Holistic Crowd Interaction Modelling for Anomaly Detection

Jiaxing Pan(✉) and Dong Liang

Collaborative Innovation Center of Novel Software Technology and Industrialization, College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, People's Republic of China
panjiaxing_happy@163.com, liangdong@nuaa.edu.cn

Abstract. Dense crowd motion analysis in surveillance scenario is a daunting task that when occlusion and low resolution happen, it is difficult to make effective use of pedestrian detection and tracking algorithms. In this study, we introduce a crowd interaction modelling framework inspired by physical and social science studies. Instead of taking the pedestrian individual as the unit of analysis, the interaction among individuals could be modeled through the social force model (SFM), and for robust representation, a modified SFM is proposed. Experiments of the visualization and anomaly detection tested on UMN and Web dataset indicate SFM-based interaction modelling outperform optical flow and particle advection.

Keywords: Dense crowd analysis · Anomaly detection · Social force model · Optical flow

1 Introduction

Crowd stampede and violence often occur worldwide. The crowd motion analysis in surveillance scenario is in urgent need for public security. Patch-based models have been proposed for tracking objects in crowd [16–18]. However, they are time-consuming. And observing pedestrian individual is difficult for the occlusions, tiny objects and low resolution in high density surveillance scenes [1]. In such case, considering macroscopic motion patterns is a wiser choice to exploit the interaction in the crowd [2], where usually the holistic properties of the scene are modeled.

High density crowd motion analysis is mainly used in anomaly detection, where the key problem is robust motion representation. Tal considered the statistics information of the magnitude of flow field vector varies as time goes by, whereby the proposed ViF descriptor realizes real-time of anomaly detection [3]; Mahadevan proposed a Mixtures of Dynamic Texture method conducting anomaly detection in time and space domain [4]. Meanwhile note that, algorithms based on deep learning is introduced in crowd scenario analysis [5,6]. However, this kind of method requires a large scale dataset to train a complex network

J. Zhou et al. (Eds.): CCBR 2017, LNCS 10568, pp. 642–649, 2017.
https://doi.org/10.1007/978-3-319-69923-3_69

containing a large number of parameters. In a recent work [7], the author proposed a method combining acceleration, body compression and aggressive drive attributes extracted from the original video to detect the violence in crowd scenes, and the experimental comparison with a pre-trained convolution neural network [5] gets a better result. Originate in the research of physics, sociology and psychology, Helbing proposed social force model [8] to model the interaction among pedestrian. Mehran et al. apply the model to anomaly detection and localization [9].

This study focuses on the analysis of robustness of social force model and improves it. The comparison of the anomaly detection results with optical flow and particle advection is conducted on UMN [10] and Web [9] benchmark. The paper is organized as follows. In the next section we will introduce related methods and their visual effectiveness. The modified model will be proposed in Sect. 3, then experiments and conclusion come in Sects. 4 and 5 respectively.

2 Crowd Motion Pattren Estimation

Most of the pixel-level motion estimation method is based on optical flow. Also, there are a large number of effective optical flow algorithm used in motion detection and segmentation tasks widely [11,12].

2.1 Particle Advection

Based on the optical flow, [13] computed the particle advection. Firstly, put some particles evenly upon the frame, and then particles move with the optical flow, so as to simulate the motion of individual changes with the people around them. Particle advection can be seen as a smoothing process of optical flow in both spatial and temporal domains. As shown in Fig. 1, the image on the left shows the optical flow field. It can be found that, local random movement happen in the leg or arm part when a person was walking, however, if we carry on the particle advection (right), this problem can be solved well, and which enables a person or area with a relative consistent motion vector.

Fig. 1. Visual effectiveness of *optical flow* (*left two columns*) and *particle advection* (*right two columns*)

In practical experiment, we will uniformly adopt Gaussian kernel with the size of 5×5 as smooth filter in the spatial domain, and in order to obtain smoother, continuous visualization, the video frames used in temporal domain are $T = 10$.

2.2 Social Force Model for Interaction Modeling

Mehran introduces social force model to computer vision field [9]. In this model, each pedestrian i with mass m_i, his/her speed can be defined as the following formula:

$$m_i \frac{dv_i}{dt} = F_a = F_p + F_{int}. \tag{1}$$

F_a is the resultant force of pedestrian, whose value equals to the sum of personal desired force F_p and interaction force F_{int}. Figure 2 is a graphical representation of social force model. Known that pedestrian in the crowd tends reach its destination in a walk. However, due to the obstacle of others and the environment barriers, the actual velocity v_i is different from the expected speed v_i^p. Consequently, one would try to exert a force F_p to achieve the expected speed:

$$F_p = \frac{1}{\tau}\left(v_i^p - v_i\right). \tag{2}$$

τ is a relaxation factor. According to (1) and (2), the key point of the work actually lies in how to represent the actual speed and expectation speed. In [9] Mehran expresses personal expectation speed and the actual speed as:

$$v_i^p = (1 - p_i)O(x_i, y_i) + p_i O_{ave}(x_i, y_i). \tag{3}$$

$$v_i = O_{ave}(x_i, y_i). \tag{4}$$

where p_i is defined as panic weight parameter, for normal scenario $p_i \rightarrow 0$; however, when the herding behavior occurs $p_i \rightarrow 1$. $O(x_i, y_i)$, $O_{ave}(x_i, y_i)$ represent optical flow and the spatio-temporal smooth optical flow in the coordinate (x_i, y_i) respectively. Therefore, we can obtain the interaction force in coordinate (x_i, y_i).

$$F_{int} = m_i \frac{dO_{ave}(x_i, y_i)}{dt} - \frac{1 - p_i}{\tau}(O(x_i, y_i) - O_{ave}(x_i, y_i)). \tag{5}$$

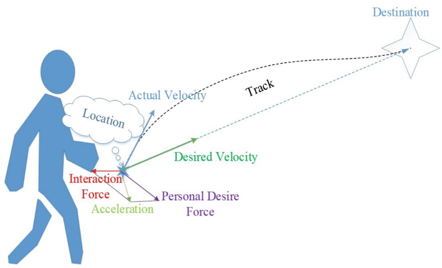

Fig. 2. A graphical representation of social force model

3 Modified Interaction Social Force Model

When solving the problem of anomaly detection using social force model, the key is how to present the particle's desired force. Mehran [9] takes expected speed described as the combination of optical flow and particle advection. However, as we discussed before, since the results of the optical flow field is unstable, a direct subtraction of these two flow field leads local popple. As for desired force, its physical meaning is the force that the individual exerts to achieve the desired speed. We assume that when a person and its surrounding people's movement speed are roughly the same, the force will be small relatively, to be the opposite, the force will shows a large value. We used desired force to measure the effect of surrounding particles put on the current particle movement, also the closer the particles locate, the stronger the force shows. As a result, on the basis of the particle advection field, we use two Gaussian kernel (G_1 and G_2) and the force flow field in convolution processing represent the particle's and its surrounding's motion respectively, so that the interaction force between particles can be expressed as formula (6):

$$F_{int} = (m\frac{dO_{ave}}{dt} - \frac{1-p}{\tau}(O_{ave} * G_1 - O_{ave} * G_2)) \times \varDelta. \tag{6}$$

$$\varDelta = (cos(O_{ave} * G_1, O_{ave} * G_2) + const)^{-1}. \tag{7}$$

The final results are multiplied by a variable $\varDelta$, where we use cosine of two different vector in the same position to measure the consistency of the object and its surroundings and $const = 2$ is added in formula (7) to ensure the denominator is positive and avoid 0.

As shown in Fig. 3, the top image shows social force calculated when a pedestrian is far away from the crowd, and the image below shows the social force calculated when a pedestrian is close to the crowd. With our method, a satisfying visual result can be obtained for retrograde of the pedestrians in the crowd. Besides, if the particles locate in a distance apart from each other, the interaction force would be small, in contrast, the closer the distance, the larger the interaction force could be. The final experiment results confirm our hypothesis.

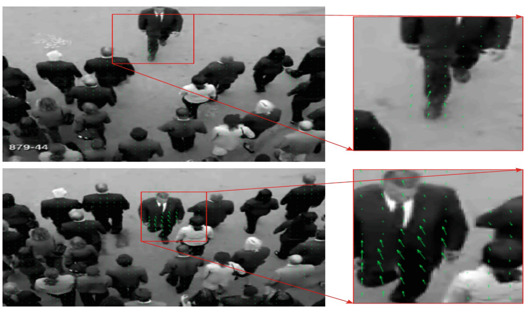

Fig. 3. Visual effectiveness of *social force* (*green arrows*) calculated by modified model (Color figure online)

4 Experiments

As introduced before, we use BOW (Bag of Words) [14] to extract a set of local patches (visual words) as patterns over the field randomly and the codebook is generated through K-means. Then the LDA (Latent Dirichlet Allocation) model [15] is trained by normal samples. Finally, in test phase the outlier samples will be regarded as anomaly.

4.1 Dataset and Experiments Setting

In this work we mainly use UMN and Web datasets (Fig. 4). UMN contains 3 scenes and 11 fragments in total, including indoor and outdoor scenes. Web dataset is a challenging sequence of videos in open environments collected from websites. There are 20 videos including 12 for normal and the rest for anomaly (bullfight, gang fight, etc.). For each video we divide them into clips (10 frames) without overlap. In the phase of particle advection, the number of particles we put over is 60×60 which are much less than the video resolution for faster calculation. We use BOW to exact the visual words through a fixed window with the size of $5 \times 5 \times 10$. Only normal samples are used to train the LDA model, so $p_i = 0$.

Fig. 4. Some samples of *UMN* (*left two columns*) and *Web* (*right two columns*) dataset including *normal* (*left*) and *anomaly* (*right*)

4.2 Experiments on UMN

There are 3 scenes in the UMN, and we train and test each scene respectively. C (codebook) contains 10 words; L (latent topic) is 30 and P (patches) extracted from a clip is 30. Figure 5 shows the result of social force model and optical flow. The method based on optical flow has more omission and miscarriage of justice compared with social force model. Also these two methods have some omissions at the end of each anomaly, as there is no person in the last few frames of each abnormal scene, so there are no motion information extracted from these frames. Social force derives from optical flow but it can capture more motion information and more robust compared with the later.

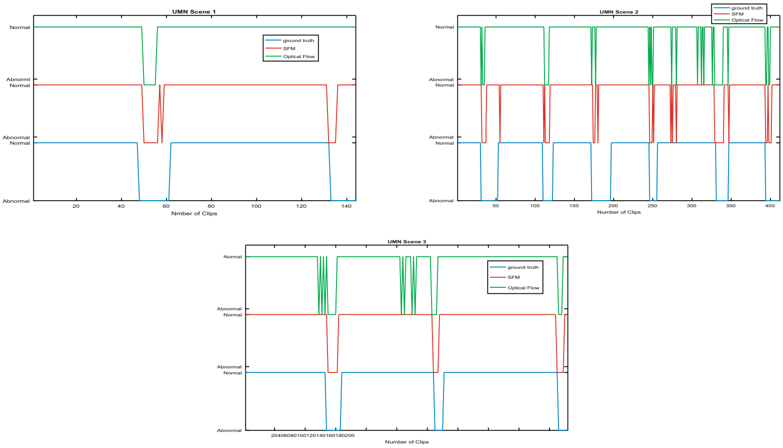

Fig. 5. Results of *SFM* (*red*) and *pure optical flow* (*green*) model (Color figure online)

4.3 Experiments on Web Dataset

Typical evaluating methods such as Accuracy and Precision may lose their efficiency for the number of each class's samples are often skew in anomaly detection problem. So in the following experiments we use the ROC and the AUC to evaluate the relevant methods. At first, we divide the 12 normal videos into 5 groups randomly, and select 4 groups to train the model, then the group left and 8 abnormal videos are chosen as test samples. The 5 groups are used for training and testing in turn and the final result achieved by averaging. Through the Fig. 6 we can find that the AUC is sensitive to the number of patches and the size of codebook, but insensitive to the number of latent topics. As the Fig. 7 (left) shows, with the same parameters setting (C = 50; L = 10; P = 40), social force model is superior to the optical flow and the particle advection based method. Out of our expectation, the particle advection based method give the worst performance. The reason can be got by analyzing the definition of particle advection. It is achieved through smoothing the optical flow field in both spatial and temporal field that it can restrain the local noise of optical flow, but the local

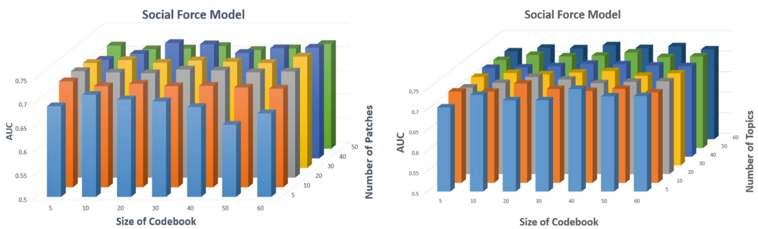

Fig. 6. A comparison of different parameters effecting the AUC

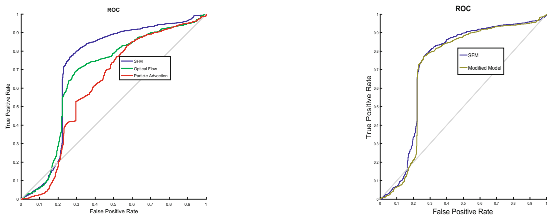

Fig. 7. The ROC curve of *SFM* (*blue*), *optical flow* (*green*), *particle advection* (*red*) and *modified model* (*golden*) (Color figure online)

anomaly is suppressed at the same time. We also apply the social force model and the modified model on the Web dataset, and compare them through AUC and standard deviation in Table 1. According the results, the modified model can achieve equivalent result as original model in average AUC and it is more robust with smaller standard deviation. Also, the reason why the performance doesn't improve is that the modified model is based on particle advection. Through the spatiotemporal smoothing process we get a more robust and successive visual representation, but a set of local information is lost during the procedure that plays an important role in anomaly detection. All of the methods discussed in our experiments can achieve real-time detection using MATLAB with Intel Xeon E5 3.5 GHz CPU and 16 GB RAM.

Table 1. The comparison of 4 methods average AUC on Web dataset and the standard deviation of SFM and the modified model

Model	AUC	Standard deviation
SFM	0.73	0.0196
Modified model	0.71	0.009
Optical flow	0.68	–
Particle advection	0.62	–

5 Discussion and Conclusion

The experiment results on different datasets confirm the superiority of social force model that, it is capable of capturing the interaction between the individual and the surrounding environment without tracking and segmentation, and can effectively deal with the crowded scene. Meanwhile, with its modification, the proposed method could obtain corresponding anomaly detection result in particular dataset, and providing better visual effect.

Our approach is based on optical flow information, there is no doubt that, only movement can produce optical flow. In the process of actual experiments, we found that when the movement speed of crowd is slow or even the moving is stop, our method cannot be used to extract optical flow then to estimate social force in the subsequent process. We will hereafter focus on combining video motion information with static information of video frames, to improve the description of motion pattern.

Acknowledgments. This work is supported by the National Key R&D Program of China under Grant 2017YFB0802300, the National Natural Science Foundation of China 61601223, Natural Science Foundation of Jiangsu Province BK20150756, and Post-doctoral Science Foundation of China 2015M580427.

References

1. Li, T., Chang, H., Wang, M., Ni, B., Hong, R., Yan, S.: Crowded scene analysis: a survey. IEEE Trans. CSVT **25**(3), 367–386 (2015)
2. Solmaz, B., Moore, B.E., Shah, M.: Identifying behaviors in crowd scenes using stability analysis for dynamical systems. IEEE Trans. PAMI **34**(10), 2064–2070 (2012)
3. Hassner, T., Itcher, Y., Klipergross, O.: Violent flows: real-time detection of violent crowd behavior. In: IEEE Conference on CVPR, pp. 1–6 (2012)
4. Mahadevan, V., Li, W., Bhalodia, V., et al.: Anomaly detection in crowded scenes. In: IEEE Conference on CVPR, pp. 1975–1981 (2010)
5. Shao, J., Kang, K., Loy, C.C., Wang, X.: Deeply learned attributes for crowded scene understanding. In: IEEE Conference on CVPR, pp. 4657–4666 (2015)
6. Shao, J., Loy, C., Kang, K., Wang, X.: Slicing convolutional neural network for crowd video understanding. In: IEEE Conference on CVPR, pp. 5620–5628 (2016)
7. Mohammadi, S., Perina, A., Kiani, H., Murino, V.: Angry crowds: detecting violent events in videos. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 3–18. Springer, Cham (2016). doi:10.1007/978-3-319-46478-7_1
8. Helbing, D., Molnár, P.: Social force model for pedestrian dynamics. Phys. Rev. E **51**, 4282–4286 (1995)
9. Mehran, R., Oyama, A., Shah, M.: Abnormal crowd behavior detection using social force model. In: IEEE Conference on CVPR, pp. 935–942 (2009)
10. A website of UMN. http://mha.cs.umn.edu/movies/crowdactivity-all.avi
11. Hu, M., Ali, S., Shah, M.: Learning motion patterns in crowded scenes using motion flow field. In: ICPR, pp. 1–5 (2008)
12. Yang, Y., Liu, J., Shah, M.: Video scene understanding using multi-scale analysis. In: ICCV, pp. 1669–1676 (2009)
13. Ali, S., Shah, M.: A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In: IEEE Conference on CVPR, pp. 1–6 (2007)
14. Feifei, L., Perona, P.: A Bayesian hierarchical model for learning natural scene categories. In: IEEE Conference on CVPR, pp. 524–531 (2005)
15. Blei, D., Ng, A., Jordan, M.: Latent dirichlet allocation. JMLR **3**, 993–1022 (2003)
16. Chen, W., Zhang, K., Liu, Q., et al.: Robust visual tracking via patch based kernel correlation filters with adaptive multiple feature ensemble. Neurocomputing **214**, 607–617 (2016)
17. Zhang, B., et al.: Adaptive local movement modeling for robust object tracking. IEEE Trans. CSVT **27**(7), 1515–1526 (2016)
18. Zhang, B., et al.: Bounding multiple gaussians uncertainty with application to object tracking. IJCV **118**(3), 364–379 (2016)