Hidden Markov Models for Optical Flow Analysis in Crowds

Ernesto L. Andrade[1], Scott Blunsden[2] and Robert B. Fisher[1]
IPAB, School of Informatics, University of Edinburgh
King's Buildings, Mayfield Road, Edinburgh, EH9 3JZ, UK
[1]eaneto,rbf@inf.ed.ac.uk, [2]S.J.Blunsden@sms.ed.ac.uk

Abstract

This paper presents an event detector for emergencies in crowds. Assuming a single camera and a dense crowd we rely on optical flow instead of tracking statistics as a feature to extract information from the crowd video data. The optical flow features are encoded with Hidden Markov Models to allow for the detection of emergency or abnormal events in the crowd. In order to increase the detection sensitivity a local modelling approach is used. The results with simulated crowds show the effectiveness of the proposed approach on detecting abnormalities in dense crowds.

1 Introduction

There is nowadays a stronger demand for automated video surveillance systems which can infer or understand more complex behaviour and scene semantics [4]. In particular large-scale video surveillance of public places, which has a huge amount of data to be monitored, would benefit a system capable of recognising hazardous and anomalous situations to alert the system operators. There are many application for such systems in emergency detection and surveillance scenarios. In this work we concentrate specifically on monitoring emergency situations in crowds by learning patterns of normal crowd behaviour in order to identify unusual or emergency events. These events are of great interest for surveillance purposes and generate disturbances in normal flow pattern. For instance, someone falling over, or a fight disruption in the middle of the crowd changes the flow pattern and might locally alter crowd flow density. Previous work in the analysis of crowds usually assumes that individuals can be tracked and identified inside the crowd [10]. Most systems only analyse crowd densities and distributions [5] aiming to derive statistics from the crowd for traffic planning. There are few publications addressing the detection of complex events within the crowd. In [8] the analysis looks for pre-defined circular flow patterns in the crowd to characterise potential emergency situations. To the best of our knowledge this work is one of the first attempts to interpret optical flow patterns of a human crowd by composing a model of crowd motion from training data. The optical flow variations of typical sequences are observed to characterise a normal crowd activity model (section 2), concentrating the analysis only on significant motion in the foreground areas. This information is used to train a Hidden Markov Model (HMM) which learns the variations in optical flow pattern allowing discrimination of abnormal behaviour. In the computer vision literature HMMs have been extensively used for gesture recognition, and interpretation of human interactions [6] and activities [3]. Our work further extends the use of HMMs to the analysis of optical flow patterns from human crowds. We show that there is sufficient perturbation in the optical flow pattern emergencies and abnormal events are detected (sections 3 and 4).

2 Crowd Flow Analysis

The flow analysis involves three phases: 1) Preprocessing and Feature Extraction: background modelling and optical flow computation; 2) HMM training: parameter estimation for a Mixture of Gaussians Hidden Markov Model in two scales; and 3) Anomaly detection: the analysis consisting of identifying unusual events in the crowd by comparing the new observations' likelihood to a detection threshold. Details of this are given in the next subsections.

2.1 Preprocessing

The change detection module starts with the adaptive mixture of Gaussians algorithm described in [9]. The resulting mask then gates with the output of the optical flow calculation. Prior to the optical flow calculation a 5x5x5 Gaussian spatio-temporal filter ($\sigma = 0.8$) is applied for noise reduction. The optical flow calculation module implements the robust dense optical flow method described in [2]. Although more computationally expensive, it provides a smooth optical flow at the motion boundaries, making it

an ideal candidate to evaluate the usefulness of flow information. The resulting optical flow is decimated using a median filter in 8x8 windows to further reduce noise and the number of flow vectors inside the model. The combination of flow information with the foreground mask allows the analysis modules to only consider flow vectors inside foreground objects, reducing observation noise. All the flow vectors outside the foreground mask are set to zero to emphasise the static areas.

2.2 Hidden Markov Models

HMMs [7] and related graphical models are a ubiquitous tool for modelling time series data. In order to encode optical flow spatio-temporal variations a HMM with mixture of Gaussians (MOGHMM) is used. The formalisation for the HMM with mixture of Gaussians output is based on [7]. The observation vector is defined as $\mathbf{O} = [O^1_{T_1}, O^2_{T_2}, ..., O^K_{T_K}]$, allowing K multiple observation sequences, each observation sample k at time t is a vector $O^k_t = (x, y, u, v)$, where x and y are the pixel position and u and v are the horizontal and vertical optical flow components. The model parameters to be determined by the Expectation-Maximisation (EM) algorithm are $\lambda = (\pi_i, a_{ij}, c_{im}, \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})$, where π_i is the prior probability for state $i = 1..N$, a_{ij} is the state transition matrix ($i = 1..N; j = 1..N$), c_{im} is the mixture coefficient, $\boldsymbol{\mu}_{im}$ is the mean vector and $\boldsymbol{\Sigma}_{im}$ is the full covariance matrix for Gaussian m in state i, with each state having a bank of M Gaussian ($m = 1..M$).

The probability of being in state i at time t is

$$\gamma_t(i) = \frac{\alpha_t(j)\beta_t(j)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)} \tag{1}$$

where α and β are the forward and backward variables [7]. The probability that an observation is generated by Gaussian m in state i at time t is

$$\kappa_t(i,m) = \left[\frac{\alpha_t(j)\beta_t(j)}{\sum_{i=1}^{N} \alpha_t(i)\beta_t(i)}\right]\left[\frac{c_{im}\aleph(\mathbf{O}; \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})}{\sum_{m=1}^{M} c_{im}\aleph(\mathbf{O}; \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im})}\right] \tag{2}$$

where $\aleph$ is the Gaussian pdf, with

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{im}\aleph(\mathbf{O}; \boldsymbol{\mu}_{im}, \boldsymbol{\Sigma}_{im}), \;\; 1 \leq j \leq N \tag{3}$$

The update equations for the EM procedure are:

$$\hat{\pi}_i = \frac{\sum_{k=1}^{K} \gamma^k_1(i)}{K} \tag{4}$$

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \xi^k_t(ij)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \kappa^k_t(i)} \tag{5}$$

where $\xi^k_t(ij)$ is the transition probability from state i to state j.

$$\hat{c}_{im} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \gamma^k_t(i,m)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\sum_{m=1}^{M} \gamma^k_t(i,m)} \tag{6}$$

$$\hat{\boldsymbol{\mu}}_{im} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \gamma^k_t(i,m).\mathbf{O}^k_t}{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \gamma^k_t(i,m)} \tag{7}$$

$$\hat{\boldsymbol{\Sigma}}_{im} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \gamma^k_t(i,m).(\mathbf{O}^k_t - \boldsymbol{\mu}_{im})(\mathbf{O}^k_t - \boldsymbol{\mu}_{im})'}{\sum_{k=1}^{K}\sum_{t=1}^{T_k} \gamma^k_t(i,m)} \tag{8}$$

In the optical flow modelling the mixture of Gaussians emissions encode the spatial distribution of motion clusters present in the training set. Whereas the coefficients in the transition matrix (a_{ij}) encode the transitions between the observed motion patterns.

3 Training

The models are trained using simulated crowd flow data for a dense crowd. The crowd simulation is based on [1]. The training sequence representing normal behaviour is composed of 6000 frames where a dense crowd moves across the scene in one direction. The original frame size is 384x288 pixels which after preprocessing results in a 48x36 optical flow field. An example of the simulated sequences used in the modelling is shown in Fig. 1. The MOGHMMs defined in the previous subsection are trained with the optical flow observations using two distinct structures. In the first structure, named *global*, only one MOGHMM is trained for the whole frame capturing the global changes in the motion patterns. The second structure, named *local*, divides the image in blocks of size b_w=4 and b_h=4 and one MOGHMM is assigned to each block. This allows the detection of smaller variations of the motion pattern. For the global structure the MOGHMM topology is ergodic with $N = 10$ states with $M = 10$ Gaussians per state. The local model has the same topology with $N = 4$ states with $M = 4$ Gaussians per state in each block. For the local model each block in the optical flow field has a size of 4x4 flow vectors segmenting the flow field into 108 blocks with one MOGHMM per block. The number of states and gaussians in the HMM is determined empirically by selecting an HMM structure with the best likelihood for the training set among a set of different configurations of number of states and Gaussians per state.

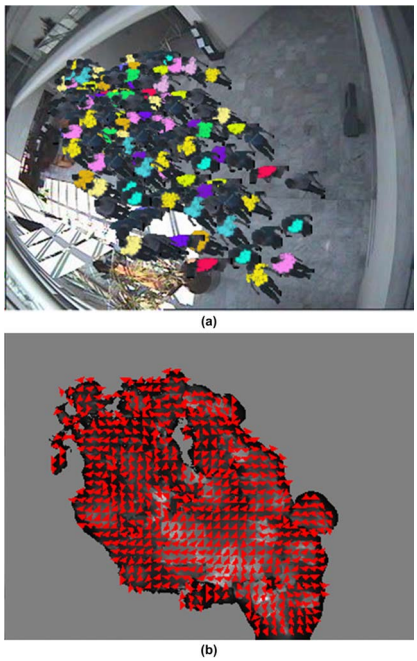

Figure 1. Crowd simulation example. (a) Simulated frame. (b) Simulated optical flow, lighter shades indicate larger displacement and arrows indicate flow direction per block.

4 Experimental Results

The experiments to evaluate the sensitivity of the HMM for abnormality detection consist of comparing the model for the normal crowd flow against two emergency scenarios. The first emergency event is the simulation of a blocked exit in the scene, where after the event people density in the scene starts to increase and the motion becomes more constrained whilst the people push each other against the blocked exit. The second emergency event is a person falling on the floor, which changes the trajectory of the other persons whilst they try to avoid stepping over the fallen person. Prior to the events the motion of the persons in the scene is similar in speed and direction to the training set. Both events occur at frame 2000 in the simulated test sequences. In all the experiments the window size to compute the model likelihood is 25 frames (1 second). For the sensitivity test five independent occurrences of each event are simulated resulting in a total of 10 test sequences of 3000 frames each.

The variations of the global model likelihood are shown in Fig. 2. For the blocked exit event Fig. 2.(a) the likelihood quickly drops to a level below the oscillations of the normal crowd behaviour. Whereas for the person falling event (see Fig. 2.(b) the global model is not able to detect the small perturbation in the model caused by the fallen person. Table 1 summarises the statistics for the five test runs of the blocked exit scenario. For comparison the likelihood statistics are computed in the intervals before the event (frame 1000 to 2000) and after the event (frame 2001 to 3000). The difference between the normal and blocked exit scenarios is easily identifiable by observing the variations in the model log-likelihood response.

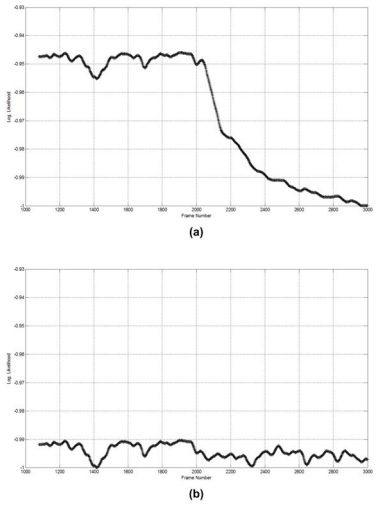

Figure 2. Global detection results. (a) Blocked exit scenario. (b) Person fall scenario.

The local model is applied to all 108 blocks in the flow field. Fig. 3 shows the model likelihood variations for the blocks adjacent to the area of the image where the person falls. We can note the sharp drop in model likelihood for the block which contains most of the person's body. No other significant likelihood drops were detected for the remaining 105 blocks. Table 2 shows an accentuated drop in likelihood for the local model in the proximity of the event location. To allow for on-line event detection the likelihood drops are measured with a simple edge filter on the likelihood function. Long lasting likelihood drops within the filter indicate the abnormal events. The filter delays are adjusted to provide the desired false alarm rate. The detection filter equation is:

$$D_e(t) = \left| \frac{\sum_{l=t-W_s/2}^{t} L(l)}{W_s/2+1} - \frac{\sum_{l=t+1}^{t+W_s/2} L(l)}{W_s/2} \right| \quad (9)$$

where t is the current frame, $W_s = 250$ is the observation window and $L(l)$ is the model log-likelihood for the l-th frame. Fig. 4 shows the response of this temporal edge filter for all five runs of the person fall event. The filter is applied to the likelihood response of each block around the area where the person falls. The only noticeable increases

	Mean	Std.
Before	-15.5178	0.0223
After	-16.1328	0.2317

Table 1. Blocked exit scenario. Loglikelihood mean and standard deviation for $R = 5$ independent simulation runs before and after the event.

	Block Position Relative to the Event					
	Left Block		Event Block		Right Block	
	Mean	Std.	Mean	Std.	Mean	Std.
Before	-1.0491	0.2168	-1.0351	0.1792	-1.1677	0.2751
After	-1.3658	0.2794	-2.3174	0.3366	-1.8767	0.2966

Table 2. Person fall scenario. Loglikelihood mean and standard deviation for $R = 5$ independent simulation runs before and after the event for the blocks around the person.

in the response are on the blocks close to the person falling and no other detections above 0.8 are present in the other blocks through the whole sequence.

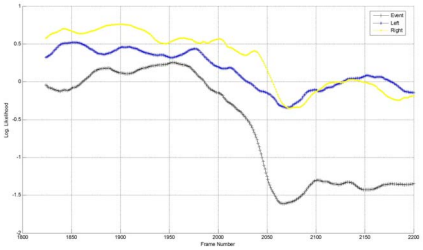

Figure 3. Local detection results.

5 Conclusions

We presented a framework for the analysis of crowd behaviour. It relies on optical flow information from video evidence to represent the crowd behaviour as optical flow variations in time. These variations are encoded in MOGHMMs, which allow detection of unusual events. Two different detection methods are implemented to detect *global* and *local* emergency scenarios. The experimental results show that MOGHMMs are able to detect emergency situations in a dense crowd. The assumption of having a dense crowd to detect enough flow changes can be lifted if the video sequences are modeled with banks of HMMs trained for normal motion in different density scenarios. This would be applicable from modelling of dense crowds to pedestrian traffic and is a subject for future work.

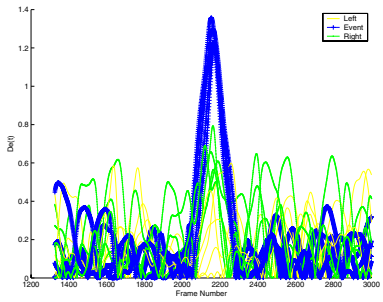

Figure 4. Person fall scenario $R = 5$ simulation runs. Response of the filter D_e with $W_s = 250$.

Acknowledgements

This work is funded by EPSRC's BEHAVE project GR/S98146.

References

[1] E. L. Andrade and R. B. Fisher. Simulation of crowd problems for computer vision. *First International Workshop on Crowd Simulation*, (3):71–80, 2005.

[2] M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. *4^{th} International Conference on Computer Vision*, pages 231–236, 1993.

[3] S. Gong and T. Xiang. Recognition of group activities using a dynamic probabilistic network. *Proceedings of the IEEE International Conference on Computer Vision*, pages 742–749, 2003.

[4] W. Hu, T. Tan, L. Wang, and S. Maybank. A survey on visual surveillance of object motion and behaviours. *IEEE Transactions on Systems, Man and Cybernetics - Part C: Applications and Reviews*, 43:334–352, 2004.

[5] B. Maurin, O. Masoud, and N. Papanikolopoulos. Monitoring crowded traffic scenes. *IEEE 5 th International Conference on Intelligent Transportation Systems*, pages 19–24, 2002.

[6] N. Oliver, A. Garg, and E. Horvitz. Layered representations for learning and inferring office activity from multiple sensory channels. *Computer Vision and Image Understanding*, 96:163–180, 2004.

[7] L. R. Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, 1989.

[8] A. L. S. A. Velastin, B. A. Boghossian. Detection of potentially dangerous situations involving crowds using image processing. *Proceedings of the Third ICSC Symposia on Intelligent Industrial Automation (IIA'99) and Soft Computing*, 1999.

[9] C. Stauffer and W. Grimson. Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):747–757, 2000.

[10] T. Zhao and R. Nevatia. Tracking multiple humans in complex situations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(9):1208–1221, 2004.