

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

# INFORMATION TECHNOLOGY JOURNAL

**ANSI***net*

Asian Network for Scientific Information  
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

## Crowd Abnormal Behavior Detection Based on Machine Learning

<sup>1</sup>Dongping Zhang, <sup>1</sup>Huailiang Peng, <sup>2</sup>Yu Haibin and <sup>1</sup>Yafei Lu

<sup>1</sup>College of Information Engineering, China Jiliang University, Hangzhou 310018, China

<sup>2</sup>College of Electronic and Information, Hangzhou Dianzi University Hangzhou 310018, China

---

**Abstract:** The detection of abnormal behavior is an important area of research in computer vision and is also driven by a wide of application domains, such as intelligent video surveillance. However, there are few detection algorithms to recognize abnormal behavior in crowds. This study proposed a novel method which can detect whether the crowd is abnormal or not in particular scene, such as stampede, fight and panic. For this purpose, a kind of feature extraction and description scheme has been put forward for particle flow information about crowd motion applying to space-time features cubes. The detection algorithm combined with space-time feature cubes and competitive neural network model is proposed to detect abnormal events in global region. The experimental results show that our approach achieves superior performance to abnormal behavior detection in crowds.

**Key words:** Abnormal behavior, machine learning, space-time features cubes, competitive neural network

---

### INTRODUCTION

In recent year, the abnormal behaviors detection in crowd is a new area of interest in the research community which could potentially lend itself to a number of new application domains proposed by Xiang and Gong (2008). Generally, there are two conventional methods used to analyze abnormal behaviors in crowds. Marques *et al.* (2003) presented the “object-based” methods which consider a crowd is made up of a collection of individuals. Hazel (2000) considers it is necessary to perform segmentation or detect objects to analyze crowd behaviors. This approach faces substantial complexity in detection of objects, recognizing activities and tracking trajectories in dense crowds where the whole process is affected by occlusion. Different from the first method, “holistic” approaches which raised by Andrade *et al.* (2006), considering the crowd as a global entity, have been used in analysis of crowd scenes.

Despite these work, researchers are still especially interested in looking for the problem of detecting about the crowd model for understanding the abnormal behavior in crowd. In related works by Wang *et al.* (2010) and Liu *et al.* (2011), crowd scene models were established to extract features for detecting abnormal behaviors. Mahadevan *et al.* (2010) used discriminator saliency and Gaussian mixture of models to detect space and temporal abnormal behaviors, respectively. Wu *et al.* (2010) achieved typical trajectories from the crowd and capture the features by calculating Lyapunov exponent and correlation dimension which is used to train a classifier to

detect anomalies behavior. Mehran *et al.* (2009) introduced a social force model to extract the force applied on the particles for describing the behavior in crowds, then exploited bag of words to detect anomalies. Khan and Sohn (2011) proposed a method for event detection in the crowd scene using HMM. Ihaddadene and Djeraba (2008) used four different modes of statistics to characterize the abnormal events.

In this study, different kinds of local motion pattern obtained from the crowd defined as features. Then by taking advantage of these features classifiers are trained. The global behavior of the crowd is modeled implicitly by the classifiers. The final performance of anomaly detection is partly determined by the choice of the classifier and the training data.

To model the global behavior of the crowd motion explicitly, this study extract the particle information from crowd motion field which is applied to space-time features cubes. Then a classifier inspired by neural network model is built to train and classified the abnormal behaviors. A detection algorithm is proposed to detect abnormal events in global region crowd which is combined with space-time feature cubes and competitive neural network model.

### CROWD REGIONS DETECTION

This section describes a crowd regions detection method proposed by Zhang *et al.* (2013) which can effectively group pedestrians in public surveillance. First of all, edge information is extracted using a method, in

which canny operator is introduced as described in the following text. Then motion edge is calculated from success frames to segment the regions of motion crowds. Furthermore particle motion information is calculated by long-range motion estimation in foreground motion regions.

**Canny edge calculation:** The process of canny edge calculation firstly is operation of Gaussian convolution smoothing, then using non-maximum gradient value to thin the edge, finally applying the non-maximum suppression to the gradient magnitude to thin the edge and the threshold operation with hysteresis to detect and link edges.

**Motion edge calculation:** The process of motion edge calculation is making difference between two successive frames. The difference of edges is defined as:

$$\Phi(f_{n-1})-\Phi(f_n) = |\theta(\nabla G * f_{n-1})| - \theta(\nabla G * f_n) \quad (1)$$

where, the edge maps  $\phi f_n$  are obtained by the canny edge detector, accomplished by performing a gradient operation  $\nabla$  on the Gaussian convoluted image  $G * f_n$ .

**Motion objects obtained:** When the edge maps are obtained, the method proposed by Kim and Hwang (2002) is used to extract the motion regions. This approach mentions that the horizontal candidates are declared to be the region inside the first and last edge points in each row and the vertical candidates for each column. After finding both horizontal and vertical candidates, the intersection regions obtained by logical AND operation is further processed by morphological operations. In order to enhance the accuracy of regions,  $\pm 45$  degree candidates are added to the method.

The ultimate detection result of motion regions is shown in Fig. 1.



Fig. 1: Detection of motion regions edge in crowd scene is marked in white color

## LOW-LEVEL VISUAL FEATURES ACQUISITION

This section defines the particle flow information extracted from crowd scene by using variational optical flow algorithm proposed by Sand (2006) as feathers which is applied to space-time features cubes. This algorithm consists of two steps. In the first step, the variational optical flow algorithm is calculated for providing initial guess of particle motion. The second step is to extract low-level visual features that contain the speed and direction information in crowd.

**Variational optical flow algorithm:** In order to avoid the large optical flow error and light sensitive trouble when the range of object motion is in a large scope, this study proposes a method based on the theory of layered pyramid optical flow estimation algorithm. Firstly, in order to enhance robustness of optical flow algorithm with variational gray in image, this study expands the gradient consistency hypothesis based on H-S differential method grey consistency hypothesis. Secondly, piecewise smooth condition is used instead of global smooth condition as smooth constraint term. Finally, a kind of multi-scale thoughts is used to calculate the optical flow with large displacement, computing optical flow field of low-resolution image and setting the optical flow field processed by Bicubic interpolation as the initial value of next optical flow field computation between two high-resolution image. So repeatedly, it stops until computation of the initial resolution optical flow is completed.

The total energy functional consists of data term and smooth term. It is shown as follows:

$$E(u,v) = E_{data}(u,v) + \alpha E_{smooth}(u,v) \quad (2)$$

where,  $\alpha > 0$  is regularization parameter. Then it could solve  $v$  and  $u$  by minimizing the energy function. The global deviations from the grey value constancy assumption and the gradient constancy assumption are measured by the energy:

$$E_{data}(u,v) = \int_{\Omega} \psi(|I(x+w) - I(x)|^2 + \gamma |\Phi \nabla I(x+w) - \nabla I(x)|^2) dx \quad (3)$$

And smooth term has to describe the model assumption of a piecewise smooth flow field:

$$E_{smooth}(u,v) = \int_{\Omega} \psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) dx \quad (4)$$

where,  $\psi(r^2) = \sqrt{r^2 + \epsilon^2}$ ,  $\epsilon = 0.001$  is a robust function used to inhibit abnormal points. According to the calculus of variations, total energy function formula should satisfy the following Euler-Lagrange equation:

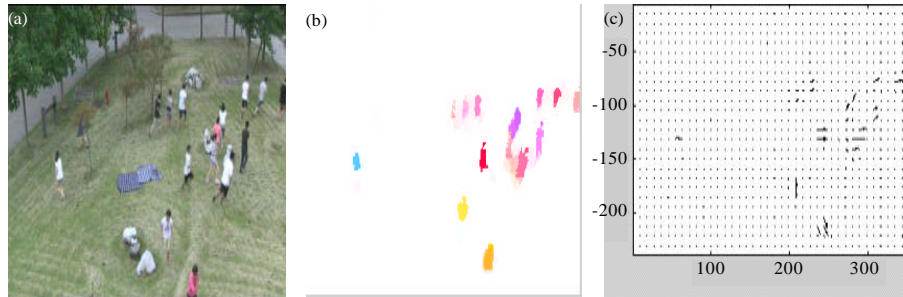


Fig. 2(a-c): Experimental results of optical flow algorithm. (a) Two random frames chose from the crowd scenem, (b) Foreground of crowd scene shown in Fig. 1a and (c) Results of optical flow algorithm

$$\begin{cases} \Psi(I_x^2 + \gamma(I_{xx}^2 + I_{yy}^2)) \cdot (I_x I_z + r(I_{xz} I_{zz} + I_{xy} I_{yz})) \\ -\alpha \text{div}(\Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) \nabla_3 u) = 0 \\ \Psi(I_x^2 + \gamma(I_{xx}^2 + I_{yy}^2)) \cdot (I_y I_z + r(I_{yz} I_{zz} + I_{xy} I_{xz})) \\ -\alpha \text{div}(\Psi(|\nabla_3 u|^2 + |\nabla_3 v|^2) \nabla_3 v) = 0 \end{cases} \quad (5)$$

The preceding Euler-Lagrange equations are nonlinear in their argument  $w = (u, v, 1)$ . However, it can be transformed into linear equation by fixed point iteration and then optical flow field can be obtained by using Successive Over Relaxation (SOR) and hierarchical policies. Figure 2 shows the experimental results diagram of optical flow algorithm.

**Low-level visual feature extraction:** This part of section extracts the particle information as a low-level visual feature from crowd motion field which is applied to space-time feature cubes. Motion area is detected in each frame. Moving pixels are detected in motion area of each frame as follows: firstly, the intensity difference is calculated between two successive frames. If the difference at a pixel is above a threshold, that pixel is detected as a moving pixel. Then the motion direction at each moving pixel is obtained by computing optical flow proposed in the section above. The moving pixels are quantized according to a space-time feature cube as follows. Each moving pixel has two features: position and direction of motion. To quantize position, the scene (360×240) is divided into cells of size 20×20 and several cells in the same position from interval frames form a space-time feature cube. The motion of a moving pixel is quantized into five directions as shown in Fig. 3. Hence, the size of the space-time feature cube is 18×12×10 (10 interval frames are selected in this study) and thus, each detected moving pixel is assigned an element from the space-time feature cube based on rough position and motion direction. Deciding the size of the space-time feature cube is a balance between the

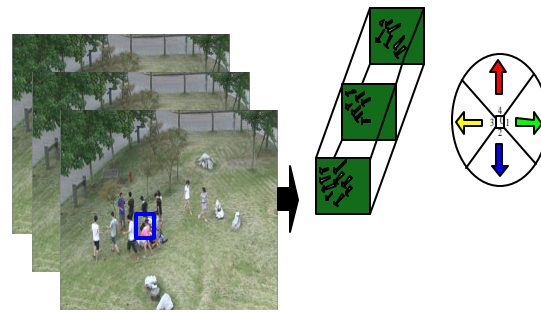


Fig. 3: Video sequence is divided into short clips as documents where, local motions are quantized into visual words based on location and motion direction

descriptive capability and complexity of the model. In our framework, the whole video sequence is uniformly divided into non-overlapping short clips, each 10 frames in length.

### CROWD ABNORMAL BEHAVIOR DETECTION

The key of designing a neural network system lies in the selection of input feature, designing of model structure and the choice of learning algorithm. By calculating the change of the video of particle motion in the motion area, each space-time feature cube is labeled by its location, direction and speed to form our basic feature set. The structure of the learning and training model is designed based on the field of study and the problem to solve. The general strategy is; (1) An appropriate model is established according to research and (2) A kind of learning algorithm is selected by the selection model. In the learning process, it needs to constantly adjust the network parameters, until the output meets the requirements of the results. This study

concerns about the small group of specific several abnormal behavior recognition of processing classification, then a hidden three layer competitive BP neural network is required so that the result are able to achieve a good performance.

The parameters setting for the network of learning and training are described as follows:

**Number of input layer neuron:** The number of input layer neurons is used to determine the size of the grid pixels characteristic information based on visual feature extraction. When training on a sample of  $18 \times 12 \times 10$ , the input layer exists the problems as follows: (1) Huge complexity of the calculation and (2) Network training convergence speed has low faults. So this study designs the sub-sample selection strategy as shown in Fig. 4. By making a synthesis of movement continuity and sample data quantity, this study extract time-space cube features in the same position of the interval 10 frames in video and select a column in the  $18 \times 12$  time-space features cubes, respectively which regard as one sample. In this study the sub-sample input layer neuron is taken for 120, making the events happened in sample period as far as possible.

**Number of hidden layer nodes:** The determination of number of hidden nodes is a difficult problem in formulation. It demands experimental personnel conduct experiments repeatedly to choose the best value under the auxiliary of empirical formula. Therefore, there is no so-called ideal selection scheme. In this article, the study set the number of input layer neuron in sub-network to 120 and the number of output layer neuron is set to 4. If

the determined number of hide layer node is too little, it has weak promotion ability for the success of training and the recognition rate is not high for the samples not in the training set. If there are large mounts of neurons of hidden layer in network, it could result in not only long time of learning but also bad fault tolerance. This study decides the number of neurons by  $n = \log_2 n_1$  and do slight adjustment in experiments.

**Number of neurons in the output layer:** The number of neural network output layer is determined by the output and decision of designed network. This designed neural network is as a pattern classifier, so the number of output layer neuron in network is defined as the number of identify class which is set to 4 and each neuron represents a recognition species.

**Training parameters:** In BP algorithm, the training rate is determined by experience which is better to select the bigger one out of the oscillation condition. In this study, the minimum training rate is set to 0.9. The choice of dynamic coefficient depends on experience selection largely, here, the study set it to 0.6 ~ 0.8. It ends when the iterative result error is less than 0.0001. The maximum iterations are set to 10000. When the iteration results do not converge, it is reasonable to treat the maximum iterations as termination conditions.

**Sample decision:** This study divided the sample into 18 sub-sample sets and training them by the same network model as is shown in Fig. 5. If half of these sets are abnormal behavior, the decision is right.

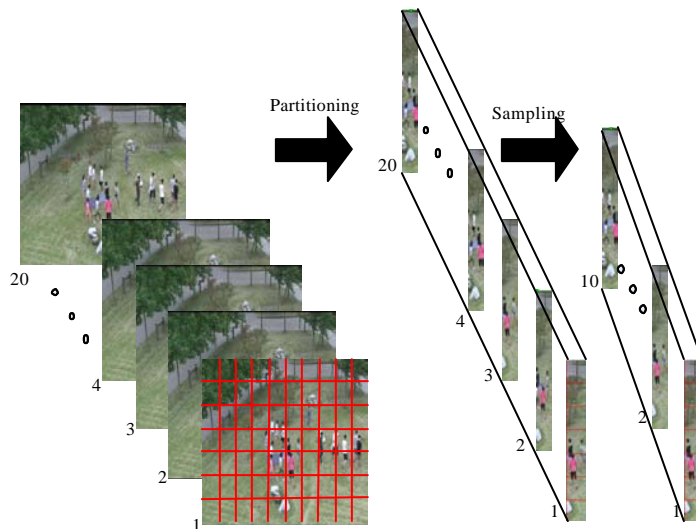


Fig. 4: Process of collecting sub-samples

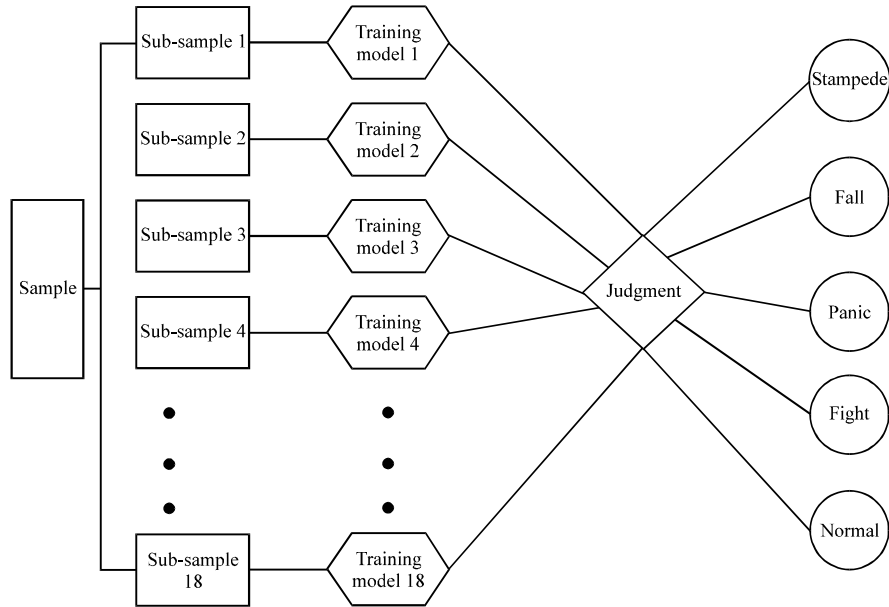


Fig. 5: Process of training samples

Table 1: Decision for stampede sample

Behaviors	Sub-sample																		Correct recognition total	Final output
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18		
Panic	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	2	
Fight	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	
Normal	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	2	
Stampede	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0	0	0	13	
Competition output	S	P	F	S	S	S	S	S	N	S	S	S	S	S	S	N	P	F	13>Thr	S

**EXPERIMENTS RESULT**

There has no database of the generic crowd abnormal behavior which is different from face and finger recognition. So in this study, a video shot by Sony HVR-V1C is used for experiment. Because of the diversity of crowd movement in direction and research still not mature, this study have made some assumptions for data acquisition, in order to reduce the influence of other interference problem. These assumptions are as follows:

- Camera motionless
- The capturing direction must have a consistency with the direction of movement of the person
- The proportion of individual must be approximately equal to each other

Test database contains three kinds of abnormal behavior data, such as panic, fighting, stampede, in relative to the camera at a fixed Angle (corresponding to the image plane is about 45 degrees), each sequence of frame number between 120 and 260. The database contains 3139 frames.

The study have tested some video scene shot in campus, the interception of video frames in Fig. 6 as follows: Fig. 6a is a group of people walking around on the lawn normally, Fig. 6b is several people fall on the lawn in a sudden, Fig. 6c shows the crowd suddenly spread around due to panic; Fig. 6d shows two groups of people are fighting on the lawn. The test result implies that the study can correctly identify different behaviors in crowd scenes and achieve the desired results.

In order to verify the validity of the method of this article, a kind of experiment is used to test the model using in the same scene which descript stampede behavior in the video sequence. Ten normal behavior samples and stampede behavior samples are extracted respectively from the ten videos. Using competitive neural network model, one result of stampede is shown in Table 1. With the fixed training methods, the video sample sequence contained 20 frames produce four decision results such as Panic, Fight, Normal and Stampede which is marked by the English initials, respectively.

In this experiment, Thr = 9 is selected as the threshold. From the Table 1, the result can clearly see that



Fig. 6(a-d): Four kind's detection results of crowd behavior scene (a) Normal, (b) Stampede, (c) Panic and (d) Fight

the right competition output value is 13 by judging in the 18 samples, thereby the sample can be judged to stampede condition.

### CONCLUSION

Using competitive neural network model, this study introduce a novel approach to detect abnormal behaviors in crowd scenes. A kind of feature extraction and description method has been put forward for particle flow information about crowd motion applying to space-time feature cubes firstly. Then a detection algorithm is proposed to detect abnormal events in global region crowd which combined with space-time feature cubes and competitive neural network model. Experimental results obtained by using our test video sequences have shown that our method is capable of detecting and locating the abnormal crowd behaviors.

### ACKNOWLEDGMENTS

This study was supported by Zhejiang Provincial NSF (Grant No. Y1110506). The work was also partially funded by the Opening Project of Shanghai Key Laboratory of Integrate Administration Technologies for Information Security and the National Natural Science Foundation of China (Grant No. 61102132).

### REFERENCES

- Andrade, E.L., S. Blunsden and R.B. Fisher, 2006. Modelling crowd scenes for event detection. Proceedings of the 18th International Conference on Pattern Recognition, Volume 1, August 20-24, 2006, Hong Kong, pp: 175-178.
- Hazel, G.G., 2000. Multivariate gaussian MRF for multispectral scene segmentation and anomaly detection. IEEE Trans. Geosci. Remote Sens., 38: 1199-1211.
- Ihaddadene, N. and C. Djeraba, 2008. Real-time crowd motion analysis. Proceedings of the 19th International Conference on Pattern Recognition, December 8-11, 2008, Tampa, FL., USA., pp: 1-4.
- Khan, Z.A. and W. Sohn, 2011. Abnormal human activity recognition system based on R-transform and kernel discriminant technique for elderly home care. IEEE Trans. Consumer Electron., 57: 1843-1850.
- Kim, C. and J.N. Hwang, 2002. Object-based video abstraction for video surveillance systems. IEEE Trans. Circuits Syst. Video Technol., 12: 1128-1138.
- Liu, Y., L.K. Cormack and A.C. Bovik, 2011. Statistical modeling of 3-D natural scenes with application to bayesian stereopsis. IEEE Trans. Image Process., 20: 2515-2530.
- Mahadevan, V., W. Li, V. Bhalodia and N. Vasconcelos, 2010. Anomaly detection in crowded scenes. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA, USA., pp: 1975-1981.
- Marques, J.S., P.M. Jorge, A.J. Abrantes and J.M. Lemos, 2003. Tracking groups of pedestrians in video sequences. Proceedings of the Conference on Computer Vision and Pattern Recognition Workshop, Volume 9, June 16-22, 2003, Madison, WI., USA., pp: 101-101.
- Mehran, R., A. Oyama and M. Shah, 2009. Abnormal crowd behavior detection using social force model. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL., USA., pp: 935-942.
- Sand, P., 2006. Long-range video motion estimation using point trajectories. Int. J. Comput. Vision, 80: 72-91.
- Wang, X., K. Tieu and E.L. Grimson, 2010. Correspondence-free activity analysis and scene modeling in multiple camera views. IEEE Trans. Pattern Anal. Mach. Intelli., 32: 56-71.
- Wu, S., B.E. Moore and M. Shah, 2010. Chaotic invariants of Lagrange particle trajectories for anomaly detection in crowded scenes. Proceedings of the 23rd IEEE Conference on Computer Vision and Pattern Recognition, June 13-18, 2010, San Francisco, CA., USA., pp: 2054-2060.
- Xiang, T. and S. Gong, 2008. Video behavior profiling for anomaly detection. IEEE Trans. Pattern Anal. Mach. Intelli., 30: 893-908.
- Zhang, D., F. Chen and H. Peng, 2013. Detecting group-level crowd using spectral clustering analysis on particle trajectories. Inf. Technol. J., 12: 174-179.