An Abnormal Event Recognition in Crowd Scene

Conference Paper · September 2011

DOI: 10.1109/ICIG.2011.66 · Source: IEEE Xplore

Jinhai Xiang
Huazhong Agricultural University

Honghong Liao
Huawei Technologies

An Abnormal Event Recognition in Crowd Scene

Honghong Liao[1], Jinhai Xiang[1,2], Weiping Sun[1], Qing Feng[1], Jianghua Dai[1]
[1] School of Computer Science and Technology
Huazhong University of Science and Technology
Wuhan, 430074, P.R. China
{hongh.liao, fengqing}@smail.hust.edu.cn, wpsun@hust.edu.cn, jhdaihust@gmail.com

[2] College of Science
Huazhong Agricultural University
Wuhan, 430070, P.R. China
jimmy_xiang@mail.hzau.edu.cn

***Abstract*—Real-world actions occur often in crowded, dynamic environments. This poses a difficult challenge for current approaches to video event detection because crowd scenes are always extremely cluttered. In this paper, we design a video content analysis method for fighting event recognition in crowd scene. Our method begins with four MPEG-7 descriptors: crowd kinetic energy, motion directions histogram, spatial distribution parameter and spatial localization parameter of two adjacent frames. Then the support vector machines (SVMs) method is introduced to train and test these descriptors for fighting event recognition. Extensive experimental results have demonstrated that our method is effective in fighting events recognition with low error rates and can be easily adopted in fixed camera environment with real-time application.**

Keywords-abnormal event recognition; crowd scene; support vector machines;MPEG-7 descriptors;

I. INTRODUCTION

Public security has become a major issue in public places such as subway stations, banks, squares, etc. A lot of researches by virtue of computer vision have been done to automatically recognize abnormal crowd events in public places. However, since crowd scenes are always cluttered, this brings great obstacle to crowd event recognition.

A. Related Work

So far, researchers in computer vision field have done much work in crowd analysis. Their work can be classified into two categories roughly [7]. The first is related to crowd density estimation and the second is focus on abnormal event detection in crowd scenes. Davis et al [2] suggested that pedestrians for a video sequence are nearly linear to the number of pixels of video foreground's thinned edges. Chow [5] improved this work by first using templates to match pedestrians, then giving a neural network with globe optimization methods for training. Marana et al. [3] assumed that images of low-density crowds tend to present coarse texture, while images of dense crowds tend to present fine textures. They used a self organizing map neural network for crowd density estimation. Many other methods have also been developed, a recent survey can be found in [4].

The work in the second category detects abnormal events in crowd scenes. Cao et al [6] detected abnormal events by combining crowd kinetic energy and direction variation. Their work aims to detect the crowd that suddenly changes their motion states, such as gathering together and scattering. However, their work needs to set special thresholds, which vary depending on camera position and illumination condition. Moreover, in some crowd scenes like sports games gathering together and scattering are not abnormal events. Andrade et al [16] used optical flows to describe crowd movement. They combined principal components analysis and spectral clustering on flow field, and then trained Hidden Markov Models (HMM) using the data of flow cluster. However, they only tested their work on simulated data sets. There are also some work based on tracking [7,8]. Ihaddadene [7] used a set of points of interest (POI) and estimate each point's motion characteristics such as density, direction and velocity by tracking for crowd density estimation and abnormal detection without training. Their work also needs to set special thresholds. Saxena et al [8] performed KLT tracking method on video frames, which, firstly, detected interesting feature points in the scene and then track these points over time. By modeling crowd behaviors, their work can recognize some crowd events such as fighting, suddenly stop, etc. However, they detected fighting just using the number of motion directions, which can not distinguish fighting from gathering together or cross-walking. Recently, Mehran et al [9] introduced social force model to abnormal detection. They first used the particle advection method to analyze videos of crowd scenarios, then computed the social force between moving particles to extract interaction forces, at last modeled the normal behaviors by the bag of words method. They tested their approach on both the UMN dataset and dataset from web. Experiments showed their work outperforms approaches based on pure optical flow. However, they did not show whether their work can detect fighting events. Chen et al [10] performed an adjacency-matrix based clustering algorithm to cluster human crowds into groups in a unsupervised manner. Then, they also used the force field model to detect unusual crowd events. They defined unusual crowd as the crowd whose orientation was abruptly changed.

978-0-7695-4541-7/11 $26.00 © 2011 IEEE
DOI 10.1109/ICIG.2011.66

B. Overview

The major work on abnormal crowd event detection paid less attention to specific event recognition, such as fighting. In this work, we concentrate on designing a video content analysis method for fighting events recognition in crowd scene. Our approach was inspired by [14]. However, our work is quite different from them. In [14] the authors just gave a simple overview of MPEG-7 motion descriptors, which include the four descriptors described in our work (will be described in detail in next section), and some application areas such as: video browsing, content-based querying and retrieval from video databases, etc. The work in [11] also extended MPEG-7 motion descriptors to detect fighting events in crowd scenes. However, their work needs to set special thresholds to each descriptor, which is infeasible for actual application. Furthermore, their motion descriptors are quite different from ours:

- *Crowd Kinetic Energy*. They averaged the motion vector magnitude of all macro blocks in a frame as their motion intensity, and didn't use the crowd kinetic energy.
- *Motion Directions Histogram*. In their work motion direction descriptor was obtained by the maximum bin of direction histogram subtract the minimum bin of direction histogram, for efficiently estimating the threshold. We set the whole direction histogram as motion directions histogram, in order to obtain more motion direction information.
- *Spatial Distribution Parameter*. Their video descriptor was just obtained by the long run length, while ours are all three run length. Different run length measures different size of motion area.
- *Spatial Localization Parameter*. This descriptor in their work was the ratio of the motion intensity of the largest connected area to that of the whole frame. They assumed that fighting events always occur in the largest motion area, which is not appropriate for real scenes. Considering their limitation, we get the largest value of the ratio of kinetic energy in connected areas to that of the whole frame as the descriptor of spatial localization parameter.

According to observations, fighting events has the following characteristics: high motion intensity, disorder motion directions and compact motion region. Fighting events is very similar to people gathering together, both of which have intense motion and disorder motion directions. Nevertheless, fighting often occurs in relative small regions. Therefore, we use four MPEG-7 descriptors to describe fighting events, which are crowd kinetic energy for motion intensity measurement, motion directions histogram for motion directions angle measurement, spatial distribution parameter for motion region compactness measurement and spatial localization parameter to suggest the kinetic energy in a small area relative to that of the whole region and locate the fighting events. After extracting the four MPEG-7 video descriptors, we introduce SVMs (support vector machines) to train these video descriptors for fighting event recognition.

This paper is organized as follows. Section 2 provides detailed explanation of our video descriptors. Followed by a brief review of SVMs and formulate our method using SVMs as well in section 3. Section 4 demonstrates effectiveness of the proposed method using real world surveillance videos collected by stationary cameras. Conclusions and future research direction in Section 5.

II. VIDEO DESCRIPTORS

In this section, we describe four video descriptors: crowd kinetic energy, motion directions histogram, spatial distribution parameter and spatial localization parameter in detail.

A. Crowd Kinetic Energy

Crowd kinetic energy is a video descriptor used to measure the intensity of motion. In our intuition, movements like sports games are motions with high intensities while human stroll with low intensive motion. In this work, we extract crowd kinetic energy like video compression using Block Matching Algorithm (BMA)[12] with three-step search for fast searching and the search window is set to be 8 × 8. And the match factor is mean squared error (MES), which is defined as:

$$\mathrm{MES}(\theta)=\sum_{x\in\Omega}\frac{1}{N_1N_2}\left(I(x+\vec{d})-I(x)\right)^2. \quad (1)$$

where N_1, N_2 are the number of pixels in reference macro block and current macro block, respectively. I(x) is image pixel intensity in position x. Ω and $\vec{d}$ is the search region and the motion vector, respectively. After the best macro block determined, the set of motion vectors, $F=\{F_1,\cdots,F_n \mid F_i=(i,j,x,y)\}$ where $id=(i,j)$ is macro-block's index and $v_{i,j}=(x,y)$ is corresponding to motion vector. Furthermore, we remove the zero blocks and noise blocks after BMA. The zero blocks are defined as macro blocks with no motion; while the noise blocks are these macro blocks whose motion vector is less than two pixels or with large angle and magnitude difference in comparison with their neighbors.

Crowd kinetic energy of a single macro-block indexed by (i, j) with motion vector $v_{i,j}=(x,y)$ is defined as $E_{i,j}=x^2+y^2$. We sum all crowd kinetic energy of macro blocks in a frame as crowd kinetic energy of the frame, which can be defined as:

$$E_f=\sum_{i=1}^{M}\sum_{j=1}^{N}E_{i,j}. \quad (2)$$

supposing block size is SIZE × SIZE, so $M=\frac{Width}{SIZE}$ and $N=\frac{Height}{SIZE}$ with the video frame size is Width × Height. Here SIZE equals to 8.

B. Motion Directions Histogram

In section 2.1, we have got the motion vectors of two consecutive frames using BMA. So, motion directions can be directly attained. For the macro block with index (i, j), its motion direction angle is

$$Dir_{i,j} = \arctan\frac{y}{x} + \pi \tag{3}$$

where y and x are components in motion vector $v_{i,j} = (x, y)$. We quantize the motion directions into a histogram with 8 equal bins, which can be described as below:

$$\begin{cases} His[1]++; \; if \quad Dir_{i,j} \in [0, \frac{\pi}{8}) \cup [\frac{15\pi}{8}, 2\pi) \\ His[2]++; \; if \quad Dir_{i,j} \in [\frac{\pi}{8}, \frac{3\pi}{8}) \\ His[3]++; \; if \quad Dir_{i,j} \in [\frac{3\pi}{8}, \frac{5\pi}{8}) \\ His[4]++; \; if \quad Dir_{i,j} \in [\frac{5\pi}{8}, \frac{7\pi}{8}) \\ His[5]++; \; if \quad Dir_{i,j} \in [\frac{7\pi}{8}, \frac{9\pi}{8}) \\ His[6]++; \; if \quad Dir_{i,j} \in [\frac{9\pi}{8}, \frac{11\pi}{8}) \\ His[7]++; \; if \quad Dir_{i,j} \in [\frac{11\pi}{8}, \frac{13\pi}{8}) \\ His[8]++; \; if \quad Dir_{i,j} \in [\frac{13\pi}{8}, \frac{15\pi}{8}) \end{cases}$$

Crowd fighting has disorder motion directions, and then the histogram can be taken as the motion direction descriptor.

C. Spatial Distribution Parameter

Fighting events often happens in a small region which is different from people gathering together. Spatial distribution parameter of a frame denotes to measure the size of motion region whose values are extracted from the thresholding kinetic energy matrix, which includes elements for each macro block indexed by (i, j). We quantize spatial distribution parameter descriptor into three levels: short, medium and long run length [14]. Each run length is obtained by recording the length of the continuous non-zero macro blocks in a raster scan order over the matrix. For each frame the "thresholding kinetic energy matrix" E_{mv}^{thresh} is defined as:

$$E_{mv}^{thresh} = \begin{cases} E_{i,j}; \; if \quad E_{i,j} > E_f^{avg} \\ 0; \; otherwise \end{cases} \tag{4}$$

where $E_{i,j}$ is the kinetic energy of macro block with index (i, j) and average kinetic energy for an M×N macro block frame are defined as

$$E_f^{avg} = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N} E_{i,j}$$

Run length always varies with different video views, so we set a threshold θ to the continuous non-zero macro blocks within a row in order to get the three levels of spatial distribution parameter, which is described as below:

$$\begin{cases} LRL^i++; \; if \quad N > \frac{2\theta}{3} \\ MRL^i++; \; if \quad \frac{\theta}{3} < N \le \frac{2\theta}{3} \\ SRL^i++; \; otherwise \end{cases}$$

where LRL^i++ represent the long run length of frame i and similarly the other two represent the middle and short run length, respectively. N is the length of the continuous non-zero macro blocks within a row in frame i.

Here θ is the only threshold we need to set. The three levels of run length are the spatial distribution parameter descriptor.

D. Spatial Localization Parameter

In fighting events, crowd kinetic energy is often with large values in some relative small area. So we have to firstly determine the connected areas in the thresholding kinetic energy matrix. After that, the largest value of the ratio of kinetic energy in connected areas to that of the whole frame as the descriptor of spatial localization parameter can be got. The spatial localization parameter suggests the intensity of motion in a connected area. If the spatial localization parameter of a connected area is large, fighting events may happen in this area. Experimental analysis shows that this descriptor is efficient enough to separate fighting from decentralized cross-walking events and gathering together.

III. Recognition of Abnormal Events

In this section, we first give a brief introduction to SVMs, and then derive our method from SVMs.

A. Support Vector Machines (SVMs)

Let (x_i, y_i) be a set of training examples, each example $x_i \in R^d$, d being the dimension of the input space, belongs to a class labeled by $y_i \in \{-1, +1\}$. The aim is to define a decision function which divides the set of examples such that all the points with the same label are on the same side of the decision function. This amounts to identify ω and b so that

$$y_i(\langle \omega, x_i \rangle + b) \ge 1, i = 1, \cdots, N. \tag{5}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product of vectors in R^d, N is the number of training data. And the decision function can be written as

$$f(x) = \langle \omega, x \rangle + b$$

When the data is not linearly separable, we introduce slack variables $\{\xi_1, \cdots, \xi_N\}$ with $\xi_i \geq 0$ such that

$$y_i(\langle \omega, x_i \rangle + b) \geq 1 - \xi_i, \quad i = 1, \cdots, N. \qquad (6)$$

to allow the possibility of examples that violate, but every violation is penalized by a certain cost term ξ_i .

When there is no hope for any good linear classifier, even if we allow for violations, we can extend SVMs for non-linear classification. First, we apply a non-linear transformation to the data. Afterwards, we again learn a single linear classifier to the resulting transformed dataset. In this case, the decision function can be written as

$$f(x) = \langle \omega, \psi(x) \rangle + b$$

where $\psi : R^d \to R^m$ is a non-linear map.

More generally, we can define the kernel function $k : R^d \times R^d \to R$ of ψ by the identity

$$k(x, x') = \langle \psi(x), \psi(x') \rangle \qquad (7)$$

By virtue of this, we obtain the classifier problem in kernelized form, which is often called *Support Vector Machines (SVMs).*

Definition 3.1 (Support Vector Machines)

Let $X = \{x_1, \cdots, x_n\} \subset R^d$ be a training set with training labels $Y = \{y_1, \cdots, y_n\} \subset \{-1, +1\}$. let $k : R^d \times R^d \to R$ be a kernel function, and let C>0. Then the decision function of the kernelized classifier problem is given by

$$f(x) = \sum_{i=1}^{n} \alpha_i k(x_i, x)$$

for coefficients $\alpha_1, \cdots, \alpha_n$ obtained by solving

$$\min_{\alpha_1, \cdots, \alpha_n \in R, i=1} \sum^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k(x_i, x_j) + \frac{C}{n} \sum_{i=1}^{n} \xi_i \qquad (8)$$

subject to

$$\begin{cases} y_i \sum_{j=1}^{n} \alpha_j k(x_j, x_i) \geq 1 - \xi_i; \quad for \quad i = 1, \cdots, n \\ \xi_1, \cdots, \xi_n \in R^+ \end{cases} \qquad (9)$$

For a more thorough introduction about SVMs see for example [1,15].

B. Recognition of Abnormal Events

In this paper, we choose the radial basis function (BRF) kernel in SVMs, and extract the four video descriptors as a feature vector with 13 dimensions (in order 8 motion directions, 3 spatial distribution parameters, 1 spatial localization parameter, and 1 crowd kinetic energy). The BRF kernel is defined as

$$k_{BRF}(x, x') = \exp(-\gamma |x - x'|^2)$$

where x and x' are feature vectors. As a result, we must solve the following optimization problem for fighting events recognition

$$\min_{\alpha_1, \cdots, \alpha_n \in R, i=1} \sum^{n} \sum_{j=1}^{n} \alpha_i \alpha_j k_{BRF}(x_i, x_j) + \frac{C}{n} \sum_{i=1}^{n} \xi_i \qquad (10)$$

subject to

$$\begin{cases} y_i \sum_{j=1}^{n} \alpha_j k_{BRF}(x_j, x_i) \geq 1 - \xi_i; \quad for \quad i = 1, \cdots, n \\ \xi_1, \cdots, \xi_n \in R^+ \\ \{x_1, \cdots, x_n\} \subset R^{13} \end{cases} \qquad (11)$$

We tune the two parameters γ and C by cross validation and the decision function can be formulated as

$$f(x) = \sum_{i=1}^{n} \alpha_i k_{BRF}(x_i, x).$$

IV. Experiments and Discussions

In this section, we evaluate our method by a set of videos provided by stationary cameras.

A. Dataset

To our best knowledge, no video database is available for benchmarking crowd behavior analysis algorithms. We evaluate our algorithm based on videos taken from the some video surveillance videos, some of which are taken in a top down view named as dataset A, while the others are named as dataset B. The video frame resolution in dataset A is 720×576 pixels while that is 352×240 for dataset B. Four different scenarios constitute the dataset B: fighting, gathering together, cross-walking and walking in the same direction as shown in Fig. 1 and Fig. 2.

(a)before fighting (b) fighting (c) fighting (d) cross-walking

Figure 1. samples frames of the dataset A

Dataset A and Dataset B include one and four video sequences, respectively. In dataset A, one person firstly walks in the scene, and then three other men come in, stop and fight him. A moment later, there are a lot of persons come in for stopping the fighting and these persons cross walking in the scene. Each video in dataset B has its own scene. The first video shows a group of people come together and then divide into two small groups. The second video shows two groups come in from different directions, and then they merge into one group leaving the scene. The third video describes a scene with several persons cross walking. The last one shows two groups come in from different direction and then they begin to fight.

(a)cross-walking (b)gathering (c) fighting (d)walking

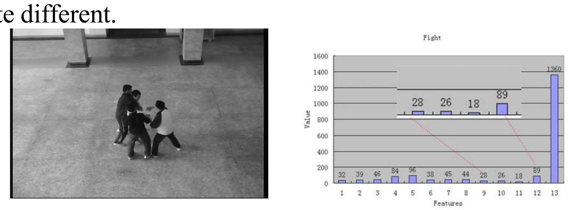
Figure 2. samples frames of the dataset B

B. Analysis on Training and Testing

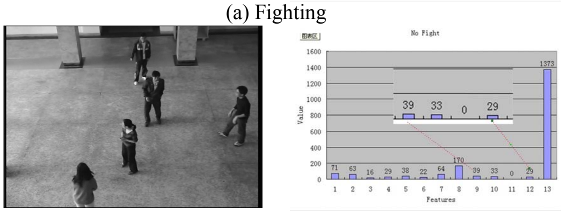
In both datasets, we extract feature vectors for a set of positive and negative examples and batch-train a discriminative classifier for the initial training algorithm. We use the LIBSVM tool [13] for binary classification (BRF Kernel) with the default classifier. We set θ to be 9, and evaluate the prediction results by

$$Recognition\ Rate = \frac{\#correctly\ \ predicted\ \ data}{\#total\ \ testing\ \ data} \times 100\%$$

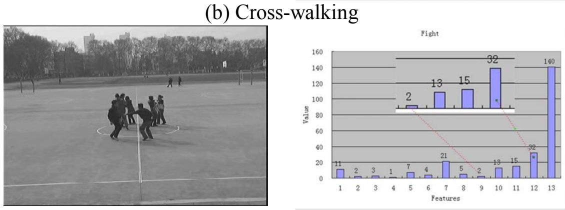
In dataset A, the total number of the video frames is 1484, with 508 positive examples. And there are totally 3917 frames in dataset B summed all videos frame number, with 256 positive examples. Fig. 3 shows the feature vector with its corresponding positive example or negative example in both datasets (Fig. 3(a) (b) for dataset A and the others for dataset B). From Fig. 3 it can be seen that fighting events and cross walking or gather together poses similar crowd kinetic energy and also disorder motion directions. However their spatial distribution and spatial localization parameters are quite different.



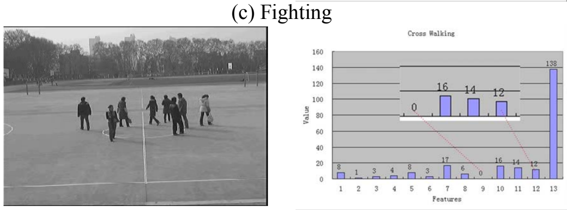
(a) Fighting



(b) Cross-walking



(c) Fighting



(d) Cross-walking

Figure 3. Examples and its corresponding feature vector

Table 1 shows our results in dataset A and some example images of recognition results show in Fig. 4. Table 2 shows our results in dataset B. As dataset B has a lot of background frames with less than two people, so we do not take all these frames as training or testing examples. As a result, only 2682 frames used for training and testing. Results shows our methods for fighting recognition can achieve fine performance, and better performance can be obtained with training example increase. (Positive) in Table 1 and Table 2 stands for the number of positive examples used for training. Videos in dataset B would lose much information, especially for the spatial distribution parameter and motion directions histogram. Moving along from near to far or oppositely, many macro blocks may be considered as noise blocks for motions in this direction are always smaller than they would really be.

Figure 4. Examples of Reconition Results

TABLE I. RESULTS FOR DATASET A

Training Size(Positive)	*Testing Size*	*Recognition Rate*
484(173)	1000	94.30%
700(242)	748	96.81%
1000(335)	484	96.90%

We compare our result with [11] in Table 3. It can be easily concluded our method achieves a better recognition rate in fighting events. The reasons are: 1) Our descriptors are more useful for recognizing fighting events than [11]; 2) More information added improves the recognition rate. The result of [11] are quoted from [11] and our average results obtained by averaging our result in both datasets.

TABLE II. RESULTS FOR DATASET B

(1) Training and Testing in Fighting Video

Training Size(Positive)	*Testing Size*	*Recognition Rate*
178(81)	400	79.25%
400(175)	178	88.89%

(2) Training and Testing in All Videos

Training Size(Positive)	*Testing Size*	*Recognition Rate*
1201(90)	1471	88.58%
1471(166)	1201	90.42%

The running configurations are: Intel Core 2 CPU at 1.86GHz, 2GB memory and OpenCV library. As training in an off-line way, the main execution cost lies in video descriptors extraction. The effective frame rate at 4-5 frames per second in dataset A and 15-16 frames per second in dataset B by our method. It suggests that our method can applicable in fixed camera environment with real-time application. Extracting crowd kinetic energy is the bottleneck in our method, as it accounts for 80% of the running time.

TABLE III. COMPARISON OF OUR RESULTS WITH [11]

	Average Recognition Rate
Result in [11]	80.00%
Our result on Dateset A	96.00%
Our result on Dateset B	86.78%

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we presented an approach to recognize fighting events in crowd scenes. As fighting have the following characteristics: high motion intensity, disorder motion directions and compact motion region, we develop four video descriptors: crowd kinetic energy, motion directions histogram, spatial distribution parameter and spatial localization parameter to identify fighting in crow scenes. Experimental results have shown that our method can detect the fighting events effectively with high recognition rates in crowd scenes and is easily applied in fixed camera environment with real-time application.

Our future work will cover fighting recognition in clutter scenes, such as in street or railway stations. To speed up computing crowd kinetic energy, we want to employ other less compute complexity approaches or some hardware with high computation power such as graphics processing units (GPU), as BMA can be easily implemented on GPU platform. Furthermore, researches on video descriptors for other crowd event detection with appropriate crowd modeling methods could also be one of our future topics.

REFERENCES

[1] John Shawe-Taylor, and Nello Cristianini, "Kernel Methods for Pattern Analysis", Cambridge University Press, 2004.

[2] A. Davies, J. H. Yin, and S. Velastin, "Crowd monitoring using image processing," Electronics Communication Engineering Journal, vol. 7(1), feb 1995, pp.37-47.

[3] A. Marana, S. Velastin, L. Costa, and R. Lotufo, "Estimation of crowd density using image processing", In Image Processing for Security Applications (Digest No.: 1997/074), IEE Colloquium on, vol. 11, march 1997, pp. 1-8.

[4] B. Zhan, et al, "Crowd Analysis: a Survey", Machine Vision and Applications, vol.19, 2008, pp. 345-357.

[5] T W S Chow, et al, "Industrial Neural Vision System for Underfround Railway Station Platform Surveillance", Advanced Engineering Informatics, 1999, pp. 16:73-83.

[6] T. Cao, et al, "Abnormal Crowd Motion Analysis", IEEE International Conference on Robotics and Biomimetics Robio, 2009.

[7] N. Ihaddadene, C. Djeraba, "Real-time crowd motion analysis", 19th International Conference on Pattern Recognition, 2008.

[8] S. Saxena, et al, "Crowd Behavior Recognition for Video Surveillance", in Lecture Notes in Computer Science, 2008.

[9] R. Mehran, A. Oyama, and M. Shah, "Abnormal Crowd Behavior Detection using Social Force Model", IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[10] D. Chen, P. Huang, "Motion-based unusual event detection in human crowds", Journal of Visual Communication and Image Re presentation, vol. 22(2, Sp. Iss. SI) , 2011, pp. 178-186.

[11] JinWang, "Study on Detection Approach of Abnormal Events in Surveillance Video", (in Chinese) Huazhong University of Science and Technology, Master Thesis, 2007.

[12] R. Li, B. Zeng, M.L Liou, "A new three-step search algorithm for block motion estimation", IEEE Transactions on Circuits and Systems for Video Technology, vol. 4(4) ,1994, pp.438-442.

[13] C.C. Chang, C.J. Lin, "LIBSVM: a library for support vector machins", 2001, Software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[14] Ajay Divakaran, "An Overview of MPEG-7 Motion Descriptors and Their Applications", In 9th Int.Conf.on Computer Analysis of Images and Patterns. Poland: Lecture Notes in Computer Science,vol. 2124, 2001, pp.29-40.

[15] C.H.,Lampert, "Kernel Methods in Computer Vision", Foundations and Trends in Computer Graphics and Vision, vol. 4(3), 2009, pp.193-285.

[16] E.L., Andrade, S. Blunsden, and R.B. Fisher, "Modelling crowd scenes for event detection", in International Conference on Pattern Recignition, vol. 1, 2006, pp. 175-178.