Abnormal Detection Using Interaction Energy Potentials

Xinyi Cui, Qingshan Liu, Mingchen Gao, Dimitris N. Metaxas
Department of Computer Science, Rutgers University, Piscataway, NJ, USA
{xycui, qsliu, minggao, dnm}@cs.rutgers.edu

Abstract

A new method is proposed to detect abnormal behaviors in human group activities. This approach effectively models group activities based on social behavior analysis. Different from previous work that uses independent local features, our method explores the relationships between the current behavior state of a subject and its actions. An interaction energy potential function is proposed to represent the current behavior state of a subject, and velocity is used as its actions. Our method does not depend on human detection or segmentation, so it is robust to detection errors. Instead, tracked spatio-temporal interest points are able to provide a good estimation of modeling group interaction. SVM is used to find abnormal events. We evaluate our algorithm in two datasets: UMN and BEHAVE. Experimental results show its promising performance against the state-of-art methods.

1. Introduction

Abnormal event detection plays an important role in video surveillance and smart camera systems. Various abnormal activities have been studied, including restricted-area access detection [9], car counting [6], detection of people carrying cases [7], abandoned objects [22], group activity detection [31, 17], social network modeling[29], monitoring vehicles [28], scene analysis [24] and so on. In this paper, we focus on modeling abnormal events in human group activities, which is a very important application for video surveillance. Fig.1 shows two sample frames. (a) shows a group of people fighting in the street, and (b) shows people running away from the scenes.

We propose a new method to detect abnormal events in group activities. We represent group activities by learning relationships between the current behavior state of a subject and its actions. Our goal is to explore the reasons why people take different actions under different situations.

In the real world, people are driven by their goals. They take into account of the environment as well as the influence of other people. We define an interaction energy potential function to represent the current state of a subject based on the positions/velocities of a subject itself as well as its neighbors. Fig.2 shows an example of interaction energy potentials and velocities. Section 2 gives the details of the definition. Social behaviors are captured by the relationship between interaction energy potential and its action, which is then used to describe social behaviors. Uncommon Energy-Action patterns indicate an abnormal activity. Experiments on two datasets UMN [1] and BEHAVE [2] show that our method is powerful to model abnormal behaviors in group activities.

Our contributions. 1) The Interaction Energy Potential is proposed to model the relationship among a group of people. 2) The relationship between the current state of a subject and the corresponding reaction is explored to model the normal/abnormal patterns. 3) Our method does not rely on human detection or segmentation technique, so it is more robust to the errors that are introduced by detectoin/segmentation techniques.

Figure 1. Abnormal event examples. (a) a group of people fighting; (b) People are panic, trying to run away from the scene.

1.1. Related Work

Human action/activity modeling in video sequences is a hot topic in the communities of computer vision and pattern recognition. In recent years, a lot of algorithms have

(a) (b)

Figure 2. Interaction energy potentials of two sample frames. Green arrow is the velocity; round dot denotes energy values. Red dot shows a low energy value and blue shows a high value.

been proposed to improve interest point detection [16], local spatio-temporal descriptors [10], building relationships among local features [23, 15]. Most of the algorithms focus on single action with one person [20](*hand-waving, running...*) or pair-wise action recognition (*answer phone*[10], *horse riding* [11]). These works do not consider interactions among multiple people. For most of the surveillance systems in public area, it is also important to identify group activities. Events like fighting or escaping often involve multiple people and their interactions.

Several algorithms for group activity modeling have been proposed in recent years. Different features are used for group activity: human body/body parts [13, 19], optical flow [4] and detecting moving regions [26]. Recently, Zhou *et al*.[31] and Ni *et al*. *et al*. [17] use trajectory analysis to describe different group activities.

Modeling social behaviors of people is an important branch to represent group activity, and it has been widely used in evacuation dynamics, traffic analysis and graphics. Pedestrian behaviors have been studied from a crowd perspective, with macroscopic models for crowd density and velocity. On the other end, microscopic models deal with individual pedestrians. A popular model is the Social Force Model [8]. In the Social Force Model, pedestrians react to energy potentials caused by other pedestrians and static obstacles through a repulsive force, while trying to keep a desired speed and motion direction. Helbing *et al*. in [8] originally introduce it to investigate people movement dynamics. It is also applied to the simulation of crowd behavior [30], virtual reality and studies in computer graphics for creating realistic animations of the crowd [25].

Social behavior analysis has also attracted much attention in the computer vision community. Ali and Shah [3] use the cellular automaton model to track in extremely crowded situations. Antonini *et al*. [5] propose a variant of Discrete Choice Model to build a probability distribution over pedestrian positions in next time step. Scovanner and Tappen [21] learns pedestrians' dynamics and motions as a continuous optimization problem. Pellegrini *et al*. [18] propose a Linear Trajectory Avoidance (LTA) method to track multiple targets. Predictions of velocities are computed by the minimization of energy potentials. Recently, Mehran *et al*. [14] propose a method to model behaviors among a group of people. It represents the abnormal patterns in a local region based on moving particles. Wu *et al*. [27] uses chaotic invariants of Lagrangian Particle Trajectories to model abnormal patterns in crowded scenes. They have been successfully used in crowded scene modeling.

Different from the above work, our method is based on the relationship between the current state of a person and his/her reactions. It fully utilizes the information of interaction energy potential and the corresponding people's reactions, which contains comprehensive information to model abnormal behavior among a group of people.

2. Methodology

We propose a new method to model group interactions for abnormal detection using interaction energy potentials. The framework is summarized in Fig.3. We first extract spatio-temporal interest points and track them. Second, an interaction energy potential is calculated for each point. Third, features are represented by relationships among interaction energy potentials and corresponding actions with a coding scheme. Finally, SVM is used to detect abnormal events.

2.1. Interest Point Detection and Tracking

The ideal case for human activity analysis is to track all the subjects and estimate their positions and velocities, but human detection and tracking is still a challenging problem. Instead we use local interest points to represent subjects in a scene. The movements of subjects can be represented by the movements of interests points associated with the subjects, and interactions among the subjects can be implicitly embodied in the interactions among interest points. We use the method proposed in [10] to detect the local spatio-temporal interest points (STIP), and then use the KLT tracker [12, 15] to track interest points. Fig.2 shows an example of interest point detection and tracking.

For each tracked interest point p_i, we record its positions $\{\mathbf{x}_i^0 ... \mathbf{x}_i^t ...\}$, where each $\mathbf{x}^t$ is a 2D vector, and its velocity $\mathbf{v}_i^t$ at time t is calculated by

$$\mathbf{v}_i^t = \frac{\mathbf{x}_i^{t+T} - \mathbf{x}_i^t}{T} \tag{1}$$

where T is the time interval. A point p_i is then modeled as $p_i = (\mathbf{x}_i^t, \mathbf{v}_i^t)$. Besides the self-representation of velocity, we also take into account of neighbor interest points,

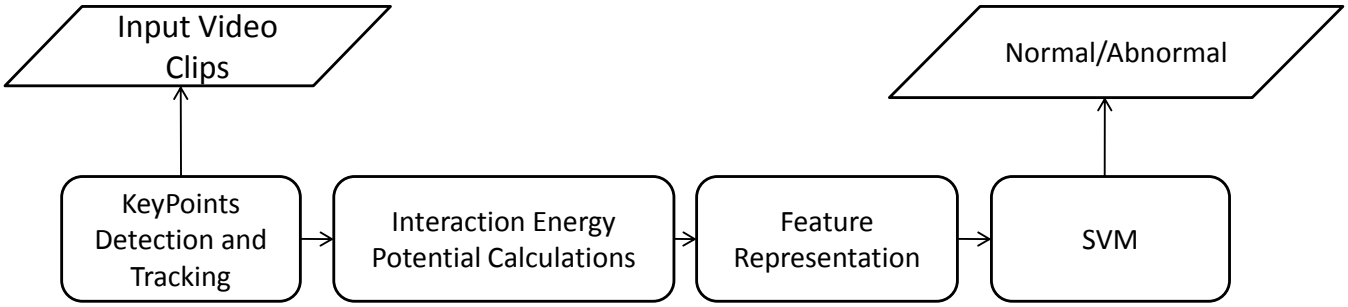

Figure 3. Flow chart: given an input clip, keypoint detection and tracking is performed first, then Interaction Energy Potential is calculated. After wrapping up with feature represention, SVM is used to find abnormal events.

which implicitly represent interactions among subjects in group activities. Interactions among subjects are modeled by interaction energy potentials, which is addressed in the following section.

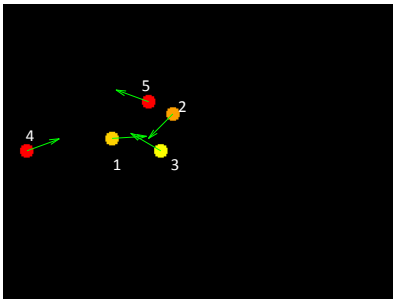

Figure 4. Toy examples. Five subjects, with their current velocities. Color denotes energy values. Red color denotes a low interaction energy potential value, while yellow denotes a high energy value. Taking perspective of subject 1, it has interactions with subject 2 and 3; ignores subject 4 subject and moves away from subject 5.

2.2. Interaction Energy Potentials

Given a set of interest points $S = \{p_i\}$ $(i = 1...n)$, energy potential E_i of p_i is calculated based on positions and velocities of its neighbor points. The calculation of the interaction energy potentials is inspired by the idea of social behaviors[18]: assuming that people are aware of the positions and velocities of other people at time t. Thus we can make a reasonable assumption that people can predict the movement of other people and have a general estimation about whether they would meet in the near future. This is also how people walk in the real world.

We first consider two subjects. Given two subjects s_i and s_j in a scene, we are now thinking from the perspective of s_i, and treating s_j as its neighbor. We define the current time as $t = 0$ and use $\mathbf{x}_i = \mathbf{x}_i^0$ for simplicity. If s_i proceeds with velocity $\mathbf{v}_i$, then it expects to have a distance $d_{ij}^2(t)$ from s_j at time t.

$$d_{ij}^2(t) = ||\mathbf{x}_i + t\mathbf{v}_i - (\mathbf{x}_j + t\mathbf{v}_j)||^2 \tag{2}$$

Minimal distance d_{ij} occurs at the time of closest point t^*, where

$$t^* = \max\{0, \arg\min d_{ij}^2(t)\} \tag{3}$$

where $\arg\min d_{ij}^2(t)$ can be obtained by setting the derivative of d_{ij} to zero with respect to time t. Then we obtain t^* as follows:

$$t^* = \max\{0, -\frac{(\mathbf{x}_i - \mathbf{x}_j) \cdot (\mathbf{v}_i - \mathbf{v}_j)'}{||\mathbf{v}_i - \mathbf{v}_j||^2}\} \tag{4}$$

By substituting t into Eq. 2, we can obtain the minimum distance d_{ij}^{*2} between subjects i and j as $d_{ij}^{*2} = d_{ij}^2(t^*)$.

d_{ij}^{*2} defines how far two subjects will meet based on the current velocities. If d_{ij}^{*2} is smaller than a distance threshold d_c, a close-distance meet would happen in the near future. If p_j has a close-distance from p_i, p_j is very likely to draw p_i's attention at this moment. We therefore build an interaction energy potential function based on their distance

$$E_{ij} = w_{ij}^c w_{ij}^\phi exp(-\frac{d_{ij}^2(t)}{2\sigma_d^2}) \tag{5}$$

$$w_{ij}^c = \begin{cases} 1 & d_{ij}^{*2} < d_c \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

$$w_{ij}^{\phi} = \begin{cases} 1 & \phi_{ij} < \phi_{view} \\ 0 & \text{otherwise} \end{cases} \quad (7)$$

where σ_d is the radius of influence of p_j. The closer of p_j, the higher attention would p_i have. ϕ is the current angular displacement of p_j from the perspective of p_i. ϕ_{view} is the field-of-view, which is the angle displacement between the current moving direction and the neighbor point direction. As people only see things in the front, ϕ_{view} controls how people see things. The Interaction energy potential E_{ij} describes the influence from p_j. E_{ij} is high when they are close, and it is minimal as their distance goes to infinity.

For the case of multiple subjects, the influence of all the other subjects can be modeled as an average of energy potential E_{ik}. The overall interaction energy for subject p_i is given by

$$E_i = \frac{1}{N} \sum_{k \neq i, E_{ik} > 0} E_{ik} \quad (8)$$

where N is the number of non-zero neighbor points. The Interaction energy potential E_i describes the current behavior state of subject p_i. Fig.4 shows an example of 5 points with their interaction energy potentials. Now we are taking perspective of subject p_1, with 4 neighboring points in the frame. p_1 is moving towards p_2 and p_3. As they are going to meet based on the current velocities, p_2 and p_3 have a high influence on p_1. p_4 is in the back, so p_1 does not see it. p_5 moves further away from p_1, so it does not draw p_1's attention at this moment. The total interaction energy potential of p_1 comes from p_2 and p_3. Next we take perspective of p_5. It moves away from all the other points, so its neighbors would not influence it at this time. It results in a low interaction energy value for p_5. The Energy potential is calculated for each subject from its own view. Then energy values are denoted by color in the figure. Yellow dot in Fig.4 shows a high energy value, while red dot shows a low energy value. In our method, E is calculated for each point. Fig.2 shows an example of detected points and corresponding interaction energy potentials.

2.3. Features Representation and Classification

The Interaction energy potential reflects the current interaction with the surrounding of a person. Different from [18], our goal is to find reasons why people take actions and what situations make them take actions. This can be modeled by relationships between current states (interaction energy potential E) and actions (velocities v).

Fig.5 shows an example of the relationship between energy changing and velocity changing over time. In Fig.5, (a) is a group of people meeting. Color lines show energy changing through time. We choose one point and its trajectory for analysis. (b) shows its energy changing over time. As people move closer, the energy increases slowly. At the same time, v_m, Δv_d and Δv_m remain stable. They are shown in (c)(d)(e) respectively. This is a common event in the real world. People have their desires to meet, and they try to remain at a constant speed and direction. In contrast, (f) shows a group of people fighting. At time 10, Δv_d changes dramatically (shown in (i)) even with low interaction energy potential (shown in (g)). This indicates an uncommon pattern. A point changes its moving direction dramatically without an obvious reason.

Each local patch around the salient points is represented by Interaction energy potentials and optical flows. Then standard bag-of-words method is used. Each video clip is represented by a bag-of-word representation. SVM is used to find the abnormal activities.

3. Experiments

We test our algorithm on two datasets: UMN dataset [1] and BEHAVE dataset [2]. Details are shown in the following sessions.

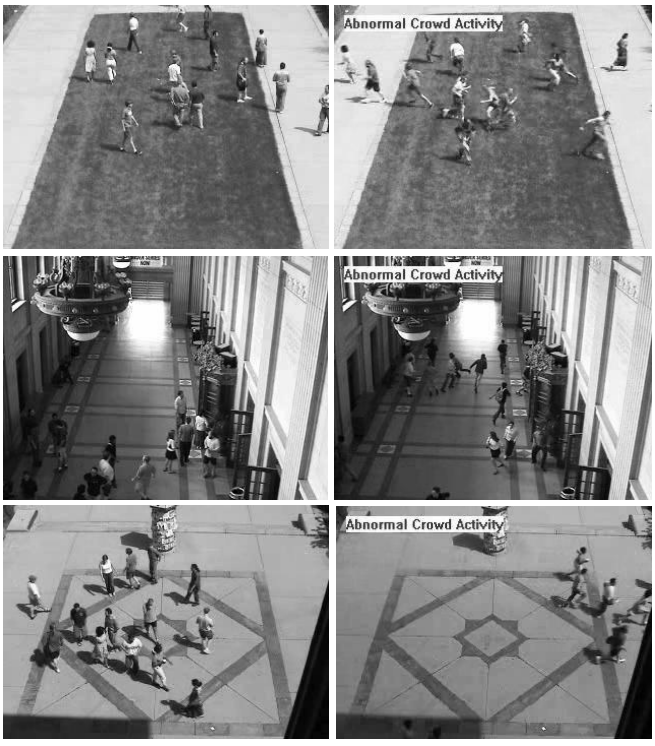

Figure 6. The UMN Dataset. Left column: samples from the normal events; Right column: samples from the abnormal events

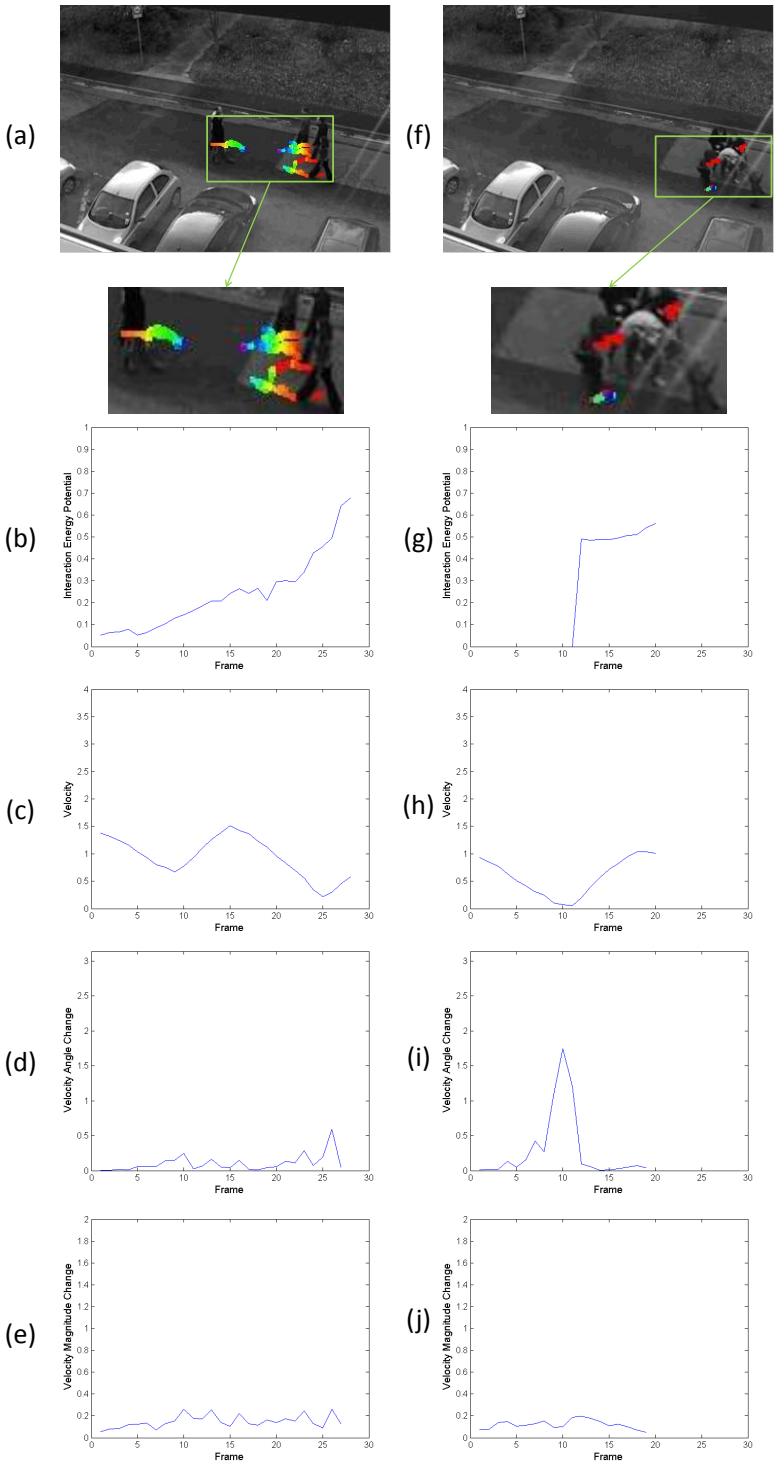

Figure 5. Two events. (a)group meeting event; (b) energy E of meeting; (c) velocity magnitude v_m of meeting; (d) velocity direction changing Δv_d of meeting; (e) velocity magnitude changing Δv_d of meeting; (f)fighting event; (g) energy E of fighting; (h) velocity magnitude v_m of fighting; (i) velocity direction changing Δv_d of fighting; (j) velocity magnitude changing Δv_d of fighting.

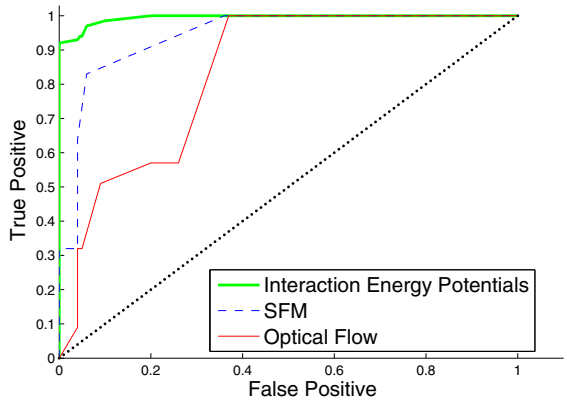

Figure 7. Results of abnormal detection in the UMN Dataset.

3.1. The UMN Dataset

This dataset is collected from University of Minnesota [1], which contains videos of 11 different scenarios of an escape event. The videos are shot in 3 different scenes, including both indoor and outdoor. Each video clip starts with an initial part of normal behaviors and ends with sequences of abnormal behaviors. Fig.6 shows some sample frames. Scenes in this dataset are crowded, with about 20 people walking around. We follow the same setup as in [14].

As in [14], we take optical flow features as the baseline. Fig.7 reports the experimental results. The results of Social Force, Optical Flow are directly obtained from their papers [14, 27]. The results show that our algorithm is competitive with these state-of-art methods.

3.2. The BEHAVE Dataset

To further demonstrate the effectiveness of our method, we conduct experiments on another dataset: the BEHAVE Dataset. We collect an abnormal activity dataset from the BEHAVE dataset [2]. The BEHAVE dataset has many complex group activities, including meeting, splitting up, standing, walking together, ignoring each other, fighting, escaping as well as running. Scenarios contain various number of participants. The dataset consists of 50 clips of fighting events, and 271 normal events. Some samples are shown in Fig.8. All the activities in this dataset are common in the real world. The scene is moderately crowded. In our experiments, the average number of interest points is 43 in each video clip. The length of tracked interest points are 27.81 frames in average.

We compare our method with the optical flow based method and Mehran *et al.*'s method [14]. Fig.9 shows our results comparing to these two methods. It shows that our

Figure 8. The BEHAVE dataset. Top row: samples from the normal events; bottom row: samples from the abnormal events: fighting

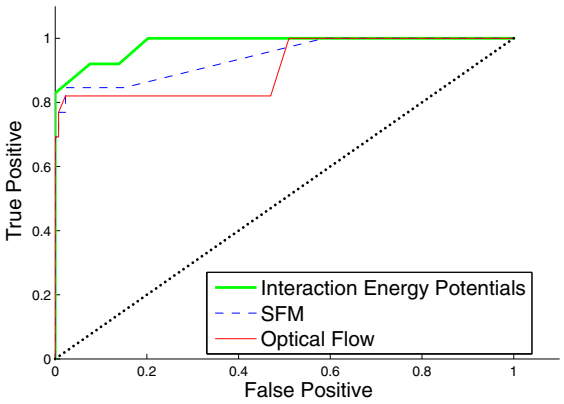

Figure 9. Results on BEHAVE abnormal dataset. Comparison of our method (green line) with Social Force [14] and Optical Flow.

Interaction Energy Potential does a better job to represent abnormal events in such complex group activities. It comes from the fact that our feature does not only consider the velocity distribution, but also utilizes the interaction among a group, which is able to improve the performance.

4. Conclusion

We proposed a method to detect abnormal events in group activities. In our algorithm, we explored the reasons why people take actions and what situations make people take actions. The relationships between the current be-

havior states and actions indicate normal/abnormal patterns. Pedestrians' environment is modeled by an interaction energy potential function. Abnormal activities are indicated in uncommon energy-velocity patterns. Our method does not depend on human detection or tracking algorithm. We conducted the experiments on the UMN dataset and the BEHAVE dataset. Results showed the effectiveness of our method, and it is competitive with the state-of-art methods.

References

[1] Unusual Crowd Activity Dataset:http://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi. 3161, 3164, 3166

[2] BEHAVE: http://homepages.inf.ed.ac.uk/rbf/BEHAVE/. 3161, 3164, 3166

[3] S. Ali and M. Shah. Floor fields for tracking in high density crowd scenes. *Proc. ECCV*, 2008. 3162

[4] E. L. Andrade, S. Blunsden, and R. B. Fisher. Modelling crowd scenes for event detection. *Proc. ICPR*, 2006. 3162

[5] G. Antonini, S. V. Martinez, M. Bierlaire, and J. P. Thiran. Behavioral priors for detection and tracking of pedestrians in video sequences. *Trans. IJCV*, 2006. 3162

[6] N. Friedman and S. Russell. Image segmentation in video sequences: A probabilistic approach. *Proc. UAI*, 1997. 3161

[7] I. Haritaoglu, D. Harwood, and L. S. Davis. W4: Real-time surveillance of people and their activities. *Trans. PAMI*, 2000. 3161

[8] D. Helbing and P. Molnar. Social force model for pedestrian dynamics. *Physical Review E*, 1995. 3162

[9] J. Konrad. Motion detection and estimation. *Handbook of Image and Video Processing, 2nd Edition*, 2005. 3161

[10] I. Laptev, M. Marszałek, C. Schmid, and B. Rozenfeld. Learning realistic human actions from movies. *Proc. CVPR*, 2008. 3162

[11] J. Liu, J. Luo, and M. Shah. Recognizing realistic actions from videos 'in the wild'. *Proc. CVPR*, 2009. 3162

[12] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc. IJCAI*, 1981. 3162

[13] G. Medioni, I. Cohen, F. Brémond, S. Hongeng, and R. Nevatia. Event detection and analysis from video streams. *Trans. PAMI*, 2001. 3162

[14] R. Mehran, A. Oyama, and M. Shah. Abnormal crowd behavior detection using social force model. *Proc. CVPR*, 2009. 3162, 3166

[15] R. Messing, C. Pal, and H. Kautz. Activity recognition using the velocity histories of tracked keypoints. *Proc. ICCV*, 2009. 3162

[16] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool. A comparison of affine region detectors. *Trans. IJCV*, 2005. 3162

[17] B. Ni, S. Yan, and A. Kassim. Recognizing human group activities with localized causalities. *Proc. CVPR*, 2009. 3161, 3162

[18] S. Pellegrini, A. Ess, K. Schindler, and L. V. Gool. You'll never walk alone: Modeling social behavior for multi-target tracking. *Proc. ICCV*, 2009. 3162, 3163, 3164

[19] A. Prati, S. Calderara, and R. Cucchiara. Using circular statistics for trajectory shape analysis. *Proc. CVPR*, 2008. 3162

[20] C. Schuldt, I. Laptev, and B. Caputo. Recognizing human actions: A local svm approach. *Proc. ICPR*, 2004. 3162

[21] P. Scovanner and M. F. Tappen. Learning pedestrian dynamics from the real world. *Proc. ICCV*, 2009. 3162

[22] K. Smith, P. Quelhas, and D. Gatica-Perez. Detecting abandoned luggage items in a public space. *PETS*, 2006. 3161

[23] J. Sun, X. Wu, S. Yan, L. F. Cheong, T.-S. Chua, and J. Li. Hierarchical spatio-temporal context modeling for action recognition. *Proc. CVPR*, 2009. 3162

[24] E. Swears and A. Hoogs. Functional scene element recognition for video scene analysis. *WMVC*, 2009. 3161

[25] A. Treuille, S. Cooper, and Z. Popovic. Continuum crowds. *ACM Trans. Graph*, 2006. 3162

[26] J. Wu, A. Osuntogun, T. Choudhury, M. Philipose, and J. Rehg. A scalable approach to activity recognition based on object use. *Proc. ICCV*, 2007. 3162

[27] S. Wu, B. Moore, and M. Shah. Chaotic invariants of Lagrangian particle trajectories for anomaly detection in crowded scenes. *Proc. CVPR*, 2010. 3162, 3166

[28] Q. Yu and G. Medioni. Motion pattern interpretation and detection for tracking moving vehicles in airborne video. *Proc. CVPR*, 2009. 3161

[29] T. Yu, S. Lim, K. Patwardhan, and N. Krahnstoever. Monitoring, recognizing and discovering social networks. *Proc. CVPR*, 2009. 3161

[30] W. Yu and A. Johansson. Modeling crowd turbulence by many-particle simulations. *Physical Review E*, 2007. 3162

[31] Y. Zhou, S. Yan, and T. Huang. Pair-activity classification by bi-trajectories analysis. *Proc. CVPR*, 2008. 3161, 3162