The Detection of Suspicious Behaviour in CCTV

Michael Leach, Heriot Watt University, Edinburgh, UK — Michael.leach@roke.co.uk
Ed Sparks, Roke Manor Research, UK — ed.sparks@roke.co.uk
Neil Robertson, Heriot Watt University, Edinburgh, UK — n.m.robertson@hw.ac.uk

Roke

Hypothesis

Behavioural anomaly detection requires the identification of behaviours which differ from a defined set of models describing what is normal. However any such set is not universally applicable over all possible surveillance scenes. For this reason it is necessary to learn to train a set of models particular to a scene. In a dynamic scene the set of normal behaviour models may not even be globally applicable over the scene. The normal behaviours which anomalies contrast are dependent upon scene region, social context, periodic events, or other scene dynamics. It is not enough to train a behaviour class model, we must also know the limitations of the model in context and the contributing factors of the environment which impact the representation of normality. Our early work looks at scene segmentation and social group detection. We learn a model of normal behaviour independently for the different partitions of the scene to enable the detection of context dependant abnormal behaviours.

Objective of the system

- Automatically detect abnormal human behaviour in a unconstrained surveillance scene
- Insert an alarm into the video stream when abnormal behaviour is detected
- Low false alarm rate—below end-user tolerance
- Detect subtly abnormal behaviour such as loitering, as well as more obvious abnormal behaviour such as running

Input

Anomaly detection: Our input data is the PETS 2007 Airport scene As a sequence of images [1]. **Social Model:** The PETS 2006 data is used to develop the social model aspect of this work.

PETS 2007 Scene 4—View 1: Image plane and ground plane

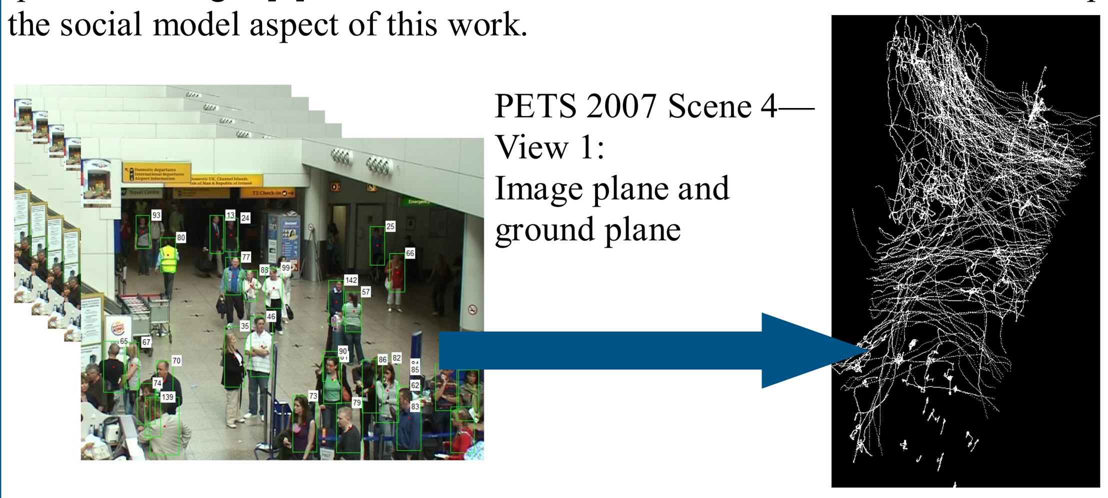


Image Plane **Ground Plane**

Pedestrian Tracking

Tracking of targets in the PETS video stream is achieved using the manually initialised TLD—Predator algorithm [2]. The output of this tracker is a bounding box at each frame for each person in the video.

Using the known camera calibration data the image plane bounding box can be translated to ground plane coordinates.

The ground plane point track forms the basis for the motion feature vector used in our behaviour ontology.

The TLD predator algorithm is a target tracker developed in Surrey University by Zdenek Kalal [3] released under the GPL license.

Scene Segmentation

Method:
We define three different scene regions [traffic, idle, and convergence/divergence] in which the scene can be segmented. Additionally the NULL region exists when there is not enough evidence to support any of the search for regions. Each region is defined by an combination of the below features.

The scene is divided into a grid in which the following features are calculated. Grid locations with enough supporting evidence for a given region classification [traffic lane, idle region or con/divergence] are classed as such.

Objective: To partition the ground plane map of the scene into regions in which the expected distribution of feature values is different for normal behaviour.

Using the scene segmentation we are able to learn a model of normality independently for each scene, allowing for regional variation in normality. For N regions we do not necessarily require N times more training data as there may be substantial mutual information between regions.

Features: The features used to define the scene regions are as follows:

Directional entropy
Speed entropy
Direction energy
Speed Energy
Temporal Persistence
Trajectory density

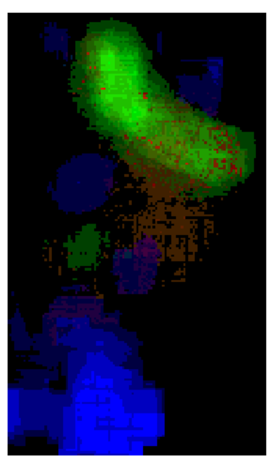
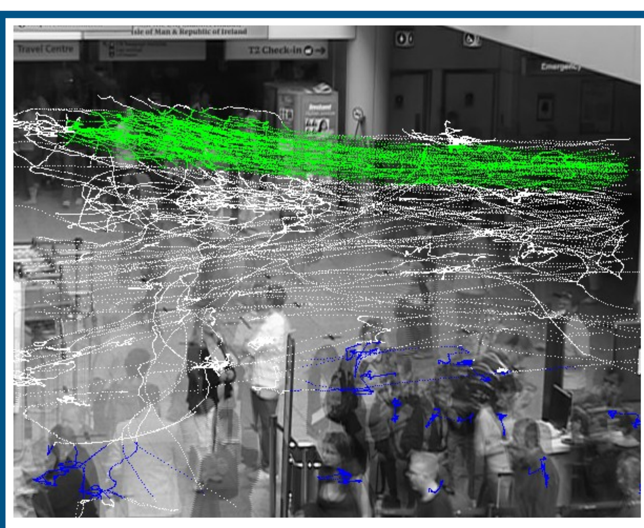


Figure: Image displaying the region intensity map. Green shows the traffic region, blue shows the idle region, and brown shows the convergence/divergence region. An intensity threshold determines the grid locations of the scene which are designated as any one of the scene locations. The convergence/divergence region is excluded due to a higher than permissible standard deviation.



Result: Image plane tracks with the scene segments. Green tracks are in the traffic region, and blue in idle areas.

Social Model

Objective: We define a metric used to measure the apparent social connectivity between any two individuals in the scene.

The strength of the connection between any two individuals allows us to identify social groupings when a cluster of individuals share a strong connection between each other.

Social groups are used to partition our behaviour observations into the feature sets observed within a social group and outside a social group, allowing the behaviour model to be incorporated with social context.

Method: The measure of social connectivity follows from similar research in computer vision, but includes a greater number of motion features. Selection and weighting of the feature weights is performed by examining the ROC curve for social grouping of ground truth data.

At each frame we calculate the social strength for all pairs of individuals using the features below, then group strong pairs into social cliques. Groups are extended to multiple people by simply locating closed groupings in which all members have one or more connections to the group above a threshold.

Result: Two illustrative social groups identified

Features: The features used in the social connectivity metric are as follows:

Feature Type	Weighting
Distance difference	0.86
Proximity difference	1.51
Speed difference	0.75
Temporal Overlap ratio	0.88

Future Work

There are three major routes forward from the current state of work:

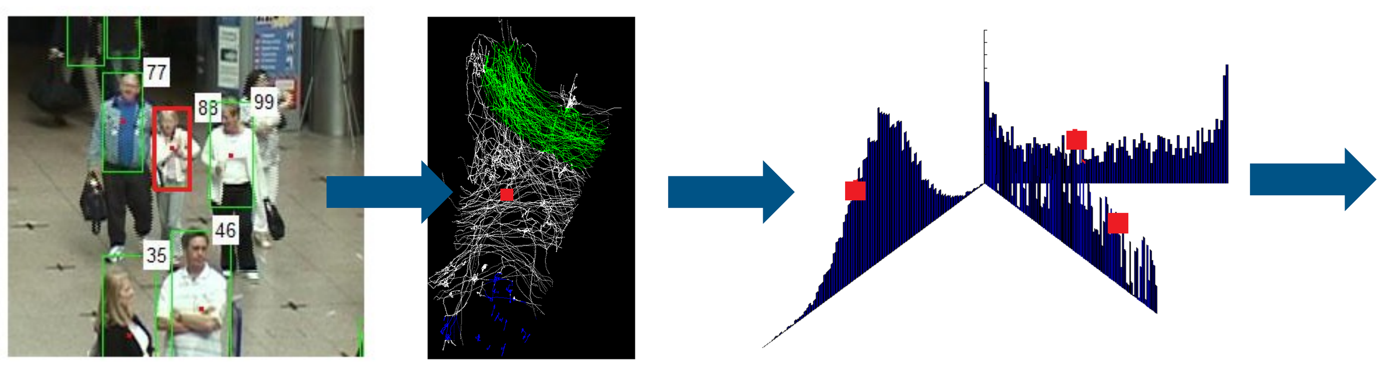
1. Implementing a automated pedestrian identification and tracking. We have initiated work combining the Felzenswalb [4] pedestrian identification and TLD tracker to automatically track multiple objects in surveillance video.
2. Inclusion of automated gaze estimation. Our current work strives to estimate the gaze direction feature at each frame. Gaze direction will be treated in our behaviour ontology in much the same way as motion features. However the gaze feature provides an indication of intention of motion, and can be used in the social modelling.
3. We will readdress the social model to take advantage of the high mutual information between social pairs. We can evaluate if an individual is included in a hypothetical social group by the increase in mutual information, or decrease in minimum description length of the scene as a whole.

References

[1] UK ESPRC REASON: *http://www.cvg.rdg.ac.uk/PETS2007/*
[2] Z. Kalal. Online Learning of Robust Object Detectors During Unstable Tracking. 3rd Online Learning for Computer Vision Workshop. 2009
[3] TLD website: info.ee.surrey.ac.uk/Z.kalal/tld.html
[4] P. Felzenswalb. Bject Detection with Discriminatively Trained Part Based Models

Anomaly Detection

Anomaly detection is performed by generating an alarm for the lowest likely behaviour observations in the video sequence. Our behaviour ontology is a feature state consisting of the current motion, gaze direction, and temporal persistence of the individual. We train our system by populating a set of histograms which record the distribution of speed, direction, persistence, and gaze features from the training data set. Separate histograms are generated for each scene segment and both in and out of social groups.



The target is tracked in the image plane and the locations translated to a ground plane point track

The scene region which the target is in is located

The targets feature vector for the current frame is compared to a learnt model for the given scene region

$$\mathrm{P(observation \mid region)} = \frac{\mathrm{P(region \mid observation)} * \mathrm{P(observation)}}{P(region)}$$

The probability of observing the given feature vector is calculated conditionally upon the scene region in which it is located. The probability of the current observation provides the basis for anomaly detection. Clusters of anomalies for a given individual are flagged as anomalies and an alarm is inserted into the video stream.

Results: We aimed to be able to detect not only globally abnormal behavioural events but events which are only apparent when seen in context.

Example 1: We found that we are able to detect the true positive instances of globally unusual behaviour. In this instance running through the scene was identified as an abnormal behaviour. High speeds have a low statistical representation for all scene segments and thus are detected as the individuals run through multiple scene segments

Example 2: In this instance two individuals are detected as anomalies due to the long persistence at slow speeds in the NULL region. This is an instance of a more subtle anomaly as the speed in itself is not abnormal, it is only when the persistence and the scene segment are taken into account that the speed is notably unusual.

Example 3: In this anomaly instance an individual is detected as abnormal whilst he stands still in a traffic region, obstructing peoples movement. This nicely illustrates the efficacy of our method as it detects a behaviour which has a high global representation (over the entire scene) but a low local representation (in a traffic region).

Our objective for the system was to take into account the contextual information in which behaviour is observed to build a more discriminative behaviour model capable of detecting anomalies which are context specific. Our system has demonstrated the detection of behaviour anomalies which are globally distinct [Example 1 above] and those which are distinct only when placed in context [Example 2&3 above]. The behaviour ontology used is a primitive example illustrating the use scene segmentation and social modelling. Alternative probabilistic models of behaviour are easily incorporated. This work will be furthered by the inclusion of other contextual scene dynamics which impact behaviour such as periodic events (a train arriving), child to adult dependencies, and staff classification.

Optics & Photonics Technologies
Industrial Doctorate Centre
University of Strathclyde Glasgow
Heriot Watt University