Characterisation of optical flow anomalies in pedestrian traffic

Andrade, E., Blunsden, S., & Fisher, B. (2005). Characterisation of optical flow anomalies in pedestrian traffic. In Imaging for Crime Detection and Prevention, 2005. ICDP 2005. The IEE International Symposium on . (pp. 73-78). doi: 10.1049/ic:20050073

Characterisation of Optical Flow Anomalies in Pedestrian Traffic

Ernesto L. Andrade, Scott Blunsden and Robert B. Fisher
School of Informatics, Edinburgh University, UK
eaneto@inf.ed.ac.uk, S.J.Blunsden@sms.ed.ac.uk, rbf@inf.ed.ac.uk

ABSTRACT – This paper applies a video modelling technique to a surveillance scenario where pedestrians are monitored to detect unusual events. The aim is to investigate the components of an automatic vision system capable of detecting normal and abnormal behaviour. Such a system has application in surveillance scenarios like town centre plazas, stadiums, train stations and shopping malls. Surveillance usually relies on tracking, but in crowded scenarios tracking is not reliable. Thus our framework for representation and analysis is based on optical flow to avoid tracking of individuals. We demonstrate that patterns derived from optical flow and encoded by a Hidden Markov Model are able to capture the dynamic evolution of normal behaviour allowing the classification of abnormal events.

INTRODUCTION

There has been increased activity in computer vision systems for surveillance applications. For a comprehensive review refer to Hu et al (1). Large-scale video surveillance of public places is a complex task aggravated by the huge amount of data to be monitored. The application of computer vision techniques simplifies monitoring in scenarios where hundreds of cameras require the attention of a few observers. There are different surveillance scenarios, which require dedicated vision systems of different complexity degrees, such as intrusion zone detection, car park security, and street monitoring. In this work we concentrate on a specific surveillance scenario, monitoring of pedestrians, aiming to learn normal patterns of pedestrian behaviour given video evidence in order to identify unusual events. These events are of main interest for surveillance purposes and represent disturbances in pedestrian flow patterns. For instance, someone falling over, or a fight disruption in the middle of a group changes the flow pattern and locally alters flow density. If such perturbations have enough resolution in the input image they can be interpreted. Previous work in the analysis of pedestrians usually assumes that individuals can be tracked and identified inside the crowd Zhao and Nevatia (2). Most systems only analyse pedestrians' densities and distributions (Maurin et al (3)) aiming to derive statistics for traffic planning. However, there is similar work in traffic monitoring where motion statistics derived from optical flow are used to characterise the typical behaviour of an intersection Brand and Kettnaker (4) and flag abnormalities in the traffic flow. Our system is based on the same assumptions as in (4), where objects can not be tracked individually either due to imprecision in tracking in their case or to the impossibility of tracking in the case of large pedestrian groups where people occlude each other. As in (4) we observe the optical flow variations of typical sequences to characterise normal pedestrian activity. Our framework also concentrates the analysis only where there is significant motion (e.g. larger than the interframe motion noise for frames with no people) of the foreground areas (people). One can adopt two distinct approaches to represent pedestrian behaviours, such as walking, running and turning, the system can either (i) explicitly model flow densities and flow changes considered to be normal, which requires a very constrained environment to be effective, or in the general case (ii) the flow pattern can be learnt by example observing footage from normal pedestrian behaviour. Here we employ learning with Hidden Markov Models (HMMs) to capture the normal variations in the input pattern of the optical flow. Allowing classification of normal and abnormal behaviour. In the computer vision literature HMMs have been extensively used in gesture recognition, and interpretation of human interactions Oliver et al (5) and activities Gong and Xiang (6). We aim to extend this analysis to human crowds. In this work we give the first step towards this goal by demonstrating that a mixture of Gaussians HMM is able to encode the optical flow dynamics of a pedestrian group.

To demonstrate this we construct a framework for feature extraction, detailed in the second section, which comprises dense optical flow calculation and foreground extraction using an adaptive Gaussian mixture model. The training and model extraction is described in the third section, where we analyse a typical surveillance footage. The model is compared against another test sequence where there are flow patterns similar and dissimilar to the model. The results obtained are promising and justify the application of an HMM framework to automatically encode dynamics in this surveillance scenario.

FEATURE EXTRACTION FRAMEWORK

When the number of people in the scene increases significantly we cannot solely rely on tracking as the input for a behavioural model of the scene. In occluded situations tracking fails due to the difficulty in resolving individuals in the scene. Another issue is how to keep the consistency of individuals' labels through time after a sequence of occlusions/de-occlusions in the crowd. This is an open problem for all the tracking algorithms in such scenario. These coherence constraints justify a more global approach for the analysis of pedestrian

groups dynamics and instead of representing the behaviour of the individuals that compose the group, we have to interpret the aggregated behaviour of the group in the scene. One technique for this kind of analysis is the optical flow computation. Initially, we investigate the applicability of dense optical flow and will not consider optimisation (e.g. search strategies for block based motion estimation). However, optical flow calculation can provide ambiguous answers for different kinds of motions in the image plane. We assume that a model incorporating flow dynamics is able to disambiguate similar patterns resulting from different motions.

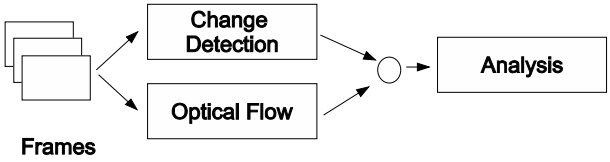


Fig. 1: Framework for behaviour characterisation in pedestrian traffic.

The modules for the feature extraction framework are shown in Fig. 1. The change extraction module starts with the adaptive mixture of Gaussians algorithm described in Stauffer and Grinson (7). The parameters of the algorithm are adjusted to have a background model memory of 500 frames and the fixed threshold for the classification of a pixel as background is set to 80%. Allowing a slow background update for the sequences and reducing the probability that a person which stops for a small period becomes part of the background model. In the composition of the background model for crowded scenes we assume that the initial model is obtained when there is no one present on the scene. Updates to this model will be done only for the pixels, which are classified as background in the subsequent crowded frames. The foreground regions are morphologically filtered to eliminate the bulk of the detection noise. The regions detected as foreground are also temporally filtered to eliminate spurious foreground detections not existing in the last frame. The resulting mask is then combined with the output of the optical flow calculation. Prior to the optical flow calculation a 5x5x5 gaussian spatio-temporal filter is applied for noise reduction. The optical flow calculation module implements the robust dense optical flow method described in Black and Anadan (8). Although more computationally expensive, it provides a smooth optical flow at the motion boundaries. This makes it an ideal candidate to evaluate the usefulness of flow information, compared to its more noisy counterparts. The resulting optical flow is threshold and decimated using an 8x8 median filter to further reduce noise and the number of flow vectors inside the model. Fig. 2 illustrates the optical calculation process. The combination of flow information with the background mask allows the analysis modules to only consider flow vectors inside foreground objects, which helps to reduce the noise in the observations, similar to the work described in Velastin et al (12).

The system in its current implementation does not perform in real time, mainly due to the computational load of the dense robust optical flow calculation since the other stages (pre-processing, change detection and analysis) operate faster than real time (25 frames per second). The modelling detailed in the next section is performed off-line.

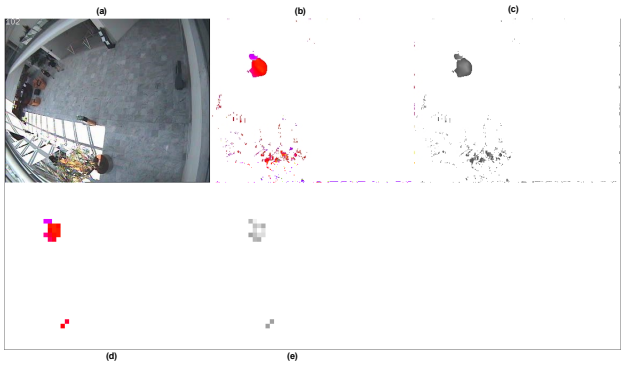


Fig. 2: Optical flow calculation. (a) Original image. Thresholded values for: (b) optical flow angle, (c) magnitude, median filtered (8x8) (d) angle and (e) magnitude.

Hidden Markov Models

HMMs Rabiner (9) and related graphical models are a ubiquitous tool for modelling time series data. They are well established in speech recognition, data compression, molecular biology, pattern recognition and artificial intelligence. Recently they are finding increasing applications in computer vision systems built for gait analysis, gesture recognition, behaviour classification and traffic monitoring. These systems use the HMM's capability to encode visual context to perform inference. Our system uses the same modelling technique to encode optical flow spatio-temporal variations. Given the continuous nature of the observed variables (flow vectors) a HMM with mixture of Gaussians is used. The formalisation for the HMM with mixture of Gaussians output is based on (9).

The set of K observations are denoted by $\mathbf{O} = [O^1, \ldots, O^K]$, where $O^k = [O_1^k, \ldots, O_T^k]$, allowing K multiple observations sequences. Each observation sample (O_t^k) is a vector $O_t^k = (x,y,u,v)$ 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 $\lambda = (\pi_i, a_{ij}, b_i(l), c_{im}, \mu_{im}, \Sigma_{im})$, where π_i is the prior probability for state $i=1..N$, a_{ij} is the state transition matrix ($i=1..N;j=1..N$), $b_i(l)$ is the emission probability of the l-th observation by the i-th state, c_{im} is

the mixture coefficient, μ_{im} is the mean vector and Σ_{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

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

The $\alpha_t(i)$ and $\beta_t(i)$ parameters are estimated using the forward-backwards algorithm described in (9). The probability that an observation is generated by Gaussian m in state i at time t is

$$\gamma_t(i,m) = \left[\frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\right]\left[\frac{c_{im}\aleph(\mathbf{O};\mu_{im}\Sigma_{im})}{\sum_{m=1}^{M}c_{im}\aleph(\mathbf{O};\mu_{im}\Sigma_{im})}\right] \quad (2)$$

where $\aleph$ specifies the Gaussian distribution function.

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm}\aleph(\mathbf{O};\mu_{jm}\Sigma_{jm})\left[\frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N}\alpha_t(j)\beta_t(j)}\right], 1 \le j \le N \quad (3)$$

The update equations for the EM procedure are given by:

$$\hat{\pi}_i = \frac{\sum_{k=1}^{K}\gamma_1^k(i)}{K} \quad (4)$$

$$\hat{a}_{ij} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\xi_t^k(i,j)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\gamma_t^k(i)} \quad (5)$$

$$\hat{b}_i(l) = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k-1}{}_{s.t.O_t=O_l}\gamma_t(i)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k-1}\gamma_t(i)} \quad (6)$$

The variable $\xi_t^t(i,j)$ is the expected number of transitions from state i to state j for the observation sequence K. The EM updates for the mixture components are:

$$\hat{c}_{im} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\gamma_t^k(i,m)}{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\sum_{m=1}^{M}\gamma_t^k(i,m)} \quad (7)$$

$$\hat{\mu}_{im} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\gamma_t^k(i,m).\mathbf{O}_t^k}{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\gamma_t^k(i,m)} \quad (8)$$

$$\hat{\Sigma}_{im} = \frac{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\gamma_t^k(i,m).(\mathbf{O}_t^k-\mu_{im}).(\mathbf{O}_t^k-\mu_{im})^{\mathrm{T}}}{\sum_{k=1}^{K}\sum_{t=1}^{T_k}\gamma_t^k(i,m)} \quad (9)$$

The HMM model is global, observing the whole frame and encoding in the Gaussian's parameters the positions and directions of flows observed in the training sequence. Such model provides a global view of the local activity in the sequence. However the model is affected by motion occurring in all the scene and is not able to fully isolate areas of influence in the image. We assume that this limitation can be alleviated by a multi-resolution approach, i.e. one model per block in each resolution level, such as the HMMs proposed for image segmentation in Li et al (10) and does not constitute a limitation of our approach. This will allow localized detection and analyses of flow anomalies facilitating the semantic labelling of image areas by an operator. Here we only concentrate on the applicability of this technique to encode global flow dynamics which can be easily escalated to more localised analysis.

TRAINING

The model is trained with the MeetCrowd video sequence from the foyer data set provided by the CAVIAR project (http://homepages.inf.ed.ac.uk /rbf/CAVIAR/)}. The sequence shows four people meeting and walking from the upper left to the lower right corner of the image. A total of 290 frames, which contain significant motion in the sequence, are used for training. Figure 3 shows an example of this sequence, together with the background model derived for the scene.

(a) (b)

Fig. 3: Training set sequence (CAVIAR Project). (a) Typical frame. (b) Most probable background.

The HMM topology (number of states (N) and number of Gaussians (M)) is arbitrarily varied to look for a model capable of better describing the training set. However, the best structure can be iteratively sought by using state splitting algorithms or entropic priors (4). Table 1 shows the EM convergence results for the tested HMM structures. The threshold for likelihood convergence is 10^{-4}. The number of (x,y,u,v) samples per frame for the training set is show in Figure 4. Figure 5

shows the number of samples per frame for the test set which contains a smaller number of people. Figures 6.a and 6.b show the motion history (accumulation of optical flow magnitude) for the test and training set as obtained after application of the foreground mask. Note that these patterns do not display the motion direction only its distribution. For the test set only a small fraction of motion in the lower right corner of Figure 6.b is in the same direction as in the training set. Figures 6.c and 6.d show the spatial distribution of model motion vectors encode by the Gaussians of the (N = 4, M = 4) and (N = 10, M = 10) HMMs respectively. Note the more effective matching of the motion history obtained with a larger number of Gaussians, which is related to a better likelihood for the (N = 10, M = 10) model obtained through EM. The likelihood between the new observations and the model is evaluated in a window of 25 frames, corresponding to 1 second for this data set. It means that the HMM is unrolled 25 times to take in to account the temporal interdependency of the flow variations.

The model learnt for the small walking crowd is compared to a new sequence consisting of two people entering and leaving the scene where just one of them moves according to the model for a while (Walk sequence from the CAVIAR dataset). The other motions in the test sequence are not encoded in the model. Only frames with significant motion are compared for testing. Based on this likelihood a simple classifier uses a global threshold to decide whether the flow vectors in the observation window are normal or abnormal. The results are compared to an annotated ground truth where frames with motions similar in direction and position to the training set are labelled as normal.

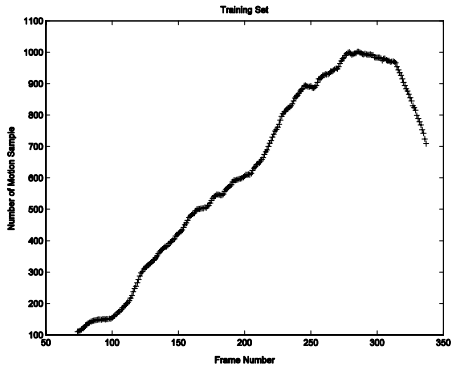


Fig. 4: Training sequence motion samples per frame.

Table 1 - EM convergence results per HMM structures (N = states, M = Gaussians per state).

NxM	Iterations	Log. Likelihood
4x4	226	-30416.6
8x6	307	-23631.6
10x10	705	-16213.4

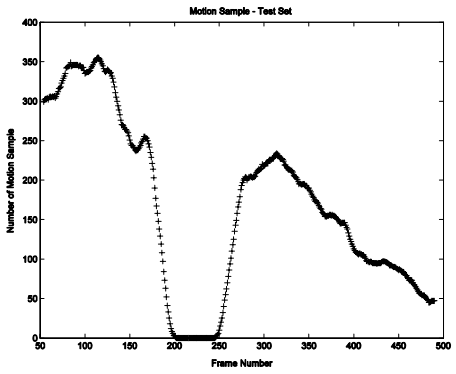


Fig. 5: Test sequence motion samples per frame.

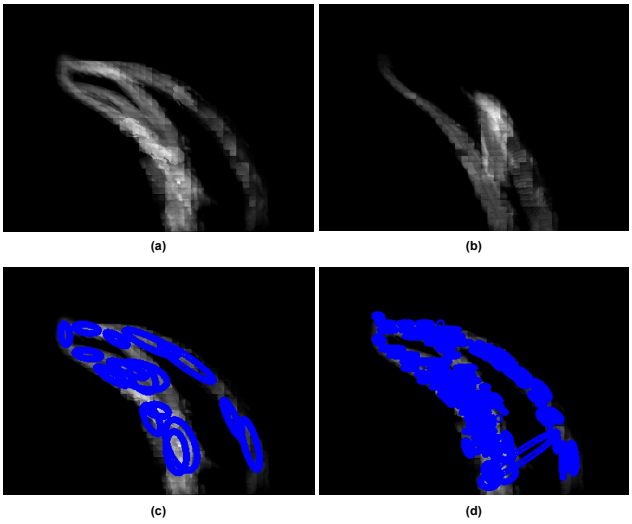


Fig. 6: Motion history for (a) training and (b) test sets. (c) Motion spatial distribution encoded in the Gaussian models for (c) (N = 4 , M = 4) and (d) (N = 10 , M = 10) HMMs superimposed on the motion history.

EXPERIMENTAL RESULTS

Figure 7 shows the likelihood results for the all frames with at last one non-null motion vector in the observation window (last 25 frames). Figure 8 shows the likelihood values for the frames with more than 150 non-null motion vectors in their observation window. The model used to produce these graphs is the HMM with (N = 10, M = 10). By correlating the likelihood variations with events in the sequence one can better understand the HMM results. Event E1 represents the entrance of an individual in the scene in the same position but moving in opposite direction to the modelled flow. The likelihood minima of Event E2 shows the point where the individual stops in a area out of the training set and starts waving its arms. Event E3 represents the individual starting movement in the same direction as in the training set correlated with an increase in likelihood. Event E4 signals the exit of the first individual from the scene. The residual increase in

likelihood after E4 is due to the trailing observation window. E5 signals the entrance of the second individual moving in a direction opposite the model. The likelihood starts to decrease as more samples of abnormal motion are gathered (note the increase in motion samples in Figure 5). E6 is the point where the second individual is misclassified because its area becomes too small but its position still agrees with the model provoking a certain level of activation. E7 marks the exit of the second individual from the scene. The likelihood response of the model is used in a simple threshold classification. The lowest likelihood value for a frame with normal behaviour is assumed as the threshold. Figure 9 shows the ground truth for classification. Frames with motion visually similar to the model (same general direction) are labelled as normal. Figure 10.a shows the results of behaviour classification using the threshold classifier. Frames above the threshold are classified as normal. The classification has a number of false negatives (abnormal frames being detected as normal) due to the vanishing observation window after E4 and to the area reduction of the second individual around frame 320. The number of false negatives (abnormal being detected as normal) can be further decreased if the system only takes in to account frames with more the 150 samples in the observation window (Figure 10.b). However, when there is enough motion in the image, the system is able to successfully identify the entrance of the first individual (E1), its arm waving behaviour (E2), and the entrance of the second individual moving in the opposite direction after its size increases. By specification (threshold choice) the system did not present any false positive results (normal frames detected as abnormal).

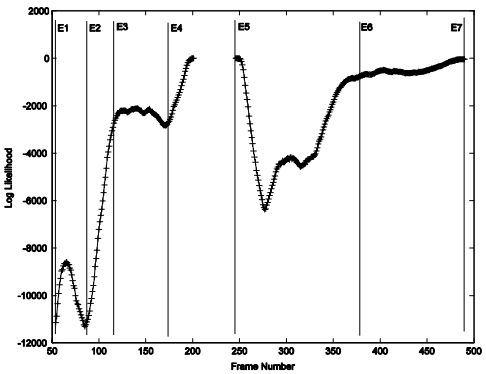


Fig. 7: Log likelihood for the test sequence for all frames with at least one motion vector in the observation window.

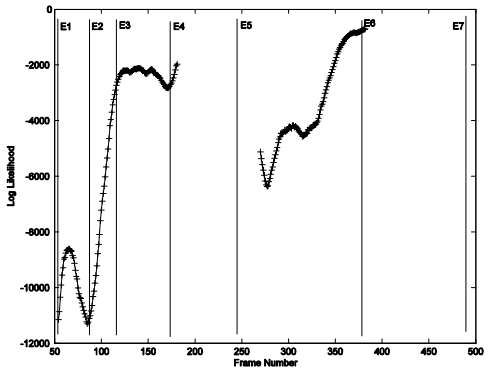


Fig. 8: Log likelihood for the test sequence for all frames with more than 150 motion vectors in the observation window.

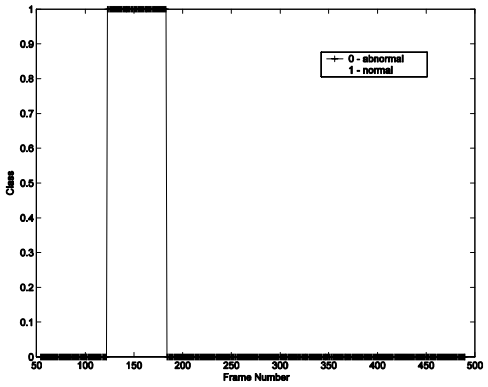
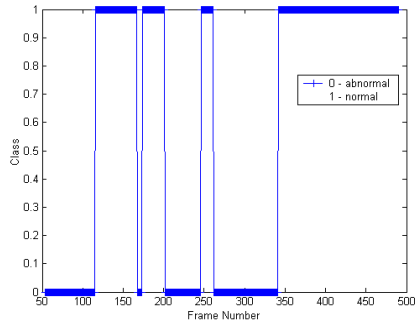
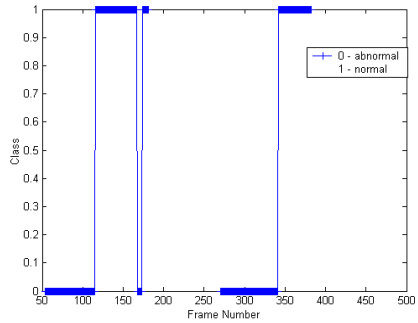


Fig. 9: Ground truth for the test sequence where normal (1) means motion visually similar to the training set.



(a)



(b)

Fig. 10: Classification results for abnormal motion detection. (a) At least one motion vector in the observation window, (b) more than 150 motion vectors in the observation window.

CONCLUSIONS

We introduced a framework for the analysis of pedestrian behaviour. It relies on optical flow information from the video evidence to represent the pedestrian group behaviour as optical flow variations in time. These variations are encoded in HMMs, which allow detection of unusual events. Although the result are shown for a limited data set they are promising because demonstrate the principle of optical flow dynamics analysis using HMM. The results suggest that such a system can display a low false positive rate suited for large-scale surveillance scenario (100 cameras and above), such tendency still needs confirmation for a larger data set.

The results demonstrate that the HMM model captures the global flow dynamics present in the training sequence. However, the model needs to be augmented to provide a better generalisation and to be able to cope with large variations in the number of samples whilst keeping a coherent likelihood output.

We are currently gathering more video evidence to study the flow patterns in crowds and extending the analysis to local and multi-resolution HMM structures. For more complex crowded scenarios HMM models derived from the distinct training sets can compose a pattern grammar. Ivanov and Bobick (11) demonstrated that such pattern grammars can improve the classification performance given the semantic output of a bank of HMM models.

REFERENCES

1. Hu, W., Tan T., Wang L., and Maybank S., IEEE Transac. on Syst. Man and Cyber. - Part C, 43, 334-352.

2. Zhao, T., and Nevatia, R., IEEE Transac. on Patt. Anal. and Mach. Intel., 26, 1208-1221.

3. Maurin, B., Masoud O., and Papanikolopoulos N., IEEE 5th Int. Conf. on Intel. Transp. Syst., 19-24.

4. Brand, M., and Kettnaker, V., IEEE Transac. on Patt. Anal. and Mach. Intel., 22, 844-851.

5. Oliver, N., Garg, A., and Horvitz, E., Comp. Vis. and Im. Underst., 96, 163-180.

6. Gong., S., Xiang, T., Int. Conf. Comp. Vis. 2002, 2, 742-749.

7. Stauffer, C., and Grimson, W., IEEE Transac. on Patt. Anal. and Mach. Intel., 22, 747-757.

8. Black, M. J., and Anandan, P., Fourth Int. Conf. on Comp. Vis., 231-236.

9. Rabiner, L. R., Proc. of the IEEE, 77, 257-286.

10. Li, J., Gray, R. M., and Olsen, R. A., IEEE Transac. on Inf. Theo., 46, 1826-1841.

11. Ivanov, Y. A. and Bobick, A. F., IEEE Transac. on Patt. Anal. and Mach. Intel., 22, 852-872.

12. Velastin, S. A., Bogossian, B. A., Lo, B. P. L., Sun, J., Vicencio-Silva, M. A., IEEE Transac. on Syst. Man and Cyber. - Part A, 35, 164-182.