An Approach for Crowd Density and Crowd Size Estimation

Ming Jiang[1,2] , Jingcheng Huang[1,2] , Xingqi Wang[1,2] , Jingfan Tang[1,2] , Chunming Wu[3]

(1. Institute of Software and Intelligent Technology, Hangzhou Dianzi University, Hangzhou, 310018, China）
(2. Zhejiang Provincial Engineering Center on Media Data Cloud Processing and Analysis, Hangzhou Dianzi University, Hangzhou, 310018, China）
(3. College of Computer Science and Technology, Zhejiang University, Hangzhou, 310027, China)

Email: jmzju@163.com, zjcnxghjc@126.com, xqiwang@163.com, tangjf@hdu.edu.cn, wuchunming@zju.edu.cn

***Abstract*—Crowd density estimation is very important in intelligent crowd monitoring. In this paper, a new approach of crowed density estimation is proposed. This method combines the advantage of pixel statistical feature and texture analysis, and reduces the impact of perspective distortion by dividing the region of interest. Moreover, we estimate the crowd size (the range of numbers) for high density and extremely high density crowd. The experiment results show that the proposed method has a relatively high accuracy in the whole density range.**

***Index Terms*—Region of interest, Crowd density estimation, Crowd size estimation, Pixel statistical, GLCM.**

I. Introduction

With the development of rapid economic, large-scale human activities have become increasingly frequent, especially some massive entertainment events, sporting events, and large exhibitions. Crowd safety has become a critical issue, causing the attention of the security sector. As we know, the terrorist attacks in large groups can result in greater lethality and social impact, large-scale group activities are now important goals of terrorist attacks. So it is very meaningful to estimate crowd density. In order to analysis group events better, we need to estimate the size of the crowd rather than the crowd density.

Traditional visual surveillance applications are based on CCTV (Closed Circuit Television) analog systems. Although such system enhances the human visual ability through the hardware device, the method for monitoring the crowd density is almost manual. This method requires a huge amount of work to monitor and the attention of surveillance personnel will distract with the growth of the monitoring time, it is likely to cause big delay and mistakes. So, all of which makes real-time monitoring is much more significant.

Crowd density estimation methods can be divided into four categories. Firstly, crowd analysis based on background removal technology. Davies[1] and Chow[2] have proposed image processing method based on pixel statistics, calculated the number of foreground pixels through background subtraction technique, and estimated crowd density by the number of pixels. Although this method is simple and effective, it is ineffective for high density crowd. Secondly, crowd analysis based on image processing and pattern recognition technology. In this category, texture features are widely used. Marana et al.[3] [4] have presented the crowd density estimation method of texture analysis. The obvious improvement of this method is that it solves the problem of overlapping, so that high density crowd can be estimated by this method. However, the disadvantage is obvious: the accuracy is very low for low density crowd. Thirdly, crowd analysis based on information fusion. Velastin et al.[5] used the Kalman filtering method to count number of people by the background removal technology and boundary technology. The last method is newly generated method. Antonio[6] and Donatell[7] used the method of feature points, and achieved better results.

For the defects of two classical algorithms: pixel statistics and texture analysis, this paper comprehensively utilizes both advantages, and reduces the impact of perspective distortion by dividing the region of interest. This method can improve the accuracy in the whole density range. Furthermore, we estimate the crowd size for the high density and extremely high density crowd. Experimental results show that the proposed method can improve the accuracy in the whole density range.

The rest of the paper is organized as follows. In section 2 we give the proposed approach. Experimental results are given in the section 3. And in the section 3, we compare our algorithm to two classic algorithms. Section 4 concludes with applications, limitations and future work.

II. Our Method

Firstly, after inputting video, we extract the foreground image through Gaussian mixture model, and then process foreground image; the processing includes filtering noise through median filtering and morphological operations. Secondly, estimate the number of people after dividing the region of interest. The specific statistical methods are: give a preliminary judgment for the crowd density

doi:10.4304/jsw.9.3.757-762

through the number of foreground pixel. For different density, we use the method of pixel statistics or GLCM texture analysis to extract crowd feature. And then use linear regression to estimate the number of the crowd. Finally, add the number of each region and then estimate crowd density. It is worth mentioning that we estimate the crowd size for the high density and extremely high density crowd.

A. Pixel feature

The property of the pixel statistic is the earliest feature to be used for crowd density estimation, and it is a very effective feature. The basic idea of this algorithm is: the denser crowd, the greater proportion of the foreground image. Researchers considered that there is a linear relationship between the number of foreground pixel and number of people in the scene. Pixel features usually are: the foreground image area, perimeter, and edge pixels, and so on.

Pixel statistical algorithm is relatively intuitive, easy to understand, low computational complexity, the relationship between the number of people and pixel feature is relatively simple after preprocessed, easy to train, and the generalization ability of classifier or function relationship is very well after training. However, the pixel statistical algorithm has some problems: foreground image segmentation algorithm is not ideal, and needs to correct the weight of extracted pixel due to the impact of perspective distortion, has bad result in high density crowd.

In this paper, this pixel statistical method is used to give the initial judgment of crowd density, and estimate the crowd density of extremely low density, low density, and medium density.

B. Texture feature

The pixel is very important feature among crowd density estimation, but the accuracy is very low for more serious occlusion area. To solve this problem, Marana proposed texture analysis algorithm. Different density crowd has different texture pattern for texture analysis. Images of low density crowds show coarse texture, while images of high density crowds show fine texture. The calculation of GLCM texture features is a common and effective method. In this paper, this texture method is used to estimate the crowd density of extremely high density and high density.

GLCM is a second-order statistical method, which can be thought of as second-order joint conditional probability density function $p(i, j \mid d, \theta)$. Describe the image along a certain direction θ, separated by a certain distance as the element d, the frequency of gray level of the pixel i and j that is elements of the matrix. Where $i, j = 0,1,2,...,N-1$ and N is gray level of the image. Generally, as data of GLCM is very large, GLCM is not used directly as texture features during the feature classification. Some researchers build some statistics based on its classification as texture features. Marana [8] used different image texture features and classifier made a detailed study for such an estimation method based on classification, drew the following conclusion: the best classification results provided by the two GLDM descriptors: Contrast 0° and Homogeneity 0°, achieved the correct classification accuracy of 85.5%. So this paper selects these two descriptors to estimate the crowd number.

Contrast: reflects the image clarity and the depth extent of texture groove. The deeper the texture grooves, the greater the contrast is and the visual effect more clearly as well. In GLCM, the larger the value drifts away from the diagonal elements, the greater the contrast is. Low density crowd has high contrast than high density crowd. So the variation of contrast can represent the density information. Contrast is defined as,

$$Con = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}(i-j)^2\, p(i, j \mid d, \theta). \qquad (1)$$

Homogeneity: reflects the homogeneity of image texture, and measures how much local texture change. The large value illustrates that texture lack of change between different regions and local is very homogeneous. Homogeneity is defined as,

$$Hom = \sum_{i=0}^{N-1}\sum_{j=0}^{N-1}\frac{p(i, j \mid d, \theta)}{1+\left|i-j\right|^2}. \qquad (2)$$

C. Definition of Classification

Polus[9] proposed crowd density from low to high is divided into five levels. In this paper, we reference to this definition, crowd density levels are defined as shown in Table Ⅰ.

D. Our Proposed Method

Figure 1 shows the details of proposed method in this paper. In this method,

TABLE Ⅰ
CROWD DENSITY OF EACH CATEGORY

Classified level	Extremely low density	Low density	Moderate density	High density	Extremely high density
Crowd boundary (people)	0-10	11-30	31-60	61-100	>100

1) Capture video, and use Gaussian mixture model to extract video foregrounds. For the foreground image, we use method of binary process, noise elimination by median filtering, and morphological operation.
2) Set the region of interest. Since the presence of abnormal projection, especially in the process of large-scale monitoring, the effect of abnormal projection is particularly evident brought by the perspective effect. To solve this problem, this paper divides into four different sub-regions for each scene image. The sub-region division effect is showed in Figure 2.

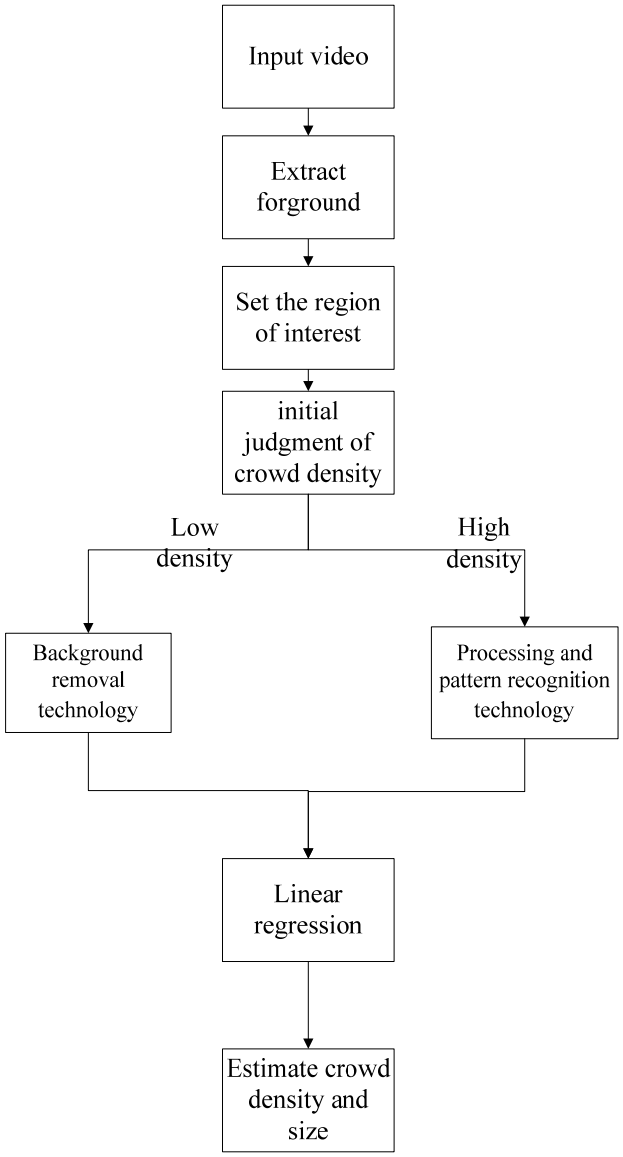

Figure 1. Our method.

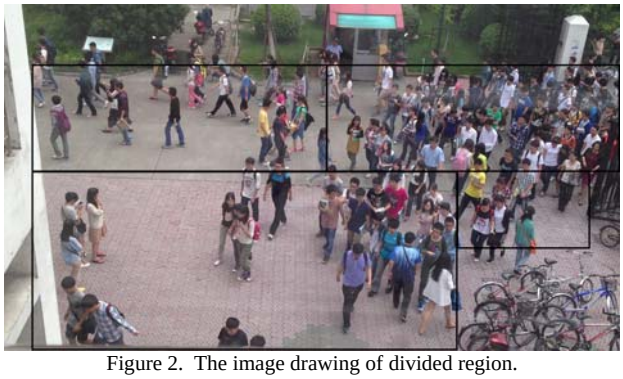

Figure 2. The image drawing of divided region.

3) The initial judgment of crowd density.

$$N_i / (W * H) \le T_i . \quad (3)$$

N_i is the total number of foreground pixels for the i-th region. W is the width and H is the height of region. T_i is the threshold of the i-th region. We set $T_1 = 0.17, T_2 = 0.23$. Since in the selected video, the number of region 3 and region 4 is very low, we do not need to give the initial judgment of crowd density. We use pixel statistical method to estimate the number of people in the two region directly.

If the i-th region satisfies equation (3), we use pixel statistical method to estimate the number of people in this region, skip to step 4). Otherwise we use texture analysis method to estimate the number of people in this region, skip to step 5).

4) Pixel statistical method. For a particular region, statistics the total foreground pixel of this region, then the number of people in this region can be calculated according to the corresponding fitting straight line.

The fitting straight line is trained by method of a linear regression. According to number of foreground pixels and count the true number of people in this region artificially for each region of training samples, using the least squares method, fitting out four straight lines that the number of pixel corresponds to the number of people.

5) Texture analysis method. For a particular region, statistics two texture descriptors (Contrast 0° and Homogeneity 0°) of this region, then the number of people in this region can be calculated according to the corresponding fitting straight line. The fitting straight line, which is trained by method of multiple linear regressions. According to the two texture descriptors and count the true number of people in this region artificially for each region of training samples, fitting out straight lines that the two texture descriptors correspond to the number of people by the least square method.

6) Add the number of the four regions through step 4 and 5; the result is the number of people in a scene. From Table Ⅰ, we can know that the number of people belongs to which density level.

If the scene density level is high density or extremely high density, we continue to estimate the crowd size of the scene. Such as high density scene is 61-80 or 81-100 people, extremely high density scene belongs to 101-120 or more than 120 people.

Ⅲ. Experimental and Comparative Results

The test video is captured in the rush hour after classes. In the course of the experiment, each region selected 40 images as training samples for different crowd densities. And 160 images of each density were selected as the test samples. The 160 images of high density include 61-80 and 81-100 people each 80 images, 101-120 and more than 120 people each 80 images in the 160 images of extremely high density. We assess different performance

characteristics by comparing our experimental result with calibration results obtained by manual.

A. The software implementation

The software was carried out in the Visual Studio 2008 platform, using C++ programming language and additional open source library of video processing: OpenCV library.

The interface can see from Figure 3.

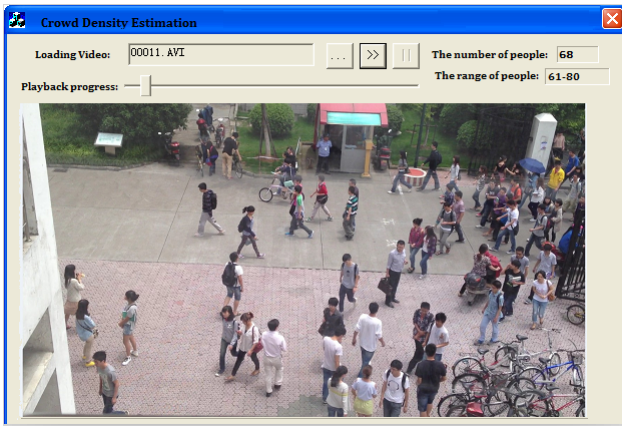

Figure 3. The software interface.

B. The sample training results of pixel statistic feature

For each region of 40 training samples, count the true number of people in this region artificially, fit out four straight lines that the number of pixel can correspond to the number of people by the least square method. Equation of fitting straight line is:

$$Y_1 = 0.0002 * N_1 + 0.7428 . \quad (4)$$

$$Y_2 = 0.0003 * N_2 - 0.1703 . \quad (5)$$

$$Y_3 = 0.0001 * N_3 + 0.732 . \quad (6)$$

$$Y_4 = 0.0002 * N_4 + 0.5562 . \quad (7)$$

N_i is the total number of foreground pixel for the i-th region, Y_i is the number of people in this i-th region. $i = 1, 2, 3, 4$.

Statistics the total foreground pixel of each region, then the number of people in this region can be calculated according to the corresponding fitting straight line.

C. The sample training results of texture statistic feature

In this paper, we select two descriptors (Contrast 0° and Homogeneity 0°) of texture feature. Count the true number of people in this region artificially for each region of 40 training samples, fit out two straight lines that the two descriptors can correspond to the number of people by the least square method. Equation of fitting straight line is:

TABLE Ⅱ
RESULTS USING PROPOSED METHOD

density level	Number of test samples	Number of test samples with correctly classified	Accuracy testing (%)
Extremely low density	160	148	92.5
Low density	160	137	85.625
Moderate density	160	133	83.125
High density	160	136	85
Extremely high density	160	143	89.375

TABLE Ⅲ
CROWD SIZE RESULTS OF HIGH DENSITY AND EXTREMELY HIGH DENSITY

Crowd size	Number of test samples	Number of test samples with correctly classified	Accuracy testing (%)
61-80	80	67	83.75
81-100	80	67	83.75
101-120	80	69	86.25
>120	80	74	92.5

TABLE Ⅳ
COMPARISON BETWEEN OUR PROPOSED METHOD AND THE METHOD OF ONLY USE PIXEL STATISTICAL FEATURE OR TEXTURE ANALYSIS FEATURE

density level	Accuracy testing of pixel statistical method (%)	Accuracy testing of texture analysis method (%)	Accuracy testing of proposed method (%)
Extremely low density	92.5	91.25	92.5
Low density	84.375	78.75	85.625
Moderate density	82.5	76.25	83.125
High density	78.125	85	85
Extremely high density	75	88.125	89.375

$$T_1 = 24.5676 * X_{11} + 576.9581 * X_{21} - 24.0584 . \quad (8)$$

$$T_2 = 38.7057 * X_{12} + 873.8693 * X_{22} - 58.1259 . \quad (9)$$

Statistics the two descriptors of each region, then the number of people in this region can be calculated according to the corresponding fitting straight line.

D. The experimental results

We select 160 test samples images for each crowd density, the estimation of the experimental results are shown in Table Ⅱ. For high density and extremely high density, the crowd size estimation can be seen from Table Ⅲ.

E. Compare with classical algorithms

We compare our method with the method of only use pixel statistical feature[1] or texture analysis feature[4]. Their experimental results are presented in Table Ⅳ.

As we can see from Table Ⅳ, compare with the simple use of pixel statistical feature or texture feature, our proposed method has a relatively high accuracy in the whole density range. The reason is that we combine the advantages of both.

Ⅳ. CONCLUSION

This paper proposes an approach for crowd density estimation, which combines the pixel statistical feature and texture feature. The proposed method removed background with Gaussian mixture model and gave a preliminary judgment for the crowd density through pixel feature, meanwhile reduced the impact of perspective distortion by dividing the region of interest. The texture features were extracted using GLCM, and selected Contrast 0° and Homogeneity 0° as texture feature. Experimental and comparative results show that the method is an effective, universal method which can be used in a real-time crowd density estimation system. And this paper estimated the crowd size for high density and extremely high density, which was more conducive to group events analysis.

Certainly, there is still much room to improve the accuracy. If the image contains crowd shadow and reflective surface, it might lead to misclassification. In the future, we will accelerate foreground detection approach and try to eliminate shadow noise in the image.

ACKNOWLEDGMENT

This work is supported by the National High Technology Development 863 Program of China (No.2011AA01A107) and the Zhejiang Provincial Technical Plan Project (No. 2011C13008).

REFERENCES

[1] A.C. Davies, J.H. Yin, S.A. Velastin. Crowd monitoring using image processing. Electronics and Communications Engineering Journal, February , pp. 37-47, 1995.

[2] Chow T. W. S. Cho, S. Y. and C. T. Leung. A neural based crowd estimation by hybrid global learning algorithm. IEEE Trans on Systems, Man, and Cybernetics, pp. 535-541, 1999.

[3] Da Fontoura Costa L. Lotufo R. Velastin S Marana, A. Estimating crowd density with Minkowski fractal dimension. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp.3521-3524, 1999.

[4] Velastin S. Costa L. Lotufo R. Marana, A. Automatic estimation of crowd density using texture. Safety Sci, 28(3):165-175,1998.

[5] Velastin, S., Yin, J., Davies, A., Vicencio-Silva, M., Allsop, R., Penn, A.: Automated measurement of crowd density and motion using image processing. Road traffic monitoring and control, 1994. In: Seventh International Conference, pp. 127–132 , 1994.

[6] Antonio Albiol, Maria Julia Silla, Alberto Albiol and Jos´e Manuel Mossi. Video Analysis using Corner Motion Statistics. Proceedings 11th IEEE International Workshop on PETS, Miami, June 25, 2009.

[7] Donatello Conte, Pasquale Foggia, Gennaro Percannella et al. A Method for Counting Moving People in Video Surveillance Videos. EURASIP Journal on Advances in Signal Processing Volume 2010.

[8] A. N. Marana, L. F. Costa, R. A. Lotufo, and S. A. Velastin. On the efficacy of texture analysis for crowd monitoring. Computer Graphics, Image Processing, and Vision, pp. 354-361,1998.

[9] Schofer J. Ushpiz A. Polus, A. Pedestrian Flow and Level of Service. J. Transportation Eng，109(1):46-56, 1983.

[10] Zi Ye, Jinqiao Wang, Zhenchong Wang, Hanqing Lu. Multiple features fusion for crowd density estimation. Proceeding ICIMCS '12 Proceedings of the 4th International Conference on Internet Multimedia Computing and Service, pp. 42-45, 2012.

[11] SUBBURAMAN V B, DESCAMPS A, CARINCOTTE C. Counting People in the Crowd Using a Generic Head Detector[C]. Proceedings of 2012 IEEE 9th International Conference on Advanced Video and Signal-Based Surveillance (AVSS): September 18-21, 2012. Beijing, China, pp. 470-475, 2012.

Ming Jiang He received the B.S. degree and M.S. degree in science in 1996 and 2001 respectively, and Ph.D. degree in Computer Science in 2004, all from Zhejiang University, China. He is currently a Professor in college of Computer Science, Hangzhou Dianzi University, China. His research interests include network virtualization, Internet QoS provisioning, and network multimedia processing.

Jingcheng Huang He received his BSc in Software Engineering from Hangzhou Dianzi University in 2011. Currently he is a master student in this university. His primary research area focuses on image and video processing, image segment.

Xingqi Wang He received his Bachelor and Master degree from Harbin Institute of Technology in 1997 and 1999, respectively, and Ph.D degree from Zhejiang University in 2002. He is an associate professor in college of Computer Science, Hangzhou Dianzi University, China. All his major are Computer Science. As a researcher, he visited CERCIA, University of Birmingham, UK from 2005 to 2006. His research interests include machine learning, data mining and multimedia content analysis.

Jingfan Tang He received the Ph.D. degree in Computer Science in 2005 from Zhejiang University, China. He is currently an Associate Professor in college of Computer Science, Hangzhou Dianzi University, China. His research interests

include network virtualization, quality assurance, process improvement and legacy system re-engineering..

Chunming Wu He received the B.S. degree, M.S. degree and Ph.D. degree in Computer Science from Zhejiang University, China, in 1989, 1992 and 1995 respectively. He is currently a Professor in college of Computer Science, Zhejiang University, and the director of NGNT laboratory. His research fields include Network Multimedia processing, reconfigurable network technology, network virtualization and artificial intelligence