determined. After that, the 64-dimensional characterization vector is formed.

In this paper, the first processing of SURF feature points extraction is that the binary foreground image is extracted from the input image by the background subtraction method and the sliding average method. Then the SURF feature is extracted from the binary foreground image. Compared with the SURF feature extraction of the overall image, the SURF feature extraction of binary foreground image reduces the computation complexity. Figure 3 shows the result of SURF feature extraction.

Figure 3. SURF feature extraction of binary foreground image

The white part of binary foreground image is the moving region. According to the principle of the SURF algorithm, most of the SURF feature points are in the motion area, but there will be a few feature points around region [10]. As shown in Figure 3, the number of feature points cannot effectively reflect the characteristics of crowd. So the feature points in the non-interest region should be rejected. This selecting process only needs to scan all feature points, which are distinguished by their value of pixel.

$$surf(x,y)=\begin{cases}1, i(x,y)=255\\ 0, i(x,y)=0\end{cases} \quad (6)$$

$surf(x,y)$ is the discriminant of *point* (x,y). When the value of $surf(x,y)$ is 1, this point will be kept; when it is 0, this point will be removed. $i(x,y)$ is the pixel values of this point.

Figure 4. After rejecting feature points in non interest region

In Figure 4, we can see that the kept SURF feature points can reflect the characteristics of crowd really and effectively.

3. Eigenvector Construction with SURF Feature and GLCM Feature

SURF feature has good invariance of scale transformation and perspective transformation, and can reflect the characteristics of crowd. But for large populations, and when there is a dense covering, the SURF feature cannot reflect the characteristics very well. Because the GLCM features can effectively overcome the occlusion problem, this paper proposes combining the GLCM feature with SURF feature. The eigenvector is composed of four uncorrelated GLCM feature vectors (energy, entropy, contrast, correlation) and the numbers of SURF feature points [11].

The four uncorrelated GLCM feature vectors are:

(1) Energy: A statistic reflects the consistency.

$$ASM=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\{p(i,j/d,\theta)\}^2 \quad (7)$$

Energy reflects the level of texture coarseness and the uniformity level of gray distribution. When the texture is coarse, the energy is high. Otherwise, the energy is low.

(2) Contrast: A statistic of contrast.

$$Con=\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2\, p(i,j/d,\theta) \tag{8}$$

The contrast reflects the sharpness of the image. When the texture is coarse, the contrast is small. Otherwise, the contrast is big.

(3) Entropy: A parameter calculating the randomness distribution of gray-level d .

$$Ent=-\sum_{i=0}^{N-1}\sum_{j=0}^{N-1}p(i,j/d,\theta)\log p(i,j/d,\theta) \tag{9}$$

Entropy indicates the level of non-uniformity texture or the complexity of the image. When the texture is coarse, the entropy is small. Otherwise, the entropy is large.

(4) Homogeneity: A correlation statistic of gray value.

$$S_H=\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\frac{p(i,j/d,\theta)}{1+(i-j)^2} \tag{10}$$

The homogeneity reflects the direction of the texture, and shows the similarity degree of rows or columns. The difference of pixel values between elements is bigger, the homogeneity value is smaller.

Through the above theory, 6-dimensional feature vector is formed in this paper, and the SURF feature is the main characteristic:

(num_{surf} , *s,* $feature_{entropy}$, $feature_{energy}$, $feature_{contrast}$, $feature_{homogeneity}$).

num_{surf} is the number of SURF points. s is the area of moving people in binary foreground image. $feature_{entropy}$ is the entropy of GLCM matrix. $feature_{energy}$ is the energy of GLCM matrix. $feature_{contrast}$ is the contrast of GLCM matrix. $feature_{homogeneity}$ is the homogeneity of GLCM matrix

4. Linear Interpolation Weights for Camera Distortion Calibration

Camera distortion calibration is caused by the increasing distance between moving objective and the camera. As we all know, the area of people near camera is bigger than the area of one far away from camera. In order to reduce the influence caused by the losing depth information of image, this paper adopts the method of linear interpolation weights for camera distortion calibration. This method has strong adaptability and high real-time. And in actual application, the researchers do not need measurement of the environment.

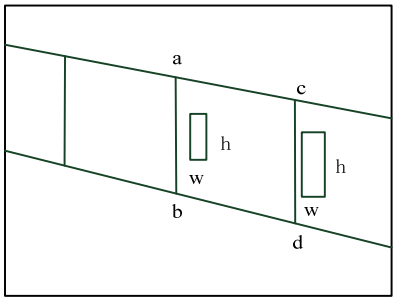


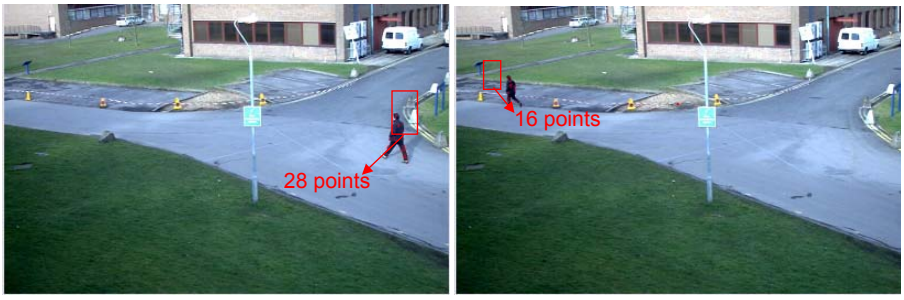
Figure 5. The theory of linear interpolation weights

In Figure 5, the image is divided into four regional grids. The width and height of object's minimum enclosing rectangle (MER) can be got in the entrance and exit of each grid. The weights of each grid can be obtained by the area change rate of the MER. As shown in Figure 5, the area change rate of grid *a, b, c and d* is as follows:

$$k = \left| \frac{h_2 \cdot w_2}{h_1 \cdot w_1} \right| \tag{11}$$

(h_1, w_1) is the width and height of object's MER in the entrance. (h_2, w_2) is the width and height of object's MER in the exit.

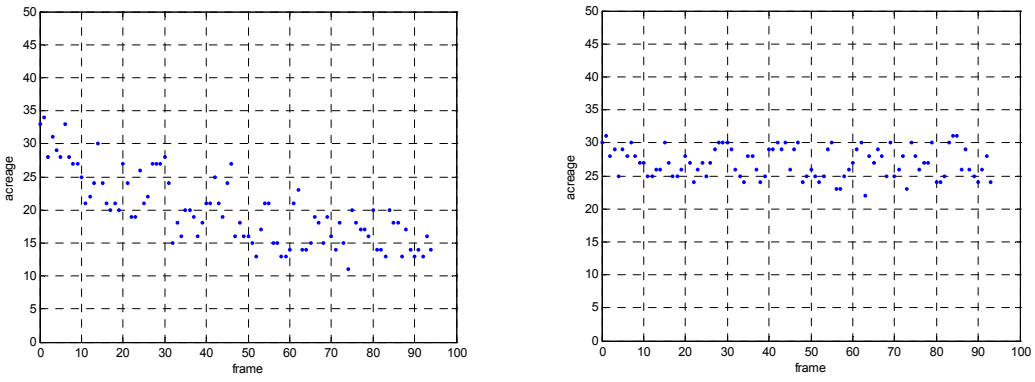
Figure 6 shows a video sequence of single people walking in PETS video sequences [12]. This paper separates the monitor space into 4 parts in each video frame. The purpose is to improve the accuracy of camera distortion calibration. It should be noted the accuracy of calibration is more accurate when the more parts separated. But the more weights interpolated, the processing of determining these weights is more complex.



(a) The 9th frame (b) The 72th frame

Figure 6. PETS2009 single people walking video sequence

For four partitions, there are four interpolation weights. As shown in Figure 6, the number of SURF feature points is different in different frames. The 4 weights can be calculated by the method of regression.



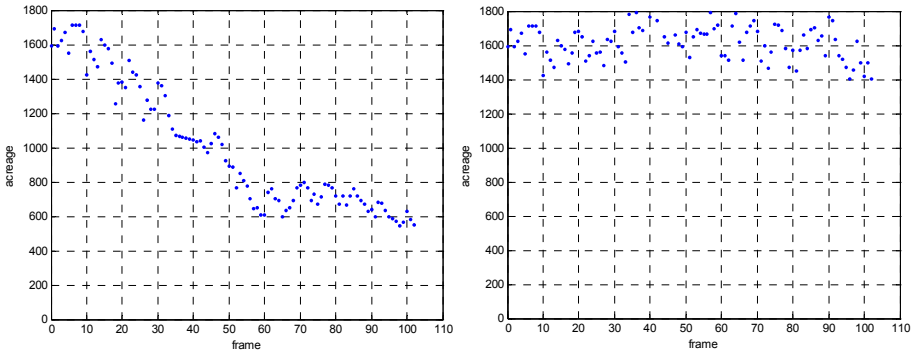
(a) Before correction (b) After correction

Figure 7. The correction effect of SURF number

Figure 7 shows obvious differences of SURF number before and after correction. The abscissa represents the number of frames, ordinate represents SURF numbers. Figure 7 (a) shows that the SURF number reduces gradually as the pedestrian walks away from the camera gradually. Figure 7 (b) shows that the SURF number remains in a stable range after interpolating the four weights for correction.

Figure 8 shows difference of foreground area before and after correction. The abscissa represents the number of frames, ordinate represents the foreground area [13]. Figure 8 (a) shows that the foreground area reduces gradually as the pedestrian walks away from the camera gradually. Figure 8 (b) shows that the foreground area remains in a stable range after interpolating the four weights for correction.

Figure 7 and Figure 8 show that the method of linear interpolation weights can solve the problem of camera distortion rapidly and effectively.



(a) Before correction (b) After correction

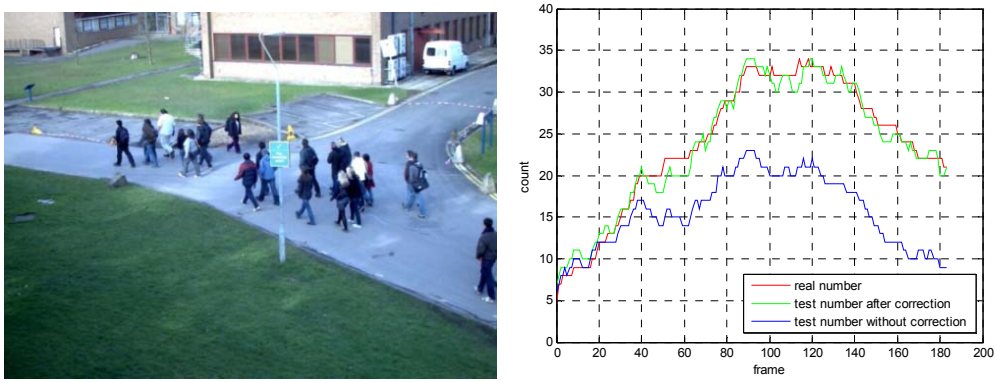
Figure 8. The correction effect of foreground area

5. The Test Results,Analysis and Comparison

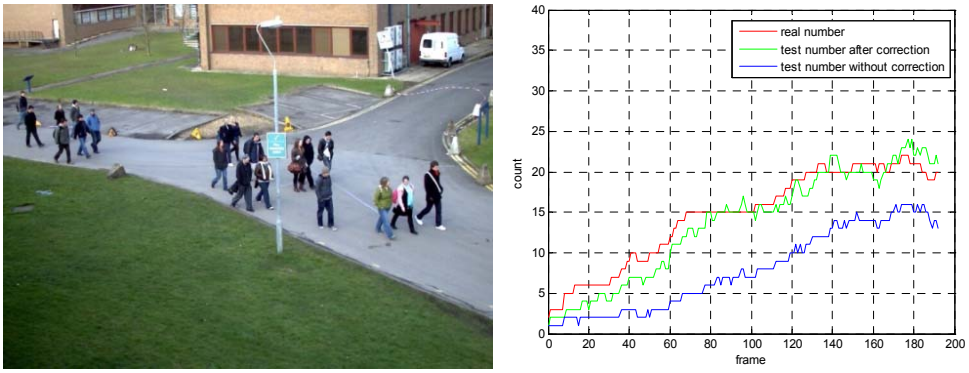
5.1 The Test Results and Analysis

This experiment uses Microsoft Visual C++ 6.0 as software development environment and OpenCV1.0 as image processing library in the operating system of Windows XP. Hardware experimental platform is a PC machine, and the PC memory is 2G. This experiment uses Matlab7.0 as the analysis tool for the conclusion and analysis.

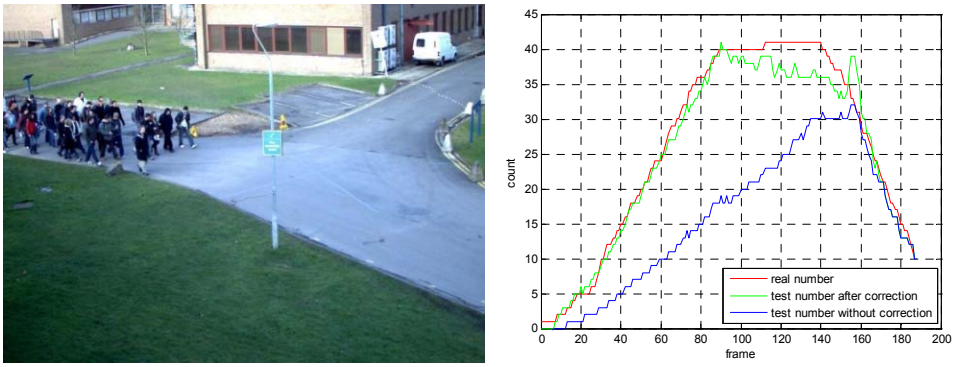
The number of crowd can be estimated after ε-SVR training. The training model can be got through training the correction feature vectors. This paper tests three video sequences in the PETS2009 video library [12]. Figure 9 shows the curve of the real crowd's number, the test crowd's number without correction and the test crowd's number after correction.



(a) Video 1



(b) Video 2



(c) Video 3

Figure 9. The test result and analysis

The three indexes of analyzing the test result are: MAE (Mean Absolute Error), MRE (Mean Relative Error) [14], and MAXE (Maximum Absolute Error).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|N(i) - N_0(i)\right| \tag{12}$$

$$MRE = \frac{1}{n}\sum_{i=1}^{n}\frac{\left|N(i) - N_0(i)\right|}{N_0(i)} \tag{13}$$

n is the frames of video. $N(i)$ is the test number of frame i. $N(i)$ is the real number of frame i.

Experimental results before and after corrections of this paper are shown in Table 1.

Table 1. The test results analysis before correction

Name	Non-correction			After-correction		
	MAE	MRE	MAXE	MAE	MRE	MAXE
Video1	8.465	32.2%	14	0.951	5%	3
Video2	6.802	51.9%	10	1.518	14.8%	3
Video3	10.654	50.1%	22	1.788	11.5%	7

Compared with the non-correction method, the accuracy after correction is much higher. It proves that the method of linear interpolation weights can solve the problem of camera distortion rapidly and effectively.

Table 2. The results of GLCM method

Name	GLCM			SURF		
	MAE	MRE	MAXE	MAE	MRE	MAXE
Video1	3.027	15.3%	7	1.02	5.1%	3
Video2	2.508	21.9%	9	1.21	9.8%	4
Video3	7.402	38.2%	15	4.47	19.4%	9

In Table 2, the GLCM method is the crowd counting method based on GLCM features. The crowd feature eigenvector of this method is only made up of the 4 GLCM feature vectors. The SURF method is the counting method based on SURF algorithm. The main vector of crowd feature eigenvector is SURF number.

The results analysis shows that this method can estimate the peoples' number in test region rapidly and accurately. Compared with the GLCM method, the accuracy of this paper increases obviously. It shows that SURF number is an important feature vector of counting crowd. Compared with the SURF method, the accuracy of video 3 increases obviously. It shows that the GLCM feature vectors can effectively overcome the occlusion problem.

5.2 Compared With The Pixel Statistic Feature Method

In many cases the pixels statistic feature can describe the population characteristics effectively, mainly including foreground feature and edge feature [15]. By introducing the perspective correction parameter, it can calculate the weight ratio calculation parameters.

$$r = \frac{\sum_i w_i \delta(i)}{\sum_i w_i} \tag{14}$$

r is the foreground pixels or edge pixels of the pixels statistic feature, w_i is the impulse response function, $\delta(i)$ is the perspective correction parameters of pixels. When it is foreground or edge pixel, the value is 1, otherwise the value is 0.

Through the comparison of the SURF algorithm and the pixels statistic feature method used in video 3, the results are shown in the Figure 10.

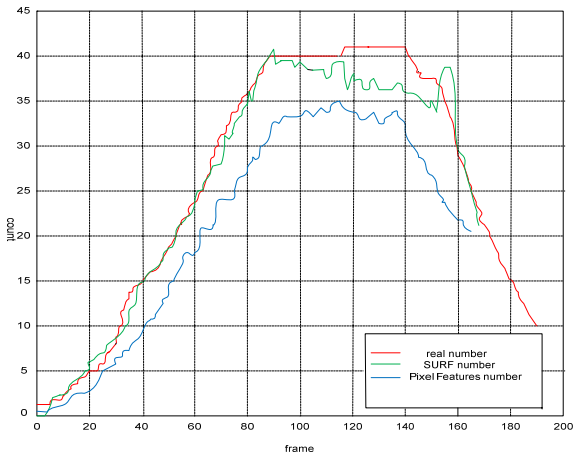


Figure 10. The result of comparison

It can be seen that the SURF algorithm method is better than the pixels statistic feature method. Overall, the deviation of experiment results of the pixels statistic feature method is bigger, while the SURF method which keeps the experiment results curve around the actual number curve all the way is more precisely.

6. Conclusion

According to the problem of counting high-density crowd, this paper proposes an improved crowd counting method based on SURF algorithm and GLCM algorithm. Experimental study finds that, the linear interpolation weights correction method is a simple and effective method for camera distortion calibration. This algorithm has strong adaptability, and can accurately estimate the number of people in each frame with the average error less than 2 people per frame. With the variety and complexity of the research environment, the method still needs further in-depth research.

References

[1] Conte D, Foggia P, Percannella G, et al. A Method for Counting Moving People in Video Surveillance Videos. *EURASIP Journal on Advances in Signal Processing.* 2010; 5(1): 1-8.
[2] Ya Huang, Su Hang, Zheng Shibao. Large-scale Crowd Density Estimation. *Video Application and Project*. 2010; 34(5): 113-116.
[3] Antoni B, Chan, Nuno Vasconcelos. Counting People With Low-Level Features and Bayesian Regression. *IEEE Transactions on Image Processing*. 2012; 21(4): 2160-2177.
[4] Wang Yalin. Research on Algorithm of Crowd Density Estimation Based on Gray Level Co-occurrence Matrix. *Master's thesis*. Xi'an: Xi'an University of Technology and Science. 2013.
[5] Bay Herbert, Ess A, Tuytelaars T, et al. Speeded-up Robust Features (SURF). *Computer Vision and Image Understanding*. 2008; 110(3): 346-359.
[6] Nan Geng, Dongjian He, Yanshuang Song. Camera Image Mosaicing Based on an Optimized SURF Algorithm.*TELKOMNIKA: Indonesian Journal of Electrical Engineering*. 2012; 10(8): 2183-2193.
[7] Wang W, Li W H, Wang C X, et al. A Novel Watermarking Algorithm based on SURF and SVD. *TELKOMNIKA: Indonesian Journal of Electrical Engineering*. 2013; 11(3): 1560-1567.
[8] K. Zhang, T.F. Xu, P. Wang, L. Feng. Real-time full-frame digital image stabilization system By SURF. *Optics and Precision Engineering*. 2011; 19(8): 1964-1972.
[9] Ryan D, Denman S, Fookes C, et al. *Crowd Counting Using Group Tracking and Local Features*. 2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance. Boston. 2010: 218-224.
[10] X. Liu, B. Dai, H.G He. *Real-time object segmentation for visual object detection in dynamic scenes*. Soft Computing and Pattern Recognition (SoCPaR). Dalian. 2011: 423-428.
[11] Sun Wenchang, Song Jianshe, Yang Meng, et al. Corner Detection Algorithm Based on Entropy and Uniqueness. *Journal of Computer Applications*. 2009; 29(2): 225-227.
[12] Choudri, S., Ferryman, JM., Badii, A. *Robust background model for pixel based people counting using a single uncalibrated camera*. Proceedings of Twelfth IEEE International Workshop on Performance Evaluation of Tracking and Surveillance (PETS-Winter). IEEE Computer Science. Miami. 2009: 1-8.
[13] Ya Lihou, Pang G KH. People Counting and Human Detection in a Challenging Situation. *IEEE Transactions on Systems*. 2011; 11(4): 24-33.
[14] K. Zhang, T.F. Xu, P. Wang, L. Feng. Real-time full-frame digital image stabilization system By SURF. *Optics and Precision Engineering*. 2011; 19(8): 1964-1972.
[15] Yang Hua, Su Hang, Zheng Shibao. Large-scale Crowed Density Estimation. *Video Engineering*. 2010; 34(5): 113-116.