

**This paper is a preprint of a paper accepted by “IET Computer Vision” and is subject to Institution of Engineering and Technology Copyright.**

# Modelling Periodic Scene Elements for Visual Surveillance

V.Leung, A.Colombo, J.Orwell and S.A.Velastin

Digital Imaging Research Centre, Kingston University,  
Penrhyn Road, Kingston upon Thames, KT1 2EE, U.K.  
{v.leung, a.colombo, j.orwell, sergio.velastin}@kingston.ac.uk

## Abstract

Some urban scenes exhibit periodic variations that can be relevant to visual surveillance applications. One example is the variation in the background elements, such as those caused by moving escalators, lights, and scrolling advertisements. When modelled correctly, the incorporation of these periodic elements as a Markov model in a foreground detection component can improve the performance significantly. Another area where the periodicity in the scene can be used is anomaly detection. In some underground metro stations where the flow of people is periodic, deviations from this periodicity can be interpreted as abnormal movements of people. This can be achieved by using a higher-dimensional model for the underlying data structure, and mapping it to a 1D signal for interpretation. This approach is tested, and the results show that abnormal behaviour can be automatically detected.

## 1 Introduction

An end-to-end visual surveillance system consists of components from foreground detection, tracking, to event or anomaly detection. In some environments where such systems are deployed, such as underground metro stations, there are periodically varying elements in the scene. Some of these elements belong to

the background, such as escalators and scrolling advertisements, and an accurate model of these elements improves the accuracy of foreground detection. There are also periodically varying elements of this foreground content, such as the flow of people which is a function of the train arrival and departure. An accurate model of this component would allow abnormal behaviour in the flow pattern to be detected.

In this paper, we address both of these cases where periodicity in the data is exploited in detecting elements of interest. Although they address different stages in the visual surveillance system, each case is concerned with the detection of signals that differ from a periodic “norm”. Similar principles can be applied to both cases.

## 1.1 Elements of interest: foreground pixels

The detection of moving objects in the scene requires a sufficiently accurate model of the background to enable them to be distinguished. This work considers the case in which the background is moving according to some repeating and predictable pattern. In data captured at underground metro stations, at least three elements of the background exhibit this property: escalators, flashing warning lights and scrolling advertisements.

The objective is to accurately and automatically model a background including periodically varying elements with an unknown time period (up to an arbitrary limit). This model can then be used to predict the appearance of these elements of the background in future frames, and thereby distinguish this changing background from foreground elements (such as moving people), allowing foreground objects on the periodic background to be estimated.

The simplest type of natural background is that observed by a fixed camera with constant illumination. In this case, process and input variability can be modelled with a Gaussian random variable for each pixel. More complex variations in the background signal can be modelled with a Gaussian Mixture Model (GMM) [1]. This can account for both signal changes caused by small camera motions and small movements in elements of the scene, e.g. windblown trees, as well as step-changes in appearance, e.g. changes in illumination caused by moving clouds, street lights, headlights, and other sources of light and shadow. For the case of significant camera motion, then this motion along with the background would need to be modelled. The GMM is also applicable to periodically varying backgrounds: for each pixel, each

component of the period is modelled by the most appropriate element in the mixture. Indeed, the regular cycle through these components ensures that the relative priors for the mixture elements can be accurately estimated. (This is in contrast to less predictable variations such as alternation between cloudy and sun-lit illumination.)

Alternatively, periodic variation in appearance can be considered as a special case of “dynamic texture”, and techniques for modelling this process have already been proposed. Soatto et al. [2] used a Kalman filter to model the evolutionary process of the dynamic texture, and determined the parameters using an iterative technique similar to Expectation-Maximisation (EM). This approach was adopted in [3], and segmentation of foreground objects from a dynamic background was achieved.

In this paper, periodic backgrounds are modelled with a Markov process. In a training phase, the periodic pixels in the image sequence are detected and characterized using Fourier analysis and used to initialise a Markov model that describes the series of observable states. This model can then be used to predict a background image, with which the input image is compared, and a foreground mask computed. The model can be incrementally updated to correct initial errors and accommodate non-stationary processes. The aim is to develop a method that can operate without human intervention.

The proposed approach is versatile in accommodating backgrounds with differing periodicity, since each pixel is modelled individually, and are only grouped into regions if they exhibit the same periodicity. In order to deal with dynamically changing periodicity, the uncertainty in the transition between the different states must be accounted for. It may be possible to employ the same approach for this problem; however it is likely that a more complex model, such as a Switching Linear Dynamic System (SLDS) [4], is required. In this paper, we focus on scenes with fixed periodicity. The topic of dynamically changing periodicity is beyond the scope of this paper.

The experiments are performed on two data sources collected as part of the CARETAKER European Project, which demonstrates technology to automatically extract knowledge from audio video sources. One sequence shows a platform and escalator in a station that forms part of the Torino Metro system. The periodic background elements are the escalator (with a period of approximately one second) and a flashing warning light (approx. two seconds). The other sequence shows the top of an escalator that is part of the Roma Metro system. It includes an advertising board that scrolls every ten seconds between four adverts, giving an overall period of about forty seconds. Foreground detection experiments have been

carried out on the Torino Metro data set and a significant improvement over the technique of Gaussian Mixture Models [1] has been shown.

## 1.2 Elements of interest: abnormal flow

The second case focuses on the detection of abnormal flow of people. This work is motivated by the requirement to automatically detect behaviour that is sufficiently different from the norm, and ultimately trigger an alarm to the operator. In cases where there is a periodic pattern to the movement of people, this requirement can be achieved by modelling the movement of people in the foreground, and detecting any deviation from this periodicity.

The detection of abnormal behaviour has received much attention recently, and has been addressed by many authors. Xiang and Gong [5] model each behaviour pattern with a Multi-Observation Hidden Markov Model (MOHMM), and an unseen behaviour pattern is detected as anomalous if the likelihood of the data given the models is below a certain threshold. A further improvement where incremental online update is incorporated is proposed in [6]. Hung and Gong [7] use the co-occurrence of salient features in both space and time for detecting anomalies. Andrade et al. [8] use the optic flow vectors as features, and train a MoGHMM (Mixture of Gaussian Hidden Markov Model) with which the likelihood of unseen data can be calculated. A significant drop in this quantity indicates an anomalous event, although this result was only illustrated on simulated data. Here, the construction of the model to describe the data is similar to the approach in [8], but the process of detecting anomalies is different.

Anomaly detection has also been an active area of research in many other disciplines such as medical signal processing [9, 22], condition-based monitoring [10, 11], as well as intruder detection in computer networks [12, 13, 14, 15] and mobile communications [16]. Of specific interest is the detection of abnormal segments in a periodic signal in medical signal processing, where a model or template for the normal process is obtained and an abnormality is detected by comparison with the normal signal. This idea is also used in our approach.

The assessment of the performance of anomaly detection applications is less straightforward than that of foreground detection. Usually, ground-truth must be provided by an expert, be it a cardiologist for the examination of ECGs (electrocardiograms), or an operator for studying CCTV footage. Here, we use data

from a Torino Metro station, and abnormal behaviour is defined as the loitering of people on sections of the platform where no persistent activity is expected. Experiments on this data shows that this behaviour can be automatically detected using the proposed method.

The remainder of the paper is structured as follows. Section 2 discusses the process of detecting periodically varying background pixels, and the generation of a Markov model to describe this variation. Foreground detection results are compared to the GMM and significant improvement is shown. In Section 3, the use of the periodic behaviour exhibited by foreground elements in anomaly detection is presented. Results show that this approach is feasible. Conclusions are drawn in Section 4.

## 2 Foreground detection

The use of periodicity in foreground detection is presented in this section. This process is divided into training and update phases, as illustrated in Fig. 1. In the training phase, the video sequence is used to generate a time series for each pixel, and the corresponding set of Fourier co-efficients provides the data to distinguish the periodic from the non-periodic elements. The Fourier analysis also gives an estimation of the number of states required in the Markov model. Neighbouring pixels of the same periodicity are grouped into regions and can be processed together in the subsequent steps of the system. The next steps in the training phase are to initialise the values of the states, and calculate the matrix of transition probabilities.

Once the training is complete, the system moves into the update phase. Here, a predicted state based on the current state is compared to a posterior state based on a Bayes update using the current state and a measurement (from the video data). Depending on whether the two states agree, the next state is determined, and the corresponding state values are used to provide a background for change detection. The mean and standard deviation of the state are then updated.

### 2.1 Training phase

This section describes the processes in the training phase of the algorithm. The training is performed once when the algorithm is first deployed.

### 2.1.1 Detection of periodic scene elements

In this section, a method is described for detecting the pixels which exhibit significant periodic characteristics and estimating this period for each of them. The effectiveness of the method is then evaluated by comparing its output with a hand-labeled map of periodic elements.

The analysis begins by creating the time series of the intensity value of each pixel over an appropriate length of training window in the time domain. The duration of the training phase is dependent on the maximum length of the period to be modelled, and was chosen to span at least five complete periods of the background signal.

It should be noted that the analysis is not limited to using pixel intensity; other pixel features can also be used. The use of illumination-invariant features will be more robust where changes in the environment are expected. In this work, the use of pixel intensity suffices as the underground environment is relatively controlled and consistent.

Figs. 2(a) and 2(b) show the scenarios in the Roma and Torino stations respectively. A typical time series (of the escalator region) is also shown in Fig. 2(c). All pixels in the sequences were included in the generation of the time series (no spatial subsampling). For the Torino dataset the temporal sampling frequency was 5 frames per second. For the Roma dataset, however, the temporal sampling frequency was much coarser at 4 frames every 10 seconds, since in this case the true period is very long (40 seconds) and the inclusion of every frame would impose an excessive computational demand (processing 200 frames at a resolution of  $720 \times 576$  requires over 1.7GB of RAM).

Next the Discrete Fourier Transform (DFT) [17] over each time series is computed. The absolute (modulus) values are used to discriminate between periodic and non-periodic background elements, and for the former, to estimate the main frequency of the periodic motion. Naturally, the periodic elements have a distinct peak at their fundamental frequency, and also at associated harmonics. The task is to discriminate between these peaks and the highest values produced by the stochastic non-periodic signals. The detection of peaks is band-limited to a set range: from above, by the Nyquist limit (or else the shortest period to be detected); from below, by the sample period (or else the longest period to be detected). In general, the spectra of natural scenes follow a  $1/f$  distribution [18] (this distribution also applies for natural scenes spectra in the spatial frequency domain). The pre-normalised spectra for the Torino data

are illustrated in Fig. 3. In order to compensate for the  $1/f$  distribution, we normalise each spectrum by multiplying each element by its corresponding frequency. The next step is to compute the mean  $\mu$  of each normalised spectrum. Periodic spectra will show a peak whose value is much higher than  $\mu$ , while non-periodic spectra will resemble white noise. The two can therefore be discriminated by thresholding each spectrum with  $K\mu$ , where  $K$  is a constant, and checking whether any component of the spectrum survives. If there is at least a surviving component, the spectrum is periodic and that component should be the main frequency (although, occasionally, harmonics are detected instead of the main frequency).

Different values of  $K$  gives different operating points for the system. It was observed that lower values of  $K$  make the algorithm more sensitive, and values between 4 and 5 have been found to give good results. All the results shown in the subsequent stages of processing have been achieved using  $K = 4.5$ . Previously [19], different thresholds were required for data with different time sampling intervals. Here, this has been improved so that the threshold varies with the mean  $\mu$  of the spectrum, meaning that the value of  $K$  can be fixed regardless of the sampling interval (e.g. the Rome and Torino sequences were sampled at 0.4Hz and 5.0Hz respectively).

The output of the thresholding procedure is a binary image where each pixel is classified either as periodic or non-periodic, and for periodic pixels, a frequency estimate is provided (corresponding to the first peak above the threshold). This output was compared to a ground truth, generated manually. (See left-hand side of Fig. 4.) The ground truth consists of a binary image where pixels corresponding to the periodic elements (i.e. the escalator and the flashing light) are white and all other non-periodic pixels are black <sup>1</sup>. It does not contain any information regarding the frequency, as no ground truth is available.

The detector algorithm was run with different values of  $K$  to investigate the space over which the different operating points lie in an ROC (Receiver Operating Characteristic) plot. A ROC curve is a plot of True Positive Rate (TPR) against False Positive Rate (FPR) and these are defined as:

$$\begin{aligned} TPR &= \frac{TP}{TP + FN} \\ FPR &= \frac{FP}{FP + TN} \end{aligned} \tag{1}$$

where TP, FP, TN, FN are the number of true positives, false positives, true negatives, and false negatives respectively [20]; more specifically,

---

<sup>1</sup>These data are available at <http://www.ist-caretaker.org/periodic>

- true positives are periodic pixels detected as periodic;
- true negatives are non-periodic pixels not detected as periodic;
- false positives are non-periodic pixels detected as periodic;
- false negatives are periodic pixels not detected as periodic.

The right-hand side of Fig. 4 gives the ROC curves by varying the value of  $K$ .

The output of the algorithm serves as a mask to identify periodic background elements. The detected frequency of each element, measured in *frames per period* (fpp) is used to choose the number of states used to model the periodic background, as explained in the following section.

### 2.1.2 Markov model

Once the pixels of the image displaying periodicity have been detected, adjacent pixels having the same periodicity are grouped into regions, and a Markov model for each region is constructed. The number of states in the model is first estimated according to the DFT analysis. In the Torino dataset, the detected frequency of the escalator pixels is 5.3 frames per period, hence both a 5-state and a 6-state model are tested.

Clearly, the number of states in the model is dependent on the duration of the period and the video sampling rate. The number of states in the Markov model is chosen to be closely related to the frames per period so that the different phases of the period are represented as different states. If more states are used, there will be redundancy in the model; whereas if fewer states are used, the different phases in the period will not be able to be distinguished. While this may not affect the change detection performance, it is decided that the representation of the underlying system is better illustrated using a 5- or 6-state model.

Using a 5-state model means there is a very low probability of the system remaining in the same state at the next time instance. For the majority of the time, the system will transit into the next state, and sometimes will skip a state and advance two states. On the other hand, for a more complex 6-state model, the system can either remain in the same state, or advance one or two states. The decision is finalised by assessing the performance of each model in the learning phase.

The learning of the transition probabilities of the Markov model is performed by assuming that the first  $N$  measurements correspond directly to the  $N$  states in the model. Subsequent measurements are assigned a state whose values are closest to the measurement in a Euclidean sense. Once the learning phase is completed (this was achieved in just over 100 frames of real data, i.e. about 20 seconds at 5fps), the transition matrix can be used in the update phase.

As mentioned, both a 5-state and a 6-state model have been tested, where different pixels on the escalator were used for the purpose of learning and validation. It was found that a 5-state model is more accurate (in predicting the subsequent state correctly); consequently this has been used in the subsequent stages of the system.

## 2.2 Update phase

Once the training phase is completed, the algorithm is run in the update phase where the objective is to determine the next state that the system will be in at each time instance. There are two estimates that can be determined: a prediction  $\hat{s}$  that is purely based on the current state and the transition matrix; and a MAP (maximum *a posteriori*) estimate  $\tilde{s}$  which incorporates a measurement. These two quantities are related mathematically as follows:

$$p(\tilde{s}_i|m) = \frac{p(m|\hat{s}_i)p(\hat{s}_i)}{p(m)} \quad (2)$$

where  $i \in I$  is the set of possible next states. The measurement  $m$  is the pixel intensity here. The measurement likelihood is calculated as:

$$p(m|\hat{s}_i) = \frac{1}{(2\pi)^{n/2}|\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2}(m - v_{s_i})^T \Sigma^{-1} (m - v_{s_i}) \right\} \quad (3)$$

where  $v_{s_i}$  is the value of state  $i$ .

Following Fig.1, the predicted state  $\hat{s}$  (without measurement) is compared with  $\tilde{s}$ , the state with the highest posterior probability given the measurement. If the predicted state and the posterior state agree (to within 1 state), then the posterior state is accepted as the next state  $\bar{s}$ . Otherwise, the predicted state is accepted as the next state (since the mismatch of  $\hat{s}$  and  $\tilde{s}$  is probably due to the presence of foreground pixels).

The corresponding mean and standard deviation of the next state  $\bar{s}$  can then be updated, using a similar scheme to that of Stauffer and Grimson's Gaussian Mixture Model [1]. An adaptive scheme is

preferred over a static approach where the state values are not updated because the Markov model is only an approximation to the periodic process (in this case, a 5-state model for 5.3 frames per period). Moreover, by calculating a standard deviation on the mean, a decision threshold for segmenting the foreground can be defined as a scalar multiple of the standard deviation. The update equations are as follows:

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho m_t \quad (4)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(m_t - \mu_t)^T(m_t - \mu_t) \quad (5)$$

where

$$\rho = \alpha p(m_t | \mu_t, \sigma_t). \quad (6)$$

Here,  $t$  denotes time, and  $\alpha$  is the learning rate of the algorithm. The conditional probability in (6) is equivalent to (3) since the state can be represented by a mean and a standard deviation.

The technique was applied to the real data collected at a Torino underground station. Since the objective of this technique is to provide a reliable periodic background from which foreground segmentation can be generated, results of foreground mask generation is given in Section 2.4 to demonstrate the efficacy of this approach .

The Markov Model described here is closely related to the Hidden Markov Model (HMM). Here, the hidden states correspond to the different phases of the period, while the observations (or measurements) are the pixel intensity. In a standard HMM, a series of measurements are used in the forwards-backwards algorithm to determine the optimal current state of the system. Here, only the previous measurement is used, because it was found that the increase in computation does not improve the final result.

Since different camera views would capture different scene elements with different periodicities, training has to be performed for each camera view in a networked system. The time required for training is dependent on the length of the maximum period we wish to detect. For the escalator here, the training process is fast (just over 100 frames, i.e. 20 seconds at 5fps); therefore if elements of similar periodicities are to be detected, the training can be achieved in a short time.

### 2.3 Application to Torino Dataset

The real data used here is the escalator sequence with people in the foreground (i.e. on the escalator), a sample frame of which is shown in Fig. 5(a). Only the area of interest has been displayed here for clarity - the section of the image with the platform has not been shown.

This sequence has approximately 300 frames at 5fps and involves 9 people using the escalator. The periodicity was determined using the process described in Section 2.1.1, and the 5-state Markov model was constructed following the approach described above. For the 5-state Markov model, the learned matrix of transitions are zero on the diagonal. (For the 6-state model, the diagonal values are between 0.0 and 0.4.)

Fig.5(b) shows the foreground mask of the proposed approach for one frame, and Fig.5(c) shows the corresponding output of the GMM. It should be noted that only the escalator region defined by the mask generated in Section 2.1.1 is being processed. It can be seen that the proposed approach seems to produce fewer false positives and the output regions are more contiguous in this frame.

For a quantitative comparison of the two algorithms, a ROC analysis is carried out and the results are presented in the next section. It should be pointed out that the images have been subsampled to facilitate the testing of many parameter combinations for the generation of the ROC curve. Nevertheless, tests on the original data, where both the learning and the operation phases were performed on the original data on a subset of the parameters, show that subsampling has not affected the validity of the results.

### 2.4 Change Detection Results

Ultimately, the efficacy of the background modelling technique is assessed by the foreground detection performance that it can provide; therefore the segmentation output of the proposed method is compared against that of the Gaussian Mixture Model (GMM) [1]. The parameters for both methods have been varied over a wide range to obtain the best performance in each case. Values for one operating point for each algorithm have been included in Table 1; these operating points have been marked by circles in Fig. 5(d). It is believed that the GMM provides the best baseline for comparison because it has the ability to model multi-modal distributions, and is adaptive. In fact, the proposed method can be viewed as a GMM with a prior on the transition probabilities. Using these outputs, a ROC curve is generated, plotting the True Positive Rate against the False Positive Rate as defined in (1).

Table 1: Parameters of the proposed method and the GMM.

Algorithm	Parameter	Description	Value for op. pt. in Fig. 5(d)
Markov	<code>alpha</code>	learning rate	0.05
	<code>numStdDev</code>	scalar determining the threshold	2.5
	<code>initVar</code>	initial variance of state model	20
GMM	<code>alpha</code>	learning rate	0.005
	<code>numModels</code>	number of Gaussians	4
	<code>initVar</code>	initial variance of Gaussians	70
	<code>initWeight</code>	initial weight of new model	0.1
	<code>numStdDev</code>	scalar determining the threshold	4
	<code>T</code>	proportion attributable to background	0.6

The dataset used in these experiments is the Torino dataset described in Section 2.3. The ground truth values have been generated by manually labelling the pixels as foreground or background<sup>2</sup>. Foreground pixels here correspond to pixels where a person is present on the escalator. The resulting ROC curve of a pixel-based analysis is shown in Fig.5(d). The blue square points correspond to the proposed algorithm, while the red triangular points are the outputs of the GMM. It can be seen that over the parameter space, the proposed method performs better than the GMM. For 80% TPR, the FPR is reduced relatively by 40%. At 20 % FPR, the number of missed detections is reduced relatively by 47%.

The implications of this improvement are not straightforward to assess. For detection and tracking of pedestrians, the increased reliability of foreground detection can be expected to lead to improved track reliability. The effect will be restricted to specific areas of the scene such as escalators and advertising boards. If these areas have a key role in the detection of important events, then the improvement will have a significant impact.

<sup>2</sup>These data are available at <http://www.ist-caretaker.org/periodic>

### 3 Anomaly detection

In this section, a periodic analysis of anomalous events is considered. In general, an anomaly is a departure from what is expected, although an “anomalous event” is highly dependent on the context and operating environment. Here, the expectation is that the scene activity is periodic in the time domain, in this case because the trains arrive at the station at regular intervals. A departure from such regularity is classified as anomalous. Indeed, the foreground detection method described in the previous section can also be viewed as low-level anomaly detection, where the foreground presence is a departure from the regularity of background.

In this application, the disruptions to the periodic scene activity could be caused by unusual human behaviour (such as not leaving the platform after the train has departed) or the failure of the train to arrive at the station on time. These types of situations are potentially of interest to surveillance operators, and the method proposed in the section below can be used to provide an indication of the anomaly. The dataset used in this section is taken from the same camera view in the Torino metro, as the data used in Section 2.

#### 3.1 Possible approaches

The periodic nature of the scene can be exploited in two ways for the detection of anomalies, namely using the amount of scene motion, and fitting a model to the underlying generative process. Both approaches aim to reduce a multi-dimensional input (i.e. colour images over time) to a one-dimensional (1D) output, allowing a direct interpretation of the data. These approaches are now discussed.

##### 3.1.1 Aggregating the amount of scene motion

The periodicity in the movement of the scene elements directly determines the amount of foreground activity. This simple observation is used here to map the input image data to a 1D signal. Foreground detection is first performed, and the amount of “foreground activity” is essentially a summation of the foreground pixels. Here the foreground map has been median-filtered and subsampled to reduce the effects of noise and to reduce the dimensionality. (This also makes the input data in both the approaches in Sections 3.1.1 and 3.1.2 consistent.) Figure 6 shows a time series of foreground activity of a 1600s (8000

frames at 5fps) sequence captured at the DOD Torino metro station. Image data corresponding to three sections of the time series are also displayed. It can be seen that the data exhibits a periodic structure: the bursts of activity correspond to a train arriving and people exiting onto the platform, while the lack of activity (sections with values close to zero) indicates an almost empty platform. By visual inspection, there seems to be two anomalies when a periodic signal is expected. First, it can be seen that the section between 240s and 320s has some activity, when none is expected. Second, the section between 680s and 860s shows a period of inactivity that is longer than usual. It is the objective in this work that these anomalies be detected automatically.

### 3.1.2 Using a higher-dimensional model

Instead of using the amount of foreground activity in the scene as a signal, the observed data (i.e. images) can be viewed as the observations of a higher-dimensional generative process. Normal data can then be used to construct such a model in a training phase, and when presented with new unseen data, the likelihood of the data given the model can be used as an indication of whether the event is anomalous.

The model is constructed based on the approach in [8], with a difference in the features used. First, foreground detection is performed, and the output is median-filtered and subsampled. These 2D foreground maps are then vectorised, and grouped into segments of  $K$  frames. Each sample therefore has a concatenation of  $K$  vectorised frames. Principal Component Analysis (PCA) is then performed on all the samples. Finally, the dimensionality-reduced dataset is clustered using spectral clustering [21] to determine the number of states in the model.

Given the model, the Bayesian framework presented in Section 2.2 can be used. However, instead of comparing the predicted state with the posterior state, here the posterior probability of the data having arisen from each state of the model is calculated directly, and this in turn is used to calculate the entropy of the system, producing a 1D signal. This signal is similar to that produced in Fig. 6, where a periodic pattern is observed; any deviation from the expected periodic trend is classified as anomalous.

Figure 7 (a) shows a 1D signal of the entropy of the system using the same dataset as in Section 3.1.1; it can be seen that the sections where anomalous events occur are the same as in Fig.6. One important difference between the two signals is that the amplitude of the activity signal has a larger variation in range than that of the entropy of the system, meaning that the latter is more suited to anomaly detection

because the expected regularities are more uniform.

### 3.2 Detecting anomaly in 1D signal

The detection of anomalies in a 1D signal has been addressed in the fields of medical signal processing and condition based monitoring. Examples include the detection of abnormal heartbeats in ECGs [22, 9], as well as the detection of mechanical failure [23]. Many of the existing techniques depend on the generation of a model which captures the evolution of the sequence. This model is then used to compare against a new sequence to detect any anomalies. Here, we use an unsupervised approach where the irregularities in a periodic signal are detected without the need for such a model. Two types of anomalies are of interest in this scenario: juxtaposition anomalies, i.e. events juxtaposed in unexpected ways; and temporal anomalies, manifesting as unexpected periodicities [24].

The method is described as follows. First, an estimate of the periodicity is found using the Fourier transform. Using the entropy time series in Fig. 7(a), it is found that the first non-DC peak in the spectrum is at a frequency of 0.0056Hz, corresponding to 177s. This allows the extraction of a pattern from the time series, which is used as a “template” for comparison with the remainder of the time series. Here, without loss of generality, the template corresponds to the first 177s of data. In an online application, this process is performed using a fixed window of samples, the duration of which should include at least 5 periods of the signal. The template can be updated online, where the samples can be selected using a sliding window. Here, the entropy time series shown is processed in its entirety to illustrate the ability of this approach in detecting anomalies.

The template is compared with each section of the time series using the KL (Kullback-Leibler) divergence. Since this metric is not symmetric, the best way to detect the anomalies is found to be:

$$D_{KL}(f_0||f_1) \times D_{KL}(f_1||f_0) \tag{7}$$

where  $f_0$  is the template, and  $f_1$  is the sub-sequence in the time series. The KL-divergence between two discrete  $n$ -point sequences  $P$  and  $Q$  is defined as

$$D_{KL}(P||Q) = \sum_n p_n \log \frac{p_n}{q_n}. \tag{8}$$

It should be pointed out that while this measure is chosen, there exist other forms of symmetrised divergence measures, such as the Jensen-Shannon divergence [25], that can be used.

Fig. 7(b) shows the “degree of anomaly” in the signal. The value at each point corresponds to the 177s of data up to that point. If a threshold is required, then the value of  $-100$  can be used to highlight the anomalies in this case. Note that the points which drop below the threshold correspond to the two anomalous events of interest: loitering, and the delay in train arrival. This approach has also been tested on an entirely normal sequence to ascertain its validity, and the results are shown in Fig. 7(c) and (d). (Note that a sequence is deemed “normal” when there are no events of interest when the video is viewed by a human inspector.) Fig. 7(c) shows the time series of the entropy, and Fig. 7(d) the abnormality score. Just over half-way, there is a slight “dip” in the time series: this is slightly abnormal as shown in the score, but compared to the example in Fig. 7(a) and (b) this is normal, which agrees with the output of human inspection.

The experiments here have proven the feasibility of using the periodicity exhibited by the data in anomaly detection. In order to fully characterise the performance of this approach, extensive performance analysis will be required.

## 4 Conclusion and Future Work

In this work, the periodicity exhibited by scene elements has been exploited for both foreground detection as well as anomaly detection. These two components of a visual surveillance system have the commonality that the signal of interest is any deviation from the expected or the norm, the norm being the background in foreground detection, and the regularity in anomaly detection.

The foreground detection process begins with the detection of periodic scene elements in the frequency domain. A frequency-normalised threshold was demonstrated to have satisfactory performance on the available test data, and a ROC analysis shows the performance of this method over a range of parameter values. The second stage models the periodic appearance with a Markov model. Each state stores an estimation of the first and second order statistics of the appearance at a given phase of the period. The transition probabilities between the states are learned, using training data, and these can accommodate both the uncertainty in periodic processes and periods that have a non-integer number of frames. The

Markov model is used to predict a background appearance, from which a segmentation of the foreground can be produced. This is compared against a Gaussian Mixture Model of the background, again using ROC analysis with hand-labelled foreground. The results show that, for this data set, there is a significant improvement in the accuracy of the foreground segmentation.

The anomaly detection procedure uses foreground detection maps as the input, and two methods of processing this data have been discussed. First, the aggregation of foreground activity can be used to map the input to a 1D signal; alternatively a model of the underlying structure of the data can be generated, and the entropy of the posterior probability over the states in the model can be used as the 1D signal for anomaly detection. In either case, the 1D signal exhibits periodicity, and a deviation from this regularity is flagged as an anomaly. Tests on a Torino dataset has shown that it is possible to automatically detect anomalies such as loitering, as well as the late arrival of a train.

Future work will include the extension of the timescale over which periodicity can be used to analyse the data. For instance, the periodicity exhibited on a daily or weekly basis can reveal interesting usage patterns in the underground stations and highlight unusual events. A challenging problem is to perform anomalous detection over a *network* of stations, correlating the different patterns exhibited in each camera view.

## Acknowledgements

This work is funded under the CARETAKER project (European Union IST 4-027231). Our thanks to *Gruppo Trasporti Torino* and *Agenzia per i Trasporti Autoferrotramviari del Comune di Roma SPA* for giving permission for images of these sequences to be used in this publication.

## References

- [1] Stauffer C. and Grimson E., “Learning Patterns of Activity Using Real-Time Tracking”, *IEEE Transactions on Pattern Recognition and Machine Intelligence*, Vol. 22. No. 8, 2000, pp 747-757.
- [2] Soatto S., Doretto G. and Wu Y., “Dynamic Textures”, in *Proc. Intl. Conf. on Computer Vision*, 2001, pp 439-446.

- [3] Zhong J and Sclaroff S., “Segmenting Foreground Objects from a Dynamic Textured Background via a Robust Kalman Filter”, in *Proc. ICCV*, 2003, pp 44-50.
- [4] Oh S.M., Rehg J.M., Balch T. and Dellaret F., “Data-Driven MCMC for learning and inference in switching linear dynamic systems”, Twentieth National Conference on Artificial Intelligence (AAAI-2005), Pittsburgh, U.S.A.
- [5] Xiang T. and Gong S., “Video Behaviour Profiling and Abnormality Detection without Manual Labelling”, in *Proc. IEEE Int. Conf. Computer Vision (ICCV)*, 2005, pp 1238-1245.
- [6] Xiang T. and Gong S., “Incremental Visual Behaviour Modelling”, in *Proc. IEEE Visual Surveillance Workshop*, Graz, May 2006, pp. 65-72.
- [7] Hung H. and Gong S., “Detecting and Quantifying Unusual Interactions by Correlating Salient Motion”, in *Proc. IEEE International Conference on Advanced Video and Signal based Surveillance*, Como, Sept 2005.
- [8] Andrade E.L., Blunsden S.J. and Fisher R.B., “Performance Analysis of Event Detection Models in Crowded Scenes”, Visual Information Engineering (VIE) Conference, Bangalore, India, 2006, pp 427-432.
- [9] Chuah M., Fu F., “ECG Anomaly Detection via Time Series Analysis”, in *Lecture Notes in Computer Science: Frontiers of High Performance Computing and Networking ISPA 2007 Workshops*, Springer, 2007, pp.123-135.
- [10] Brophy B., Kelly K. and Byrne G., “AI-based condition monitoring of the drilling process”, *Journal of Materials Processing Technology*, 124(3), 2002, pp.305-310.
- [11] McArthur S.D.J., Booth C.D., McDonald J.R. and McFadyen I.T., “An agent-based anomaly detection architecture for condition monitoring”, *IEEE Trans. Power Systems*, 20(4), 2005, pp.1675-1682.
- [12] Burgess M., “Probabilistic anomaly detection in distributed computer networks”, *Sci. Comput. Program.*, 60(1), 2006, pp 1-26.

- [13] Liao Y., Vemuri V.R. and Pasos A., “Adaptive anomaly detection with evolving connectionist systems”, *Journal of Network and Computer Applications* 30(1), Jan 2007, pp. 60-80.
- [14] Ostaszewski M., Seredynski, F. and Bouvry, P., “A nonself space approach to network anomaly detection”, *20th Int. Parallel and Distributed Processing Symposium (IPDPS), NIDISC*, Greece, April 2006.
- [15] Thottan M. and Ji C., “Anomaly detection in IP networks”, *IEEE Trans. Sig. Pro.*, 51(8), 2003, pp.2191-2204.]
- [16] Frota R.A., Barreto G.A. and Mota J.C.M., “Anomaly detection in mobile communication networks using the self-organizing map”, *Journal of Intelligent and Fuzzy Systems*, 18(5), 2007, pp.493-500.
- [17] Bracewell R., “The Fourier Transform and Its Applications”, McGraw-Hill, New York, 1986 (ISBN: 978-0070070158).
- [18] van Hateren J.H. and van der Schaaf A., “Temporal properties of natural scenes”, *IS&T/SPIE Proceedings Vol. 2657 (Human Vision and Electronic Imaging)*, Jan. 1996, pp.139-143.
- [19] Colombo A., Leung V., Orwell J., Velastin S.A., “Markov Models of Periodically Varying Backgrounds for Change Detection”, *Visual Information Engineering (VIE) 2007*, IET, July, London, 2007.
- [20] Fawcett T., “ROC Graphs: Notes and Practical Considerations for Researchers”, Tech report HPL-2003-4. HP Laboratories, Palo alto, CA, USA, 2004.
- [21] Zelnik-Manor L. and Perona P., “Self-Tuning Spectral Clustering”, *Neural Information Processing Systems (NIPS)*, 2004, pp 1601-1608.
- [22] Wei L., Kumar N., Lolla V., Keogh E., Lonardi S., Ratanamahatana C.A., “Assumption-Free Anomaly Detection in Time Series”, in *Proc. 17<sup>th</sup> Int. Scientific and Statistical Database Management Conference (SSDBM)*, Santa Barbara, USA, June 2005, pp 237-242.
- [23] Mahoney M.V. and Chan P.K., “Trajectory Boundary Modelling of Time Series for Anomaly Detection”, *Workshop on Data Mining Methods for Anomaly Detection, SIGKDD Conf.*, Chicago, IL, USA, 2005, pp 32-40.

- [24] Maxion R.A. and Tan K.M.C., “Benchmarking Anomaly-Based Detection Systems”, in *Proc. 1<sup>st</sup> Int. Conf. Dependable Systems and Networks*, New York, USA, June 2000, pp 623-630.
- [25] Fuglede B. and Topsøe, “Jensen-Shannon Divergence and Hilbert space embedding”, in *Proc. Int. Symp. Information Theory*, Chicago, 2004, p31.

## List of Figures

1	System diagram for the detection and update of periodic background elements. . . . .	22
2	Periodic scene elements in the Rome and Torino underground data. . . . .	23
3	Normalised spectra of the pixels from the escalator scene in Fig. 2. . . . .	24
4	Comparison between the ground truth and the detection output of the periodic background detection algorithm applied to the escalator and scrolling advertisement scenes. . . . .	25
5	Experimental results. . . . .	26
6	A time series corresponding to the amount of foreground activity in the scene. . . . .	26
7	Two time series and their anomaly scores for the detection of anomalous events. . . . .	27

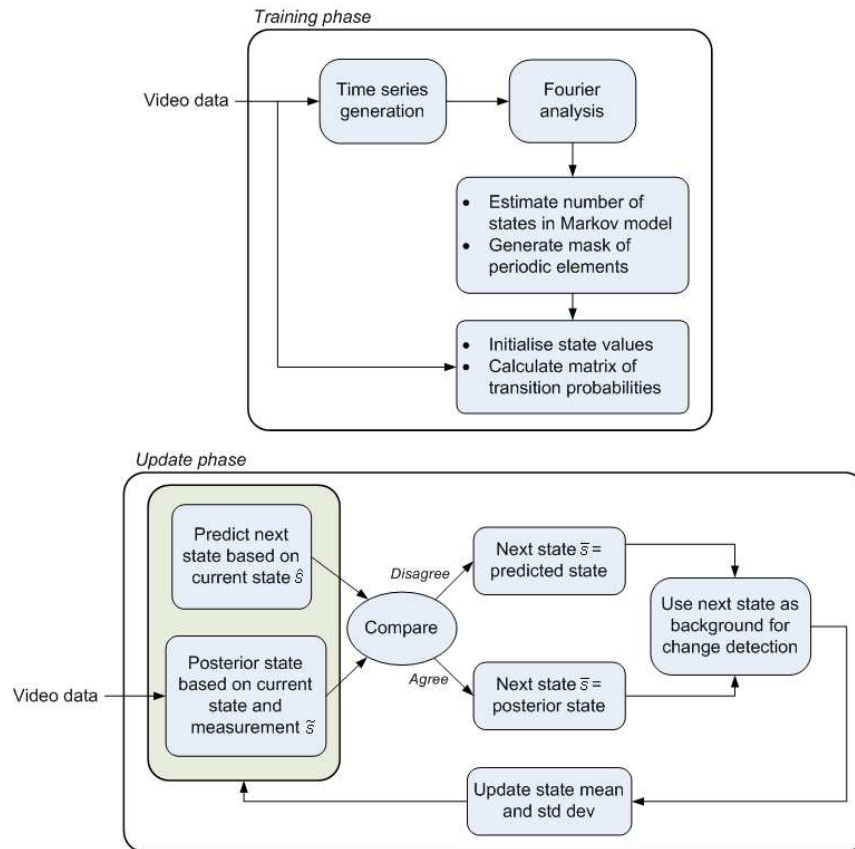
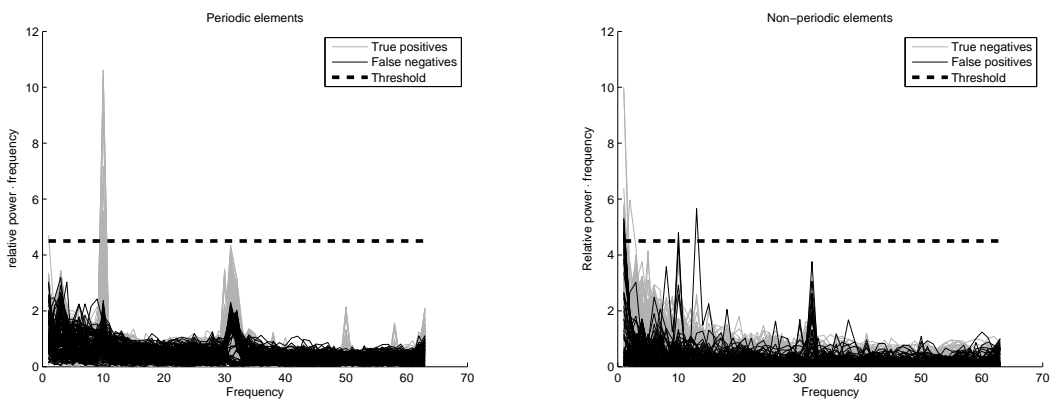


Figure 1: System diagram for the detection and update of periodic background elements.



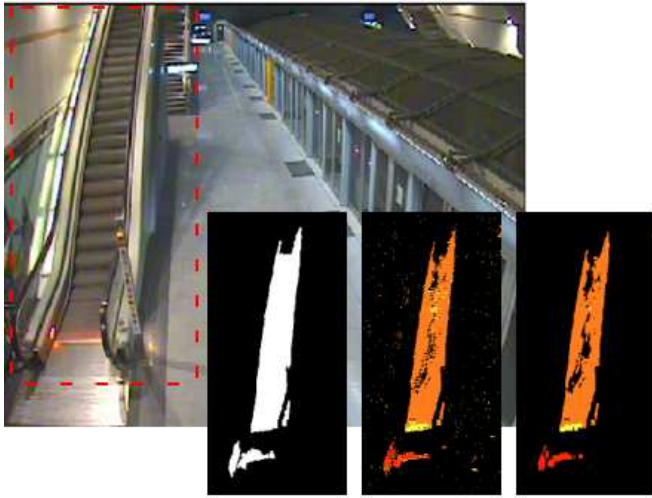
Figure 2: Periodic scene elements in the Rome and Torino underground data. (a) The scrolling advertisement scene from Rome Underground. The advertisement presents three different posters, A, B, and C, repeated as A-B-C-B. Each poster remains visible for approximately 10 seconds (including scrolling time), yielding a main period of about 40 seconds. (b) The escalator scene from Torino Underground (c) A time series of 12 pixels sampled from the escalator for 100 frames.



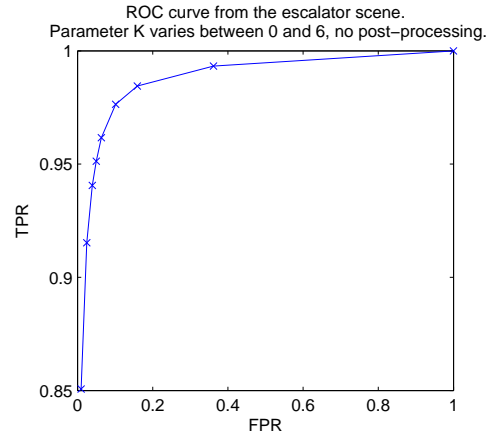
(a) Periodic elements

(b) Non-periodic elements

Figure 3: Normalised spectra of the pixels from the escalator scene (Fig. 2). For display purposes, the DC component has been removed and the amplitude has been normalised by the mean  $\mu$ , so that only  $K$  could be plotted (since now all spectra have  $\mu = 1$ ). For a definition of *true positives*, etc., see text after Eq. 1.



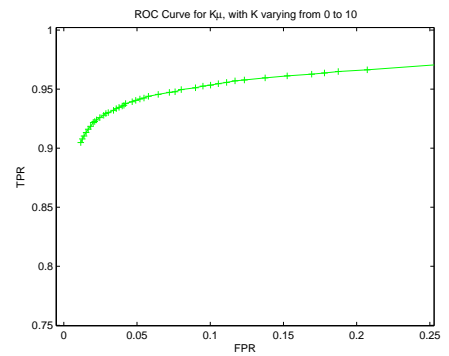
(a)



(b)



(c)



(d)

Figure 4: Comparison between the ground truth and the detection output of the periodic background detection algorithm applied to the escalator and scrolling advertisement scenes. Images (a) and (c) show the scene and, inset, the ground truth, the detection output, and the output after the application of an dilate/erosion operation (coloured pixels are periodic, and different colours indicate different frequencies). Images (b) and (d) are the corresponding ROC curves.

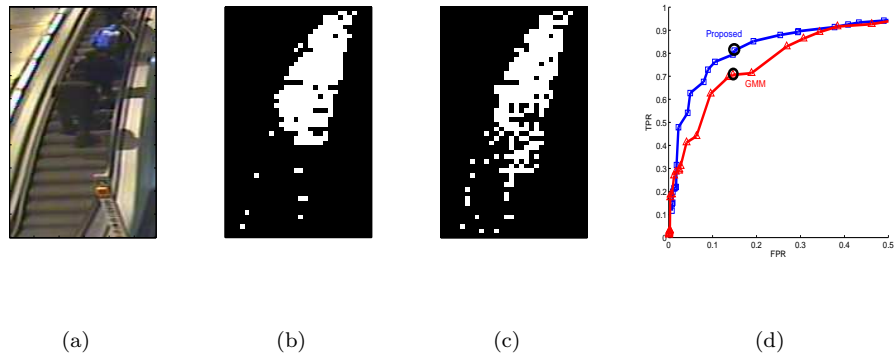


Figure 5: Experimental results. (a) A sample data frame where the humans on the escalator are the foreground components that we wish to detect. Note that here only the escalator region is processed. (b) Foreground map using the proposed method. (c) Foreground map using GMM. (d) ROC curve of the pixel-wise segmentation results.

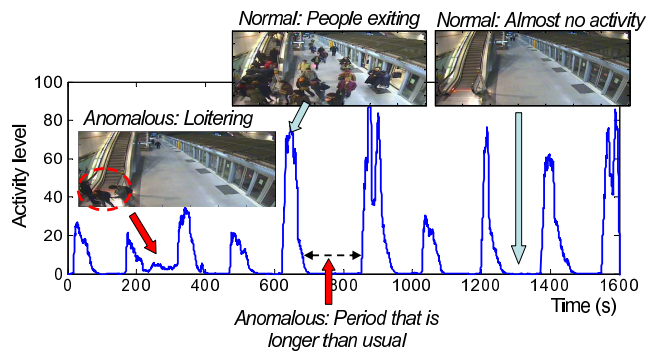


Figure 6: A time series corresponding to the amount of foreground activity in the scene.

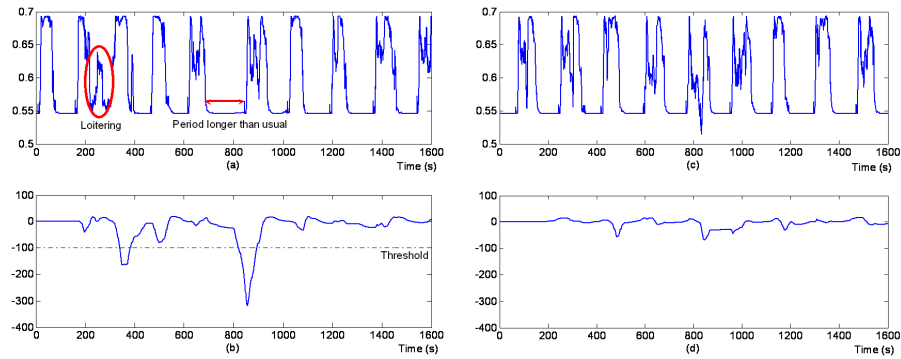


Figure 7: (a) A time series corresponding to the entropy of the system in the scene where two anomalies are present. (b) The metric with which the anomalies are detected. (c) An entropy time series in a normal scenario (with no annotated anomalies). (d) The abnormality score shows that no anomalies have been detected. See text for more description.