Detection of Anomalous Crowd Behaviour Using Hyperspherical Clustering

Aravinda S. Rao, Jayavardhana Gubbi, Sutharshan Rajasegarar, Slaven Marusic, and Marimuthu Palaniswami
ISSNIP, Department of Electrical and Electronic Engineering, The University of Melbourne,
Parkville Campus, VIC - 3010, Australia.
Email: aravinda@student.unimelb.edu.au, {jgl, sraja, slaven, palani}@unimelb.edu.au

***Abstract*—Analysis of crowd behaviour in public places is an indispensable tool for video surveillance. Automated detection of anomalous crowd behaviour is a critical problem with the increase in human population. Anomalous events may include a person loitering about a place for unusual amounts of time; people running and causing panic; the size of a group of people growing over time etc. In this work, to detect anomalous events and objects, two types of feature coding has been proposed: spatial features and spatio-temporal features. Spatial features comprises of contrast, correlation, energy and homogeneity, which are derived from Gray Level Co-occurrence Matrix (GLCM). Spatio-temporal feature includes the time spent by an object at different locations in the scene. Hyperspherical clustering has been employed to detect the anomalies. Spatial features revealed the anomalous frames by using contrast and homogeneity measures. Loitering behaviour of the people were detected as anomalous objects using the spatio-temporal coding.**

I. Introduction

Analysis of crowd behaviour in public places is an indispensable tool for video surveillance. Automated analysis of crowd behaviour has the potential to detect and alert the anomalous crowd behaviour in almost real-time. Most of the public places such as stadia, bus stops, movie theaters, shopping malls are nowadays equipped with the surveillance cameras to monitor and observe the crowd activities. However, automated detection of anomalous crowd behaviour is still in its infancy. Surveillance officials are overloaded with plenty of cameras to monitor anomalous behaviour using an array of screens that leads to inefficiency and high miss rates in in addition to viewer fatigue experienced by them. Hence, an automated system is central to video surveillance as it both helps the operators to reduce the work overload errors and human errors. However, the detection of anomalous crowd behaviour is challenging in itself. As humans, we have the ability to interpret the scene correctly using our biological visual system, whereas developing such an equivalent engineering system is nontrivial.

Anomalous crowd behaviour is a highly relevant issue given the context of human population growth. With over 7 billion people currently on this Earth, congregation of few thousands for any popular social event in public places is a natural and common phenomenon. However, there will also be anomalous events happening within the crowd. Anomalous events may include a person loitering about a place for unusual amounts of time; people running and causing panic; the size of a group of people growing over time at a particular point of entry or exit etc. Anomalous behaviors are those considered to be unusual. When such behaviors are observed, there is a need to address the issue immediately. Automated analysis provides an unbiased interpretation of the scene, which is regarded as an important tool in making the right decisions by the event managers. For instance, consider a stadium like Melbourne Cricket Ground (MCG), which can accommodate nearly $100,000$ spectators for any sporting event. If there is any unforeseen circumstance, where people panic and cause a commotion, the stadium management would face a daunting task in calming the people down or even attending to the people. To avoid any untoward issues within the spaces of the stadium, continuous monitoring of the behaviour of the people moving within the limits of stadium is of utmost importance.

According to Barnett and Lewis [1], an anomaly is defined as “an observation (or subset of observations), which appears to be inconsistent with the remainder of that set of data”. In this work, to detect anomalous frames and objects, two types of feature coding been have been proposed keeping the newly developed clustering scheme: (1) the spatial features comprising of contrast, correlation, energy and homogeneity, and (2) the spatio-temporal features. Spatial feature are derived from Gray level Co-occurrence Matrix (GLCM). These features are then encoded using blocks with the frame to create a feature matrix with rows representing the frames. Likewise, spatio-temporal feature matrix is designed with objects indicating the rows. Spatio-temporal features encode the time take by the an object in the scene. These feature matrices are then used to detect the anomalous frames and anomalous objects. Elsewhere, distributed hyperspherical clustering has been applied for anomaly detection in large datasets [2]. In this work, drawing motivation from [2] a new scheme to automatically identify anomalies has been proposed. This algorithm detects the anomalies by first clustering similar frames (in case of spatial features for detecting anomalous frames) or objects (in case of spatial-temporal features for detecting anomalous objects) and then classifying the clusters as normal or anomalous.

In this work, six cameras were installed inside the corridors of the Melbourne Cricket Ground (MCG). The data was collected on four different dates, when Australian Football League (“footy”) matches were played at MCG totalling to approximately 31.05 hours of data. The cameras were named C1-C6, and in this work, only C5 is used for validating our algorithm, which was determined based on the domain knowledge that best captures the crowd movements from a visual surveillance perspective. The sample frames of the camera C5 are shown in Fig. 1. The Fig. 1 also shows the mask used for the analysis. The main contributions of this work are : (1) this is the first work in detecting the anomalies

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

Fig. 1. Sample frames from MCG dataset for camera C5. (a) scene with no objects, (b) scene with few people coming out from the gamin bowl and exiting, (c) scene with many people moving around, standing and exiting, and (d) the mask used for detecting anomalies in this video sequence (Note: mask is used in regions where people movements are observed and is of particular interest depending on the end applications).

using the spatial and spatio-temporal feature coding, and (2) additionally, the hyperspherical clustering [2] is for the first time being used in detecting the crowd behaviour anomalies.

II. Literature Review

Anomaly detection mainly deals with the detection (locating) of events that are abnormal or unusual. In other words, the behaviors of the people that are not in sync with rest of the crowd are considered to be an anomalous crowd events [3]. In [4], one can find a comprehensive study of vision-based anomaly detection methods. Anomaly detection, in general, operates on the temporal domain data to identify the events. To handle the emergency events in crowded scenarios, Andrade *et al.* [5], [6], [7] proposed the spectral clustering of optical flow as features. An automatic model was extracted by fitting a Hidden Markov Model (HMM) for each of the video segments. Mehran *et al.* [8] proposed particle advection and social force model computation to find the interacting forces for every pixel. Then, the bag-of-words approach was applied to classify the events as normal or abnormal. Jiang *et al.* [9] used the hierarchical clustering to cluster the object trajectories governed by HMM, where the idea was to detect the unusual trajectories to identify unusual events.

In order to handle the unusual events, Adam *et al.* [10] used small regions of the scene to monitor the flow of objects using optical flow. Based on the flow probability matrix, the unusual events (running speeds) were recognized. Chen *et al.* [11] used the optical flow (Lucas-Kanade) to establish correspondence between feature points. Later, the binned orientations of optical flow about feature points were used to cluster the groups. Force-field model was applied to determine the dominant forces and their directions. A transient appearance of force was considered to be an indication of the anomalous event. Mahadevan *et al.* [3] used mixture of dynamic textures as representative features. Temporal anomaly was detected using the Gaussian Mixture Model (GMM), whereas the spatial anomaly was detected using the discriminant saliency and threshold. The net abnormality map was the combination of temporal and spatial anomalies.

Wang and Miao [12] extracted the Kanade-Lucas-Tomasi (KLT) corners indicating the moving objects and optical flow for tracking the feature points. Identical motion patterns from different blocks were clustered to generate a model and was classified as normal or abnormal based on the amount of deviation from the trained model. Liao *et al.* [13] proposed anomaly detection (detection of fighting events) by four descriptors: crowd kinetic energy (motion intensity), histogram of motion directions, spatial distribution of motion intensity and localization between two frames. Radial Basis Function (RBF) kernel with 13-dimensional feature vector is input to SVM for training and classification of abnormal events.

Tziakos *et al.* [14] proposed to detect the abnormal events by detecting the motion vectors and then classifying them on a low-dimensional manifold using the Laplacian Eigenmaps. Xu *et al.* [15] extracted the bag of Local Binary Patterns (LBP) from Three Orthogonal Planes (LBP-TOP) descriptor and applied the hierarchical Bayesian models to locate the regions of unusual events. Zhao *et al.* [16] learned the Space-Time Interest Points (STIPs) from videos using sliding windows. The unusual events were detected using sparse coding based on the reconstructible query of the learned events. Thida *et al.* [17] used the blocks of Histogram of Optical Flow (HOOF) for each frame and compared with the neighboring frames. Based on the spatial and temporal distances, low-dimensional embedding was found motivated by Laplacian Eigenmaps termed as Spatio-Temporal Laplacian Eigenmap (ST-LE). Based on the similarity of trajectories between the known event and the test event in the embedding, the abnormal event was detected. A Novelty classifier was trained to classify the abnormal events. To detect the anomalous motion patterns in groups (of people), Andersson *et al.* [18] proposed the $K-$means clustering and semi-supervised HMM, where it was assumed that a reliable people detection algorithm exists. Cong *et al.* [19] proposed the region-based descriptor that characterizes both motion and appearance (in spatial and temporal directions) and the anomaly detection was posed as a matching problem.

Anomaly detection is widely used in automatically detecting the unexpected behaviour in a set of data in many applications [20], [21], [22], [23]. Several supervised and unsupervised algorithms have been proposed in the literature [24], [22], [25], [26]. In the case of the video analytics, unsupervised schemes are required as the emergence of suspecting behaviour in the crowded scene are highly non-anticipatory and does not follow a set pattern, and hence having a pre-labelled data set of those cases are impractical in larger scenarios, especially crowds. The aforementioned methods in the literature either use a supervised or semi-supervised training to detect anomalous behaviour. Clustering based anomaly detection algorithms, such as [27], [28], [29], provides the means to detect such new anomalies in the data. In this work, we utilise a computationally efficient hyperspherical clustering based algorithm [2] for detecting unusual behaviors in the crowd.

III. Approach

Due to the nature of video analytics, it is hard to distinguish between normal and abnormal activities based on a set of rules. For instance, running is something most of us perform, however when someone is running in the middle of a walking crowd, it becomes an abnormal event and there is a need to flag such an activity. As a result, supervised training may not be the best solution for generalized performance. Considering one camera view, the detection of anomalies can be accomplished at three different levels — spatial (within frame), temporal (across frames but restricted to single object) and spatio-temporal, which is a combination of the other two. Although spatio-temporal approach is the ultimate target, spatial and temporal anomalies are quite critical in video surveillance applications. For instance, an anomaly in the density of a crowd is very useful in planning alternative exit path strategy in emergency situations. Likewise, an unruly person in the crowd has to be tracked over certain amount of time to avoid injuries and causing discomfort to the other people. Crowd density anomalies can be found using the spatial anomaly detection and the unruly person in the crowd can be found by detecting the temporal anomalies. In the case of large crowds, people usually move in clusters and the behaviour analysis must include spatial as well as temporal aspects for anomalous crowd behaviour detection. Another advantage of the spatio-temporal anomaly detection is that it enables the system to capture the true anomalies for a given period of time.

Due to the nature of human behaviour and its understanding, categorizing an event as abnormal based on a set of predefined characteristic features is infeasible. The system that flags the unusual event amidst a set of usual event (majority presumed normal) is more useful for surveillance applications. Further, the type of feature representation used for detecting the anomalies affects the overall performance. In the proposed scheme, the video obtained from the cameras are subjected to preprocessing followed by foreground object detection. The detected objects are then tracked using a Kalman filter and the features representing behaviour are extracted using the novel coding schemes. The feature vectors are then classified as normal and abnormal using the hyperspherical clustering [2]. The feature coding and anomaly detection are the two key contributions of this paper in addition to the detection of abnormal crowd behaviour.

A. Preprocessing and Object Detection

Most CCTV cameras installed at large arenas produce low-to-medium quality frames and are usually not the best fit for automated video analysis unless some preprocessing is performed. High-frequency noise is often critical to be identified and filtered. So this is accomplished by first converting the video frames to grayscale images. Then the grayscale images are filtered through Gaussian low-pass filter with $\sigma = 0.5$ and a block size of 5×5 to remove the high-frequency spatial noise. The choice of the filter is based on the crowd monitoring work presented in [30]. Objects of interest can be detected in three ways (categorized based on computational complexity): (1) frame differencing, (2) motion estimation using the optical flow approach, and (3) by modeling the background and then subtracting the background model from the incoming frames. These methods have several advantages and disadvantages. Due to the complexity of the data being handled, in this work, the GMM [31] is utilized for the background modeling. The background model is learned from the video scene when the foreground objects were absent. The objects were then detected by subtracting the model from the incoming frames. Shadows were handled using the texture-based method [32], which is shown to have markedly improved results in the literature.

B. Tracking

For tracking the recognized objects, Kalman filter [33] is used. The motion model is given by:

$$x_t = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} + \mathcal{N}(0, Q) \tag{1}$$

and the measurement model is given by

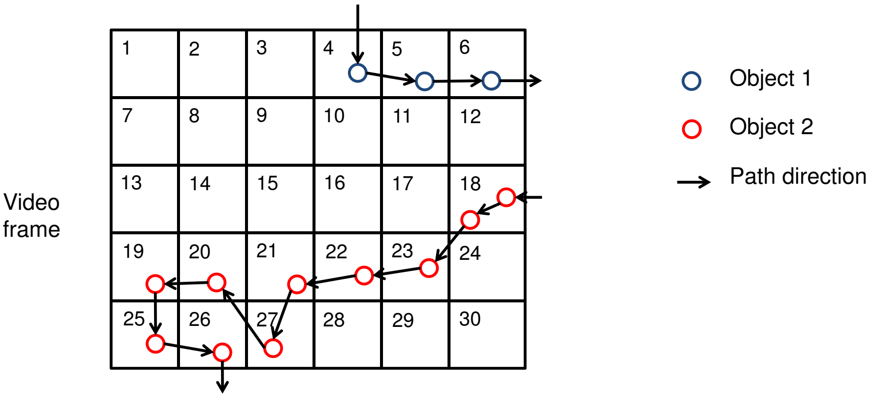
$$\begin{bmatrix} x_p \\ y_p \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ \dot{x} \\ \dot{y} \end{bmatrix} + \mathcal{N}(0, R), \tag{2}$$

where, the random variables w_t and v_t correspond to the process noise and the measurement noise; Q is the process noise covariance and R is the measurement noise covariance; x_p and y_p are the observed objects positions; object positions (x, y). The Hungarian cost algorithm [34] is used for track associations of the objects in different frames to handle tracks from multiple objects.

C. Features extraction and representation

In an unsupervised learning, the algorithm intends to separate the usual and unusual behaviour through an appropriate metric. Hence, choosing the correct features, encoding scheme and the metric play a vital role. In this work, several features have been extracted and a new way of coding the crowd anomalies is defined. Spatial and spatio-temporal representations are used as features to demonstrate the detection of anomalous behaviour. The details of the features and the encoding scheme is given in this section.

1) Spatial features: Gray level Co-occurrence Matrix (GLCM) was extracted for each frame corresponding to the object bounding box. Four types of statistical texture information was utilized to extract spatial features. Contrast, correlation, energy and homogeneity were extracted. These features have been shown to be of statistical importance by Haralick [35]. Chan *et al.* [36] have used 12 texture features for people counting application. Let I_t denote a frame at time t and (x, y) denote the pixel locations in the frame giving rise to the notation of $I_t(x, y)$ for a frame. For each frame in the video sequence, four features consisting of texture information were extracted. Let m be the number of rows and n be the number of columns of the video frame. The input frame is divided into $b_r \times b_c$ blocks, where $b_r, b_c \in \mathbb{R}$ and $1 < b_r \leq m, 1 < b_c \leq n$.



Blocks

Objects	1	2	3	4	5	6	....	18	19	20	21	22	23	24	25	26	27	28	29	30
○	0	0	0	1	2	3		0	0	0	0	0	0	0	0	0	0	0	0	0
○	0	0	0	0	0	0		1	7	6	4	3	2	0	8	0	5	0	9	0

Spatio-temporal feature matrix

Fig. 2. Depiction of spatio-temporal feature coding for two objects. There are two objects (1 and 2) that have traced their path in the scene. Object 1 enters and exists the scene relatively quickly. Object 2 loiters around at a place for more time compared to Object 1. The result of the coding mechanism described in Section III-C2 for the two objects are shown in the feature matrix. Object 2 enters the block 18 and then traces the route via the blocks 23, 22, 21, 27, 20, 19, 25 and 26 before exiting. The feature coding using the proposed scheme will be 1 for 18th block, 2 for 23rd block, 3 for 22nd block and so on, in the order the object 2 traverses through the blocks.

The edges of the frame, where the remainder of the blocks are incomplete, were padded with zeros. A feature matrix is created with rows equal to the number of frames and columns equal to the total number of blocks $\frac{m}{b_r} + R_m \times \frac{n}{b_c} + R_n$, where R_m and R_n are the remainder blocks added when m and n are not multiples of b_r and b_c. Depending on the location of the object (based on centroid), the block in which the object is residing is updated with four texture features. For a given bounding box region, the texture features were computed as given by [35]:

$$\text{Contrast} = \sum_{i,j} p(i,j)|i-j|^2 \tag{3}$$

$$\text{Correlation} = \sum_{i,j} p(i,j)(i-\mu i)(j-\mu j) \tag{4}$$

$$\text{Energy} = \sum_{i,j} p(i,j)^2 \tag{5}$$

$$\text{Homogeneity} = \sum_{i,j} \frac{p(i,j)}{1+|i-j|} \tag{6}$$

where p is the un-normalized distribution of the measure, i and j correspond to the pixel locations.

2) Spatio-temporal features: Spatio-temporal features proposed are one of the most basic features that encode the location of an object with respect to time. The output of the Kalman filter is critical in this phase as the correct object identifications are required for the encoding process. First, the video frames are divided into blocks as described previously. Then, for each object in the video sequence, a feature representation consisting of spatio-temporal information is extracted. For each object, the block corresponding to the initial location in a frame is assigned with numerical value 1. If the object remains in the same block, then other blocks are not updated and value of the block in which it is present remains the same. When the object moves to another block, the numerical value of the block that it takes is the incremented value from the previous block for the same object. This process continues for all the objects. In essence, blocks are coded with values starting from 1 and is incremented when it moves to another block. Fig. 2 shows an example of coding two objects in the video scene.

IV. Hyperspherical Cluster Based Behaviour Anomaly Detection

The proposed scheme performs anomaly detection iteratively on smaller time windows thereby effectively capturing the anomalous crowd behaviour more precisely depending on the situation. In this work, we propose to use a recent development in anomaly detection applied for environmental monitoring [2] for detecting behaviour anomalies at all three levels. The method involves three main components: (1) a fixed width clustering applied on the motion data, (2) merging of closely placed clusters of crowd behaviour, and (c) identification of anomalous clusters using the K nearest neighbor (K-NN) approach. This scheme helps us to find out similar activities within the data in addition to identifying the unusual crowd events.

The schematic of the flow is shown in Fig. 3 and the details of the algorithm is briefly presented here. The tracks are obtained from the previous processes. Once the tracks are obtained, they are appropriately coded as discussed in Section III-C. Fig. 3 provides the schematic for temporal anomaly detection. For illustration, different tracks are named

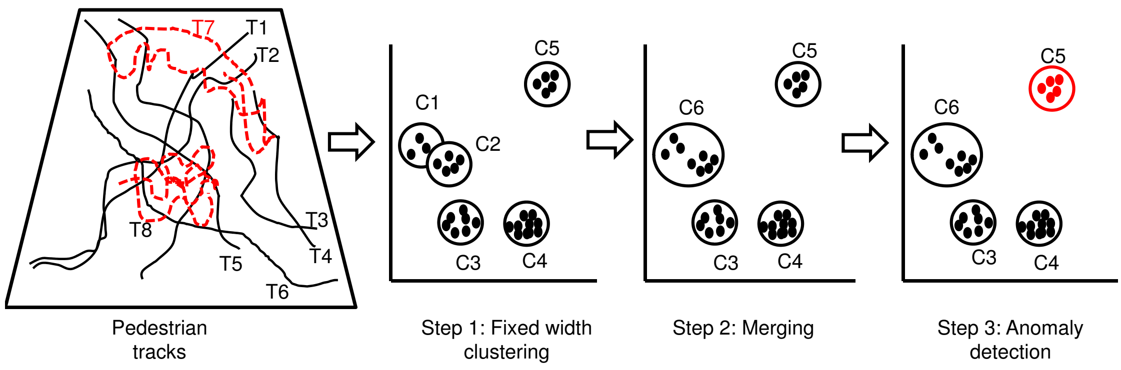


Fig. 3. Schematic of anomaly detection algorithm used for behaviour analysis. T1 to T8 are tracks. C1-C6 are clusters. Loitering anomaly is shown in red (dotted track). X and Y axes are some attributes shown for explaining the concept.

from T1 to T8 with T7 and T8 being the loitering anomalies within the field of view. The remaining tracks are entry and exit paths of the scene at some given point. As a first step, a fixed width clustering [37] is used for representing the similar tracks. This results in hyperspherical clusters with fixed width w. For instance, as shown in the figure, tracks similar to T1 and T2 form cluster C4; T3 and T4 form C3; T5 forms C1; T6 forms C2; T7 and T8 form C5. As a second step, similar clusters are merged based on the distance between clusters. If the inter-cluster (centroid) distance is less than the threshold τ, the clusters are merged. In the Fig. 3, C1 and C2 clusters are merged together to form C6. Finally, a K nearest neighbor (K-NN) based algorithm is used to find the anomalous cluster based on the average inter-cluster distances. If the average inter cluster distance is more than ψ number of standard deviations of the inter-cluster distance from the mean inter-cluster distance, a cluster is declared as anomalous. The details of the algorithm and the calculation of inter-cluster distance is covered in detail in [2]. In the Fig. 3, the result of the K-NN is shown as C5, referring to the tracks T7 and T8. The parameter w influences the number of clusters produced; the parameter ψ determines the sensitivity of the anomaly detector and is usually selected from a set of values $S = \{1, 2, 3\}$. The K value of the K-NN scheme can be selected as a given percentage of the number of clusters produced in the system [2].

V. Results and Discussion

Keeping the flow of the paper consistent, in this section, we discuss the dataset used followed by the results of the various encoding schemes proposed. The University of Minnesota (UMN) dataset [38] is a benchmark dataset used for unusual crowd activity recognition. In the UMN dataset, events where people are walking are considered as normal and running events are considered as abnormal. Our objective is not to detect events based on walking and running, but inclusive of all possible crowd events. The dataset used in this work was collected at the Melbourne Cricket Ground (MCG). Five cameras were installed in the selected areas of the corridor and another camera at the gaming bowl. The cameras were named C1 to C6. A total of 31.05 hours of data was collected at 25 frames per second. For this particular work, as a novel study in this area, only C5 (23-September-2011) was used, which is of length 22 minutes and 1 second. In this view, people come out to corridor after watching a game. Some of the people will be standing in the scene for a long time (5 minutes) while others are exiting and some others loitering. The details of the dataset are provided in Table I. The vision algorithms were implemented in MATLAB 8.0 (R2012b) and the anomaly detection in Java on Windows 7 (64 bit) comprising of an Intel® i7 $-$ 2600 CPU running at 3.4 GHz with 4 GB RAM. The system also included 512 MB ATI Radeon™ HD 5450 Graphics card.

TABLE I. Dataset details collected at Melbourne Cricket Ground (MCG).

Date	Camera	Length (hh:mm:ss)
16-September-2011	C1-C2	18:03
	C3-C6	18:01
23-September-2011	C2-C6	22:01
24-September-2011	C2-C6	14:01
01-October-2011	C2-C6	5:15:01

A. Performance

The results for the spatial and spatio-temporal features have been tabulated in Table II and Table III respectively. In Table II, the anomalous frames based on spatial features (contrast and homogeneity) are derived from GLCM. The frames were divided into 16×16 and 32×32 blocks. For a particular frame, if an object was present, then the contrast and homogeneity were computed and coded as spatial features into the feature matrix. The feature matrix consisted of frames as rows and blocks as columns. The cluster width w was set to 1, merging threshold $\tau = \frac{1}{2}w$, the number of K nearest neighbors equal to 1 and standard deviation $\psi = 1$ for all the spatial features and block sizes. It is evident from the Table II that for different block sizes and different features, the anomalous frames change.

Table III provides the anomalous objects detected by following the spatio-temporal coding scheme. Again briefly mentioning, the frames were divided into different block sizes as tabulated in the Table III. The feature input matrix consisted of object identification numbers as corresponding rows and blocks as the columns. If an object enters (centroid location) a particular block, then that particular block is incremented by value one and if the object stays in the same block, the value for that block is unchanged. More specifically, the maximum value of the entire row matrix is computed and

TABLE II. THE TABLE LISTS THE ANOMALOUS FRAMES DETECTED BY HYPERPHYSICAL CLUSTERING (DESCRIBED IN SECTION IV) USING THE SPATIAL FEATURES (CONTRAST AND HOMOGENEITY) EXTRACTED BY COMPUTING GLCM FOR EACH OF THE FRAME. THE FRAMES WERE DIVIDED INTO 16×16 AND 32×32. TOTAL NUMBER OF FRAMES IN THE VIDEO SEQUENCE WERE 31375

Spatial features			
Contrast		Homogeneity	
16×16	32×32	16×16	32×32
$2276 - 2300$	$2276 - 2300$	$9468 - 9475$	12800
$11201 - 11225$	$5557 - 5575$	$16526 - 16550$	$28467 - 28475$
$17601 - 17625$	$5726 - 5750$	$26851 - 26875$	$30276 - 30300$
$18101 - 18125$	$5851 - 5873$	$27751 - 27775$	$31026 - 31050$
$18251 - 18275$	$5901 - 5925$		
$22999 - 23000$	$6001 - 6025$		
$29976 - 30000$	$11201 - 11225$		
	$17601 - 17625$		
	$18101 - 18125$		
	$18251 - 18275$		
	$18451 - 18475$		
	$18801 - 18825$		
	$19401 - 19425$		
	$22999 - 23000$		
	$23501 - 23525$		
	$23626 - 23650$		
	$29976 - 30000$		

the resulting value is incremented by one. If the object is revisiting the same block, the values are left unchanged. For spatio-temporal features, the clustering parameters were set as: $w = 5$, $\tau = \frac{1}{2}w$, K-NN$= 3$ and $\psi = 1$. From the Table III, it is clear that the encoding scheme generates a feature matrix that enables to detect different anomalous objects for different block sizes.

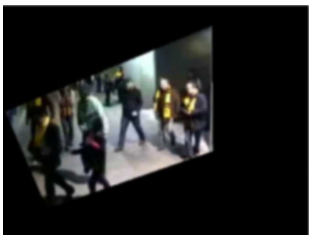
TABLE III. THE TABLE LISTS THE ANOMALOUS OBJECT IDENTIFICATION NUMBER DETECTED BY HYPERPHYSICAL CLUSTERING (DESCRIBED IN SECTION IV) USING THE SPATIO-TEMPORAL FEATURES CODED FOR EACH OF THE FRAME. THE FRAMES WERE DIVIDED INTO 8×8, 16×16, 32×32 AND 48×64. TOTAL NUMBER OF UNIQUE OBJECTS IN THE VIDEO WERE 1163. THE UNIQUE OBJECT IDENTIFICATION NUMBERS START FROM 1.

Spatio-temporal features			
8×8	16×16	32×32	48×64
163	62	78	62
336	78	163	78
638	96	336	443
677	163	366	541
780	336	459	783
783	677	503	784
784	780	541	911
830	783	783	919
831	784	784	1025
911	830	813	1046
921	831	830	1072
922	911	831	1086
932	919	911	1110
982	921	919	
1005	922	921	
1025	982	1072	
1086	1025	1109	
1087	1061	1110	
1109	1086		
1110	1087		
1136	1109		
	1110		
	1136		

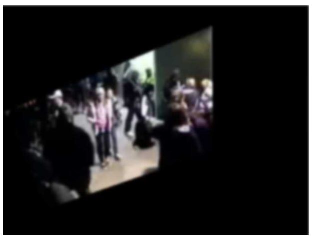
In the case of detection of anomalous frames, the spatial features not only included contrast and homogeneity, but also correlation and energy. Furthermore, the block sizes included 8×8 for each of the features. However, there were no anomalous frames detected by the hyperphysical clustering algorithm. The reason behind this phenomenon can be explained in terms of the spatial features. GLCM is calculated by measuring the frequency of occurrence of pixel (grayscale) i which is horizontally adjoining to j. This provides us a gray tone co-occurrence matrix. The Contrast provides us a measure of intensity difference $\sum_{i,j} |i-j|^2$ between a pixel and its neighbor that is being measured using the Euclidean distance factored by the normalized joint probability function $p(i, j)$ (derived from GLCM). Likewise, correlation provides a measure of correlation of a pixel with its neighbor; energy is simply the sum of squared values of normalized joint probability matrix elements $p(i, j)$; and the homogeneity captures the degree of closeness of the distribution of the GLCM matrix elements with respect to diagonal elements of the GLCM matrix. Due to the nature of these measures, the feature input matrix using correlation and energy resulted such that the hyperspherical clustering algorithm was unable to detect the anomalous frames. In other words, the feature input matrix was identical (or closely similar) for all the frames. Thus, the hyperspherical clustering was unable to detect anomalies. Fig. 4 shows the sample anomalous frames representing the Table II.

In the case of detection of anomalous objects, due to the nature of the spatio-temporal features proposed, the coding mechanism resulted in detection of loitering objects in the video scene. Loitering objects are those that spend significant amount of time in the scene (not necessarily at a single space). The identified objects were verified using the ground truth generated by annotating the video files. All the objects detected as anomalous were loitering in the scene compared to the other objects in the scene (both spatially and temporally). In the feature matrix, the maximum value of the row vector corresponding to each object increases for anomalous objects compared to other usual objects. Additionally, the feature vector length for the anomalous object increases as the time spent by the loitering object changes its spatial location but remain within the scene, the values of the columns (representing blocks) increases for the same object. This makes coding efficient for hyperspherical clustering to detect these objects as anomalous. This is a significant result, since, an efficient coding scheme can effectively uncover the underlying crowd dynamics, and in particular, the pedestrian loitering using the texture features.

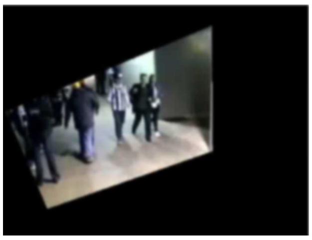
The selection of the block sizes affects the detection of anomalous frames and/or anomalous objects. For instance, in the case of spatio-temporal approach, the larger block size results in less coding (the maximum value for a particular object is small) and hence the number of detected anomalous objects are less. This is not necessarily true in all the cases. In Table III, 16×16 yielded the maximum number of anomalous objects as opposed to 8×8. However, as expected, the 32×32 and 48×64 block sizes provided less number of anomalous objects, 18 and 13 respectively. 8×8 produced 21 anomalous objects, whereas 16×16 resulted in 23. Additionally, some of the the anomalous objects (783, 784 and 911) detected are common although the block sizes are different. The calculation of feature matrix for 8×8 is computationally intensive as



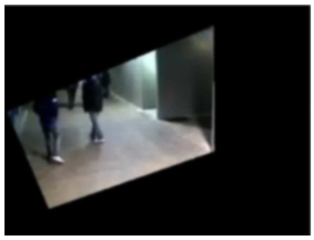
(a) Frame number: 2206



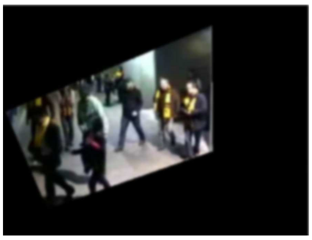
(b) Frame number: 11201



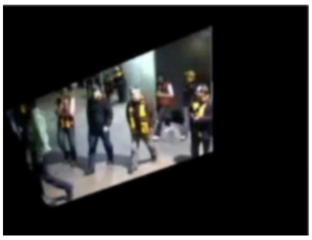
(c) Frame number: 23000



(d) Frame number: 30000

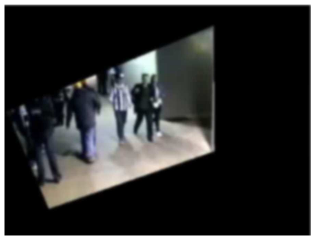


(e) Frame number: 2276



(f) Frame number: 5557

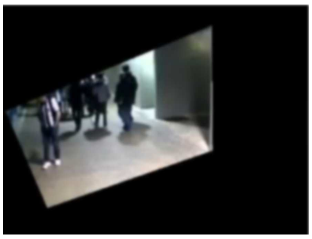
(g) Frame number: 5726



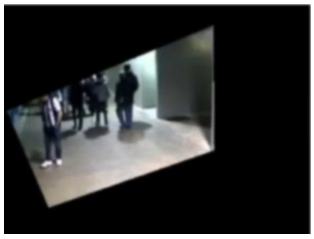
(h) Frame number: 23626

(i) Frame number: 9468

(j) Frame number: 5557

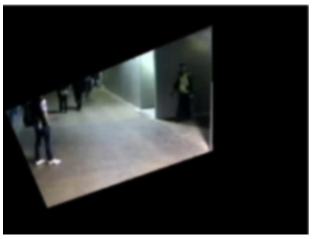


(k) Frame number: 26875

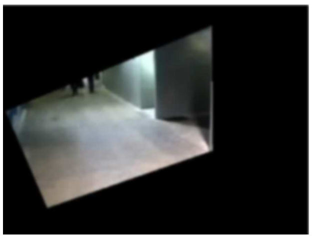


(l) Frame number: 27775

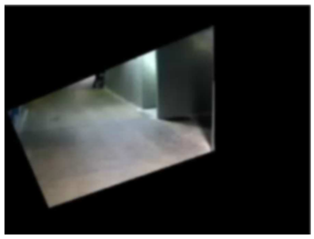
(m) Frame number: 12800



(n) Frame number: 28470



(o) Frame number: 30276



(p) Frame number: 31050

Fig. 4. The figures show the frames detected as anomalous. The mask defined in Fig. 1 is used. Anomalous frames are ordered as follows: Contrast (16×16) (a)—(d); Contrast (32×32) (e)—(h); Homogeneity (16×16); (i)—(l), and Homogeneity (32×32) : (m)—(p). The frame numbers are representative frames from Table II. The black polygonal region in the figures surronding the objects are due the mask used in this view of camera. The mask is provided in Fig. 1-(d). Only the interior (white region) in the mask is considered for analysis.

compared to the others. We encountered memory problems and were unable to detect the anomalous frames (using the spatial features) for 8×8. For the same reason, this has not been reported in Table II.

The advantage of the co-occurrence matrix is that the spatial inter-relationships of the gray tones remain invariant under transformation. On the other hand, the co-occurrence matrix loses its strength in capturing the shape information of the objects [35]. There is a need for more research on how to effectively code the spatial and spatio-temporal features and the optimal block sizes. Furthermore, the parameters chosen for the hyperspherical clustering were based on empirical knowledge of the video and the clustering algorithm. Estimation of optimal parameters for clustering requires additional work in this direction.

VI. Conclusion

Analysis of crowd behaviour in public places is an indispensable tool for video surveillance. Detection of automated anomalous crowd behaviour is a critical problem with the increased human population and surveillance applications. In this work, the anomalous frames and objects in a video were detected using the new encoding schemes for spatial and spatio-temporal features. Spatial features revealed the anomalous frames by using contrast and homogeneity measures. Loitering behaviour of the people were detected as anomalous objects using the spatio-temporal features. Hyperspherical clustering algorithm was used to detect anomalies with excellent results.

VII. Acknowledgement

This work is partially supported by the ARC linkage project LP100200430, partnering the University of Melbourne, Melbourne Cricket Club and ARUP. Authors would like to thank representatives and staff of ARUP and MCG.

References

[1] V. Barnett and T. Lewis, *Outliers in statistical data*. Wiley New York, 1994, vol. 3.

[2] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Hyperspherical cluster based distributed anomaly detection in wireless sensor networks," *Journal of Parallel and Distributed Computing*, vol. 74, no. 1, pp. 1833 – 1847, 2014.

[3] V. Mahadevan, L. Weixin, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2010, pp. 1975–1981.

[4] O. P. Popoola and W. Kejun, "Video-based abnormal human behavior recognition—a review," *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, vol. 42, no. 6, pp. 865–878, 2012.

[5] E. L. Andrade, S. Blunsden, and R. B. Fisher, "Modelling crowd scenes for event detection," in *18th International Conference on Pattern Recognition (ICPR 2006)*, vol. 1. IEEE, 2006, pp. 175–178.

[6] E. L. Andrade, R. B. Fisher, and S. Blunsden, "Detection of emergency events in crowded scenes," in *The Institution of Engineering and Technology Conference on Crime and Security*. IET, 2006, pp. 528–533.

[7] E. L. Andrade, O. J. Blunsden, and R. B. Fisher, "Performance analysis of event detection models in crowded scenes," in *IET International Conference on Visual Information Engineering (VIE 2006)*. IEEE, 2006, pp. 427–432.

[8] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2009, pp. 935–942.

[9] F. Jiang, W. Ying, and A. K. Katsaggelos, "A dynamic hierarchical clustering method for trajectory-based unusual video event detection," *IEEE Transactions on Image Processing*, vol. 18, no. 4, pp. 907–913, 2009.

[10] A. Adam, E. Rivlin, I. Shimshoni, and D. Reinitz, "Robust real-time unusual event detection using multiple fixed-location monitors," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 555–560, 2008.

[11] D.-Y. Chen and P.-C. Huang, "Dynamic human crowd modeling and its application to anomalous events detcetion," in *2010 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2010, pp. 1582–1587.

[12] S. Wang and Z. Miao, "Anomaly detection in crowd scene," in *2010 IEEE 10th International Conference on Signal Processing (ICSP)*. IEEE, 2010, pp. 1220–1223.

[13] H. Liao, J. Xiang, W. Sun, Q. Feng, and J. Dai, "An abnormal event recognition in crowd scene," in *2011 Sixth International Conference on Image and Graphics (ICIG)*. IEEE, 2011, pp. 731–736.

[14] I. Tziakos, A. Cavallaro, and L.-Q. Xu, "Event monitoring via local motion abnormality detection in non-linear subspace," *Neurocomputing*, vol. 73, no. 1012, pp. 1881–1891, 2010. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0925231210001487

[15] J. Xu, S. Denman, C. Fookes, and S. Sridharan, "Unusual event detection in crowded scenes using bag of lbps in spatio-temporal patches," in *2011 International Conference on Digital Image Computing Techniques and Applications (DICTA)*. IEEE, 2011, pp. 549–554.

[16] B. Zhao, L. Fei-Fei, and E. P. Xing, "Online detection of unusual events in videos via dynamic sparse coding," in *2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2011)*. IEEE, 2011, pp. 3313–3320.

[17] M. Thida, H.-L. Eng, M. Dorothy, and P. Remagnino, *Learning Video Manifold for Segmenting Crowd Events and Abnormality Detection*, ser. Lecture Notes in Computer Science. Springer Berlin Heidelberg, 2011, vol. 6492, book section 34, pp. 439–449. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-19315-6_34

[18] M. Andersson, F. Gustafsson, L. St-Laurent, and D. Prevost, "Recognition of anomalous motion patterns in urban surveillance," *IEEE Journal of Selected Topics in Signal Processing*, vol. 7, no. 1, pp. 102–110, 2013.

[19] Y. Cong, J. Yuan, and Y. Tang, "Video anomaly search in crowded scenes via spatio-temporal motion context," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 10, pp. 1590–1599, 2013.

[20] S. Rajasegarar, C. Leckie, and M. Palaniswami, "Anomaly detection in wireless sensor networks," *IEEE Wireless Communications*, vol. 15, no. 4, pp. 34–40, 2008.

[21] C. OReilly, A. Gluhak, M. Imran, and S. Rajasegarar, "Anomaly detection in wireless sensor networks in a non-stationary environment," *IEEE Communications Surveys and Tutorials*.

[22] S. Rajasegarar, J. C. Bezdek, C. Leckie, and M. Palaniswami, "Elliptical anomalies in wireless sensor networks," *ACM Transactions on Sensor Networks (TOSN)*, vol. 6, no. 1, p. 7, 2009.

[23] J. C. Bezdek, S. Rajasegarar, M. Moshtaghi, C. Leckie, M. Palaniswami, and T. C. Havens, "Anomaly detection in environmental monitoring networks [application notes]," *Computational Intelligence Magazine, IEEE*, vol. 6, no. 2, pp. 52–58, 2011.

[24] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009.

[25] S. Rajasegarar, C. Leckie, J. C. Bezdek, and M. Palaniswami, "Centered hyperspherical and hyperellipsoidal one-class support vector machines for anomaly detection in sensor networks," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 3, pp. 518–533, 2010.

[26] A. Shilton, S. Rajasegarar, and M. Palaniswami, "Combined multiclass classification and anomaly detection for large-scale wireless sensor networks," in *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE, 2013, pp. 491–496.

[27] S. Rajasegarar, C. Leckie, M. Palaniswami, and J. C. Bezdek, "Distributed anomaly detection in wireless sensor networks," in *10th IEEE Singapore International Conference on Communication systems*. IEEE, 2006, pp. 1–5.

[28] S. Rajasegarar, A. Gluhak, M. Ali Imran, M. Nati, M. Moshtaghi, C. Leckie, and M. Palaniswami, "Ellipsoidal neighbourhood outlier factor for distributed anomaly detection in resource constrained networks," *Pattern Recognition*, vol. 47, no. 9, pp. 2867–2879, 2014.

[29] M. Moshtaghi, C. Leckie, S. Karunasekera, and S. Rajasegarar, "An adaptive elliptical anomaly detection model for wireless sensor networks," *Computer Networks*, vol. 64, pp. 195–207, 2014.

[30] A. S. Rao, J. Gubbi, S. Marusic, P. Stanley, and M. Palaniswami, "Crowd density estimation based on optical flow and hierarchical clustering," in *2013 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. IEEE, pp. 494–499.

[31] C. Stauffer and W. E. L. Grimson, "Learning patterns of activity using real-time tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 747–757, 2000.

[32] A. Sanin, C. Sanderson, and B. C. Lovell, "Improved shadow removal for robust person tracking in surveillance scenarios," in *2010 20th International Conference on Pattern Recognition (ICPR)*. IEEE, 2010, pp. 141–144.

[33] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Journal of Fluids Engineering*, vol. 82, no. 1, pp. 35–45, 1960.

[34] H. W. Kuhn, "The hungarian method for the assignment problem," *Naval research logistics quarterly*, vol. 2, no. 1-2, pp. 83–97, 1955.

[35] R. M. Haralick, "Statistical and structural approaches to texture," *Proceedings of the IEEE*, vol. 67, no. 5, pp. 786–804, 1979.

[36] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, 2008, pp. 1–7.

[37] E. Eskin, A. Arnold, M. Prerau, L. Portnoy, and S. Stolfo, "A geometric framework for unsupervised anomaly detection," in *Applications of data mining in computer security*. Springer, 2002, pp. 77–101.

[38] University of Minnesota. (2006) Detection of Unusual Crowd Activity. http://mha.cs.umn.edu/proj_events.shtml. [Online; accessed 25-June-2014].