

# Abnormal Crowd Motion Detection with Hidden Conditional Random Fields Model

Dongping Zhang, Kaihang Xu, Yafei Lu, Chen Pan and Huailiang Peng

College of Information Engineering China Jiliang University, Hangzhou 310018,  
China,  
[silenttree\\_zju@cjl.u.edu.cn](mailto:silenttree_zju@cjl.u.edu.cn)

## Abstract

*Crowd motion analysis in public places is an important research subject in the monitoring field. This paper proposes an approach for detecting abnormal crowd motion using Hidden Conditional Random Fields Model (HCRF). This approach derives variations of motion patterns from direction distribution of the crowd motion obtained by the optical flow and these variations are encoded with HCRF to allow for the detection of abnormal crowd motion. Modeling the temporal neighborhood relations in a video sequence based on HCRF can incorporate hidden states and label the video depending on long range observations. The experimental results show that this proposed algorithm can achieve better results than HMM and CRF.*

**Keywords:** *Abnormal Crowd Motion Detection, Hidden Conditional Random Fields Model, Optical flow, HMM*

## 1. Introduction

Crowd security management in public places is increasingly important in social activities [1-2]. Monitoring crowd through video surveillance system is a direct and extensive way in our social life. But it often needs artificial identification in the existing monitoring system. An intelligent monitoring system which can be used to identify abnormal crowd behavior automatically is needed in future development. This is also a main research task in the computer vision and pattern recognition field. The motion of crowd is a complex one which consists of many people and may exhibit long-range dependence. So this is a big challenge to make abnormal crowd motion detection in computer vision.

Since several years ago, many state-space models have been proposed for modeling crowd motion in video sequences, due to the continuous and time-ordered property of movement [3], such as MRF, HMM, CRF, and so on.

As a statistical model with finite state, HMM (Hidden Markov Model) is the most representative algorithm in state-space models [4]. It is a generative model that includes hidden state structure. HMM has been widely applied to voice processing [5]. Similar to the voice signal, crowd motion is also a time-series signal which is smooth in short-term and stable in long-term. So it is feasible to model and analyze crowd motion using HMM. But HMM assumes that the observations are conditionally independent. This restriction makes it difficult to model long-range dependencies among observations or multiple overlapped features of these observations.

To overcome the shortcoming of HMM, Conditional Random Fields (CRF) was first proposed by Lafferty [6]. The main advantage of the CRF framework is that it relaxes the restriction of conditional independence of the observations. CRF can choose contextual features to solve the problem of the tag bias, and can get the global optimal value through

globally normalizing all the characteristics. Because of these good characteristics, CRF has been widely used in the vision community [7-8].

Quattoni [9] introduced hidden variable into CRF, which is Hidden Conditional Random Fields Model (HCRF). The structure of HCRF may be more expressive than CRF which is completely observable [10]. The experiments by Quattoni proved the superiority of HCRF than the generative hidden state model and discriminant completely observable model.

In this paper, we propose an approach for abnormal crowd motion detection using HCRF. Our approach derives variations of motion patterns through direction distribution of the crowd motion obtained by optical flow. The detection can be seen as a problem of binary classification (*i.e.* normal and abnormal cases), of which an abnormal motion is discriminatively trained against all other normal motion. HCRF incorporates long-range dependence and can describe the dependencies between crowd motions effectively.

In Section 2, the methodology is describe in detailed. Experiments in Sections 3 show that the proposed method can detect the abnormal crowd motion effectively, and then present a comparative evaluation of different models.

## 2. Motion Feature Extraction

In the pre-processing, the mask of motion region is obtained using the Mixture of Gaussians background model (GMM) per frame. A Gaussian filter is used to smooth the frame sequence to reduce the acquisition noise prior to the optical flow computation.

This paper extracts optical flow as motion feature to build observation vectors. The optical flow method is often used as a description of the events in complex scenes nowadays [11]. Optical flow vectors have a good description of the change rate and direction of the pixel gray values and indirectly reflect the target moving speed and direction. So we can use optical flow features to build different motion patterns of th crowded motion. In our approach, an observation vector is built from the motion vectors of each frame. We divide the angle  $2\pi$  into 12 levels, and each level has an increase of  $\frac{\pi}{6}$ . Each observation vector has 12 dimensions on behalf of these 12 directions, respectively. The direction probability distribution of each frame can be directly estimated from the direction. We compute the velocity and angle of every motion vector of a certain frame and then put the total number of motion vectors with the same change in the angle range into the corresponding dimension of the observation vector. And the velocity is weighed to the number of certain angle. The observation vector of a frame is obtained by this method, and the same processing is performed to all frames which we will train and test. The calculated observation vector indicates the direction distribution and direction tendency of the frame. To have an intuitive representation of observation vector, we draw a direction histogram of the vector field. The direction histograms are displayed in Figure 1.

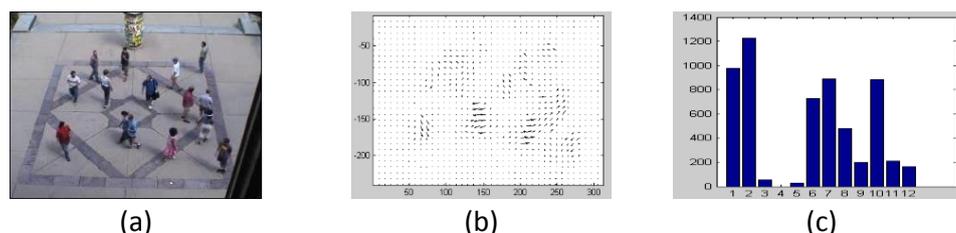


Figure 1. (a) Original Image (b) Vector Field (c) Direction Histograms

### 3. HCRF Model

Leave two blank lines after the title. Leave two blank lines after the title. Suppose there are in training video sequences  $X = \{x_1, x_2, \dots, x_n\}$ , where  $x_i$  is the  $i$ -th video sequence containing  $m$  frames, and each frame is of  $d$  dimensions, *i.e.*  $x_i = \{x_{i,1}, x_{i,2}, \dots, x_{i,m}\}$ , where  $x_{i,j} \in R^d (j=1, \dots, m)$ .

Meanwhile,  $Y = \{y_1, \dots, y_n\}$  is the corresponding label set for  $X$ , where  $y_i \in \{0,1\}$ .  $y_i = 0$  represents  $x_i$  belonging to the normal case, while  $y_i = 1$  stands for the abnormality.

We assume a set of invisible latent variables  $H = \{h_1, h_2, \dots, h_m\}$  for each video sequence. In HCRF and HMM, the element  $h_j$  corresponds to the frame  $x_{i,j} (i=1, \dots, n)$ .

Similar to CRF proposed by Lafferty in [6], in the training process, we estimate parameters using the following objective function:

$$L(\theta) = \sum_{i=1}^n \log P(y_i | x_i, \theta) - \frac{1}{2\sigma^2} \|\theta\|^2 \quad (1)$$

The first term in (1) is the log-likelihood function, *i.e.*

$$P(y | x, \theta) = \sum_h P(y, h | x, \theta) = \frac{\sum_h e^{\psi(y, h, x; \theta)}}{\sum_{y', h} e^{\psi(y', h, x; \theta)}} \quad (2)$$

The second term in (1) is the log of a Gaussian prior, *i.e.*  $p(\theta) = \exp(-\frac{1}{2\sigma^2} \|\theta\|^2)$ .

Based on this standard, we search the optimal parameter values using gradient descent method,  $\theta^* = \arg \max_{\theta} L(\theta)$ . As other models with hidden states (*e.g.*, HMM), the hidden states make the optimization non-convex. We search for parameters and local optimal solution through initializing multiple random starting points and calculate the mean value of the parameters.

We represent the structural constraints with the structure of an undirected graph, in which hidden variables  $\{h_1, \dots, h_m\}$  stand for vertexes of a graph  $E$ . And  $E$  can be arbitrary, but it should capture the structure of  $H$  in any domain. A set of edges  $(j, k) \in E$  represents links connecting hidden variables  $h_j$  and  $h_k$ . In the task of abnormal crowd motion detection,  $E$  is a chain which can capture the temporal dynamics.

In our proposed HCRF model, window function  $\varpi$  is used for representing the number of observations in the past and the future which predicts the state at frame  $x_j$ .

No long-range dependence case (*i.e.*  $\varpi = 0$ )

In Eq.(2)  $\psi(y, h, x; \theta) \in R$  is a potential function parameterized by  $\theta$ , *i.e.*

$$\psi(y, h, x; \theta) = \sum_{j=1}^m \sum_l f_l^1(j, y, h_j, x) \theta_l^1 + \sum_{(j,k) \in E} \sum_l f_l^2(j, k, y, h_j, h_k, x) \theta_l^2$$

Where  $f_l^1, f_l^2$  are functions defining the features in the model, and  $\theta_l^1, \theta_l^2$  are the components of  $\theta$ . The  $f_l^1$  feature depends on single hidden variable values in the model, and the  $f_l^2$  features depend on pairs of values. If a feature does depend on variables  $h_j$  and  $h_k$ , there must be an edge  $(j, k)$  in  $E$ .

Next step is for test. Given a query video sequence  $x$  and the parameter  $\theta^*$  derived from training sequences, we can make the judgment that  $x$  is normal if  $P_{y=0}(y|x, \theta^*) > P_{y=1}(y|x, \theta^*)$ , and vice versa.

Long-range dependence case (*i.e.*  $\varpi = 1, 2$ )

We define a potential function  $\psi(y, h, x; \theta, \varpi) \in R$  which contains parameters  $\theta$  and  $\varpi$ :

$$\psi(y, h, x; \theta, \varpi) = \sum_{j=1}^m \varphi(x, j, \varpi) \cdot \theta_h[h_j] + \sum_{j=1}^m \theta_y[y, h_j] + \sum_{(j,k) \in E} \theta_e[y, h_j, h_k] \quad (3)$$

$\varphi(x, j, \varpi)$  is a vector that can include any feature of the observation sequence for a specific window size  $\varpi$  (*e.g.* observations from  $j - \varpi$  to  $j + \varpi$  is used for feature computing for window size  $\varpi$ ). Parameter vector  $\theta$  is consist of three parts:  $\theta = [\theta_e, \theta_y, \theta_h]$ . Symbol  $\theta_h[h_j]$  stands for parameter  $\theta_h$  and its related state  $h_j \in H$ . Similarly,  $\theta_y[y, h_j]$  represents the parameter corresponding label  $y$  and state  $h_j$ ,  $\theta_e[y, h_j, h_k]$  represents parameter corresponding label  $y$  and a pair of states  $h_j$  and  $h_k$ .

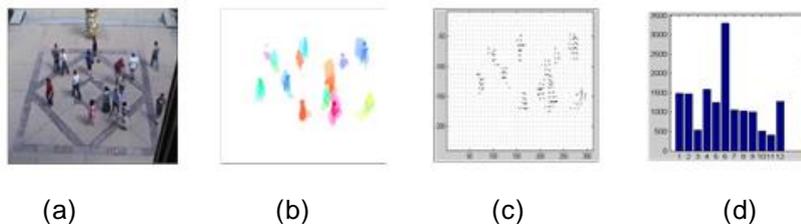
Inner product  $\varphi(x, j, \varpi) \cdot \theta_h[h_j]$  can be interpreted as a measure of compatibility between observation sequence and state at frame  $j$  under the window size  $\varpi$ . The parameter  $\theta_y[y, h_j]$  can be interpreted as a measure of compatibility between a state  $k$  and crowd motion  $y$ . Finally, every parameter  $\theta_e[y, h_j, h_k]$  measures the compatibility between a pair of coherent states  $h_j$  and  $h_k$  and crowd motion  $y$ .

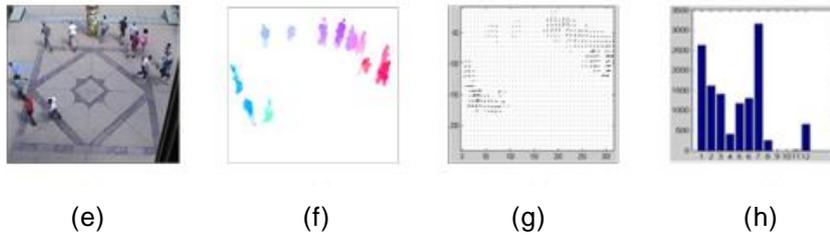
Given a new test sequence  $x$  and parameter  $\theta^*$  learned from training samples, very similar to the no long-dependence case,  $x$  is considered normal if  $P_{y=0}(y|x, \theta^*, \varpi) > P_{y=1}(y|x, \theta^*, \varpi)$ , and vice versa.

As  $E$  is a chain, there exists a certain method to inference and parameter estimation. The objective functions of HCRF, *i.e.* Eq.(1) can be described as marginal distributions of hidden state variables. These distributions can be computed using belief propagation.

## 4. Experimental Results

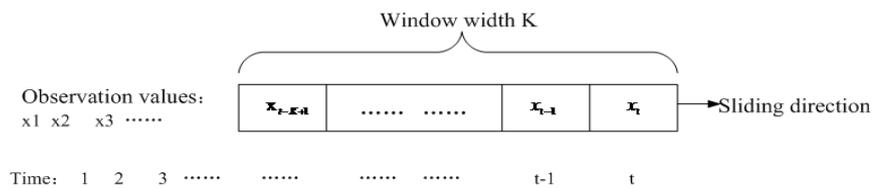
In order to evaluate the efficiency of our proposed method, the experiment has been done on the UMN dataset [12]. This dataset includes some scenes of crowded scatter events. Each video consists of two cases, *i.e.* normal and abnormal situations, such as the crowd suddenly changes their motion state from walk to scatter in different directions. Figure 2 (a) and Figure 2 (e) are example frames which represent normal and abnormal behaviors. Figure 2 (b) and Figure 2 (f) are paintings drawn according to the different speed of movement of the persons. Figure 2 (c) and Figure 2 (g) are vector fields of two frames. Figure 2 (d) and Figure 2 (h) are the direction histograms.





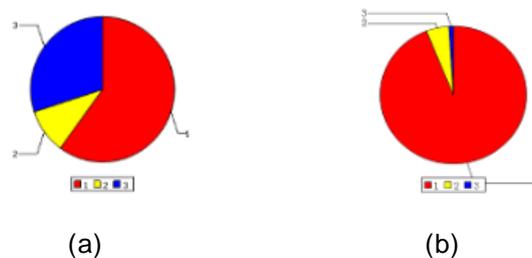
**Figure 2. (a) and (e) are Example Frames which Represent Normal and Abnormal Behaviors. (b) and (f) are Paintings Drawn According to the different Speed of Movement of the Persons. (c) and (g) are Vector Fields. (d) and (h) are the Direction Histograms**

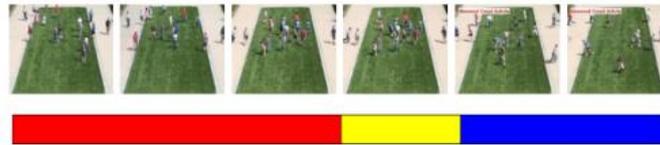
As the needed to give real-time test results rather than analyze the whole video before the final result is given, we select a specific length of the observed sequence and update the values of the sequence in real time. Here we give a length of window  $K = m$  of observed sequence, the  $K$  values in the window is initialized to 10. The process is equivalent to slide window on the observation sequence, the newest observation value enters the window and the oldest observation value slide out of the window on the arrival of the new frame, shown in Figure 3. A constantly updated finite-length sequence of observations has gotten in this way.



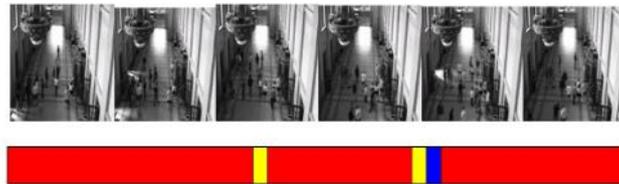
**Figure 3. Build and Update the Observation Sequence**

HCRF model is trained to discriminate the normal motion with the abnormal motion. Three hidden states are used in the training process. In order to incorporate long range dependencies in HCRF model, we change the value of  $\omega$  from 0 to 2 to including past and future history. The parameters of HCRF are trained using maximum likelihood. Figure 4 is the distribution of hidden states in normal and abnormal video sequence, of which (a) and (c) are distributions of normal case, (b) and (d) are distributions of abnormal case. The distribution is obtained by searching the Viterbi path. The number 1, 2 and 3 represent the three hidden states assigned by HCRF model and the area of each colour represents its corresponding proportion. As shown in (a) and (b), the HCRF model can find a unique distribution of states in normal and abnormal situations. The state assigned for each video clip is illustrated in (c) and (d), in which visually unique crowd motion is assigned distinct states.





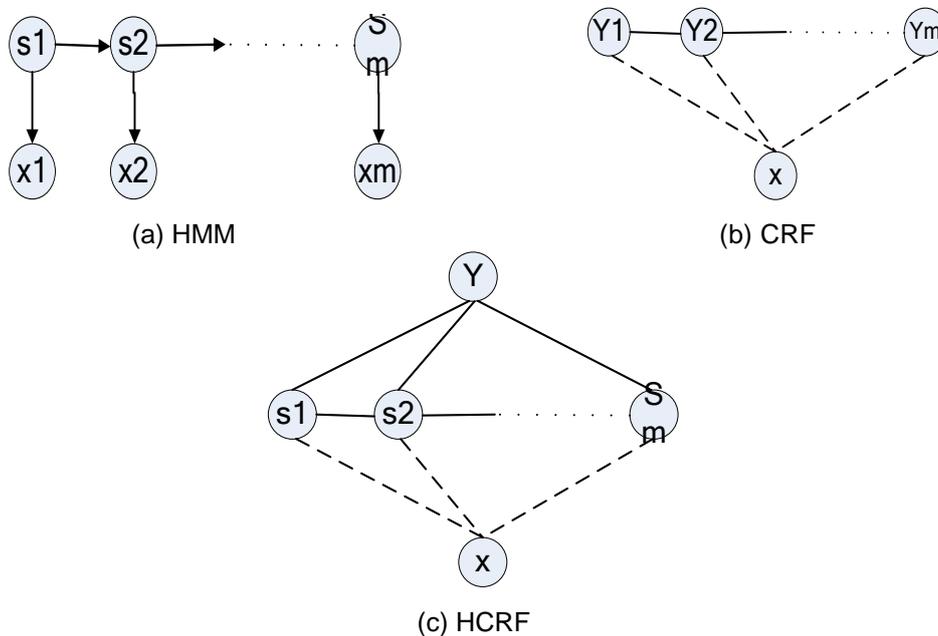
(c)



(d)

**Figure 4. This Graph Shows the Distribution of Hidden States in Video Sequence, (a) and (c) are Distributions of Normal (b) and (d) are Distributions of Abnormal the Number 1, 2 and 3 Represent the Category of Hidden State Assigned by HCRF Model and the Area of Each Category Represents the Proportion**

In order to compare the effect of different models, we take the same dataset for experiments using HMM model and CRF model, and statistic average accuracy. Figure 5 shows the graph of these models used in our experiment. The Gaussian cluster method is used by Shu Wang [13]. The accuracies of different models are shown in Table 1. CRF has a better performance than HMM when window size  $\omega = 0$  and has a lower performance when window size  $\omega = 1$ . HCRF has a great improvement in performance than HMM and CRF. And the increase of  $\omega$  shows the long-range dependencies is useful.



**Figure 5. The State Transition of a Single Video Sequence in (a) HMM (b) CRF and (c) HCRF**

## 5. Conclusion

In this paper, we have presented an approach to estimate the abnormality of a crowd scene. The method is based on Hidden Conditional Random Fields Model (HCRF) and represents different motion patterns with direction distribution. The experiments on two different scenarios show that our method is able to detect the abnormal crowd motion effectively. In future work, we will involve more motion characteristics such as velocity and acceleration in the abnormality detection of a crowd scene.

## Acknowledgements

This work was supported by Zhejiang Provincial NSF (Grant No.LY15F020021) and Zhejiang Provincial Science & Technology Research Program (Grant No.2013C33046, 2014C33075).

## References

- [1] Pan X., Han C. S., Dauber K. and Law K. H., "Human and social behavior in computational modeling and analysis of egress", *Automation in construction*, vol. 15, no. 4, (2006), pp. 448-461.
- [2] W. Lijun and M. Dong, "Real-time detection of abnormal crowd behavior using a matrix approximation-based approach", *Image Processing (ICIP)*, 19th IEEE International Conference on. IEEE, (2012).
- [3] Zhan B., Monekosso D. N., Remagnino P., Velastin S. A. and Xu, L. Q., "Crowd analysis: a survey", *Machine Vision and Applications*, vol. 19, no. 5-6, (2008), pp. 345-357.
- [4] Yamato J., Ohya J. and Ishii K., "Recognizing human action in time-sequential images using hidden markov model", *Computer Vision and Pattern Recognition*, 1992. Proceedings CVPR'92. 1992 IEEE Computer Society Conference on. IEEE, (1992), pp. 379-385.
- [5] Bui H. H., Phung D. Q. and Venkatesh S., "Hierarchical hidden Markov models with general state hierarchy", *Proceedings of the National Conference on Artificial Intelligence*. Menlo Park, CA; Cambridge, MA; London; AAAI Press; MIT Press; vol. 1999, (2004), pp. 324-329.
- [6] Lafferty J., McCallum A. and Pereira F. C. N., "Conditional random fields: Probabilistic models for segmenting and labeling sequence data", (2001).
- [7] Bousmalis K., Zafeiriou S. and Morency L. P., "Infinite hidden conditional random fields for human behavior analysis", *Neural Networks and Learning Systems*, IEEE Transactions on, vol. 24, no. 1, (2013), pp. 170-177.
- [8] Bousmalis K., Zafeiriou S. and Morency L. P., "Variational hidden conditional random fields with coupled Dirichlet process mixtures", *Machine Learning and Knowledge Discovery in Databases*. Springer Berlin Heidelberg, (2013), pp. 531-547.
- [9] Quattoni A., Wang S. and Morency L. P., "Hidden conditional random fields", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 10, (2007), pp. 1848-1852.
- [10] Wang S. B., Quattoni A., Morency L. P., Demirdjian D. and Darrell T., "Hidden conditional random fields for gesture recognition", *Computer Vision and Pattern Recognition*, 2006 IEEE Computer Society Conference on. IEEE, vol. 2, (2006), pp. 1521-1527.
- [11] Ince I. F., Socarras G. M and Yang T. C., "Hand mouse: real time hand motion detection system based on analysis of finger blobs", *International Journal of Digital Content Technology and its Applications*, vol. 4, no. 2, (2010).
- [12] "Unusual crowd activity dataset of University of Minnesota", available from <http://mha.cs.umn.edu/movies/crowd-activity-all.avi>.
- [13] Cao T., Wu X. and Guo J., "Abnormal crowd motion analysis", *Robotics and Biomimetics (ROBIO)*, 2009 IEEE International Conference on. IEEE, (2009), pp. 1709-1714.

## Authors



**Dongping Zhang**, He was born in 1970. He received the PhD in Information & Communication Engineering from Department of Information Science & Electronic Engineering, Zhejiang University, Hangzhou, China, in 2006. Since 2006, He is an associate professor at College of Information Engineering, China Jiliang University in Hangzhou. His research interests include image processing and pattern recognition, computer vision and videos.



**KaiHang Xu**, He was born in 1991, and graduated at the China Jiliang University in Hangzhou in 2013. Currently he is a MSc. student at the College of Information Engineering of China Jiliang University. His research interests are oriented to image processing and image compression for security and multimedia applied imaging systems.



**YaFei Lu**, He was born in 1988, and graduated at the China Jiliang University in Hangzhou in 2011. Currently he is a MSc. student at the College of Information Engineering of China Jiliang University. His research interests are oriented to image processing and image compression for security and multimedia applied imaging systems.



**Pan Chen**, He was born in 1966. He received the Ph.D. degree in biomedical engineering from Xi'an Jiaotong University in 2005. Now he is a professor at China Jiliang University. His research interests include machine vision, image processing and pattern recognition, *sec.*



**HuaiLiang Peng**, He was born in 1988, and graduated at the China Jiliang University in Hangzhou in 2011. Currently he is a MSc. student at the College of Information Engineering of China Jiliang University. His research interests are oriented to image processing and image compression for security and multimedia applied imaging systems.