http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (16): 3678-3684, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.3678.3684

High-density Crowd Behavior Recognition Based on Improved Harris

Cheng Xu, Xianxian Tian, Wenxian Zhao, Hongzhe Liu and Hong Bao
Beijing Key Laboratory of Information Service Engineer,
Beijing Union University, Beijing, China

Abstract: The study uses Gaussian mixture model to build background model of video sequence, then we segment the foreground image and detect the corners' characteristics of the crowd, finally we use the covariance matrix of of feature points' coordinates sets to do optimization calculation and we use the changes of the determinant value to describe the changes of corner area, the curve of corner areas changing with time can be use to describe the behavior of the crowd, we can use this curve to distinguish the behavior of the crowd. We test PETS standard data sets and self-shooting images. The results show that the proposed algorithm has better realistic effects than the original Harris.

Key words: Crowd behavior improved harris covariance matrix

INTRODUCTION

In public places in the state study of the crowd, the use of video supervision and management by more and more researchers of all ages. Outdated video surveillance mainly relies on manual to complete, because people will appear decreased attention and eye strain, etc. often occurs false alarm and other issues, more often monitoring information is recorded only as a tool for the incident. Drawbacks of being safe for the formation of high-density crowd risk, so the development of intelligent video surveillance has become an inevitable trend shift.

Although a lot of researches have been commercialized but there are still a lot of complex scenes which are lack of stability and robustness methods. For indoor the crowd intelligent monitoring environmental impacts are mainly illumination, camera jitter and offset camera occlusion or other issues, the environmental impact of the outdoor crowd, mainly rain, snow, fog, haze and other inclement weather, night low illumination, etc.

How to reduce the environmental impact on the image quality, faster significant the crowd divided area, solve block caused false alarms and other issues found in the camera offset state of intelligent Research of crowd key technical points, so that the various application scenarios stable output under all the desired results.

Based on Actual requirements and long-term Experimental Research proposed a method of judgment the crowd behavior can be real-time monitoring and video applications

RELATED WORK

In numerous international universities and research institutes, crowd research is an important part of the video monitoring technology. In 2003, Europe has launched a PRISMATICA and ADVISOR system. In 2013, China's large-scale separation and portraits retrieval intelligent character recognition system achieved a breakthrough, capable of processing large-scale portraits, proactively identify key populations, as an important means of providing security (Bae *et al*., 2013).

On the study of the crowd behavior analysis, numerous literatures are generally based on two methods. Direct detection methods, the main goal of a single individual as the rapid detection and segmented and then the crowd of density estimates, people counting or behavior analysis. Another method is indirect detection methods, the highest crowd through appropriate algorithm to be processed as a whole, this has the advantage of do not need to detect separately from the each scene. Methods are known in recent years (Brostow and Cipolla, 2006; Zhao *et al*., 2008).

Based on a single or a small amount of direct detection of the abnormal behavior target analysis of method. Body structure analysis based on the approach is to use different postures the human body model posture extracted from each frame of video, with the sequence of human posture by analyzing the changes to the frame to describe the operation of various human behavior. Pramod *et al*. (2009) extracted the edge features on the head and shoulder of the person and then estimates the number of people. The direct detection method based on individual can have segmented each target individual but in high-density crowd suffers more severe occlusion occurs when large errors in the crowd and high density shielding serious cases, for a sole individual segmentation very difficult.

Based on analysis method spatial-temporal characteristics of the model first extract the image of the

Corresponding Author: Hong Bao, School of Beijing Union University, China

human body area and then use the contours individual characteristics of each frame to achieve changes in the description of human behavior and identifications.

Tracking based behavior analysis method is mainly divided into three steps: the first step on the current image frame individual target detection; the second step of the detected target tracking; The third step is to extract the trajectory, the use of machine learning and data mining Technical analysis target specific behaviors. The best type is the optical flow method, the (Duan *et al*., 2012; Bae *et al*., 2013) combined with spectral clustering, principal component analysis and other methods optical flow capacity crowd images analysis, testing crowd's behavior. Nallaivarothayan *et al*. (2012) used the Markov model to analyse and determine the anomalous behavior of the crowd, the scene in the direction of unity application better. Optical flow analysis to determine the behavior of the crowd is mainly reflected in the calculation of velocity, due to complexity of the algorithm and compared with additional algorithm needs more time and therefore cannot meet the needs of real-time aspects.

The test data is PETS (IEEE International Workshop on Performance Evaluation of Tracking and Surveillance) professional crowd analyze data sets.

HIGH DENSITY CROWD BEHAVIOR RECOGNITION

For high-density population, it is always a difficult problem to find a solution to solve the occlusion problem effectively. This study proposes a mixture Gaussian model to build a background model of the video sequence and segments the foreground image using a different method. The study uses the improved Harris operator to detect corners in the image, finally the study conduct the covariance matrix of the coordinates of the feature point set to represent the the corner area value of the crowd, in another word, the curve of the corner area with time is the characteristic curve of the movement people, we can identify the crowd behavior with this curve. As Fig. 1 shows.

Gaussian mixture model of crowd: There are many mature background modeling method, such as pixel estimation method and Gaussian modeling method, time averaging, etc. The gaussian model is the most successful one, it can meet real-time requirements, so we will use the Gaussian mixture model as the crowd background model. It is similar between the law of Multi-Gaussian model and the principle of a single Gaussian model method. The difference is that Multi-Gaussian model uses the K Gaussian model for modeling for each pixel. Sequence of video frames is as following equation:

$$\{X_1, X_2, \ldots\ldots, X_t\} = \{I(\varphi, i), 1 \le i \le t\} \tag{1}$$

Then the probability of each pixel in X_t frame is computed in the following equation:

$$P(X_t) = \sum_{i=1}^{K} \omega_{i,t} \bullet G_i(X_t, \vec{\mu}_{i,t}, U_{i,t}) \tag{2}$$

The above equation, K is the number of the mixed Multi-Gaussian model, The greater the K value is, the stronger the ability to process the fluctuations of pixel values is but the processing efficiency is reduced accordingly, K is generally 3-5, ω is the weight of the i-th Gaussian model at time t, $\vec{\mu}_{i,t}$ and $U_{i,\ t}$ is the mean and variance of the i-th Gaussian model at time t, G is a Gaussian probability density function which is as the following equation:

$$G_i(X_t, \vec{\mu}_{i,t}, U_{i,t}) = \frac{1}{(2\pi)^{\frac{n}{2}} \left|U_{i,t}\right|^{\frac{1}{2}}} e^{-\frac{1}{2}(X_i - \vec{\mu}_{i,t})^T U_{i,t}^{-1} (X_i - \vec{\mu}_{i,t})} \tag{3}$$

For gray scale images, n is 1 in the above equation. After initialization task of the model is completed, model parameter will be updated. At time t, we compare the value of each pixel in X_t with its corresponding Gaussian mixture model, if the distance between the pixel values and the i-th Gaussian distribution mean $\vec{\mu}_{i,t}$ is less than the 2.5 times of standard deviation U_i, then we define the pixel values matching the Gaussian distribution, the threshold effectively reduces the cyclical changing impact of light.

If the pixel matches with at least one of the Gaussian distribution of Gaussian mixture model, then the parameter

Fig. 1: Flow chart of the high-density crowd behavior recognition algorithm

update rule of Gaussian mixture model is: For the pixels not matching any of a Gaussian distribution, their mean $\vec{\mu}_{i,t}$ and covariance matrix H_i, t will keep still; for the pixels matching Gaussian distribution, their mean and covariance matrix H_i, t will respectively changing as the following equation:

$$\vec{\mu}_{i,t} = (1-\rho)\vec{\mu}_{i,t-1} + \rho X_t$$

$$H_{i,t} = (1-\rho)H_{i,t-1} + \mathrm{diag}[\rho(X_t - \vec{\mu}_{i,t})^T(X_t - \vec{\mu}_{i,t})] \tag{4}$$

where, $\rho = aG_i(X_t|\vec{\mu}_{i,t-1}, H_{i,t-1})$, a learning rate of parameter estimation, the value of a is usually 0.002.

If the pixel does not match any of a Gaussian distribution, then the process will reassignment the most unlikely to represent the background Gaussian distribution, the assignment principle is: Treat the current pixel value as the mean of the Gaussian distribution, as well as the Gaussian distribution will take a larger variance and smaller weight values. Then, the k-Gaussian distribution will be update the weight at time t as equation:

$$\omega_{i,t} = (1-\alpha)\omega_{i,t-1} + \alpha(M_{i,t}) \tag{5}$$

The above equation, if a Gaussian distribution matches a pixel value at time t, then the value of M_i, t is 1, otherwise values of M_i, t is 0. The simulation results are shown as Fig. 2-5.

Recognition method of crowd behavior: Crowd burst always occurrence in a public place, after splitting crowd prospects through background modeling, it is very necessary to try to effectively identify the acts of the crowd, detect security risks and conduct appropriate measures well in advance.

Antonio *et al*. (2009) proposed the method about statistical the dynamically changes of the corner feature, they use the variation area of the corners as a basis to identify the conduct of the crowd, the advantage of this method is not easy to change with illumination impact but if the people of the scene get together for a long time, it will cause the corner of the target keeping still, those stationary target points will be deleted, in addition if there is a small amount of targeted individuals appearing with a few meters from the gathered people, this small amount of dynamic corner points will cause the entire dynamic fluctuations of the corner area, then there will be false positives. Conte *et al*. (2010) improved Antonio *et al*. (2009) method, it experimental results show that this method has improved accuracy compared with Antonio while it maintained the robustness of the indirect method.

Fig. 2: Raw image of PETS dataset

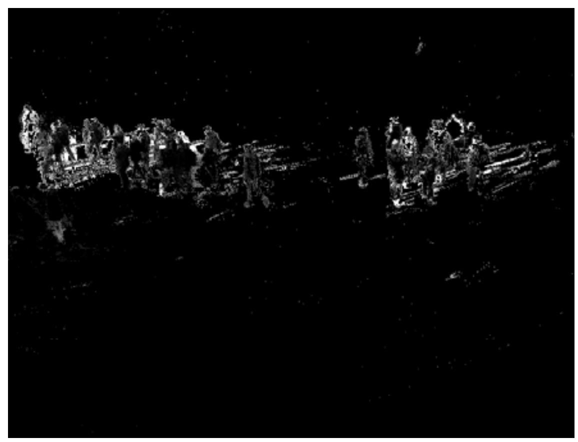

Fig. 3: Prospects image of the Gaussian mixture modeling of Fig. 2

Fig. 4: Raw image of self-shooting dataset

Fig. 5: Background image of the Gaussian mixture modeling of Fig. 4

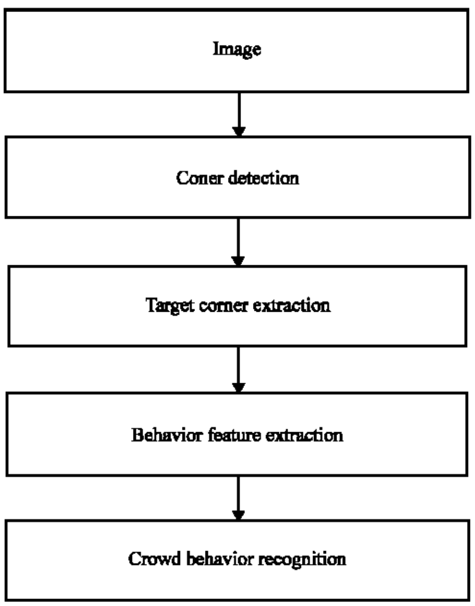

Fig. 6: Flow chart of the algorithm about the crowd behavior

Conte *et al*. (2013) proposed a real time and efficient monitoring system which has high accuracy and high speed, it can be used for real-time monitoring applications, the main method is based on the extraction of a suitable scale-invariant corner feature, the associated corner features and moving continuous after dividing horizontal region. The method uses simple training program demonstrated a good real-time performance and robustness, the program can automatically evaluate the system parameters, it is suitable for the application of the crowd counting and the crowd behavior analysis, it can also be segmented regions at different levels of identification.

Lots of literatures have made different kinds of improvements on how to effectively calculate the statistical characteristics of the different corners but they didn't have reliable effects on crowd behavior classifying. The improved Harris operator extracts the corner features, then we use the covariance matrix for all corner coordinates to do statistical analysis, we propose specific identification scheme, the specific implementation process is shown in Fig. 6.

Improved Harris corner detection operator: Harris operator depends on the Gaussian convolution function, by the use of the linear differential image of each frame to extract feature points, the corner detection is based on the gradient changes of the gray image, the corner point of gray image are the pixel values changing greater than the pixels of non-corner region where its gradient is the same. If the absolute value of the difference between the gray value of neighborhood of the center pixel and the gray value of the center pixel I (i, j) is in the range of a threshold value t, let t = 20, then we treat this pixel as a similar pixel with the center pixel. At the same time, counter m for the similarity of the center pixel I (i, j) will be implemented by one:

$$m = \sum R(i+x, j+y)(-1 \le x \le 1, -1 \le y \le 1, x \ne 0, y \ne 0) \tag{6}$$

Wherein:

$$R(i+x, j+y) = \begin{cases} 1, \Delta(i+x, j+y) \le t \\ 0, \Delta(i+x, j+y) > t \end{cases} \tag{7}$$

When 8 neighborhood points of I (i, j) are traversed, we can get similar statistical value m between the center pixel and 8-neighborhood points. The value of m, if m = 8 which means that the current central pixel has 8 similar neighborhood points; m = 0, shown in Fig. 7, it represents the current central pixel has no similar pixels with the neighborhood pixels, so the pixel is an isolated pixel or noise points. If $1<m<7$, the pixel are included in the candidate corner points and then determine whether the current center points is a corner or not by calculating CRF.

Feature extraction of crowd behavior: After utilizing improved Harris operator to obtain the coordinates of the target feature corner, we use the determinant value of the covariance matrix of the target coordinate datasets to describe the area of target points. Each element of the covariance matrix is the variance between the elements of

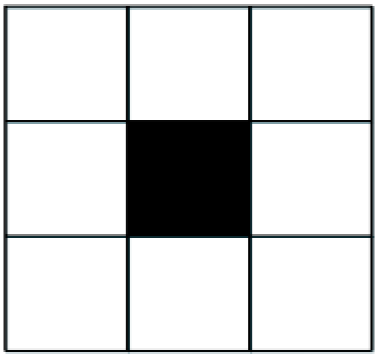

Fig. 7: The similarity between the center pixel and the surrounding pixels

the various vectors. The vector of covariance matrix is composed of the horizontal and vertical coordinates of corners, the precise calculation procedure of the determinant of the covariance matrix is as following.

Let we assume that D describes the coordinates of corner points, D is a n×2 matrix, each row is the horizontal and vertical coordinates of the corners, we set the coordinates of corner points as (x, y), then the corner covariance matrix of the corner points can be defined as:

$$\sum(x,y)=\begin{bmatrix} S_{xx}(x,y) & S_{xy}(x,y) \\ S_{xy}(x,y) & S_{yy}(x,y) \end{bmatrix} \quad (8)$$

where S_{xx} is the variance of x-coordinate in D, S_{yy} is the variance of the vertical coordinate in D, S_{xy} is the covariance of corner coordinates in D. The determinant value H is as following:

$$H=\left|S_{xx}(x,y)\times S_{yy}(x,y)-S_{xy}(x,y)^2\right| \quad (9)$$

The determinant value of the H describes the area of the corner in the scene. We utilize the frame sequence as a horizontal axis. The determinant value H is the vertical axis. The resulting curve will reflect the curve of the changing area of the crowd.

EXPERIMENT AND ANALYSIS

Figure 8-10, respectively shows the curves of the changing area of crowd, the three kinds of crowds are the gathered crowds, walk normally crowds and crowds with a small gathered crowd.

Through the Table 1, we can get the following conclusion through the experiments, when the aggregation behavior and evacuation behavior of crowds happens, the difference determinant values of the

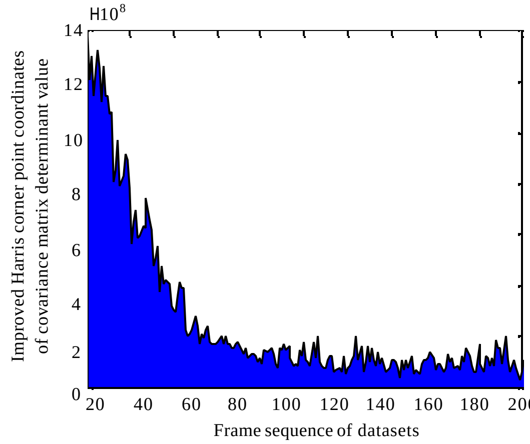

Fig. 8: Curve of gathered crowds

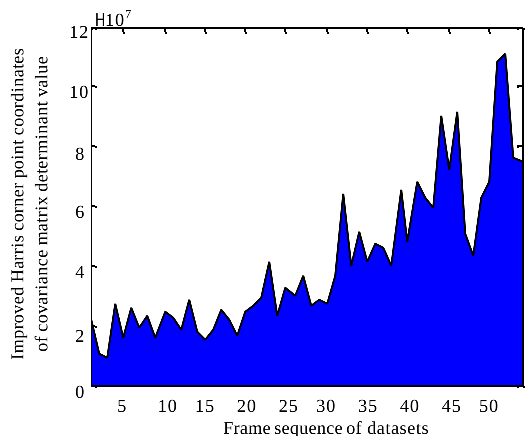

Fig. 9: Curve of the normal walking crowds

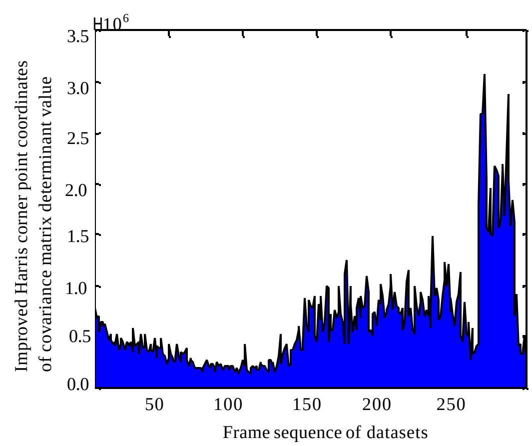

Fig. 10: Curve of the crowds with small crowds around

covariance matrix of multi-frame ΔH changes obviously in magnitude than the normal walking crowds, so you can take advantage of the optimal threshold to tell from the

Table 1: Results of max amplitude variation for the determinant value H of the covariance matrix of corners with different frame numbers

	Dataset	10 frame (ΔH)	30 frame (ΔH)	50 frame (ΔH)
PETS	Gathering dataset 1	1.4	4.9	12.0
	Gathering dataset 2	2.6	6.5	15.0
	Normal walking dataset 1	1.5	4.0	3.5
	Normal walking dataset 2	2.2	5.2	2.6
	Evacuation dataset 1	11.0	90.0	100.0
	Evacuation dataset 2	9.0	60.0	58.0
Self-shooting	Gathering dataset 1	5.0	9.0	12.3
	Gathering dataset 2	6.0	7.6	10.0
	Normal walking dataset 1	2.0	1.8	2.6
	Normal walking dataset 2	2.8	2.2	2.3
	Evacuation dataset 1	6.0	7.0	15.0
	Evacuation dataset 2	8.0	6.0	12.0

Table 2: Different methods through threshold value to recognize the crowd behavior

	Recognition number			Accuracy		
Video type and quantity	Antonio (2009)	Donatello (2013)	Our	Antonio (2009)	Donatello (2013)	Our (%)
7-segment data sets on normal walking	6	7	7	85.7	100%	100
7-segment data sets on crowd gathered	7	7	7	100.0	100%	100
7-segment data sets on evacuation	7	7	7	100.0	100%	100
4-segment after gathering activities in situ	1	2	3	25.0	50%	75

behavior of the different crowds. Through the experiments of the self shooting scene data sets and through different experimental analysis and the optimal threshold is 30, the change rate of the curve can achieve optimal results by using threshold 30.

Through Fig. 8 and 9 we can see that slope can be used to distinguish the behavior of gathered crowds and the behavior of dispersal crowds, according to the results of Table 1, ΔH can be set to 1.8 to distinguish the normal walking behavior and the behavior of situ activity and ?H can be set to 4.5 to distinguish the behavior of the gathered crowds, normal walking crowds and evacuation crowds.

The threshold of the different methods of the literature on crowd behavior recognition is shown in Table 2.

In Table 2, Antonio *et al.* (2009) has successfully judged on the PETS dataset S1_L3 in Time14-33, it is mainly because that there is another little crowds gathered away from the target crowds, so there is wrong judgment about the corners of the crowds and it has wrong predicting results as the behavior of evacuation crowds. Conte *et al.* (2013) can not have good results on the behavior of crowds of far scene by using horizontal region segmentation.

On Time_14-16's view002 scene of the S1_L2, because of the special placed camera, the camera's horizontal angle is small, the scene for the gathered crowd walking from far place to the camera, the closer the crowd to the camera is, the more the detected crowd feature area of corners, the curve of the corner area changes obviously.At that time there is always misjudgment but the same self-shooting scene can be correct judge, it is mainly due to the high resolution data of the self-shooting images. In addition, the image sequence is different, the self-shooting image is 50 ms per frame.

CONCLUSION

It is necessary to study the behavior of the crowd in the public place, this study use the Gaussian mixture background model to segment the crowds, it use the improved Harris operator to detect the corner features of the crowd.The study conduct the determinant value's changes of covariance matrix of target corner coordinates set as features to do recognition, finally we set the threshold ÄH to identify the behavior of different groups.

With the deepening study of crowd behavior, the next step will be to learn how to semi-supervised automatic learning crowd behavior, how to effectively segment the different target's behavior through intelligent analysis the camera's the parameters by machine learning to provide help for crowd safety.

ACKNOWLEDGMENT

This project was supported by the Project of Construction of Innovative Teams and Teacher Career Development for Universities and Colleges under Beijing Municipality (CIT&TCD20130513), and the National Natural Science Foundation of China (grant No.61271370 and grant No.61271369), and the project No. 201311417SJ056. We thank PETS for sharing its source images and the reviewers for insightful comments.

REFERENCES

Antonio, A., M.J. Silla and J.M. Mossi, 2009. Video analysis using corner motion statistics. Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, December 7-9, 2009, Miami, FL., USA., pp: 31-38.

Bae, G.T., S.Y. Kwak and H.R. Byun, 2013. Motion pattern analysis using partial trajectories for abnormal movement detection in crowded scenes. Electronics Lett., 49: 186-187.

Brostow, G.J. and R. Cipolla, 2006. Unsupervised bayesian detection of independent motion in crowds. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 17-22, 2006, Portland, OR., USA., pp: 594-601.

Conte, D., P. Foggia, G. Percannella, F. Tufano and M. Vento, 2010. A method for counting people in crowded scenes. Proceedings of the 7th IEEE International Conference on Advanced Video and Signal Based Surveillance, August 29-September 1, 2010, Boston, MA., pp: 225-232.

Conte, D., P. Foggia, G. Percannella and M. Vento, 2013. Counting moving persons in crowded scenes. Machine Vision Appl., 24: 1029-1042.

Duan, G., H. Ai, J. Xing, S. Cao and S. Lao, 2012. Scene aware detection and block assignment tracking in crowded scenes. Image Vision Comput., 30: 292-305.

Nallaivarothayan, H., D. Ryan, S. Denman, S. Sridharan and C. Fookes, 2012. Anomalous event detection using a semi-two dimensional hidden markov model. Proceedings of the International Conference on Digital Image Computing Techniques and Applications, December 3-5, 2012, Fremantle, WA., pp: 1-7.

Pramod, K.S., H. Chang and N. Ram, 2009. Evaluation of people tracking, Counting and density estimation in crowded environments. Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, December 7-9, 2009, Newyork, USA., pp: 39-46.

Zhao, T., R. Nevatia and B. Wu, 2008. Segmentation and tracking of multiple humans in crowded environments. IEEE Trans. Pattern Anal. Mach. Intell., 30: 1198-1211.