Asia-Pacific Journal of Science and Technology: Volume: 22. Issue: 03. Article ID.: APST-22-03-07. Research Article

APST

Asia-Pacific Journal of Science and Technology

https://www.tci-thaijo.org/index.php/APST/index

Published by the Faculty of Engineering, Khon Kaen University, Thailand

Anomaly detection in a crowd scene using the interaction force model

Chonthisa Wateosot[1, *], Nikom Suvonvorn[1]

[1] Department of Computer Engineering, Faculty of Engineering, Prince of Songkla University
*Corresponding author: chonthisa.w@gmail.com

Abstract

Anomaly detection in crowded scenes is an important issue in the field of computer vision. Many researches have studied and tried to describe crowd behavior. In this paper, we introduce a novel social-based method for detecting abnormal events in crowded scenes called Interaction Energy Force. The method is designed for low-level features without object extraction and tracking. The force modeling is based on optical flow fields, and its interactions are defined by an energy force inspiring energy propagation phenomena that depend on directions and velocities. An energy map is designed to represent the interaction forces corresponding to events, where the abnormal events are detected using a thresholding method. Our method was evaluated using the well-known UMN dataset. The results show the efficiency of our approach with high accuracy in various conditions. It is a technique that is competitive with the state-of-the-art methods currently employed.

Keywords: computer vision, optical flow, energy force

1. Introduction

In dense crowd situations, such as walking streets, concert halls, markets, ritual places and etc., interactions among people in a large group may increase the probability of abnormal events. These events may suddenly disturb the convergence or divergence of the crowd's flow, which may result in uncontrollable security issues. A feasible solution to this problem is to use automated visual surveillance systems for monitoring the entire scene via video camera and analyzing any abnormal events in the crowd. This involves the automatic detection and localization of abnormal events within crowds in order to prevent hazardous accidents in real time, which is an extremely challenging problem.

Crowd analysis can be divided in two levels: macroscopic and microscopic [1]. The macroscopic level focuses on the global motion of a group of people and ignores the movement of any single individual. The microscopic level concerns the movement of each person within a crowd.

Analysis of crowded scenes could be based on knowledge derived from both vision techniques and dynamic models of crowds. Feature extraction using computer vision plays an important role in the crowd scene analysis. Motion features are the basis of crowd analysis and can be classified into three types: flow-based features, local spatio-temporal features, and tracklet features. Flow-based features are extracted densely at the pixel level. The optical motion flow fields between consecutive frames are computed pixelwise instantaneously [2], which is a robust method to ascertain the motion of multiple objects. This method is popular for motion detection and segmentation [3-5]. Spatio-temporal features, such as the 3D gradient detectors [6 & 7] and HOF descriptors [8], are generally used to characterize the structure of object movement in the scene. These features are widely used in the analysis of extremely crowded scenes. Tracklet features are defined by the trajectories of objects and are suitable for scenes with low crowd density and high resolution. For example, trajectories are used for recognizing the human action [9] and learning semantic regions [10].

Dynamic crowd modeling is used by many researches to study how/where crowds form and if they exceed a critical level. This technique is also used in the modeling of crowd behavior. Dynamic crowd modeling can be divided in two major categories: continuum-based approaches and agent-based approaches [11]. The first kind of approach uses physics-inspired models [12] in which the crowd is treated as a physical fluid with particles, and physics principles, such as statistical mechanics, thermodynamics, etc., are applied. However, although these methods are able detect the group behavior, the movements of each individual in the crowd is ignored. Agent-based approaches, on the other hand, consider the movements of and interactions among individuals in the crowd. A novel model, called the Social Force Model (SFM), has been proposed to define the behavior of crowded scenes by focusing on the interactions among individual people [13]. Recently, many SFM-based methods have been adopted as basic models for crowd behavior analysis [14] & [15]. However, these SFM methods are developed based on physical parameters and do not account for the velocity field of people's interactions. As a result, they cannot detect abnormal events on both the local and global levels simultaneously.

In this paper, we propose a novel social-based method, called Interaction Energy Force (IEF) inspired by energy propagation phenomena, which define interaction force depending on both physical variables and velocity fields of people's interactions. This model is featured from optical flow fields, and its energy map visually represents the interaction forces that are later used for detecting abnormal events.

The rest of this paper is organized as follows: in Section 2, we describe the idea of Energy Force Modeling by mathematical formalization; in Section 3, we explain an overview of abnormal detection in crowded scenes; in Section 4, we illustrate the experiments and results; finally, we conclude this paper in Section 5.

2. Energy Force Modeling

2.1 Optical Flow Estimation

The optical flows are estimated to define the motion of objects between any two consecutive frames, which represents the transition from time *t-1* to *t*. Its quantities, magnitude and direction, can be defined as vectors. Thus, the motion of the object with respect to time can be described as follows:

$$\vec{\mathbf{r}}(t) = \mathbf{x}(t) - \mathbf{x}(t-1) \tag{1}$$

where $\mathbf{x}(t) = \begin{bmatrix} x \\ y \end{bmatrix}_t$ and $\mathbf{x}(t-1) = \begin{bmatrix} x \\ y \end{bmatrix}_{t-1}$

Pedestrians usually walk in a targeted direction $\mathbf{x}(t)$with a certain desired speed $\vec{v}(t)$. The acceleration can, thus, be described by the equation:

$$\vec{a}(t) = \frac{\vec{v}(t) - \vec{v}(t-1)}{\Delta t} \tag{2}$$

2.2 Distance Energy

Distance energy describes the propagation of velocity as energy propagation from its origin with regard to distance. The closer to the origin, the more energy is gained, and the farther from the origin the less energy is gained. For example, if people were to walk into a scene, their distance energy would be low. By contrast, if they were to run into the scene, their distance energy would be high. The concept is implemented into a model as expressed by the equation below:

$$\mathbf{V}(\|\vec{\mathbf{r}}(t)\|) = k_1 \frac{1}{\sqrt{2\pi}} e^{-\frac{(\|\vec{\mathbf{r}}(t)\|)^2}{2\sigma}} \tag{3}$$

where $\|\vec{\mathbf{r}}(t)\| = \sqrt{(x_t - x_{t-1})^2 + (y_t + y_{t-1})^2}$, $\mathbf{V}(\|\vec{\mathbf{r}}(t)\|)$ is the distance energy with respect to $\vec{\mathbf{r}}(t)$, and k_1 is constant value. The σ value is estimated at double the size of the pedestrian in the scene at $\mathbf{V}(\|\vec{\mathbf{r}}(t)\|) = 0.1$. The distance energy forms a Gaussian distribution, in which the maximum value is one.

2.3 Spreading Energy

Spreading energy is defined as the propagation of velocity from its origin into the surroundings regarding the principal direction of the movement. Our model is inspired by the propagation of energy from its source. The spreading direction of energy's ray depends on its velocity. Lower speeds of movement influence a wider angle of energy propagation, conversely higher speeds produce a closer angle. Spreading energy is defined as:

$$\beta(v(t)) = k_2 e^{-k_3 v(t)} \tag{4}$$

where $\beta(v(t))$ is the angle of spreading energy ranging from 0 to 180 degrees, k_2, k_3 are constant values, and $v(t)$ is the velocity of movement.

The propagation of energy will be spread out uniformly over the direction of pedestrian flow, depending on the degree of angle and velocity, which is defined as the Gaussian distribution shown in Figure 1.

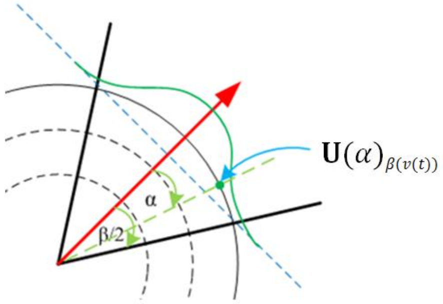


Figure 1. The energy distribution of a flow.

The energy distribution of flow is defined as follows:

$$\mathbf{U}(\alpha)_{\beta(v(t))} = k_4 \frac{1}{\sqrt{2\pi}} e^{-\frac{(r\cos\frac{\beta}{2}\tan\alpha)^2}{2r\sin\frac{\beta}{2}}} \tag{5}$$

where $\mathbf{U}(\alpha)_{\beta(v(t))}$ is the energy distribution of flow with respect to $\beta(v(t))$ at angle $\alpha = [0 \ldots \beta]$, and k_4 is constant value.

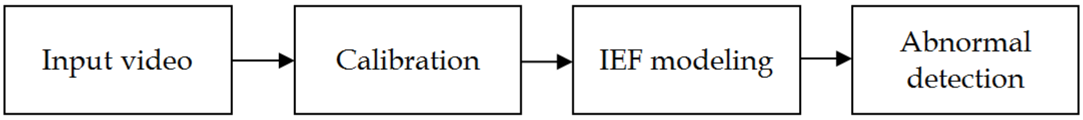


Figure 2. Overview of anomaly detection in crowded scenes.

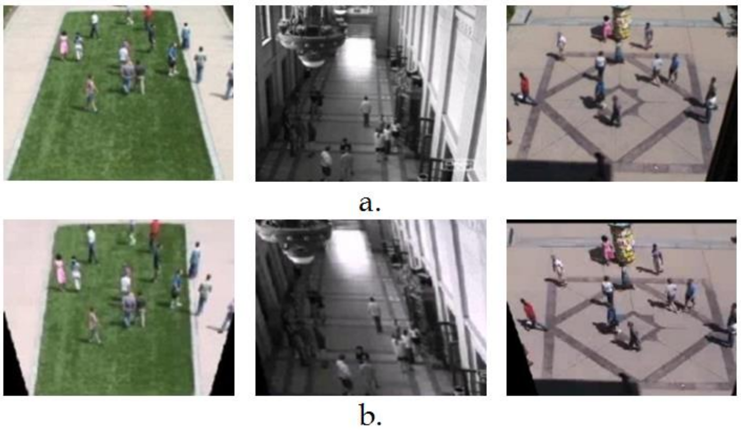


Figure 3. Perspective transformation: a) the original scene, b) the scene after using the perspective transformation

2.4 Interaction Energy Force

In the previous sections, the distance and spreading energies are defined for an optical flow in the field. Each optical flow will have a total energy called Interaction Energy Force (IEF), as defined in the following equation:

$$\mathbf{F} = \mathbf{V}(\|\vec{\mathbf{r}}(t)\|)\mathbf{U}(\alpha)_{\beta(v(t))} \tag{6}$$

A pedestrian in the field could be represented by more than one flow, and our model must be able to unify the energy of these flows. On the other hand, the optical flow from different pedestrians must be distinctive and able to characterize possible conflict or abnormal events. We define the combination of the Interaction Energy Force (IEF) at a certain point in the scene by the following equation:

$$\mathbf{T} = \sum_{i=1}^{N} \mathbf{F}_i \tag{7}$$

where N is the number of optical flows.

In our abnormal detection, the optical flow will be estimated in a grid-based manner. Then, the Interaction Energy Force (**T**) will be computed at every grid.

3. Abnormal Detection using IEF Model

In this section, we introduce an abnormal detection technique using the IEF model described above. The system is summarized as consisting of three main steps: calibration, IEF modeling, and abnormal detection, as shown in Figure 2.

The perspective distortion due to the angle of the camera has a significant effect on the images captured, which eventually influence the IEF modeling and the abnormal detection process. We fix this problem by using the homography technique (Figure 3), which is able to reconstruct the images using perspective distortion. The 2D homography as homographic transformation is defined by a 3×3 homogeneous matrix (**H**) that maps any points $p(x, y)$ on plane π to their corresponding points $p'(x', y')$ on π'as follows:

$$p' = \mathbf{H} \cdot p \tag{8}$$

$$\begin{pmatrix} wx' \\ wy' \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \tag{9}$$

After applying the distortion removal to each frame, we used the IEF model to represent the interaction between people in the crowded scene. For each frame, the motion information was estimated via optical flow field using grid-based particle advection. The optical flows are necessary to calculate the energy force for every grid of the frame. The size of the grid is an important factor that needs to be chosen carefully in order to reduce time-consumption and increase the accuracy of the system. We intuitively applied the Shannon sampling theory [17] to estimate the maximum of sample rate at which motion information can be extracted with low error probability. Then, the number of optical flows were extracted and limited to twice of channel bandwidth, assumed to be twice the size of the relevant moving object. If the grid size is greater than the size of people being tracked, motion information will be lost. When the energy force of each frame is ready, abnormal detection can proceed. In this study, the decision method used for detecting abnormal events is based on a thresholding technique in order to emphasize on the capacity of the IEF model. Thus, a frame is detected as containing an abnormal event only if its maximum IEF energy is greater than the threshold. In the next section the optimal thresholds are empirically estimated based on experiments.

4. Experimental Results

Our method was implemented using C++ and open source library for Computer Vision (OpenCV). All experiments were tested on a PC with 4GB RAM and 3.10GHz CPU.

4.1 Simulation of the IEF Model

Figure 4 shows the simulation results of the IEF model relating to the objects in the scene moving at different speeds and directions, computed among consecutive frames. This figure represents the possible directions of movement and depicts the energy field obtained from our method at a specific angle and various speeds, where its velocities are 5, 10, 15, and 20 pixels/frame. A color map of is used to represent the energy values. Blue color represents low energy whereas green, yellow, and red colors each represent consecutively higher energy levels. We can observe that an energy force with low speed can have a wide spread and propagate at a short distance. Conversely, higher speeds will have higher energy that propagates at a narrow angle and long distance.

4.2 Simulation of IEF Model Experimentation with the UMN Dataset

We evaluate the performance of our technique for abnormal event detection in crowds using the UMN dataset, which is well-known and publicly available as a benchmark [16]. This dataset consists of eleven videos taken at three different locations, including indoor and outdoor scenes. The scenario is set up as follows: initially the pedestrians are walking slowly, and then they suddenly panic and begin running in different directions. This dataset contains 7739 frames.

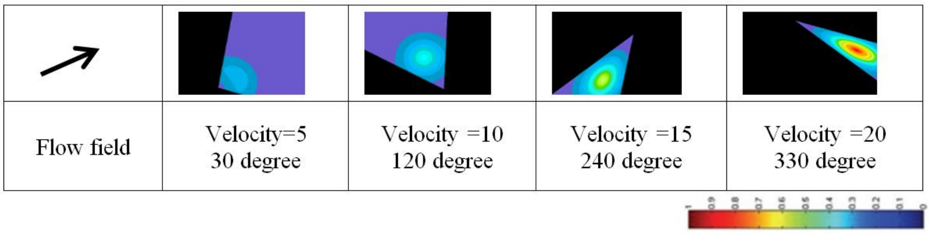


Figure 4. The simulation results of IEF energy model.

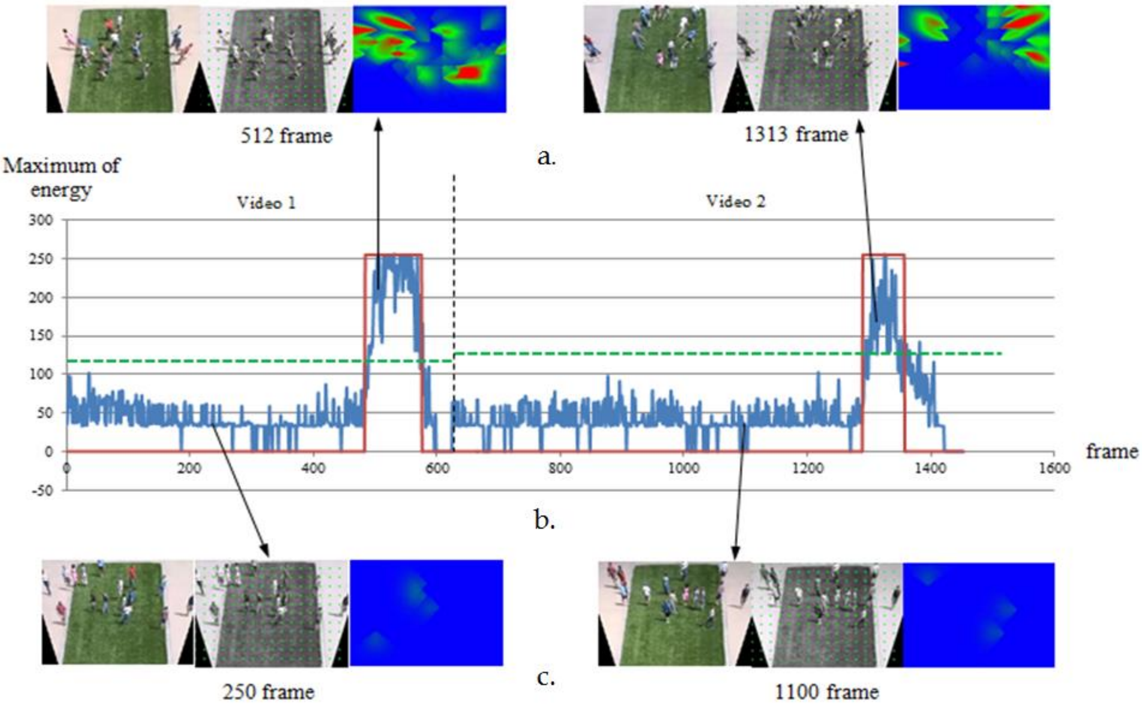


Figure 5. Results of scene 1: a) Examples of abnormal frames, b) Maximum energy, c) Examples of normal frames

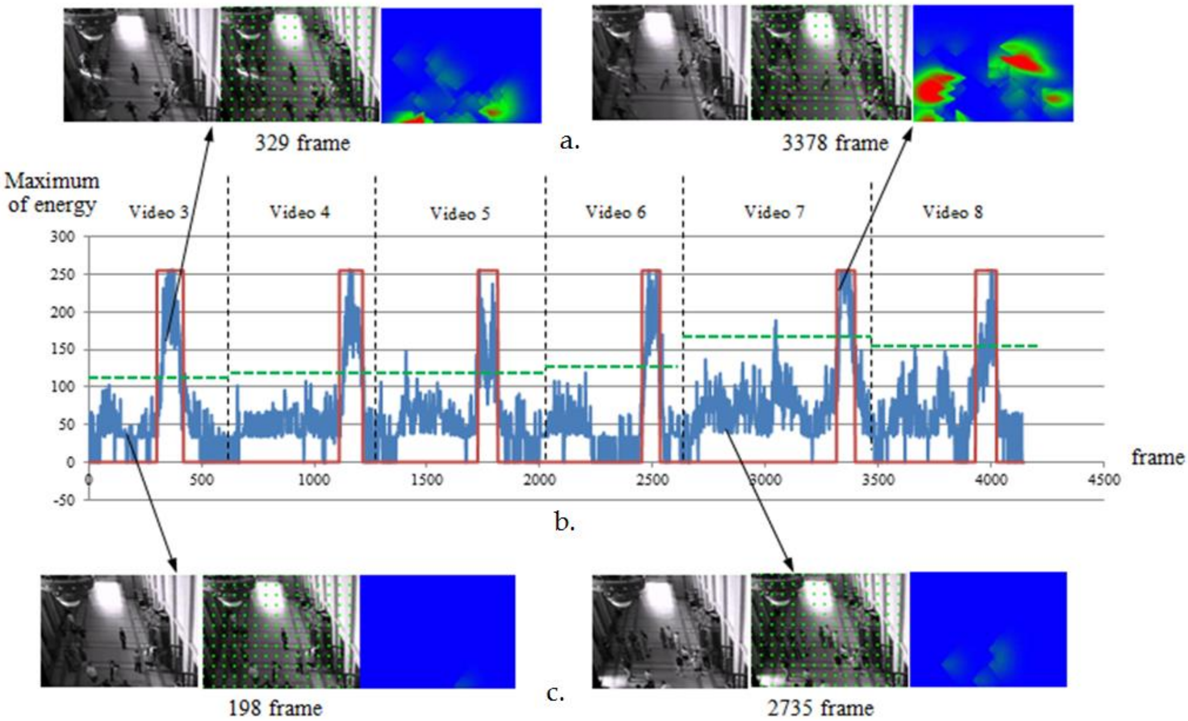


Figure 6. Results of scene 2: a) Examples of abnormal frames, b) Maximum energy, c) Examples of normal frames

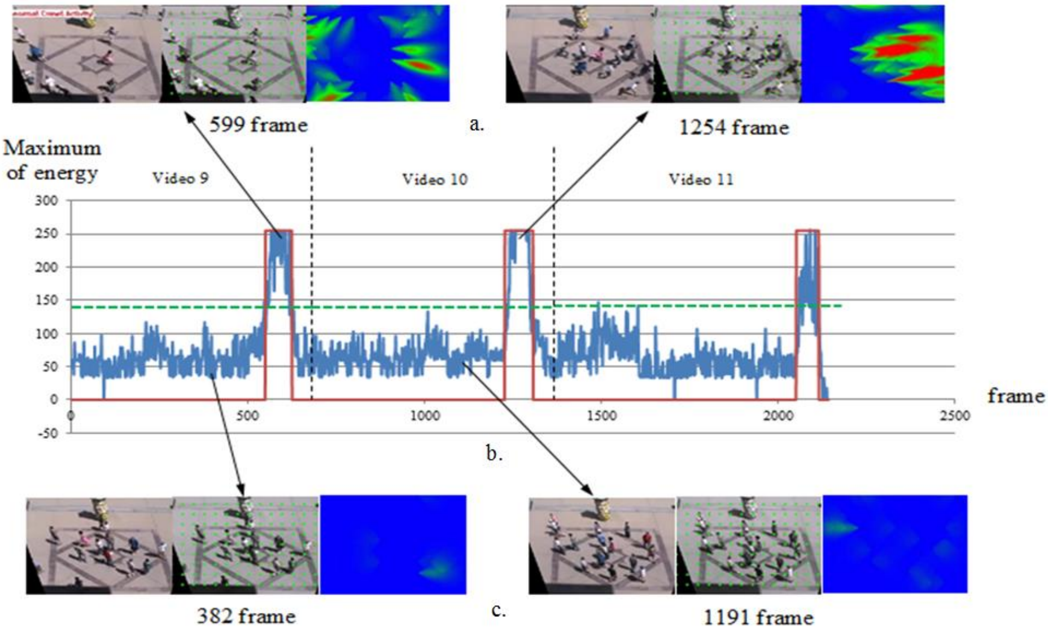


Figure 7. Results of scene 3: a) Examples of abnormal frames, b) Maximum energy, c) Examples of normal frames

In our experiments, we used the optimal constant values of k_1 to k_4 as follows: 2.5, 180, 0.25, and 2.5. Figures 5a, 6a, and 7a show the experimental results of abnormal events with their energies. Figures 5c, 6c, and 7c depict examples of frames and interaction energy when there are no abnormal events. In Figures 5b, 6b, and 7b, the blue curve shows the maximum energy of every frame, representing the behavior of peoples in the scene. The values of maximum energy are normalized, ranging from 0 to 255. The red curve is the ground truth representing abnormal events (at energy level 255) and normal events (at energy level 0). The optimal threshold value to classify the frame as containing either a normal or abnormal event is represented by the green line. For example, in Figure 5 (videos 1 and 2 of scene one), the 250th and 1100th frames are normal events, where people walk slowly and have very few interactions among them, which means their maximum energies are low. By contrast, in a panic event, the speeds and interaction forces among people in the crowd will be higher than in a normal event. In this manner, the 512th and 1313th frames were detected as panic events.

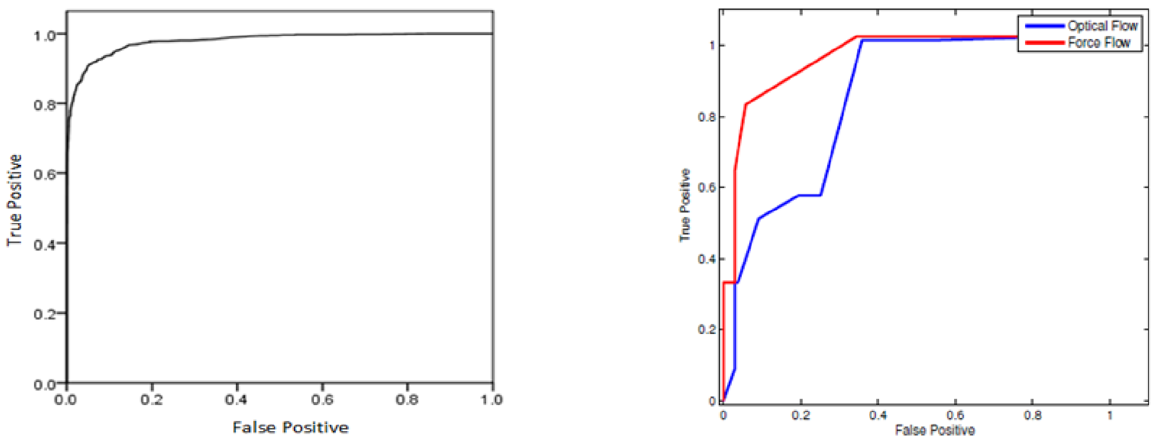
We used the standard measurement Receiver Operating Characteristic (ROC) for evaluating our abnormal event detection technique in experiments. The ROC curve was computed at a frame-level measurement in the three scenes in the UMN dataset. Table 1 shows the experimental results in terms of ROC curve based on energy force.

Table 1. The comparison of our technique with the state-of-the-art methods for anomaly detection in the UMN Dataset.

Method	Area under ROC
Optical Flow [14]	0.84
Social Force [14]	0.96
Proposed method in scene 1	
• Video 1	0.985
• Video 2	0.976
Proposed method in scene 2	
• Video 3	0.97
• Video 4	0.971
• Video 5	0.962
• Video 6	0.978
• Video 7	0.969
• Video 8	0.96
Proposed method in scene 3	
• Video 9	0.986
• Video 10	0.976
• Video 11	0.98
Proposed method in all scenes	0.974

We noticed that the performance of the proposed method was better on scenes 1 and 3 than on scene 2. This is likely because scene 1 and 3 are outdoor scenarios with crowds of pedestrians whose movements are mostly localized and clear. On average, the accuracy of this method in both scenes was about 0.98. The highest levels of accuracy were in video 1 of scene 1, and video 9 of scene 3, at 0.985 and 0.986 respectively. We found that the inaccurate results in some videos of scene 1 and scene 3 were due to the shadows and slow velocities at the end of the video, where the ground truth is always defined as abnormal events. The overall accuracy of our proposed method is 0.974 (the ROC is illustrated in Fig. 8), which is slightly higher than the conventional methods (0.84 for the optical flow and 0.96 for the social force). With scene 2, our method was almost 1.3% less accurate than on scenes 1 and 3, at a 0.968 in average, but still boasted a 12% improvement over optical flow, and was comparable to the social force method. The lower accuracy on scene 2 was caused by some aspects of the indoor setting, such as shadow, low contrast, changing illumination due to door opening, cloth color (with regard to the background where the motion detection is erroneous), and unexpected individual movements.

Finally, we found that the accuracy of our proposed method depends on grid sizes and threshold values. Experimentally, the appropriate grid size is fixed proportionally to the size of pedestrians in the scene. Consequently, the suitable grid sizes for scene 1, scene 2, and scene 3 are 20, 16, and 20 pixels, respectively. We defined three different threshold values for the three scenes by practices in order to obtain good results. The environment of the scene affected the selection of parameters.



a. b.

Figure 8. The ROC for abnormal detection in the UMN dataset: a) our method, b) the state-of-the-art method [14]

5. Conclusion

We proposed a new social-based method, called an IEF model, to be used for anomaly detection in crowded scenes. Our technique is inspired by energy propagation phenomena. The model of interaction force and its characteristics are thoroughly described. We use a thresholding method for anomaly detection in order prevent bias in the efficiency of our proposed model. We tested our method on a UMN dataset, by which abnormal events can be detected at a high accuracy of around 0.98 on average, regardless of conditions such as occlusion or interaction among pedestrians. Our technique is competitive with the state-of-the-art methods.

6. Acknowledgements

The authors would like to thank the Office of the Higher Education Commission, Thailand, for supporting this work under the Strategic Scholarships Fellowships Frontier Research Networks (specific for Thailand's southern region) for the Thai Ph.D. degree program. This work was also partially supported by Graduate school thesis grant from Prince of Songkla University.

7. References

[1] Li, T., Chang, H., Wang, M., Ni, B., Hong, R., Yan, S., 2015. Crowded scene analysis: a survey. IEEE Transactions on Circuits and Systems for Video Technology 25, 367-386.
[2] Brox, T., Bruhn, A., Papenberg, N., Weickert, J., 2004. High accuracy optical flow estimation based on a theory for warping. Proc European Conference on Computer Vision 25-36.
[3] Jodoin, P.-M., Benezeth, Y., Wang, Y., 2013. Meta-Tracking for video scene understanding. Proc Advanced Video and Signal Based Surveillance IEEE 1-6.

[4] Su, H., Yang, H., Zheng, S., Fan, Y., Wei, S., 2013. The large-scale crowd behavior perception based on spatio-temporal viscous fluid field. IEEE Transactions on Information Forensics and security 8, 1575-1589.
[5] Benabbas, Y., Ihaddadence, N., Djeraba, C., 2011. Motion pattern extraction and event detection for automatic visual surveillance. Journal on Image and Video Processing, 1-15.
[6] Kratz, L., Nishino, K., 2012. Tracking pedestrians using local spatio-temporal motion patterns in extremely crowded scenes. IEEE Transactions on Pattern Analysis and Machine Intelligence 34, 987-1002.
[7] Kratz, L., Nishino, K., 2009. Anomaly detection in extremely crowded scenes using spatio-temporal motion pattern models. Proc Computer Vision and Pattern Recognition IEEE., 1446-1453.
[8] Cong, Y., Yuan, J., Liu, J., 2013. Abnormal event detection in crowded scenes using sparse representation. Pattern Recognition 46, 1851-1864.
[9] Bak, S., Chau, D.-P., Badie, J., Corvee, E., 2012. Multi-target tracking by discriminative analysis on Riemannian manifold. Proc Image Processing IEEE., 1605-1608.
[10] Chongjing, W., Xu, Z., Yi, Z., Yuncai, L., 2013. Analyzing motion patterns in crowded scene via automatic tracklets clustering. China Communications 10, 144-154.
[11] Still, G.K., 2000. Crowd dynamics. [Ph.D. dissertation]. London: Univ. Warwick. UK.
[12] Hughes, R.L., 2002. A continuum theory for the flow of pedestrians. Transportation Research Part B: Methodology. 36, 507-535.
[13] Helbing, D., Molnar, P., 1995. Social force model for pedestrian dynamics. Physical Review 51, 4282-4286.
[14] Mehran, R., Oyama, A., Shah, M., 2009. Abnormal crowd behavior detection using social force model. Proc Computer Vision and Pattern Recognition IEEE., 935-942.
[15] Raghavendra, R., Bue, A.D., Cristani, M., Murino, V., 2011. Abnormal crowd behavior detection by social force optimization. Proc Human Behavior Understand., 134-145.
[16] Unusual Crowd activity dataset of University of Minnesota., 2006. [WWW Document]. URL http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi (accessed 1. 7. 14).
[17] Bell, D.A., 1962. Information Theory and its Engineering Applications (3rd ed.). New York: Pitman.