Global Abnormal Behaviour Detection Using a Network of CCTV Cameras

Emanuel E. Zelniker, Shaogang Gong, Tao Xiang

Global Abnormal Behaviour Detection Using a Network of CCTV Cameras

Emanuel E. Zelniker, Shaogang Gong, Tao Xiang
Queen Mary University of London
Department of Computer Science
Mile End Road, E1 4NS, London, United Kingdom
(zelniker,sgg,txiang)@dcs.qmul.ac.uk

Abstract

This paper investigates the detection of global abnormal behaviours across a network of CCTV cameras. Although the problem of multiple camera tracking has attracted much attention recently, little work has been done on modelling global behaviours of objects monitored by a network of CCTV cameras with disjointed camera views, and no effort has been taken to tackle the challenging problem of detecting abnormal global behaviours, which are only meaningful and recognisable when observed over space and time across multiple camera views. To that end, we propose a novel framework, which consists of object tracking across camera views, global behaviour modelling based on unsupervised learning, and probabilistic abnormality inference. The effectiveness of the framework is demonstrated with experiments on real-world surveillance video data.

1 Introduction

There are now large networks of CCTV cameras collecting colossal amounts of image data for safety and security purposes. A multiple camera and multi-sensor system has the potential for gaining better viewpoints resulting in both improved imaging quality and more relevant details being captured. However, more is not necessarily better. Such a system can also cause overflow of information and confusion if data content is not analysed in real-time. There is an urgent need for the development of automated systems to monitor cooperatively the behaviours of objects across a distributed network of cameras in order to enable on-the-fly decision-making.

Despite the wide use of networked CCTV systems, there is relatively little effort made worldwide on the development of effective and robust algorithms for utilising/sharing object behaviour information across multiple cameras for abnormal behaviour detection, as the problem is potentially very difficult and unconstrained. Recent studies on behaviour profiling and abnormality detection have been focused on a single fixed camera view. Among the existing efforts using multiple cameras, almost all have been focused on tracking. However, what has emerged to be of critical importance and urgently required for assisting decision making is the capability for monitoring and correlating local behaviours in order to infer global abnormal behaviours in a wider context beyond any single camera view. To the best of our knowledge, none of the existing work addresses the problem of correlating behaviours across different camera views and detecting subtle abnormalities that are only meaningful and recognisable when observed over space and time across multiple camera views.

To that end, we propose a novel framework, which consists of object tracking across camera views, global behaviour modelling based on unsupervised learning, and probabilistic abnormality inference. In particular, objects are tracked within each camera view and their trajectories are then connected to form a representation of the global behaviour pattern. The connected global trajectories are then profiled using a clustering technique, upon which a model for the normal behaviours are constructed. The abnormality is detected when an unseen behaviour pattern has a low likelihood of being generated by the learned normal global behaviour model.

The literature is rich in trajectory clustering methods. A recent review paper discussing various methods of trajectory clustering in outdoor surveillance scenes can be found in [23]. Much work has been invested in clustering trajectories in a viewfield of a single camera [18, 4, 12]. The paper in [18] uses two similarity measures. One similarity measure is used to cluster trajectories into vehicles or pedestrians based on the size of the targets. Another similarity measure is used to cluster vehicles and pedestrian trajectories into different clusters using spectral clustering based on the spatial and velocity information from the trajectory of the targets. Abnormal trajectories are detected as outlier tracks, *i.e.*, those which do not match very well to any of the vehicle or pedestrian clusters learnt. The aim of this paper

is to show that this can be done with multiple cameras.

In order to extend clustering approaches to a network of CCTV cameras, we must link tracks obtained in different fields of view. The correspondence (or tagging) task aims at matching the appearance of the same object in different cameras [7, 8, 9]. This is not a trivial task as often the color histograms of the same object in different cameras are quite different. Also, different objects often have similar appearance across different camera views. The main reasons for this are due to different lighting conditions in each camera, different orientation of the object to each camera and different characteristics of the cameras themselves. As such, the correspondence problem is still very much an open ended problem and is yet to be solved well. The paper in [19] recognises the challenges of the correspondence problem and bypasses it by imposing less constraints on the topology of the camera network. The authors build a graph where each node in the graph represents a trajectory. The start and end time of trajectories are known. Then, if two trajectories are in different cameras but their temporal extents are close, within a set positive temporal threshold, it is likely that the two trajectories belong to the same object.

The input into the clustering algorithms that we will be discussing here is usually in the form of observations on trajectories, *i.e.*, tracking, whether it be for a single field of view or multiple fields of view. The paper in [17] looks at feature matching for tracking applications in a camera network. It looks at three main issues; treating feature similarities as random variables when feature similarities are uncertain, looking at finding features over space and time, *i.e.*, long term interdependence of features and showing how the first two issues can be addressed by treating the tracking problem in a camera network as an optimisation problem in a stochastic adaptive framework. This is achieved by learning the probability distributions between observations at two camera nodes by treating target features (normalised color and gait width vector [11, 10]) as random variables and using dynamic programming [16] for finding optimal paths in graphs with stochastic weights. The drift problem is dealt with by looking at long-term interdependences between features over space and time for correction and by proposing a path smoothness function to recalculate the weights for the dynamic programming procedure.

In this paper, we use temporal links [19, 14] in conjunction with trajectory color signatures to link tracks across cameras and detect abnormal global trajectories within a network of multiple cameras. We have a statistical representation of where normal tracks occur globally in the camera network from a training phase. More specifically, using our training data set of local trajectories within each camera, we first learn the entry and exit zones for each camera via Gaussian Mixture Models [13] and then we learn the camera tempo-topographical model to obtain temporal links between exit and entry zones in different cameras [14]. We use these temporal links in conjunction with a tagger in order to link tracks of the same object across cameras to obtain global trajectories.

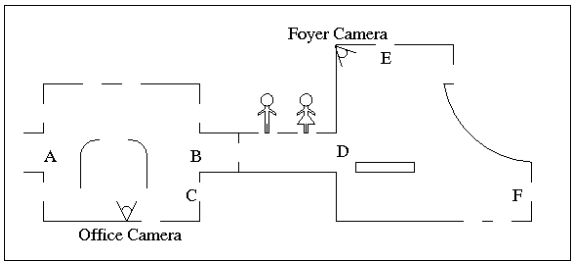


Figure 1. Schematic of the office and foyer cameras.

Using the global trajectories obtained from the training phase, we obtain similarities between global tracks using spatial and velocity information and we use spectral clustering methods in order to group global tracks into clusters. For each global cluster, we learn its semantic scene model. This model can be viewed as a likelihood of a point on a trajectory at a certain global position. Then, for each global track in the testing phase, we can look at the ensemble of points which make up its trajectory, find their corresponding likelihoods from the semantic model for each cluster and determine if the trajectory is likely to fit well with at least one global cluster. If this is the case, then that global trajectory is classified as being normal but if the global track does not match well with any of the global clusters, then it is classified as an abnormal track.

2 Multicamera Behaviour Profiling

We start by obtaining local trajectories for training within each camera. In our experiments we used two cameras, which we call 'office' and 'foyer' respectively, as seen in Figure 2. A schematic of how the two cameras are positioned with respect to each other can be seen in Figure 1.

After the background/foreground segmentation, we use a particle filter in our experiments [1]—a multiple hypothesis tracker—to track blobs, and we group blobs into targets by employing similar techniques to those in [15, 6]. The number at one extremity of a local trajectory in Figure 2 is a unique identifier for each track. At the same time, it also marks the beginning of a track and provides information regrading the track direction.

Using the local tracks, we are able to obtain entry and exit zones into each scene using Gaussian Mixture Models (GMMs) [13], depicted in Figure 3 (an '**s**' is an entry or start zone and an '**e**' is an exit zone). It can be seen from Figure 1 and Figure 3 that the entry and exit zones of interest

are $1 \rightarrow 2$ for tracks going from the office to the foyer and $2 \rightarrow 1$ for tracks going from the foyer to the office.

Using tempo-topographical cross-correlation functions described in [14], we can learn the most popular transition time across the blind spot for each entry/exit zone combination of interest by constructing a cross-correlation function for each combination, $R_{1\rightarrow 2}^{\mathrm{o}\rightarrow\mathrm{f}}$ and $R_{2\rightarrow 1}^{\mathrm{f}\rightarrow\mathrm{o}}$ where o $\rightarrow$ f and f $\rightarrow$ o means 'office to foyer' and 'foyer to office' respectively. The cross-correlation plots can be seen in Figure 4.

From each cross-correlation function, we extract two pieces of information, μ and σ, where μ is the time of the peak in the cross-correlation function and σ is the standard deviation of the times in the discrete buffer. We use this information in conjunction with color signatures in order to associate tracks and obtain global trajectories across our two cameras. We now briefly discuss how we do this.

2.1 Track Association

Our aim is to associate tracks so that we can cluster them globally. In this section, we discuss the feature similarity scores that we use.

Since we know the start times, t^s, and end times, t^e, of local trajectories in each camera and we have clustered the start and end points of tracks into entry and exit zones, for each track ending in an exit zone of interest, we look at all tracks starting in an entry zone of interest and assign for each combination a temporal transitional score using

$$S_t(e_p, s_q) = \begin{cases} \exp\left(-\frac{\left[\left(\left|t_p^s - t_q^e\right| - \mu\right)^2 + \lambda\right]}{\sigma}\right) & \text{if } t_q^s - t_p^e \leq 0 \\ \exp\left(-\frac{\left(\left|t_p^s - t_q^e\right| - \mu\right)^2}{\sigma}\right) & \text{if } t_q^s - t_p^e > 0 \end{cases}, \quad (1)$$

where $p = 1, \ldots, P$ is the number of points ending in an exit zone cluster, $q = 1, \ldots, Q$ is the number of points starting in an entry zone cluster and μ and σ were defined above. If the absolute time difference between an ending track and starting track pairing, $\left|t_p^s - t_q^e\right|$, is close to the peak time, μ, in the corresponding cross-correlation function, the temporal transitional score will be high (1 at the most) and as this difference starts to become larger, the temporal transitional score will become smaller following an exponential rule. If however a track in the entry zone of interest starts before a track ends in the exit zone of interest for a particular pairing of tracks, it is very unlikely that the two tracks should be associated so a positive penalty term, λ, is introduced in the equation.

We also assign for each combination a color score using the BHATTACHARYYA distance [2]

$$S_c(e_p, s_q) = \sum_{i=1}^{N} \sqrt{H_p^e(i) H_q^s(i)}, \quad (2)$$

where H_p^e is the color histogram of a target disappearing in an exit zone of interest and H_q^s is the color histogram of a target appearing in an entry zone of interest. N is the number of bins used in the histograms. In our experiments, we set $N = 256$. Geometrically, the BHATTACHARYYA distance is a dot product between two vectors. If the two vectors (in this setting, color histograms) are normalised so that

$$\sum_{i=1}^{N} H_p^e(i) = 1, \quad \text{and} \quad \sum_{i=1}^{N} H_q^s(i) = 1, \quad (3)$$

the BHATTACHARYYA distance will have a minimum value of 0 if the two vectors are orthogonal and a maximum value of 1 if the two vectors are coincident.

An overall score is computed for each combination by performing a multiplication of the temporal transitional score and the BHATTACHARYYA distance, *i.e.*

$$S_o(e_p, s_q) = S_t(e_p, s_q) S_c(e_p, s_q). \quad (4)$$

Then, only if

$$S_o(e_p, s_q) > \rho, \quad (5)$$

where ρ is a positive threshold between 0 and 1, the ending track in the exit zone of interest is assigned to the starting track in the entry zone of interest with the highest overall score, *i.e.*, $\max\left(\{S_o(e_p, s_q)\}\right) \forall S_o(e_p, s_q) > \rho$ where $p = 1, \ldots, P$ and $q = 1, \ldots, Q$. If the constraint in (5) is not met for a given track ending in an exit zone of interest and all tracks starting in the entry zone of interest, no track association will occur.

2.2 Clustering Global Trajectories

We employ the similarity measures in [18] with automatic selection of scaling constants and the rationale in [3] to construct a similarity matrix in order to cluster global tracks. We construct a symmetric $M \times M$ affinity matrix, $\mathbf{A}$, which is made up of symmetric similarity scores, $A_{I,J}$, between global tracks I and J where $I, J = 1, \ldots, M$ and M is the number of global tracks in the training data set. Using the rationale in [3], each entry in $\mathbf{A}$ is evaluated using position and velocity information of global tracks, so that

$$A(I, J) = \begin{cases} a_{\min}(I, J) & \text{if } a_{\min}(I, J) < \tau \\ 1 - \frac{a_{\max}(I,J) - a_{\min}(I,J)}{a_{\max}(I,J) + a_{\min}(I,J)} & \text{if } a_{\min}(I, J) \geq \tau \end{cases}, \quad (6)$$

where

$$a_{\min}(I, J) = \min(a_{I\rightarrow J}, a_{J\rightarrow I}), \quad (7)$$

$$a_{\max}(I, J) = \max(a_{I\rightarrow J}, a_{J\rightarrow I}). \quad (8)$$

The parameter τ is a chosen similarity threshold and $a_{\cdot\rightarrow\cdot}$ is a directed similarity score between global tracks I and J or J and I. Looking at (6), if the directed similarities $a_{I\rightarrow J}$ and $a_{J\rightarrow I}$ are asymmetric or both low, we bound them by their lowest value, $a_{\min}(I,J)$. If however, $a_{\min}(I,J)$ is above or equal to τ, the symmetric similarity is defined as per (6) where we have some confidence that the two trajectories I and J are similar spatially and velocity wise, as we shall now see. The directed similarity score $a_{I\rightarrow J}$ is calculated as follows

$$a_{I\rightarrow J} = \frac{\sum_{\mathbf{i}_k \in I} \gamma(\mathbf{i}_k, \mathbf{j}_{\psi(k)}) \alpha(\mathbf{i}_k, \mathbf{j}_{\psi(k)})}{\sum_{\mathbf{i}_k \in I} \gamma(\mathbf{i}_k, \mathbf{j}_{\psi(k)})}, \tag{9}$$

and similarly for $a_{J\rightarrow I}$, where $\mathbf{i}_k$ are observations on trajectory I and $\mathbf{j}_{\psi(k)}$ is the closest observation on trajectory J to observation $\mathbf{i}_k$ on trajectory I, *i.e.*,

$$\psi(k) = \arg\min_{m \in J} \|\mathbf{i}_k - \mathbf{j}_m\|. \tag{10}$$

The function $\gamma(\mathbf{i}_k, \mathbf{j}_{\psi(k)})$ is a comparison confidence between the two observations $\mathbf{i}_k$ and $\mathbf{j}_{\psi(k)}$ and $\alpha(\mathbf{i}_k, \mathbf{j}_{\psi(k)})$ is a feature similarity between the two observations $\mathbf{i}_k$ and $\mathbf{j}_{\psi(k)}$ such that

$$\gamma(\mathbf{i}_k, \mathbf{j}_{\psi(k)}) = \exp\left(-\frac{\left\|\mathbf{i}_k - \mathbf{j}_{\psi(k)}\right\|}{\sigma_1}\right), \tag{11}$$

$$\alpha(\mathbf{i}_k, \mathbf{j}_{\psi(k)}) = \exp\left(-\frac{\delta(\mathbf{v}^{\mathbf{i}}_k, \mathbf{v}^{\mathbf{j}}_{\psi(k)})}{\sigma_2}\right), \tag{12}$$

where

$$\delta(\mathbf{v}^{\mathbf{i}}_k, \mathbf{v}^{\mathbf{j}}_{\psi(k)}) = 1 - \frac{\mathbf{v}^{\mathbf{i}}_k \cdot \mathbf{v}^{\mathbf{j}}_{\psi(k)}}{\left\|\mathbf{v}^{\mathbf{i}}_k\right\| \left\|\mathbf{v}^{\mathbf{j}}_{\psi(k)}\right\|}, \tag{13}$$

and $\mathbf{v}^{\mathbf{i}}_k$ and $\mathbf{v}^{\mathbf{j}}_{\psi(k)}$ are the velocities at observations $\mathbf{i}_k$ and $\mathbf{j}_{\psi(k)}$. For each observation $\mathbf{i}_k$ on trajectory I, we find its spatially nearest observation $\mathbf{j}_{\psi(k)}$ on J, compute the comparison confidence $\gamma(\mathbf{i}_k, \mathbf{j}_{\psi(k)})$ and feature similarity $\alpha(\mathbf{i}_k, \mathbf{j}_{\psi(k)})$. Along trajectory I, feature similarities of observations are averaged and weighted by the comparison confidences to get the directed similarity score $a_{I\rightarrow J}$. The similarity of observations close in space has larger weight for computing global trajectory similarity. The comparison confidence indicates how far apart I is from J.

We found that the choice of scaling constants σ_1 and σ_2 in (11) and (12) is crucial for the clustering stage. As discussed in [22], the selection of scaling constant is commonly done manually, as is the case in [18]. However, we found that manual selection of such quantities can have a very big impact on the performance of the clustering algorithm. We therefore calculate our scaling constants automatically as a function of spatial and velocity information from observations on global trajectories as follows

$$\sigma_1 = \max\left(\text{mean}_{m \in J}\left(\|\mathbf{i}_k - \mathbf{j}_m\|\right), \zeta\right), \tag{14}$$

where ζ is a small predefined constant. We take the maximum of the two arguments in (14) in case an observation lies on both global tracks. And

$$\sigma_2 = \text{std}_{m \in J}\left(1 - \frac{\mathbf{v}^{\mathbf{i}}_k \cdot \mathbf{v}^{\mathbf{j}}_m}{\left\|\mathbf{v}^{\mathbf{i}}_k\right\| \left\|\mathbf{v}^{\mathbf{j}}_m\right\|}\right). \tag{15}$$

Our aim is to group global trajectories which have similar spatial and velocity characteristics. Once we calculate the affinity matrix $\mathbf{A}$, we apply a standard normalised cut clustering algorithm [21, 20].

2.3 Learning Path Regions for Global Clusters

Once we have trajectories grouped into clusters, we detect path regions in the scene by thresholding a density distribution proposed in [18]. The difference here is that we have clustered global trajectories and the density distribution, which can be viewed as a likelihood function, is for global path regions. For each global cluster, Ω, the density at each position, $\mathbf{p}$, in the global scene is

$$L_\Omega(\mathbf{p}) = \sum_{\mathbf{i}_k \in I} \sum_{I \in \Omega} \exp\left(-\frac{\|\mathbf{p} - \mathbf{i}_k\|^2}{\sigma_3}\right), \tag{16}$$

and here, I and $\mathbf{i}_k$ represent global tracks in global clusters and the observations on global clusters respectively. The scaling constant σ_3 is automatically set to the standard deviation of all the standard deviations of global tracks in the cluster, *i.e.*, for each global cluster

$$\sigma_3 = \text{std}_{I \in \Omega}\left(\text{std}_{k \in I}\left(\|\mathbf{i}_k - \mathbf{c}\|\right)\right), \tag{17}$$

where $\mathbf{c}$ is the centroid of each global track in the global cluster

$$\mathbf{c} = \frac{1}{K}\sum_{k=1}^{K} \mathbf{i}_k. \tag{18}$$

The density for each cluster is shown in Figure 7 (7(a),7(c),7(e) and 7(g)). The path regions are obtained by thresholding the density distributions using $\max L_\Omega(\mathbf{p})/10$ as in [18] (see Figures 7(b),7(d),7(f) and 7(h)).

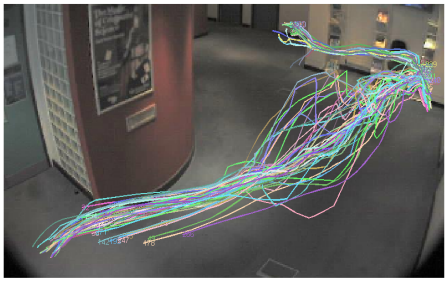
2.4 Abnormality Detection

Abnormal trajectories are detected as outlier tracks. If a global trajectory deviates from all global path regions in all global clusters, there will be observations on that trajectory with low corresponding density values from the cluster distributions. If a certain percentage, ϕ, of the length of a global track deviates from the path regions, the global track is classified as an abnormal trajectory.

3. Experiments

A plot of the local trajectories used for training within each camera is shown in Figure 2.

(a) Office tracks.



(b) Foyer tracks.

Figure 2. Local tracks in the office and foyer cameras.

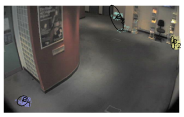
Using the local trajectories, we learn entry and exit zones into each scene using Gaussian mixture models [13] and the expectation-maximisation algorithm [5]. The clusters can be seen in Figure 3.

The entry and exit zones of interest are $1 \rightarrow 2$ for tracks going from the office to the foyer and $2 \rightarrow 1$ for tracks going from the foyer to the office. We use these entry and exit zones to construct tempo-topographical cross-correlation functions described in [14] and learn the most popular transition time across the blind spot for each entry/exit zone combination of interest, $R_{1\rightarrow 2}^{\mathrm{o}\rightarrow\mathrm{f}}$ and $R_{2\rightarrow 1}^{\mathrm{f}\rightarrow\mathrm{o}}$ where $\mathrm{o} \rightarrow \mathrm{f}$ and $\mathrm{f} \rightarrow \mathrm{o}$ means 'office to foyer' and 'foyer to office' respectively. In our experiments, we used a time-search window, T, of 30 seconds and the cross-correlation functions are plotted in Figure 4. It can be seen from the plots that there are peaks at 6 seconds in $R_{1\rightarrow 2}^{\mathrm{o}\rightarrow\mathrm{f}}$ and 8 seconds in $R_{2\rightarrow 1}^{\mathrm{f}\rightarrow\mathrm{o}}$ respectively, which are the transition times from the respective exit zone to the entry zone. Of course, the more tracks that we have during training, the more robust the cross-correlation function will be. Furthermore, cross-correlation functions can have several modes, depending on what people do in the blind spots between cameras, however, the data that we collected for training did not exhibit this statistically.

(a) Office entry zones.

(b) Office exit zones.

(c) Foyer entry zones.



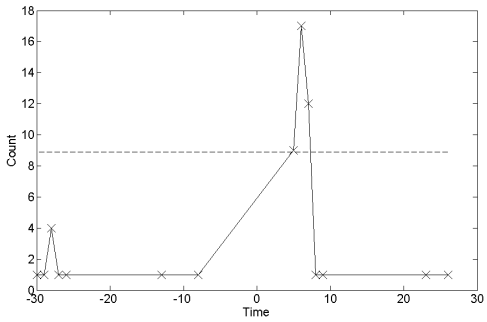
(d) Foyer exit zones.

Figure 3. Entry and exit zones in the scenes. 's' stands for start and 'e' stands for end.

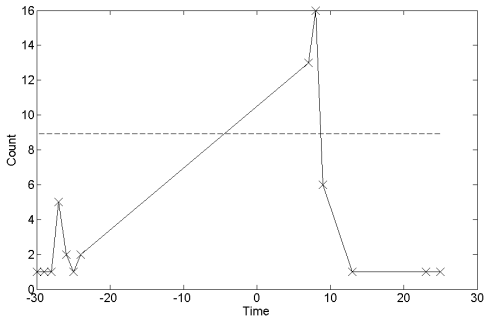
Using cross-correlation functions and color information of targets, we can associate tracks as discussed in Section 2.1. Looking at Figure 1 and the local tracks in Figure 2, for the training set, we obtained global trajectories going from the office camera to the foyer camera through $ABDE$ and $CBDF$ and global trajectories going from the foyer camera to the office camera through $EDBA$ and $FDBC$. There were 40 local tracks going from the office camera to the foyer camera and 37 local tracks going from the foyer camera to the office camera, *i.e.*, 77 global tracks. 76 global tracks were linked correctly. The only global track that was not associated in the training set belonged to a person that went from the foyer to the office through EDBA via the male toilets in the blind gap between the cameras. That person spent more time in the blind gap than the learnt temporal connection between the respective exit zone in the foyer camera and the entry zone in the office camera (which was 8 seconds), the overall score (see Equation 5) was very low and hence, no track association occurred.

A plot of all the global associated tracks going from the office to the foyer and from the foyer to the office is in Figure 5. The numbers in brackets at one extremity of a global track in Figure 5 is a unique identifier for each trajectory. The two numbers in the brackets $(I - J)$ specify which local tracks were associated together to obtain the global track. Our $(I - J)$ notation also marks the beginning of a global track and that together with the order of the local scenes in the global scene provides information regarding the track direction.

Having obtained the global tracks, we are able to con-



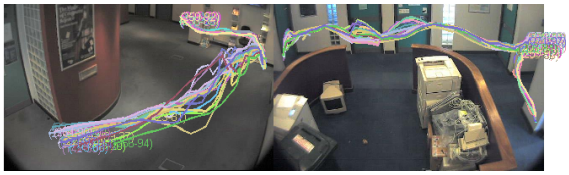
(a) $R_{1\rightarrow 2}^{\mathrm{o}\rightarrow \mathrm{f}}$.



(b) $R_{2\rightarrow 1}^{\mathrm{f}\rightarrow \mathrm{o}}$.

Figure 4. Cross-correlations functions for entry/exit zones of interest.

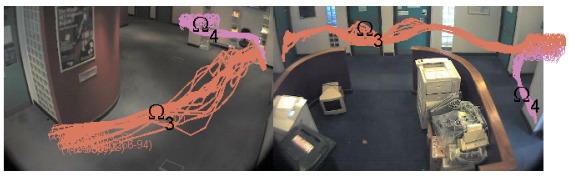
(a) Global tracks from the office to the foyer camera.



(b) Global tracks from the foyer to the office camera.

Figure 5. Global tracks across the cameras.

(a) Clustered global tracks from the office to the foyer camera.



(b) Clustered global tracks from the foyer to the office camera.

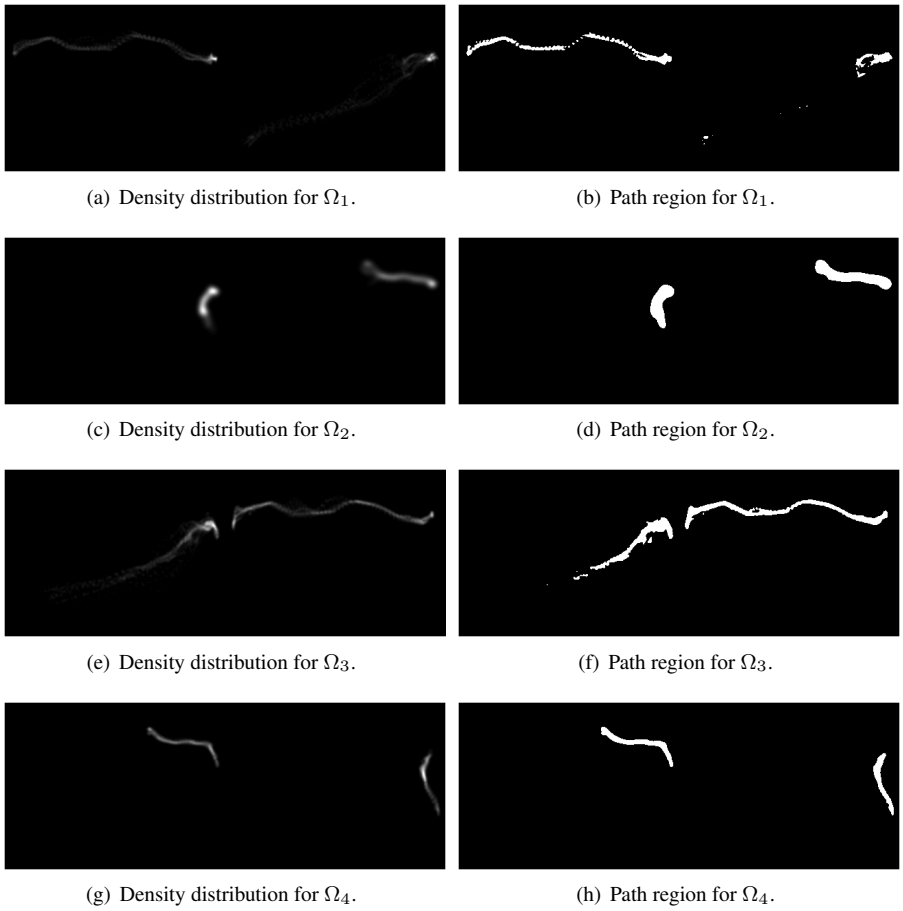
Figure 6. Clustered global tracks across the cameras.

struct the affinity matrix, $\mathbf{A}$, using (6) as described in Section 2.2 and use that as input to a normalised cut clustering algorithm [21, 20]. The results are shown in Figure 6. As expected, we obtain four global clusters of trajectories going from the office to the foyer through $CBDF$ (global cluster Ω_2) and $ABDE$ (global cluster Ω_1), and trajectories going from the foyer to the office through $FDBC$ (global cluster Ω_4) and $EDBA$ (global cluster Ω_3).

The density for each cluster can be seen in 7(a),7(c),7(e) and 7(g) and the path regions are obtained by thresholding the density distributions using $\max L_{\Omega}(\mathbf{p})/10$ as in [18] (see Figures 7(b),7(d),7(f) and 7(h)). Given the path regions, it is possible to use some logical operations in order to detect global anomalous trajectories in the scene. This can be done using many heuristics depending on the application at hand. It might be sufficient to observe individual positions on tracks and see if their probability is low. However, because the thresholding might not be perfect at all sections of a global path region, as can be seen in Figure 7(b), a more robust approach to detect anomalous tracks is to check if all observations on global trajectories are sufficiently close to path regions of global clusters. Statistically, this approach becomes increasingly more robust when the global scene is comprised of a larger set of local scenes, *i.e.*, global path regions are longer and occur across more cameras.

If a global trajectory deviates from all global path regions in all global clusters, there will be observations on that trajectory with low corresponding density values from the cluster distributions. In our experiments, we set ϕ defined in Section 2.4 to 20%, *i.e.*, if more than 20% of the total length of a global track deviates from the path regions, that track is classified as an abnormal trajectory. We tested 14 global trajectories, 2 were abnormal and 12 were normal. The global test trajectories are shown in Figure 8(a) and 8(b).

The trajectory going through $FDBA$ and the one going through $EDBC$ (refer to Figure 1) should be detected as abnormal trajectories as they do not fit well with any of the training path regions. The rest of the trajectories should be



(a) Density distribution for Ω_1. (b) Path region for Ω_1.

(c) Density distribution for Ω_2. (d) Path region for Ω_2.

(e) Density distribution for Ω_3. (f) Path region for Ω_3.

(g) Density distribution for Ω_4. (h) Path region for Ω_4.

Figure 7. Global density distributions and path regions for clusters.

detected as normal as they conform well with a learnt path region. The results can be seen in Figure 8(c) and 8(d). As expected, the normal global tracks were labeled in green and the abnormal global tracks were labeled in red.

As another experiment, we let ϕ vary from $0\% - 100\%$ in increments of 2.5% between $0\% - 20\%$ and in increments of 10% between $20\% - 100\%$ and plotted the ROC curve to get a relationship between the true-positive and false-positive prediction rate. The results are plotted in Figure 9. The ROC curve shows that the abnormal global tracks can be easily distinguished from the normal ones resulting in excellent detection performance.

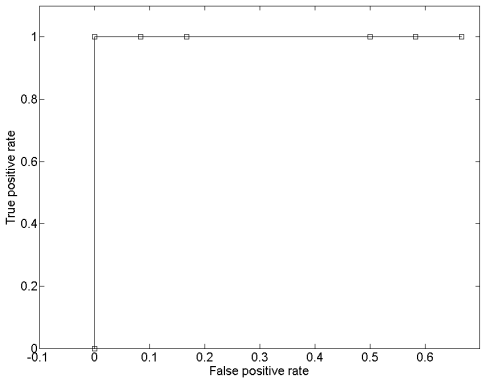
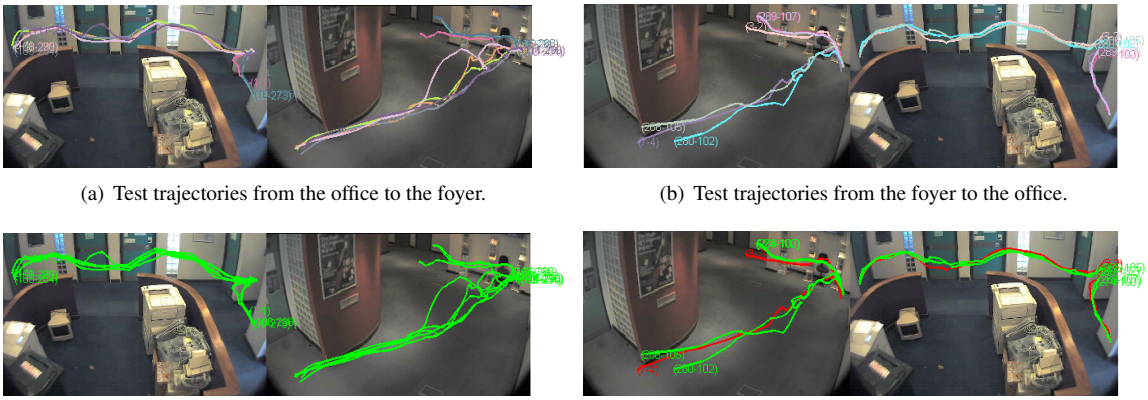


Figure 9. Abnormality detection performance measured using a ROC curve.

4 Conclusions

We have presented a novel framework for addressing the problem of correlating behaviours across different camera views and detecting subtle abnormalities that are only meaningful and recognisable when observed over space and time across multiple camera views. The effectiveness of the framework has been demonstrated with experiments on real-world surveillance video data. Future work will involve the use of more cameras in the system and experi-



(a) Test trajectories from the office to the foyer.

(b) Test trajectories from the foyer to the office.

(c)

(d)

Figure 8. Test trajectories in the global scenes. The green global test trajectories in 8(c) and 8(d) were detected as normal tracks while the red global test trajectories were detected as anomalous tracks.

menting with multi-modal temporal cross-correlation functions and tagging routines for global trajectory association across many disjoint cameras. Also, further experiments are required with more complicated activity across the camera network.

References

[1] S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for on-line non-linear/non-gaussian bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–188, Feb. 2002.

[2] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[3] E. D. Cheng and M. Piccardi. Mathching of objects moving across disjoint cameras. In *International Conference on Image Processing: ICIP*, 2006.

[4] H. Dee and D. Hogg. Detecting inexplicable behaviour. In *Proc. of BMVC*, 2004.

[5] 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.

[6] L. M. Fuentes and S. A. Velastin. People tracking in surveillance applications. In *Proc. of 2nd IEEE Int. Workshop on PETS*, 2001.

[7] N. Gheissari, T. B. Sebastian, J. Rittscher, and R. Hartley. Person reidentification using spatiotemporal appearance. In *Proc. of CVPR*, 2006.

[8] O. Javed, Z. Rasheed, K. Shafique, and M. Shah. Tracking across multiple cameras with disjoint views. In *Proc. of ICCV*, 2003.

[9] O. Javed, K. Shafique, and M. Shah. Appearance modeling for tracking in multiple non-overlapping cameras. In *Proc. of CVPR*, 2005.

[10] A. Kale, A. Rajagopalan, A. Sundaresan, N. Cuntoor, A. Roy-Chowdhury, A. Krueger, and R. Chellappa. Identification of humans using gait. *IEEE Trans. on Image Processing*, pages 1163–1173, Sep 2004.

[11] A. Kale, A. Roy-Chowdhury, and R. Chellappa. Towards a view invariant gait recognition system. In *Advanced Visual Surveillance Systems: IEEE AVSS*, 2003.

[12] D. Makris and T. Ellis. Path detection in video surveillance. *Image Vision and Computation*, 20(1):859–903, 2002.

[13] D. Makris and T. Ellis. Automatic learning of an activity-based semantic scene model. In *Proc. of AVSS*, 2003.

[14] D. Makris, T. Ellis, and J. Black. Bridging the gaps between cameras. In *Proc. of CVPR*, 2004.

[15] S. J. McKenna, S. Jabri, Z. Duric, H. Wechsler, and A. Rosenfeld. Tracking groups of people. *Computer Vision and Image Understanding: CVIU*, 80(1):42–56, 2000.

[16] R. Sedgewick. *Algorithms*. Addison-Wesley, Reading, Mass., 2 edition, 1988.

[17] B. Song and A. K. Roy-Chowdhury. Stochastic adaptive tracking in a camera network. In *International Conference on Computer Vision: ICCV*, 2007.

[18] X. Wang, K. Tieu, and E. Grimson. Learning semantic scene models by trajectory analysis. In *Proc. of ECCV*, 2006.

[19] X. Wang, K. Tieu, and W. Grimson. Correspondence-free multi-camera activity analysis and scene modeling. In *Proc. of CVPR*, 2008.

[20] S. X. Yu and J. Shi. Grouping with bias. In *Proc. of NIPS*, 2001.

[21] S. X. Yu and J. Shi. Segmentation given partial grouping constraints. *IEEE Transactions on Pattern Analysis and Machine Intelligence: PAMI*, 26(2):173–183, 2004.

[22] L. Zelnik-Manor and P. Perona. Self-tuning spectral clustering. In *Proc. of NIPS*, 2004.

[23] Z. Zhang, K. Huang, and T. Tan. Comparison of similarity measures for trajectory clustering in outdoor surveillance. In *Proc. of ICPR*, 2006.