Proceedings of the 2009 IEEE
International Conference on Robotics and Biomimetics
December 19 -23, 2009, Guilin, China

Abnormal Crowd Motion Analysis

Tian Cao*†, Xinyu Wu*, Jinnian Guo*, Shiqi Yu* and Yangsheng Xu*
*Shenzhen Institute of Advanced Integration Technology
Shenzhen Institute of Advanced Technology
The Chinese University of Hongkong
tian.cao@sub.siat.ac.cn, {xy.wu, jn.guo, sq.yu}@siat.ac.cn, ysxu@mae.cuhk.edu.hk
†Department of Computer Science
Sichuan University, Chengdu, China

***Abstract*— Video surveillance in crowded areas is becoming more and more significant for public security. This paper presents a method for the detection of abnormality in crowded scenes based on the crowd motion characteristics. These characteristics includes the crowd kinetic energy and the motion directions. This approach estimates the crowd kinetic energy and the motion directions based on the optical flow techniques. The motion variation is derived from the crowd kinetic energy of two adjacent frames, and the motion direction variation is estimated using mutual information of the direction histograms of two neighboring motion vector fields. The proposed method combines crowd kinetic energy, motion variation and direction variation for the abnormality detection. The experiments on the video data which captured by ourselves demonstrate that our method can detect the abnormal behaviors effectively.**

***Index Terms*— Crowd analysis, video surveillance, abnormal detection**

I. INTRODUCTION

Public security has become a major issue in public places such as subway stations, banks, malls, airports, etc. The video surveillance systems can provide massive real-time informative data to people. However, it is nearly impossible to interpret all the data manually for abnormal and emergency events. Meanwhile, dealing with the crowded scenes is one of the most challenging problems in computer vision field.

A. Related Work

More and more intelligent video surveillance systems have been developed in recent years, and crowd analysis in computer vision has become a hot research topic in numerous countries. Crowd information extraction and crowd modeling for event inference are two main components of crowd analysis from computer vision perspective [1]. Crowd information extraction includes crowd density measurement and tracking [1]. Crowd density is an significant feature in crowd and the different density level should indicate different crowd status. Marana et al. [2] estimated the crowd density based on the texture information. Ma et al. [3] established a mathematical relation between the foreground pixels and the number of people. Tracking in crowd scenes is complex due to the existence of multiple moving objects in the scene. Some tracking methods based on interest points were presented in [4]. Cupillard et al. [5] also provided a tracking approach using multiple cameras.

Crowd modeling is often used for abnormal events detection. Andrade et al. [6], [7] combined spectral clustering, Principal Components Analysis (PCA) and Hidden Markov Model (HMM) to detect the crowd emergency scenarios. But this approach was only tested in simulated data. Ali et al. [8] applied the Lagrangian Particle Dynamics to high density crowd flow segmentation, and this method can also detect the flow instabilities. [9] detected the abnormal events in crowd flows by considering simultaneously density, direction and velocity in the motion regions. This method does not need learning process and training data, and it also applied to some collapsing events. But this method was only suitable for really crowded scenes. [10] defined the crowd energy based on the crowd kinetic features and used wavelet analysis to tackle the problem of the detection of abnormality. This method was tested in a railway station surveillance system. Recently, Mehran et al. [11] presented a novel method to detect and localize abnormal behaviors in crowd videos using Social Force model. Their experiment conducted on both publicly available dataset and crowd videos taken from web. The result shows that their approach outperforms similar approach based on pure optical flow.

B. Overview

Our approach to detect the abnormal behaviors was inspired by [9], [10]. However, our method is quite different from them. In [9], the authors only considered the variance of motion characteristics such as direction variance and motion magnitude variance, but the author ignored the total value of the motion magnitude in a crowd event. The author also included direction histogram peaks in their system, but in most crowd scenes, the direction histogram peaks are hard to obtain. In [10], the author defined the crowd energy based on kinetic energy, but the author neglected the direction information.

In this paper, a grid of points is firstly generated as the feature set. We do not use any feature extraction methods in this paper, because no methods can guarantee to obtain a regular distributed feature set and most feature detection methods are time consuming when the feature number is large which makes the real time application impossible. Then we use adaptive gaussian mixture modeling (GMM) method to extract the foreground of the video scene, and

978-1-4244-4775-6/09/$25.00

exclude the noise by applying the foreground as a mask. The motion vectors are captured using optical flow technique, and the features can be further refined by removing these features with large variation in magnitude and direction of their motion vectors to their neighboring motion vectors. We also introduce mutual information method to measure the similarity between two direction histograms derived from three adjacent frames. We modified the traditional kinetic energy by multiplying the foreground to background ratio. This ratio is a rough estimate of the crowd density. Both the modified kinetic energy and mutual information of direction histograms contribute to the detection of the abnormal events.

This paper is organized as follows. Section 2 presents the algorithm, which contains optical flow computation and mutual information measurement. Section 3 discusses how to detect the abnormality using the computed results. In Section 4, the proposed algorithm is applied to the video data captured by ourselves. Finally, we draw some conclusions of our work and point out the future research direction in section 5.

II. METHODOLOGY

The proposed method involves several steps: background modeling, feature tracking, direction histogram building, mutual information computation and finally the anomaly detection based on the calculated crowd motion characteristics.

A. Optical Flow Computation

In most crowd analysis applications, the first step of optical flow computation is feature extraction. But in our approach, we did not use any feature extraction methods because that none of these methods can provide an equally distributed feature set. We use a grid of points as features of each frame. These features can be refined by using a mask which represents the crowd region. We generate this mask from the foreground which is detected by using the GMM described in [12]. All the motion vectors outside the foreground mask are excluded from the feature set to emphasize the motion area.

Once we generate a set of features for a frame, the optical flow approach is applied to these features to track them in the next frame. In our approach, we choose Kanade-Lucas-Tomasi (KLT) optical flow method for feature tracking [13].

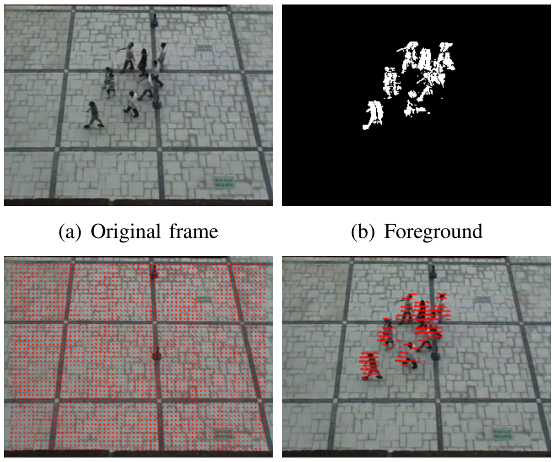
The result of the feature tracking is a set of vectors,

$$F = \{F_1 \dots F_n \mid F_i = (x_i, y_i, u_i, v_i)\}$$

where $P_i = (x_i, y_i)$ are the coordinates of feature i in a certain frame, $V_i = (u_i, v_i)$ is the motion vector. n is the feature number. We can obtain the magnitude and angle of the motion vector from V_i.

We can also further remove the noise features after we applied the optical flow method to the refined feature set. The noise features are these features whose motion is less than two pixels and have a large angle and magnitude difference with their neighbor features [9]. These noise features often come from calculation errors. We apply a 3×3 median filter to the feature set and then use a 3×3 mean filter to smooth the motion field.

Images in Fig. 1 show the results of optical flow computation in a frame. Fig. 1(a) shows the original frame, Fig. 1(b) is the foreground detected by using GMM, Fig. 1(c) shows the feature set and Fig. 1(d) illustrates the purified vector flow in which the calculation noise has been removed.



(a) Original frame (b) Foreground

(c) Grid of features (d) Refined vector flow

Fig. 1. Example of optical flow computation

B. Mutual Information of Direction Histogram

Mutual information is a concept in information theory. It measures the mutual dependance of two random variables. Mutual information has been applied to many computer vision applications, such as used as a criterion for feature selection, feature transformation in machine learning and image registration in medical imaging [14].

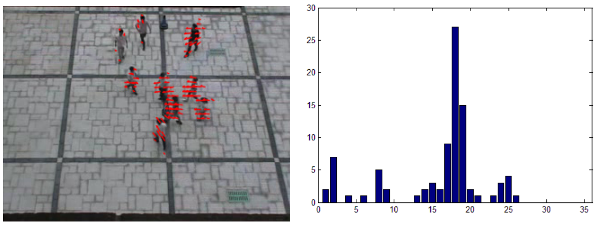
1) Direction Histogram: In our approach, we build a direction histogram from the motion vectors in a certain frame. The histogram h is defined as follows,

$$h(i) = \{k_i, 0 < i \leqslant n\} \tag{1}$$

where n denotes the number of histogram bins, k_i implies the number of motion vectors at a certain angle range which is related to the ith bin for each frame. In this paper, we divide 2π into 36 bins, which means each histogram bin contains 10 degrees.

The calculated histogram indicates the direction distribution and direction tendency of the frame. Fig. 2 is an example of a direction histogram.

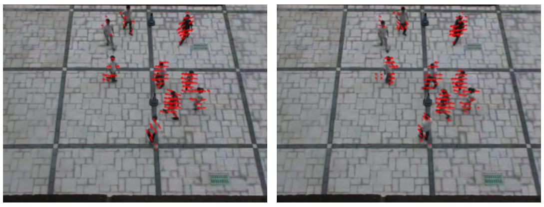
We can also compute the joint direction histogram $h(i, j)$ of two successive vector fields similarly. The two vector fields derived from three continuous frames. Each entry in the joint direction histogram indicate the number of features whose motion vector derived from the first two frames is at ith bin while the motion vector obtained from the last two frames is at jth bin of the direction histogram correspondingly. Here is an example of joint direction histogram in Fig. 3. Fig. 3(a) and Fig. 3(b) show the two successive vector



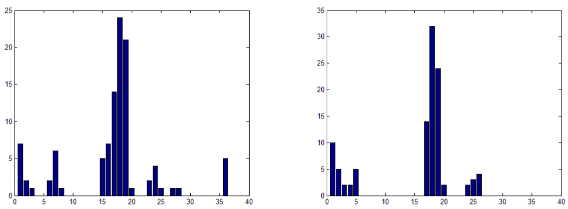
(a) Vector field (b) Direction histogram

Fig. 2. Example of direction histogram

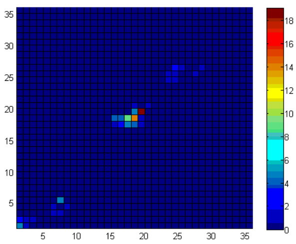
fields separately. The two direction histograms are displayed in Fig. 3(c) and Fig. 3(d) respectively. The joint direction histogram is displayed in Fig. 3(e) where different colors refer to different numbers.



(a) Previous vector field (b) Next vector field



(c) Direction histogram of first vector field (d) Direction histogram of second vector field



(e) Joint direction histogram of two vector fields

Fig. 3. Example of joint direction histogram

2) Direction Probability Distribution: The direction probability distribution for each frame can be directly estimated from the direction histogram by dividing the entries by the total number of the vectors,

$$p(i) = \frac{h(i)}{m}, 0 < i \leqslant n, i \in N \tag{2}$$

where m is the total vector number in a frame. A large number of vectors in a histogram bin will result in a large probability value which means that the people in a frame are mainly moving at that direction.

In order to compute the mutual information of two adjacent vector fields, we also need to estimate the joint probability distribution $p(i,j)$ of them. For joint probability distribution of two vector fields, the most straightforward way to estimate it is to be derived from joint direction histogram $h(i,j)$ using the method discussed above.

3) Direction Entropy: Entropy is a concept stems from information theory [15], which has three interpretations: the amount of information an event gives when it take place, the uncertainty about an event and the dispersion of the probabilities with which the events take place [16]. There are many kinds of definition of entropy, and the most widely used one is defined as,

$$H(X) = \sum_{i=1}^{n} p(x_i) \log(\frac{1}{p(x_i)}) \tag{3}$$

where p is a probability distribution. Similarly, the joint entropy is defined as

$$H(X,Y) = \sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log(\frac{1}{p(x_i, y_i)}) \tag{4}$$

In this paper, we use entropy to describe the dispersion of direction probability distribution. For example, if the direction probability is 1 at angle bin a, then $p(a) = 1$, so the entropy of this direction probability distribution is $H = 1 \cdot \log 1 = 0$. If the direction probability is equally distributed on all the possible angle bins, we get $H = n \cdot 1/n \cdot \log(1/(1/n)) = \log n$. Therefore, a distribution with a single sharp peak yields to a low entropy value, whereas a dispersed distribution corresponds with a high entropy value.

4) Mutual Information: The definition of mutual information is based on the entropy,

$$I(X,Y) = H(X) + H(Y) - H(X,Y) \tag{5}$$

The last term $-H(X,Y)$ means that maximizing mutual information is related to minimizing joint entropy. We have described that joint entropy is a measure of dispersion of the joint direction probability distribution. Given a direction probability distribution X, when the direction probability distribution Y which has the same dispersion as X, the joint entropy of X and Y will be minimized, therefore the mutual information will be maximized.

The mutual information is calculated from two probability distributions. In our application, we need a metric for all the video frames, so we introduce normalized mutual information (NMI). The NMI is defined as follows,

$$NMI(X,Y) = 1 - \frac{H(X,Y)}{\max(H(X), H(X))} \tag{6}$$

In our approach, the NMI is used to detect the abnormal behaviors, and the details will be discussed in the next section.

III. ABNORMALITY DETECTION

In this part, we define our detection method of abnormal event using both the motion vectors and the statistic measure described in previous section.

A. Crowd Kinetic Energy

The calculation of crowd kinetic energy is based on the motion vectors which we captured before. The features in this paper are a grid of points. We did not use any feature extraction methods, because none of these methods can ensure that the number of features for a person is constant in a few consecutive frames. In this application, we construct the features using a grid of points, and the features are constant among the video sequence. Therefore, we can assume that the feature number for a person is relatively stable in adjacent frames. Under this assumption, we compute the crowd kinetic energy in (7). The experiments in next section also show that our algorithm is robust enough for abnormal detection.

The crowd kinetic energy of each frame is as follows,

$$E_i = \sum_{j=1}^{n} m_j v_j^2 \tag{7}$$

where E_i is the crowd kinetic energy of the ith frame in a video sequence, v_j is the magnitude of the jth motion vector in the ith frame, and n is the total number of motion vectors in the ith frame respectively. For a given scene with similar sizes of objects, we assume that $m_i = 1$. From (7) we can obtain the crowd kinetic energy for each video frame which can be used to indicate the global status of the crowd scene.

However, the crowd kinetic energy cannot provide any information about the crowd density, which is actually closely related to abnormal situations. So we define the modified crowd kinetic energy (MCKE) as follows,

$$ME_i = \rho_i \sum_{j=1}^{n} v_j^2 \tag{8}$$

where ρ_i is the ratio of foreground area to background area. In this application, we do not need accurate crowd density information. Thus the foreground to background ratio, which is a rough estimate of crowd density, is appropriate for abnormal events detection.

B. Normalized Mutual Information

After generating a grid of features in each frame, we track these features in three adjacent frames. So we get two sets of feature vectors which can be used to compute the probability of direction distribution. We calculate the mutual information based on the entropies of two probability distributions and the joint entropy.

Mutual information is a measure of the similarity of two vector fields in our application. However, the mutual information value is not a metric itself because it is dependant on the two vector fields. We have defined *NMI* in (6). Low *NMI* means high similarity of the two vector fields, i.e. small variation on motion directions, and high *NMI* means low direction similarity. *NMI* is a universal metric [14], thus we can apply it to our abnormality detection system.

C. Detection of Abnormality

The abnormality description is not an easy task for most crowd events. Because the definition of abnormality varies a lot under different situations. In this paper, we consider two types of crowd abnormal behaviors.

1) Static Abnormality: The first type is called static abnormal behaviors which is similar to the definition in [10]. The abnormality is detected in a static method by comparing the MCKE of a frame to a specific threshold. The threshold estimation is necessary for each crowd scene because that the threshold varies depending on camera position and the time when the crowd events happen.

2) Dynamic Abnormality: The second type is dynamic abnormality where the sudden changes of the MCKE in a video sequence have been measured. The calculation of MCKE variance is not sufficient for abnormal detection because it only considers the magnitudes of motion vectors while the directions of motion vectors have been ignored. Thus we involve normalized mutual information of direction histograms in our application. As we discussed above, normalized mutual information of direction histograms can indicate the similarity of two neighboring vector fields.

The dynamic abnormality is very important in crowd abnormality detection due to the fact that abnormality is often associated with severely change in the motion state, such as gathering, scattering and chaos situations.

IV. RESULTS AND DISCUSSIONS

In order to evaluate the effect of our algorithm, we use a set of videos provided by a stationary camera installed on building B of Shenzhen Institute of Advanced Technology in a topdown view, to monitor the situation of the square near this building. The dataset is consisted of 4 different scenarios. Each video consists of two kinds of situations: normal situation and abnormal situation. Fig. 4 shows the sample frames of these scenes.

The original video frame size is 320×240 pixels. The first crowd scene is that the crowd suddenly change their motion state from walk to run, and the second crowd scene is from run to walk. In the third scene, people are gathering from many directions and then walk together in one direction, and the fourth scene is about a group of people scattering to various directions.

Fig. 5 shows the results of our method. We display the results of scene 1 to scene 4 and the corresponding frames which the abnormal behaviors happened of each scene.

From Fig. 5(a) and Fig. 5(c), we can find a wave crest clearly. The energy curve begins to increase at about $320th$ frame in Fig. 5(a) when comparing to the video data, the crowd of people starts to enter the scene. The energy curve suffers a steep jump at about $460th$ frame and reaches its peak at about $500th$ frame, while in the video data, the crowd starts to run at about $460th$ frame and all the people reache their highest speed in the scene at about $500th$ frame. After $500th$ frame, people begin to leave out the scene while the energy curve starts to decrease since that frame. However, Fig. 5(b) does not show this trend, because the motion

(a) Scene 1

(b) Scene 1

(c) Scene 2

(d) Scene 2

(e) Scene 3

(f) Scene 3

(g) Scene 4

(h) Scene 4

Fig. 4. Sample frames of video dataset

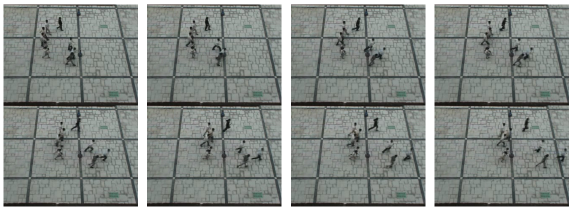
direction does not change much along the motion magnitude. The situation is quite similar in Fig. 5(c) and Fig. 5(d).

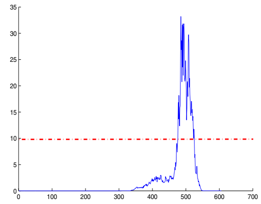
Fig. 5(e) shows the MCKE curve of scene 3. There is also a change in this curve. However, this change is relatively low when it compared with the change in Fig. 5(a) and (c), and the MCKE value is also lower than the threshold (the threshold is much larger than the MCKE value in Fig. 5(e) and Fig. 5(g), so we did not plot it under the same coordinate system in the two figures). Thus the abnormality detection method used in Fig. 5(a) and (c) is no longer appropriate. We can use NMI curve instead. In Fig. 5(f), we can find the curve encounters a great variation at about $220th$ frame while In video data, crowd of people began to meet together at about $220th$ frame. Fig. 5(h) shows very similar result like Fig. 5(f).

From Fig. 5, our approach successfully detected most of the abnormality by comparing with the results annotated manually. However, in order to get a desirable output, it is necessary to choose the favorable threshold carefully.

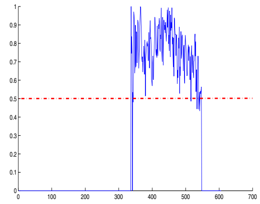
V. CONCLUSIONS AND FUTURE WORKS

In this paper, we presented an approach to estimate the abnormality of a crowd scene. We defined two types of abnormality which are static and dynamic abnormality respectively. The measure of the abnormality is based on the crowd

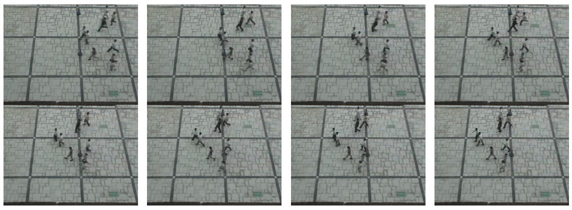


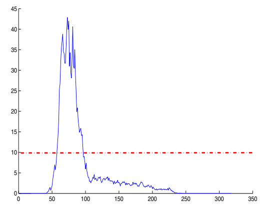


(a) MCKE curve of scene 1

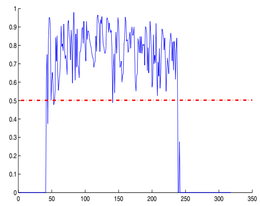


(b) NMI curve of scene 1

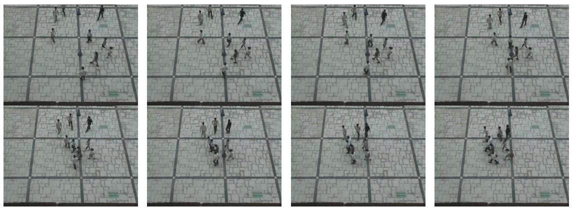


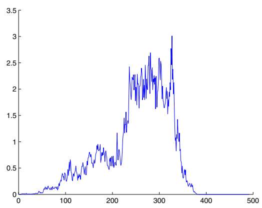


(c) MCKE curve of scene 2

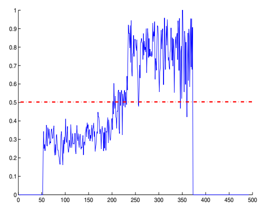


(d) NMI curve of scene 2

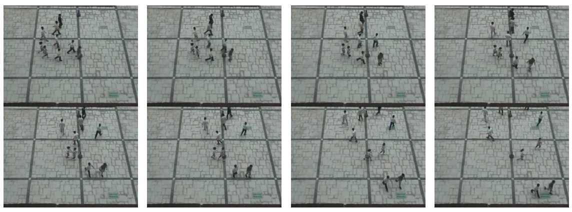


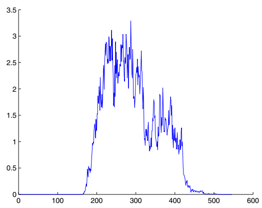


(e) MCKE curve of scene 3

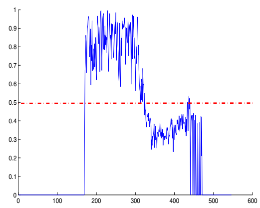


(f) NMI curve of scene 3





(g) MCKE curve of scene 4



(h) NMI curve of scene 4

Fig. 5. Results of Our Method

kinetic energy and the normalized mutual information of the direction histograms which are sensitive to the velocity and direction. This method has been applied to detect abnormal events in four different scenarios. The results show that our method can detect the abnormal events effectively.

Our future work will cover how to involve more crowd characteristics into the abnormality detection, such as the number of crowd groups, acceleration of the motion. To improve the accuracy of our system, we also want to employ more accurate tracking methods to capture some small local motion in the crowd scene. Some appropriate crowd modeling methods are also very significant. In order to develop more intelligent system and to tackle more complicated situations, some machine learning methods will be included in this application. But more complicated algorithm will need more computation time which make the real time application impossible. So some hardwares with high computation power such as graphics processing unit(GPU) can also be involved in our system.

VI. Acknowledgments

The work described in this paper is partially supported by the grants from the Ministry of Science and Technology, the People Republic of China (International Science and Technology Cooperation Projects 2006DFB73360).

References

[1] B. Zhan, D. Monekosso, P. Remagnino, S. Velastin, and L. Xu, "Crowd analysis: a survey," *Machine Vision and Applications*, vol. 19, no. 5, pp. 345–357, 2008.

[2] A. Marana, S. Velastin, L. Costa, and R. Lotufo, "Automatic estimation of crowd density using texture," *Safety Science*, vol. 28, no. 3, pp. 165–175, 1998.

[3] R. Ma, L. Li, W. Huang, and Q. Tian, "On pixel count based crowd density estimation for visual surveillance," in *2004 IEEE Conference on Cybernetics and Intelligent Systems*, vol. 1, 2004.

[4] P. Gabriel, J. Verly, J. Piater, and A. Genon, "The state of the art in multiple object tracking under occlusion in video sequences," in *Advanced Concepts for Intelligent Vision Systems*. Citeseer, 2003, pp. 166–173.

[5] F. Cupillard, F. Brémond, and M. Thonnat, "Group behavior recognition with multiple cameras," in *Proceedings of the Sixth IEEE Workshop on Applications of Computer Vision*. Citeseer, 2002, pp. 177–183.

[6] E. Andrade, S. Blunsden, and R. Fisher, "Modelling crowd scenes for event detection," in *ICPR 2006. 18th International Conference on Pattern Recognition*, vol. 1, 2006.

[7] E. Andrade., S. Blunsden, and R. Fisher, "Hidden markov models for optical flow analysis in crowds," *Pattern Recognition*, pp. 460–463, 2006.

[8] S. Ali and M. Shah, "A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis," in *IEEE CVPR*, 2007.

[9] N. Ihaddadene and C. Djeraba, "Real-time Crowd Motion Analysis," in *ICPR 2008. 19th International Conference on Pattern Recognition*, 2008, pp. 1–4.

[10] Z. Zhong, W. Ye, S. Wang, M. Yang, and Y. Xu, "Crowd Energy and Feature Analysis," in *IEEE International Conference on Integration Technology, 2007. ICIT'07*, 2007, pp. 144–150.

[11] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection using Social Force Model," in *CVPR*, vol. 1, 2009, p. 7.

[12] C. Stauffer and W. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[13] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International joint conference on artificial intelligence*, vol. 81, 1981, pp. 674–679.

[14] Wikipedia, "Mutual information — wikipedia, the free encyclopedia," 2009. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Mutual_information

[15] Wikipedia., "Entropy (information theory) — wikipedia, the free encyclopedia," 2009. [Online]. Available: http://en.wikipedia.org/w/index.php?title=Entropy_(information_theory)

[16] J. Pluim, J. Maintz, and M. Viergever, "Mutual-information-based registration of medical images: a survey," *IEEE transactions on medical imaging*, vol. 22, no. 8, pp. 986–1004, 2003.