Online Video Behaviour Abnormality Detection Using Reliability Measure

Tao Xiang and Shaogang Gong
Department of Computer Science
Queen Mary, University of London, London E1 4NS, UK
{txiang,sgg}@dcs.qmul.ac.uk

Abstract

An approach is proposed for robust online behaviour recognition and abnormality detection based on discovering natural grouping of bebaviour patterns through unsupervised learning and a time accumulative reliability measure. A novel behaviour learning model and a run-time accumulative reliability measure are introduced to determine both the natural groupings of possible normal behaviour classes without manual labelling and when sufficient visual evidence has become available for differentiating ambiguities among different behaviour classes observed online. This ensures behaviour recognition at the shortest possible time and robust abnormality detection.

1 Introduction

One of the critical functionalities of an automatic video-based behaviour monitoring system is to detect abnormal behaviour and recognise normal behaviour reliably *on-the-fly*. A novel behaviour modelling approach is proposed in this work based on discovering natural grouping of bebaviour patterns through unsupervised learning and by introducing a time accumulative reliability measure on visual features available at a given time. Our approach differs from previous approaches in the following aspects: (1) Different classes of behaviour patterns are discovered automatically. This is to avoid the laborious process of manual labelling and the bias in manual labelling caused by the inconsistency of human interpretation of behaviour. (2) A novel relevance learning algorithm is employed for clustering behaviour patterns using the eigenvectors of the behaviour affinity matrix. The number of behaviour classes is automatically determined using *only* the relevant eigenvectors. Unlike previous unsupervised feature relevance learning algorithms such as [4, 2], our algorithm is specially tailored for fast and robust selection of relevant eigenvectors of the behaviour affinity matrix. (3) A novel time-accumulated reliability measure is introduced to determine when sufficient visual features have become available in order to overcome any ambiguity among different behaviour classes observed online due to insufficient visual evidence at a given time instance. This ensures robust behaviour recognition and abnormality detection at the shortest possible time, as opposed to previous work such as [14, 3, 6] which requires completed behaviour patterns. Our approach is also advantageous over previous approaches using the *Maximum Likelihood* (ML) method [13, 3, 6]. Such as a ML based approach makes a forced decision on behaviour recognition at each time instance without considering the reliability and sufficiency of the accumulated visual evidence. Consequently, it can be error prone. The effectiveness and robustness of our approach is demonstrated through experiments using noisy and sparse data sets collected from both indoor and outdoor surveillance scenarios.

2 Behaviour Modelling

A continuous video $\mathbf{V}$ is segmented into N segments $\mathbf{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_n, \ldots, \mathbf{v}_N\}$ so that each segment contains approximately a single behaviour pattern [10, 14]. A discrete event based approach is adopted for behaviour representation [11]. First, an adaptive Gaussian mixture background model is used to detect foreground pixels. Second, the foreground pixels in a vicinity are grouped into a blob using the connected component method. Each blob with its average pixel-change-history value greater than a threshold is then defined as an event. An event is represented as a 7-dimensional feature vector capturing location, shape and motion information. Third, classification is performed in a 7D feature space using a Gaussian Mixture Model (GMM). The number of event classes K_e is determined automatically using Bayesian Information Criterion (BIC) [7]. The learned GMM is used to classify each detected event into one of K_e event classes. Finally, the behaviour pattern captured by the nth video segment $\mathbf{v}_n$, consisting of T_n image frames, is represented as a behaviour pattern feature vector $\mathbf{P}_n$=$[\mathbf{p}_{n1}, \ldots, \mathbf{p}_{nt}, \ldots, \mathbf{p}_{nT_n}]$, where the tth element $\mathbf{p}_{nt}$ is a K_e dimensional variable: $\mathbf{p}_{nt} = [p_{nt}^1, ..., p_{nt}^k, ..., p_{nt}^{K_e}]$; $\mathbf{p}_{nt}$ is computed from the tth image frame of $\mathbf{v}_n$ where p_{nt}^k is the posterior probability that an event of the kth event class has occurred in the frame given the learned GMM.

2.1 Behaviour Affinity Matrix

Consider a training data set $\mathbf{D} = \{\mathbf{P}_1, \ldots, \mathbf{P}_n, \ldots, \mathbf{P}_N\}$ consisting of N behaviour patterns, where $\mathbf{P}_n$ is the nth behaviour pattern feature vector as defined above. We aim to first discover the natural grouping of the training behaviour patterns upon which a behaviour model can be based. This is an unsupervised clustering problem with the number of clusters unknown. However, there are two characteristics of the behaviour feature vectors that make the clustering problem challenging: (1) Each feature vector can be of different length therefore requires dynamic warping before they can be compared with. Conventional clustering approaches such as K-means and mixture models thus cannot be applied directly. (2) A definition of a distance/affinity metric among these variable length feature vectors is not simply Euclidean therefore requires a nontrivial string similarity measure.

We propose to utilise Dynamic Bayesian Networks (DBNs) to provide a dynamic representation of each behaviour pattern feature vector in order to both address the need for dynamic warping and provide a string similarity metric. More specifically, each behaviour pattern in the training set is modelled using a DBN. To measure the affinity between two behaviour patterns represented as $\mathbf{P}_i$ and $\mathbf{P}_j$, two DBNs denoted as $\mathbf{B}_i$ and $\mathbf{B}_j$ are trained on $\mathbf{P}_i$ and $\mathbf{P}_j$ respectively using the EM algorithm [1, 5]. The affinity between $\mathbf{P}_i$ and $\mathbf{P}_j$ is then computed as: $S_{ij} = \frac{1}{2}\left\{\frac{1}{T_j}\log P(\mathbf{P}_j|\mathbf{B}_i) + \frac{1}{T_i}\log P(\mathbf{P}_i|\mathbf{B}_j)\right\}$, where $P(\mathbf{P}_j|\mathbf{B}_i)$ is the likelihood of observing $\mathbf{P}_j$ given $\mathbf{B}_i$, and T_i and T_j are the lengths of $\mathbf{P}_i$ and $\mathbf{P}_j$ respectively. DBNs of different topologies can be used. In this work, we employ a Multi-Observation Hidden Markov Model (MOHMM) [3]. The number of hidden states for each hidden variables in the MOHMM is set to K_e, i.e. the number of event classes [1].

The eigenvectors of an affinity matrix $\mathbf{S} = \{S_{ij}\}$ can then be employed directly for data clustering. However, it has been shown in [9, 8] that it is more desirable to perform clustering based on the eigenvectors of the normalised affinity matrix $\bar{\mathbf{S}}$, defined as $\bar{\mathbf{S}} = \mathbf{L}^{-\frac{1}{2}}\mathbf{S}\mathbf{L}^{-\frac{1}{2}}$ where $\mathbf{L}$ is an $N \times N$ diagonal matrix with $L_{ii} = \sum_j S_{ij}$. The

[1] K_e reflects the complexity of the behaviour patterns, so is the number of hidden states. So it is appropriate to set these two to be equal.

remaining problem is to first determine the number (order) of behaviour classes K before clustering the behaviour patterns in the training set.

2.2 Selecting Relevant Eigenvectors for Behaviour Clustering

We assume that the number of clusters K is between 1 and K_m, a number considered to be sufficiently larger than the true value of K. We set $K_m = \frac{1}{5}N$ where N is the number of training samples [2]. The training data set is now represented using the K_m largest eigenvectors, denoted as $\mathbf{D_e} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n, \ldots, \mathbf{x}_N\}$, with the nth behaviour pattern being represented as a K_m dimensional feature vector $\mathbf{x}_n = [e_{1n}, \ldots, e_{kn}, \ldots, e_{K_m n}]$, where e_{kn} is the nth element of the kth largest eigenvector $\mathbf{e_k}$.

Because only the first K largest eigenvectors are needed for grouping K clusters [12, 9], there are certainly redundant/irrelevant eigenvectors among the K_m largest eigenvectors. It is important to identify those irrelevant but large eigenvectors because that (1) irrelevant features degrade the accuracy of learning, and (2) the dimension of the features space (K_m) is high compared to the sample size (N) resulting in learning subject to the curse of dimensionality. To overcome these problems, we derive here a novel eigenvector relevance learning algorithm. Specifically, we proposed to measure the relevance of an eigenvector according to how well it can separate a data set into separate groups.

We denote the likelihood of the kth eigenvector $\mathbf{e_k}$ being relevant as $R_{\mathbf{e_k}}$. Apparently, we have $0 \leq R_{\mathbf{e_k}} \leq 1$. We assume that the elements of $\mathbf{e_k}$, e_{kn} follow two different distributions depending on whether $\mathbf{e_k}$ is relevant. The probability density function (pdf) of e_{kn} is thus formulated as a mixture model of two components: $P(e_{kn}|\theta_{e_{kn}}) = (1 - R_{\mathbf{e_k}})P(e_{kn}|\theta^1_{e_{kn}}) + R_{\mathbf{e_k}}P(e_{kn}|\theta^2_{e_{kn}})$ where $\theta_{e_{kn}}$ are the parameters describing the distribution, $p(e_{kn}|\theta^1_{e_{kn}})$ is the pdf of e_{kn} when $\mathbf{e_k}$ is irrelevant/redundant and $P(e_{kn}|\theta^2_{e_{kn}})$ otherwise. $R_{\mathbf{e_k}}$ acts as the weight or mixing probability of the second components. The distribution of e_{kn} is assumed to be a single Gaussian to reflect the fact that $\mathbf{e_k}$ cannot be used for data grouping when it is irrelevant: $P(e_{kn}|\theta^1_{e_{kn}}) = \mathcal{N}(e_{kn}|\mu_{k1}, \sigma_{k1})$ where $\mathcal{N}(.|\mu, \sigma)$ denotes a Gaussian of mean μ and covariance σ. We assume the second component of $P(e_{kn}|\theta_{e_{kn}})$ as a mixture of two Gaussians to reflect the fact $\mathbf{e_k}$ can separate one group of data from others when it is relevant: $P(e_{kn}|\theta^2_{e_{kn}}) = w_k\mathcal{N}(e_{kn}|\mu_{k2}, \sigma_{k2}) + (1 - w_k)\mathcal{N}(e_{kn}|\mu_{k3}, \sigma_{k3})$ where w_k is the weight of the first Gaussian in $P(e_{kn}|\theta^2_{e_{kn}})$. There are 8 parameters required for describing the distribution of e_{kn}: $\theta_{e_{kn}} = \{R_{\mathbf{e_k}}, \mu_{k1}, \mu_{k2}, \mu_{k3}, \sigma_{k1}, \sigma_{k2}, \sigma_{k3}, w_k\}$. The maximum likelihood (ML) estimate of $\theta_{e_{kn}}$ can be estimated using the following algorithm. First, the parameters of the first mixture component $\theta^1_{e_{kn}}$ are estimated as $\mu_{k1} = \frac{1}{N}\sum_{n=1}^{N} e_{kn}$ and $\sigma_{k1} = \frac{1}{N}\sum_{n=1}^{N}(e_{kn} - \mu_{k1})^2$. The rest 6 parameters are then estimated using EM.

Since our relevance learning algorithm is essentially a local searching method, it could be sensitive to parameter initialisation especially in the presence of noise [1]. To overcome this problem, our *a priori* knowledge on the relevance of each eigenvector is utilised to set the initial value of $R_{\mathbf{e_k}}$. Specifically, we set the initial value of $R_{\mathbf{e_k}}$, $\tilde{R_{\mathbf{e_k}}} = \bar{\lambda}_k$ where $\bar{\lambda}_k \in [0, 1]$ is the normalised eigenvalue for $\mathbf{e_k}$ with $\bar{\lambda}_1 = 1$ and $\bar{\lambda}_{K_m} = 0$.

The estimated $\hat{R_{\mathbf{e_k}}}$ provides a continuous-value measurement of the relevance of $\mathbf{e_k}$. Since a 'hard-decision' is needed for dimension reduction, we simply eliminate the kth eigenvector $\mathbf{e_k}$ if $\hat{R_{\mathbf{e_k}}} < 0.5$ and weight the relevant eigenvectors using $\hat{R_{\mathbf{e_k}}}$. This gives us a new data set denoted as $\mathbf{D_r} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n, \ldots, \mathbf{y}_N\}$. We model the distribution of $\mathbf{D_r}$ using a Gaussian Mixture Model (GMM) for behaviour pattern clustering. The Bayesian

[2] As a rule of thumb, if $K > \frac{1}{5}N$, the training data set would be too sparse for model training.

Information Criterion (BIC) is then employed to select the optimal number of components K, denoted as K_o, corresponding to the number of behaviour classes. Each behaviour pattern in the training data set is then labelled as one of the K_o behaviour classes using the learned GMM. It is found by our experiments (see Section 4) that the number of behaviour classes could be severely under-estimated without relevant eigenvector selection.

2.3 A Composite Behaviour Model using Mixture of MOHMMs

To build a model for the observed/expected behaviour, we first model the kth behaviour class using a MOHMM $\mathbf{B}_k$. The parameters of $\mathbf{B}_k$, $\theta_{\mathbf{B}_k}$ are estimated using all the patterns in the training set that belong to the kth class. A behaviour model $\mathbf{M}$ is then formulated as a mixture of the K_o MOHMMs. Given an unseen behaviour pattern, represented as a behaviour pattern feature vector $\mathbf{P}$ as described in Section 2, the likelihood of observing $\mathbf{P}$ given $\mathbf{M}$ is $P(\mathbf{P}|\mathbf{M}) = \sum_{k=1}^{K} \frac{N_k}{N} P(\mathbf{P}|\mathbf{B}_k)$, where N_k is the number of training behaviour patterns that belong to the kth behaviour class.

3 Abnormality Detection with Reliability Measure

An unseen behaviour pattern of length T is represented as $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_t, \ldots, \mathbf{p}_T]$. At the tth frame, the accumulated visual information for the behaviour pattern, denoted as $\mathbf{P}_t = [\mathbf{p}_1, \ldots, \mathbf{p}_t]$, is used for online reliable abnormality detection and behaviour recognition. First, the normalised log-likelihood of observing $\mathbf{P}$ at the tth frame given the behaviour model $\mathbf{M}$ is computed as $l_t = \frac{1}{t} \log P(\mathbf{P}_t|\mathbf{M})$. l_t can be computed by extending the forward part of the forward-backward procedure [5] for HMM to MOHMM [3]. We then measure the abnormality of $\mathbf{P}$ at each frame t using Q_t:

$$Q_t = \begin{cases} l_1 & \text{if } t = 1 \\ \\ (1-\alpha)Q_{t-1} + \alpha(l_t - l_{t-1}) & \text{otherwise} \end{cases} \tag{1}$$

where α is an accumulating factor determining how important the visual information extracted from the current frame is for abnormality detection. Compared to l_t as an indicator of normality/abnormality, Q_t could add more weighting to more recent observations. Abnormality is detected at frame t if $Q_t < Th_A$ where Th_A is a threshold [4]. Note that it takes a time delay for Q_t to stabilise at the beginning of evaluating a behaviour pattern due to the nature of the forward-backward procedure. The length of this time period, denoted as T_w is related to the complexity of the MOHMM used for behaviour modelling [5].

Beyond abnormality detection, our model is also employed to perform normal behaviour classification. At each frame t a behaviour pattern needs to be recognised with a reliability measure as one of K_o behaviour classes when $Q_t > Th_A$. To this end, we measure the reliability of a decision on $\mathbf{P}$ belonging to the kth behaviour class as:

$$r_k = \frac{\frac{N_k}{N} P(\mathbf{P}_t|\mathbf{B}_k)}{\sum_{i \neq k} \frac{N_i}{N} P(\mathbf{P}_t|\mathbf{B}_i)} \tag{2}$$

r_k is the ratio of the probability of $\mathbf{P}_t$ belonging to the kth behaviour class and that of $\mathbf{P}_t$ belonging to the other $K_o - 1$ classes. It is a function of t. $\mathbf{P}_t$ is reliably recognised as

[3] The complexity of computing l_t is $\mathcal{O}({K_e}^2)$ and does not increase with t.

[4] Th_A is determined in practice according to the detection and false alarm rate required by each particular surveillance application.

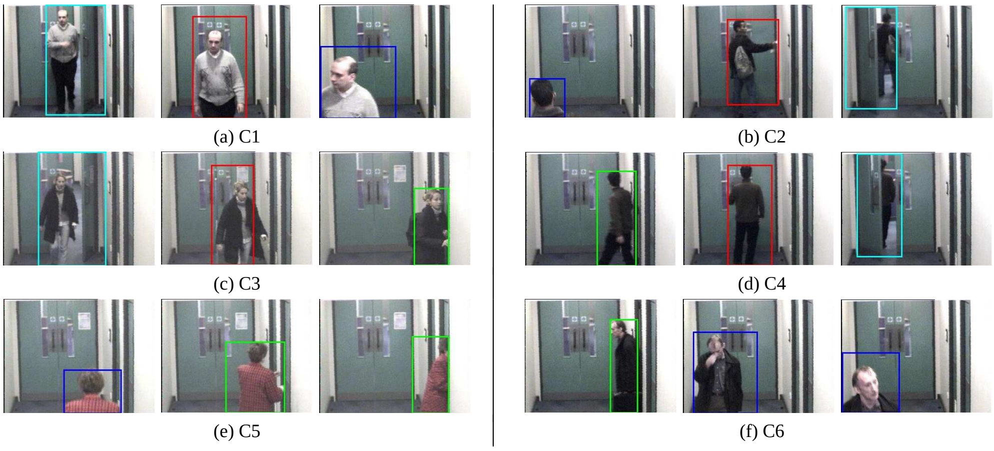
[5] We set $T_w = 3K_e$ in our experiments reported later in the experiment section.

the kth behaviour class only when $r_k > Th_r$, a threshold [6]. When there is more than one r_k greater than Th_r, the behaviour pattern is recognised as the class with the largest r_k.

For comparison, the commonly used *Maximum Likelihood* (ML) method recognises $\mathbf{P}_t$ as the kth behaviour class when $k = \arg\max_k \{P(\mathbf{P}_t|\mathbf{B}_k)\}$. Using the ML method, recognition has to be performed at each single frame without considering how reliable and sufficient the accumulated visual evidence is. This often causes errors especially when there are ambiguities among different classes (e.g, a behaviour pattern can be explained away equally well by multiple plausible behaviours at its early stage). Compared to the ML method, our approach holds the decision on behaviour recognition unless sufficient evidence has been accumulated to overcome ambiguities. The recognition results obtained using our approach are thus more reliable compared to those obtained by ML.

4 Experiments

4.1 Corridor Entrance/Exit Human behaviour Monitoring


(a) C1

(b) C2

(c) C3

(d) C4

(e) C5

(f) C6

Figure 1: Behaviour patterns in a corridor scene. (a)–(f) show image frames of typical behaviour patterns belonging to the 6 behaviour classes listed in Table 1. Events detected during each behaviour pattern are shown by colour-coded bounding boxes in each frame.

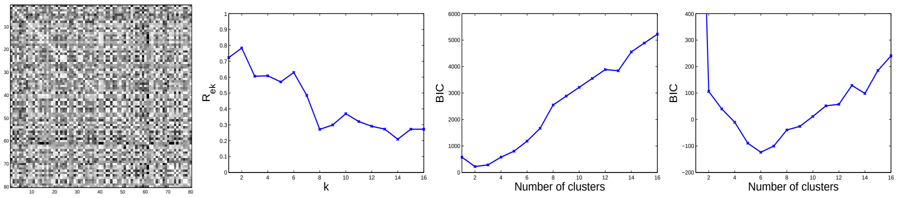
A CCTV camera was mounted on the ceiling of an office entry corridor, monitoring people entering and leaving the office area (see Fig. 1). The office area is secured by an entrance-door which can only be opened by scanning an entry card on the wall next to the door (see middle frame in Fig. 1(b)). Two side-doors were also located at the right hand side of the corridor. Typical behaviours occurring in the scene would be people entering or leaving either the office area or the side-doors, and walking towards the camera. For this experiment, a data set was collected over 5 different days consisting of 6 hours of video totalling 432000 frames captured at 20Hz with 320×240 pixels per frame. This data set was then segmented into sections separated by any motionless intervals lasting for more than 30 frames. This resulted in 142 video segments of actual behaviour pattern instances. Each segment has on average 121 frames.

Model training — Each training set consist of 80 randomly selected video segments without any behaviour class labelling of the video segments. The remaining 62 segments were used for testing later. This model training exercise was repeated 20 times and in each

[6] Th_r should be greater than one. In our experiments we found that $Th_r = 10$ led to satisfactory results.

C1	Office area to near end of the corridor	C2	Near end of the corridor to office area
C3	Office area to side-doors	C4	Side-doors to office area
C5	Near end of the corridor to side-doors	C6	Side-doors to near end of the corridor

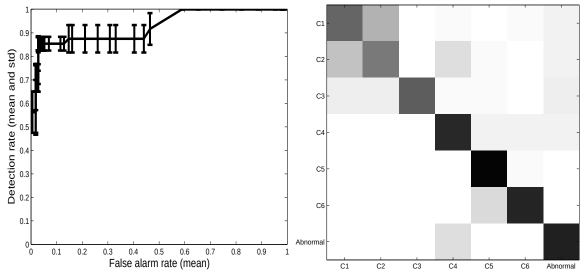
Table 1: Six classes of commonly occurred behaviour patterns in the corridor scene.



(a) (b) (c) (d)

Figure 2: An example of model training. (a): The normalised behaviour affinity. (b): the learned relevance for the K_m largest eigenvectors. The first 7 largest eigenvectors were determined as relevant features for clustering. (c) and (d) show the BIC model selection results without and with relevant eigenvector selection respectively.

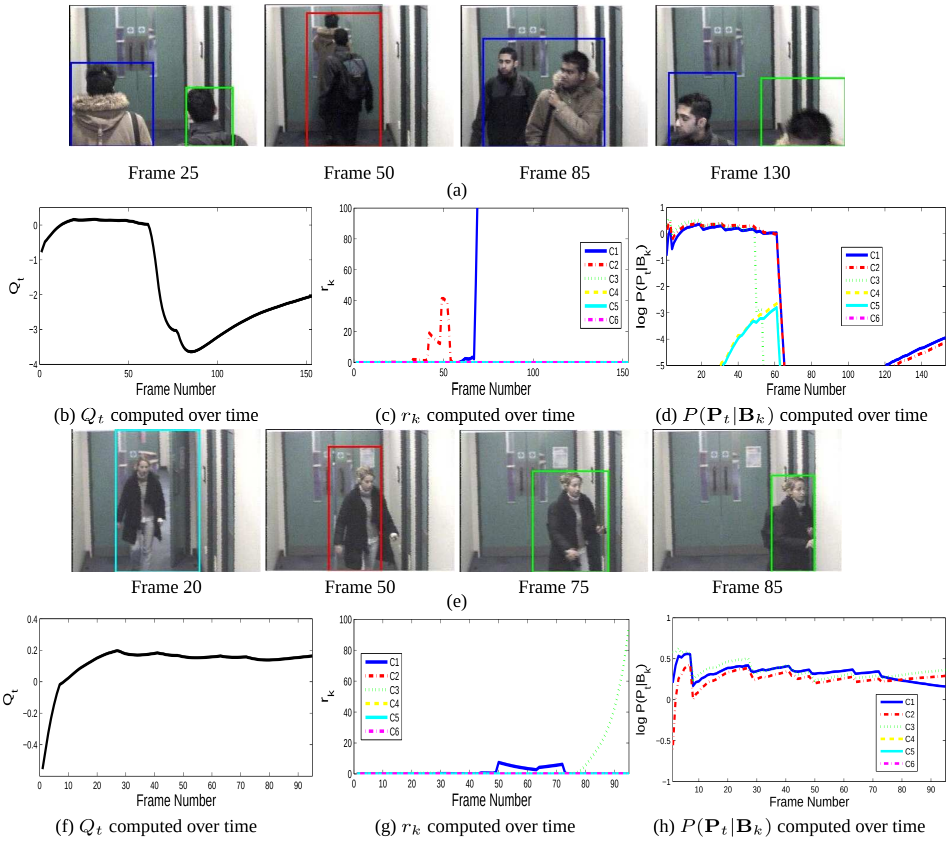
trial a different model was trained using a different random training set. Given each training set, 4 classes of discrete events were detected and classified using automatic model order selection in clustering (see Figure 1). Over the 20 trials, on average 6 eigenvectors were automatically determined as being relevant for clustering with smallest 4 and largest 9. The number of clusters for each training set was determined automatically as 2 and 6 in every trial without and with relevant eigenvector selection respectively (see Fig. 2(c)&(d)). By observation, each discovered data cluster mainly contained samples corresponding to one of the 6 behaviour classes listed in Table 1.



(a) (b)

Figure 3: The performance of abnormality detection and behaviour recognition for the corridor scene. (a): The mean and ±1 standard deviation of the ROC curves for abnormality detection obtained over 20 trials. (b): Confusion matrix for behaviour recognition. Each row represents the probabilities of that class being confused with all the other classes averaged over 20 trials. The main diagonal of the matrix shows the the fraction of patterns correctly recognised and is as follows: [.68 .63 .72 .84 .92 .85 .85].

Abnormality detection — To measure the performance of the learned models on abnormality detection, each behaviour pattern in the testing sets was manually labelled as normal if there were similar patterns in the corresponding training sets and abnormal otherwise. On average, there were 7 abnormal behaviour patterns in each testing set.



(b) Q_t computed over time

(c) r_k computed over time

(d) $P(\mathbf{P}_t|\mathbf{B}_k)$ computed over time

(f) Q_t computed over time

(g) r_k computed over time

(h) $P(\mathbf{P}_t|\mathbf{B}_k)$ computed over time

Figure 4: (a): An abnormal behaviour pattern where two people attempted to enter an office area without an entry card. It resembles C2 in the early stage. (b): The behaviour pattern was detected as abnormality from Frame 62 till the end based on Q_t. (c): The behaviour pattern between Frame 40 to 53 was recognised reliably as C2 based on r_k before being detected as an abnormality. (d) The behaviour pattern was wrongly recognised as C3 before Frame 20 using ML. (e): A normal C3 behaviour pattern. Note that it can be interpreted as either C1 or C3 before the person entered the sidedoor. (f): The behaviour pattern was detected as normal throughout using Q_t. (g): It was recognised reliably as C3 from Frame 83 till the end based on r_k. (h): The behaviour pattern was recognised prematurally and unreliably as either C1, C2, or C3 before Frame 83 using ML.

The detection rate and false alarm rate of abnormality detection are shown in the form of a ROC curve. Fig. 3(a) shows that high detection rate and low false alarm rate can be achieved. Th_A was set to -0.2 in the rest results unless otherwise specified, which gave an abnormality detection rate of $85.4 \pm 2.9\%$ and false alarm rate of $6.1 \pm 3.1\%$. Fig. 4(b)&(f) show examples of online reliable abnormality detection results obtained by monitoring the value of Q_t over time. α was set to 0.1 for computing Q_t.

Recognition of normal behaviours — To measure the performance of behaviour recognition results, the normal behaviour patterns in the testing sets were manually labelled into different behaviour classes. A normal behaviour pattern was recognised correctly if it was detected as normal and classified into the right behaviour class. The behaviour recognition results is illustrated as a confusion matrix shown in Fig. 3(b). Overall, the recognition rates had a mean of 77.9% and standard devation of 4.8% for the 6 behaviour classes over 20 trials. Examples of online behaviour recognition are shown in Fig. 4. Based on r_k, normal behaviour patterns were reliably and promtly recognised after sufficient visual evidence was available (see Fig. 4(c) & (g)). On the contrary, based on the ML method decisions on behaviour recognition were made prematurally and unreliably due to the ambuiguities among different behaviour classes (see Fig. 4 (d)&(h)).

4.2 Aircraft Docking Area Behaviour Monitoring

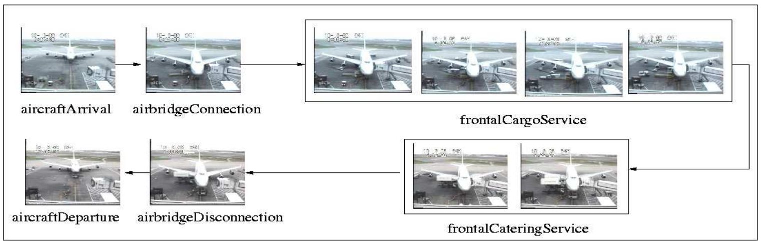


Figure 5: Typical, visually detectable behaviour patterns in an aircraft docking scene.

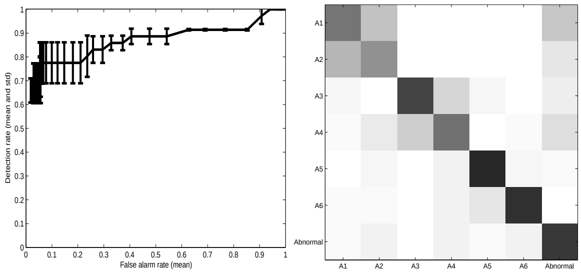
Now we consider an outdoor scenario. A fixed CCTV camera was mounted at an aircraft docking area, monitoring the aircraft docking procedure. Typical visually detectable behaviour patterns in this scene involved the aircraft, the airbridge and various ground vehicles (see Fig. 5(a)). The captured video sequences have a very low frame rate of 2Hz which is common for CCTV surveillance videos. Each image frame has a size of 320×240 pixels. Our database for the experiments consists of 72776 frames of video data (around 10 hours of recording) that cover different times of different days under changing lighting conditions.The video was segmented automatically using an online segmentation algorithm proposed in [10], giving 59 video segments of actual behaviour pattern instances. Each segment has on average 428 frames.

A1	Aircraft arrives	A2	Airbridge connected	A3	Frontal cargo service
A4	Frontal catering service	A5	Aircraft departs	A6	Airbridge disconnected

Table 2: Six classes of commonly occurred behaviour patterns in the airport scene.

Model training — A training set now consisted of 40 video segments and the remaining 19 were used for testing. 20 trials were conducted, each of which had a different random training set. Given each training set, eight classes of discrete events were detected and classified automatically (see Fig. 7(a)&(e)). On average 7 eigenvectors were automatically determined as being relevant for clustering with smallest 4 and largest 10. The number of clusters for each training set was determined automatically as 2 and 6 in every trial without and with relevant eigenvector selection respectively. By observation, each discovered data cluster mainly contained samples corresponding to one of the 6 behaviour classes listed in Table 2.

Abnormality detection — Fig. 6(a) shows that high detection rate and low false alarm rate can be achieved. Th_A was set to -0.5 in the rest results unless otherwise specified, which gave an abnormality detection rate of $79.2\pm8.3\%$ and false alarm rate of $5.1\pm3.9\%$. Fig. 7(b)&(f) show examples of online reliable abnormality detection.



(a) (b)

Figure 6: The performance of abnormality detection and behaviour recognition for the airport scene. (a): The mean and ± 1 standard deviation of the ROC curves. (b): Confusion matrix for behaviour recognition. The main diagonal of the matrix is: [.65 .58 .83 .72 .87 .80 .79].

Recognition of normal behaviour patterns — Overall, the recognition rates were $72.1 \pm 5.4\%$ for the 6 behaviour classes over 20 trials (see Fig. 6(b)). Examples of online reliable behaviour recognition are shown in Fig. 7. Again, the results show that our approach is superior to the ML based approach in that normal behaviour patterns can be reliably and promptly recognised after sufficient visual evidence has become available to overcome the ambiguities among different behaviour classes.

5 Conclusions

Compared to the corridor scene experiments, our results on the airport scene were obtained using much more noisy and sparse data sets. These results further demonstrate the effectiveness and robustness of our algorithm. In conclusion, we presented a novel approach for robust online behaviour recognition and abnormality detection based on discovering natural grouping of bebaviour patterns through unsupervised learning and a time accumulative reliability measure. Our approach is advantageous over previous approaches, such as [14, 13, 3, 6], in that it is online, robust and capable of dealing with ambiguities among different behaviour pattern classes.

References

[1] A. Dempster, N. Laird, and D. Rubin. Maximum-likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.

[2] J. Dy, C. Brodley, A. Kak, L. Broderick, and A. Aisen. Unsupervised feature selection applied to content-based retrival of lung images. *PAMI*, pages 373–378, 2003.

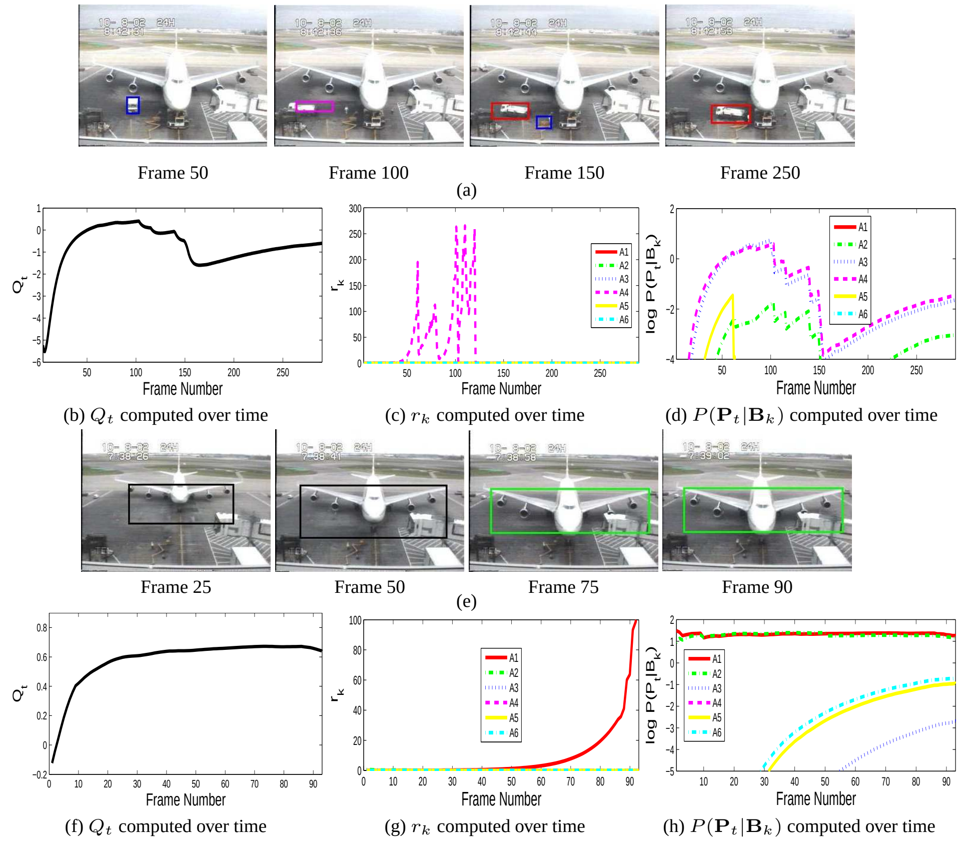
[3] S. Gong and T. Xiang. Recognition of group activities using dynamic probabilistic networks. In *ICCV*, pages 742–749, 2003.

[4] M. Law, M.A.T. Figueiredo, and A.K. Jain. Simultaneous feature selection and clustering using mixture model. *PAMI*, 26(9):1154–1166, 2004.

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

[6] N. Oliver, B. Rosario, and A. Pentland. A bayesian computer vision system for modelling human interactions. *PAMI*, 22(8):831–843, August 2000.

[7] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.



Frame 50 Frame 100 Frame 150 Frame 250

(a)

(b) Q_t computed over time (c) r_k computed over time (d) $P(\mathbf{P}_t|\mathbf{B}_k)$ computed over time

Frame 25 Frame 50 Frame 75 Frame 90

(e)

(f) Q_t computed over time (g) r_k computed over time (h) $P(\mathbf{P}_t|\mathbf{B}_k)$ computed over time

Figure 7: (a): An abnormal behaviour pattern where a truck brought engineers to fix a ground power cable problem. It resembled A4 in the early stage. (b): It was detected as abnormality from Frame 147 till the end based on Q_t. (c): The behaviour pattern between Frame 53 to 145 was recognised reliably as A4 using r_k before becoming abnormal and being detected. (d) The behaviour pattern was wrongly recognised as A3 before Frame 80 to 98 using the ML method. (e): A normal A1 behaviour pattern. (f): The behaviour pattern was detected as normal throughout based on Q_t. (g): It was recognised reliably as A1 from Frame 73 till the end using r_k. (h): It was wrongly recognised as A2 between Frame 12 to 49 using ML.

[8] J. Shi and J. Malik. Normalized cuts and image segmentation. *PAMI*, 22(8):888–905, 2000.

[9] Y. Weiss. Segmentation using eigenvectors: a unifying view. In *ICCV*, pages 975–982, 1999.

[10] T. Xiang and S. Gong. Activity based video content trajectory representation and segmentation. In *BMVC*, pages 177–186, 2004.

[11] T. Xiang, S. Gong, and D. Parkinson. Autonomous visual events detection and classification without explicit object-centred segmentation and tracking. In *BMVC*, pages 233–242, 2002.

[12] S. Yu and J. Shi. Multiclass spectral clustering. In *ICCV*, pages 313–319, 2003.

[13] L. Zelnik-Manor and M. Irani. Event based video analysis. In *CVPR*, 2001.

[14] H. Zhong, J. Shi, and M. Visontai. Detecting unusual activity in video. In *CVPR*, pages 819–826, 2004.