Crowd Gathering Detection Based on the Foreground Stillness Model

Chun-Yu LIU†, Wei-Hao LIAO†, Nonmembers, and Shanq-Jang RUAN†, Member

SUMMARY The abnormal crowd behavior detection is an important research topic in computer vision to improve the response time of critical events. In this letter, we introduce a novel method to detect and localize the crowd gathering in surveillance videos. The proposed foreground stillness model is based on the foreground object mask and the dense optical flow to measure the instantaneous crowd stillness level. Further, we obtain the long-term crowd stillness level by the leaky bucket model, and the crowd gathering behavior can be detected by the threshold analysis. Experimental results indicate that our proposed approach can detect and locate crowd gathering events, and it is capable of distinguishing between standing and walking crowd. The experiments in realistic scenes with 88.65% accuracy for detection of gathering frames show that our method is effective for crowd gathering behavior detection.

key words: crowd gathering detection, surveillance application, abnormal crowd event detection, image recognition

1. Introduction

Recent technology developments on digital surveillance systems have emerged because of the steep demands on the security solutions in order to tackle the increasing numbers of various threats in public areas. In general, practical applications of the surveillance system demand a substantial amount of cameras. To completely supervise the video feeds in real-time, numerous operators are required. Therefore, automated surveillance schemes are developed to detect patterns that do not conform to an established normal behavior, which is applicable in a variety of applications, e.g., intrusion detection, fault detection, and event detection in sensor networks.

One of the abnormal behaviors is crowd gathering. The abnormal crowd gathering behavior which is dangerous at the crowded place makes people fall or be injured by pushing. Several researchers investigated the concepts of abnormal crowd behaviors: how to analyze people behavior correctly in crowded places? For example, Zhan et al. and Li et al. have written a widely survey about the crowd analysis in [1] and [2]. In [3], they divided the research on crowd video analysis into three broad categories: microscopic modeling, macroscopic modeling and crowd event detection based on the microscopic and macroscopic. Microscopic analysis and modeling [4] depends on the analysis of trajectories of moving objects. However, it is hard to detect and trajectory objects in crowd videos due to the severe occlusions. The macroscopic utilizes holistic properties of the scene to analyze the crowd events. Ramin et al. proposed a novel method to detect the abnormal crowd behavior using the social force model. The method illustrated that a grid of particles is placed over the image, and it is advected with the spacet ime average of optical flow. However, the interaction force of the social force model does not present any information when the crowd is static. In [6] and [7], they used the energy model to detect the crowd gathering. Although they can detect crowd gathering by potential energy and kinetic energy, the method cannot distinguish the stationary group of people from the crowded place. They are not able to locate the crowd gathering position on the surveillance image. In addition, their definition of crowd gathering is weak because they just estimate the distribution of people which do not include gathering behavior.

In this letter, a novel method for crowd gathering behavior detection is proposed based on the foreground stillness model. The foreground stillness model is designed specifically for estimating the stillness level of the observed crowd. Further, we obtain long-term crowd stillness level by the leaky bucket model. The proposed approach is tested on the public dataset PET2009 [8]. The experimental results indicate that our proposed approach can detect and locate crowd gathering events robustly, and it is capable of distinguishing between the standing and walking crowd. Figure 1 shows an example of the crowd gathering behavior detection using the proposed scheme.

The rest of this letter is organized as follows. Section 2 describes the proposed method which detects the crowd gathering behavior. The experiments and demonstrations are presented in Sect. 3. Finally, Sect. 4 provides the conclusions of this letter.

Manuscript received January 11, 2018.
Manuscript revised March 6, 2018.
Manuscript publicized March 30, 2018.
†The authors are with the Department of Electronic and Computer Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan.
a) E-mail: sjruan@mail.ntust.edu.tw
DOI: 10.1587/transinf.2018EDL8005

Fig. 1 An example of the crowd gathering behavior detection using the proposed scheme. (a) Normal scene. (b) Crowd gathering behavior.
2. Proposed Method

The proposed method is comprised of three integrated modules as shown in Fig. 2. The first module measures the instantaneous crowd stillness level by the foreground stillness model, of which foreground segmentation is performed by Adaptive Gaussian Mixture Model [9]. The second module using the leaky bucket model to detect the objects which are always motionless. The threshold analysis determines whether the crowd gathering event occurs. These three modules are elaborated on the rest of this section.

2.1 Foreground Stillness Model

Compared with the crowd gathering detection method which uses the energy model[6] to calculate the crowd distribution for a whole image, our foreground stillness model can directly measure the instantaneous crowd stillness level for each pixel by the proposed foreground stillness model. To measure the stillness state of the observed crowd, only comparing the pixel information (including the luminance and chrominance channels) of the successive frames is not enough since the individuals inside the crowd may not stand on their positions always. In that situation, only comparing the pixel information may fails in recognizing the unusual activities such as crowd commotion and pedestrian occlusion. Therefore, in the proposed model, the pixel of the static object is assigned a higher value namely higher stillness level while the pixel of the moving object is assigned a lower value, which can clearly distinguish different stillness states and can solve the problem mentioned above by applying the Leaky Bucket Model which will be discussed later. The foreground stillness model $S_{i,j}(F,m_{i,j})$ which is designed specifically for measuring the stillness level of the observed crowd by the maximum acceptability motion complement is formulated as

$$S_{i,j}(F,m_{i,j}) = \begin{cases} \text{Max} - |m_{i,j}| \times n(i), & F = 1 \\ 0, & F = 0 \end{cases}$$ (1)

where Max denotes the maximum acceptability motion value, $m_{i,j}$ is the motion vector obtained from the dense optical flow, and $F$ is the foreground label from the foreground segmentation. The linear transformation $n(i)$ in (1) is required to adjust object motion size due to the phenomenon of perspective. This $n(i)$ transformation is expressed as

$$n(i) = (i - AH_m) \times \frac{H_m - 1}{h} + 1,$$ (2)

where $H_m$ and $H_M$ denote the pixel height of the closest and the farthest calibration object respectively. $AH_m$ is the $y$-coordinate of the farthest calibration object. The $h$ is the distance of the two calibration objects on the images. The variable $i$ is the $y$-coordinate of the object.

2.2 Leaky Bucket Model

Through the proposed foreground stillness model, the pixel of the still object presents the higher value of stillness level. To make the stillness measurement obtain the time domain information, the leaky bucket model is employed. The leaky bucket model $LBM_{i,j}(t)$ is expressed as

$$LBM_{i,j}(t) = \begin{cases} 0, & L_{i,j}(t) < 0 \\ L_{i,j}(t), & 0 \leq L_{i,j}(t) < \text{Ceiling} \\ \text{Ceiling}, & L_{i,j}(t) > \text{Ceiling}, \end{cases}$$ (3)

with

$$L_{i,j}(t) = LBM_{i,j}(t-1) + S_{i,j} - E,$$ (4)

where $\text{Ceiling}$ is the ceiling value of accumulated value, the variable $t$ is the frame number of the video stream, and $E$ is the constant value for decreasing the accumulated value until the value becomes zero.

The leaky bucket is a model that is used to accumulate the specific feature, and the accumulated value decreases over time in the limited scope. In term of the normal ways such as majority voting and temporal averaging, it is hard to decide how many frame information needed to be stored for the determination of the person stillness state, especially in the situation of commotion and occlusion. In contrast, the Leaky Bucket Model can control the accumulated value easily to determine the personal stillness state in every different place and in every different situation. In Eq. (4), it shows that the model only needs the information of the last frame LBM, which indicates that the model uses less memory resource to achieve the same result.

2.3 Threshold Analysis

Through the leaky bucket model, each pixel presents the long-term stillness level of the observed crowd. To store the locations of pixels that represent the long-term static region on $LBM_{i,j}(t)$, the labeling function $LSP_{i,j}(t)$ based on $th_{LS}$ is formulated by the following inequalities:

$$LSP_{i,j}(t) = \begin{cases} 1, & LBM_{i,j}(t) > th_{LS} \\ 0, & \text{otherwise} \end{cases}$$ (5)

The labeling function presents the position of long-term stillness objects. To analyze the stillness object, each pixel of long-term stillness objects is clustered by the connected-component analysis. Thus, the crowd gathering behavior alarm $A_C$ can be decided by the cluster area. The

![Fig. 2](image-url) The flow diagram of the proposed scheme for crowd gathering detection.
alarm expression is formulated as
\[ A_G = \begin{cases} 
1, & \text{if } \sum SC(k) > 0 \\
0, & \text{otherwise}
\end{cases}, \] (6)

with
\[ SC(k) = \begin{cases} 
1, & \text{if } SCA(k) > th_A \\
0, & \text{otherwise}
\end{cases}, \] (7)

where the variable \( k \) is the long-term stillness object, the \( SCA(k) \) is the stillness cluster area, and the \( th_A \) is the event alarm threshold.

3. Experimental Results

The publicly available dataset of crowd gathering videos from PET2009 [8] is chosen for evaluating the performance of the proposed method. The dataset comprises of the different scenario videos, and there are four crowd gathering videos in the different scenes. Each video for testing successively consists of sequences of normal behavior, sequences of the crowd gathering behavior and sequences of the crowd running. Each tested video resolution is 768*576, and the frame rate is close to 7fps.

3.1 Foreground Stillness Model

A sequence of video is tested to estimate the instantaneous stillness level by the proposed foreground stillness model. In the foreground stillness model, the maximum acceptability motion is set to the average speed of people walking. The average speed of walking is about 5km/h (1.4m/s). Because the pedestrian moving distance between each two frame is 20cm which maps to the image is 5 pixels, the maximum acceptability motion value \( MAX \) is set to 5. The instantaneous stillness level in Fig. 3 is visualized by different colors. The color presents the level from higher to lower is: red, green, and blue. The 50th frame shows the lower instantaneous stillness level because the crowd is moving. In the 100th frame, the foreground stillness model distinguishes clearly between the still and moving crowd. The left walking people present the lower stillness level because of the higher motion. The right stand people present the higher stillness level because of the lower motion. And the 150th and 200th frames demonstrate the still crowd presents the higher value. The results show that the foreground stillness model can correctly estimate the instantaneous stillness level.

3.2 Leaky Bucket Model

Figure 4 shows the results that illustrate the long-term stillness level on a sequence of real video. The parameters for the leaky bucket model is present: the decreasing value \( E \) is set to 60 percent of the \( MAX \), and the ceiling value of accumulated value ceiling is 105. It means that the maximum accumulated value can be decreased to zero in 5 seconds. As the figure shows, the long-term standing crowd presents the higher long-term stillness level, and the long-term stillness level increases correctly from the crowd gathering at the regions of gathering. The long-term stillness level is capable of locating the crowd gathering.

3.3 Threshold Analysis

The long-term still objects are distinguished by (5). The long-term stillness threshold \( th_{LS} \) is set to 49. The pixels of still objects are able to emerge in 3.5 seconds. The maximum long-term stillness blob area is recorded in Fig. 5. The maximum long-term stillness blob area of each frame is increasing when the crowd starts gathering, and the maximum long-term stillness blob area of each frame is decreasing when the crowd starts running. From the line graph, the maximum long-term stillness blob area is capable of detecting crowd gathering behavior. Finally, the gathering event threshold \( th_A \) for the tested video is set to the area of people. The people number is decided by the user, and \( th_A \) is set to the area of five people here.

Figure 6 shows the results for the detection of crowd gathering behavior. In each row, the figure shows the first frame of the sequence on the left and the detected crowd gathering frame on the right. The black triangles on the hor-

![Fig. 3](https://via.placeholder.com/150)  
**Fig. 3** The instantaneous stillness level tested on a sequence of real images in the PET2009 dataset.

![Fig. 4](https://via.placeholder.com/150)  
**Fig. 4** The long-term stillness level tested on a sequence of real images in the PET2009 dataset.

(a) Sequence 1  (b) Sequence 2  (c) Sequence 3  (d) Sequence 4

![](https://via.placeholder.com/150)

**Fig. 5** The line graph presents the maximum area of the long-term stillness object of each video frame on different tested videos. The vertical axis is the maximum area of the stillness person and the horizontal axis is the video frame number.
Fig. 6  The experimental results of the gathering behavior detection for four sample videos of the PET2009 dataset. The detection bar represents the labels of each frame for that video. Green color represents the normal frames and red corresponds to gathering frames. The left image shows the first frame of the video, and the right image is the first frame of the detected gathering position (black triangles).

Table 1  The confusion matrix for detection of gathering frames in PET2009.

| Sequence | TP  | TN  | FP  | FN  | TPR | FPR | ACC  |
|----------|-----|-----|-----|-----|-----|-----|------|
| 1        | 209 | 130 | 28  | 10  | 95.43% | 17.72% | 89.92% |
| 2        | 215 | 129 | 25  | 8   | 96.41% | 16.23% | 91.24% |
| 3        | 207 | 124 | 34  | 12  | 94.52% | 21.52% | 87.79% |
| 4        | 206 | 117 | 33  | 21  | 90.74% | 22%    | 85.67% |

Fig. 7  The ROC curve for gathering detection rate of four test sequence in PET2009.

Fig. 8  The gathering detection result (a) Crowd gathering detection resulting image by Xiong et al. [6]. (b) Crowd gathering detection resulting image by proposed scheme.

from [6] does not detect the crowd gathering position on the image, and our proposed method in Fig. 8 (b) is able to mark the location of the crowd gathering.

4. Conclusion

In this letter, we proposed a novel approach to detect the crowd gathering behavior by measuring the crowd stillness level. First, the proposed foreground stillness model measures the instantaneous crowd stillness level. Moreover, the leaky bucket model is presented to obtain the long-term crowd stillness level. The stationary group of people can be distinguished from the sequence of videos. The proposed method has been validated on the PET2009 dataset, and the results of our method indicate that the method is effective in detection and localization of crowd gathering behavior.

References

[1] B. Zhan, D.N. Monekosso, P. Remagnino, S.A. Velastin, and L.-Q. Xu, “Crowd analysis: A survey,” Machine Vision and Applications, vol.19, no.5-6, pp.345–357, Oct. 2008.
[2] T. Li, H. Chang, M. Wang, B. Ni, R. Hong, and S. Yan, “Crowded scene analysis: A survey,” IEEE Trans. Circuits Syst. Video Technol., vol.25, no.3, pp.367–386, March 2015.
[3] M. Thida, Y.L. Yong, P. Climent-Pérez, H.-L. Eng, and P. Remagnino, “A literature review on video analytics of crowded scenes,” Intelligent Multimedia Surveillance — Current Trends and Research, ed. P.K. Atrey, M. Kankanhalli, and A. Cavallaro, pp.17–36, Springer-Verlag Berlin, 2013.
[4] S.M. Khan and M. Shah, “Tracking multiple occluding people by localizing on multiple scene planes,” IEEE Trans. Pattern Anal. Mach. Intell., vol.31, no.3, pp.505–519, March 2009.
[5] R. Mehran, A. Oyama, and M. Shah, “Abnormal crowd behavior detection using social force model,” Proc. IEEE Conf. on Computer Vision and Pattern Recognition, Miami, FL, pp.935–942, 2009.
[6] G. Xiong, X. Wu, Y.-L. Chen, and Y. Ou, “Abnormal crowd behavior detection based on the energy model,” Proc. IEEE International Conf. on Information and Automation, Shenzhen, pp.495–500, 2011.
[7] G. Xiong, J. Cheng, X. Wu, Y.-L. Chen, Y. Ou, and X. Xu, “An energy model approach to people counting for abnormal crowd behavior detection,” Neurocomputing, vol.83, pp.121–135, 2012.
[8] J. Ferryman and A. Shahrokni, “PETS2009: Dataset and challenge,” Proc. 12th IEEE International Workshop on Performance Evaluation of Tracking and Surveillance, Snowbird, UT, pp.1–6, 2009.
[9] Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction,” Proc. 17th International Conference on Pattern Recognition (ICPR 2004), Aug. 2004.