J.Natn.Sci.Foundation Sri Lanka 2016 44 (4): 385-397
DOI: http://dx.doi.org/10.4038/jnsfsr.v44i4.8021

RESEARCH ARTICLE

Online tracking and event clustering for vision systems

P.H. Perera*, H.M.S.P.B. Herath, W.S.K. Fernando, M.P.B. Ekanayake, G.M.R.I. Godaliyadda and J.V. Wijayakulasooriya
Department of Electrical Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya.

Revised: 22 February 2016; Accepted: 28 April 2016

Abstract: This paper proposes a comprehensive method for online-event clustering in videos. Adaptive Gaussian mixture model was modified to obtain consistent foreground estimates for object tracking by introducing shadow filtering, stillness handling, visual impulse removal and visual distortion filtering. Object-events were defined in terms of feature trajectories of foreground and they were modelled using the time series modelling technique. A cross-substitution based model comparison method was employed to compare the disparity between events. Spectral clustering (SC) was utilised to cluster events, and methods for SC initial parameter selection have been proposed. A method for cluster identity assignment in consecutive clustering iterations is also utilised to handle the evolving nature of the unsupervised learning methodology adopted. The proposed method is capable of producing reliable clustering results online, amidst a number of complications including dynamic backgrounds, object shadows, camera distortions, sudden foreground bursts and inter-object interactions.

Keywords: Event clustering, foreground estimation, trajectory modelling.

INTRODUCTION

Humans are capable of identifying similarities and dissimilarities between the events they perceive. The objective of event clustering is to produce a set of event clusters to follow human heuristics through an automated process. Such systems are widely applied in traffic surveillance (Rowe, 1991), automated surveying (Gowsikhaa *et al*., 2014), bio vision applications (Jiang *et al*., 2013; Aqel *et al*., 2016) and abnormal event detection (Archetti *et al*., 2006 ; Guo *et al*., 2013; Ranjith *et al*., 2015).

An object event can be broadly defined as a collection of activities performed by an object (Porikli & Haga, 2004; Utsumi *et al*., 2013). An event is described using a set of temporal features of the relevant object such as its location, velocity and size. Therefore, obtaining various feature variations of objects under consideration becomes one of the primary exercises in event clustering. In order to construct the necessary feature space, the desired foreground objects need to be segmented and tracked throughout the video.

Therefore, any event detection practice comprises a series of sub processes, namely, foreground estimation, foreground refining, object tracking for feature space construction and clustering. The paper proposes a comprehensive methodology, which addresses each of these sub processes in a sequential manner to arrive at a robust event clustering system.

Most of the previously proposed foreground estimation methods utilise the principle of background subtraction (Zang & Klette, 2004; Qin *et al*., 2013). Employment of a single background has long gone obsolete due to its inability to capture gradual changes in the background. Rahman *et al*. (2010) and Yang *et al*. (2005) have proposed remedies to absorb the background illumination change into the background reference model. But both these methods are ineffective when external objects are introduced to the background. The method proposed in Stauffer and Grimson (2000) is capable of tackling the issue of gradual background changes by modelling each pixel as a mixture of Gaussian distributions.

* Corresponding author (pramuditha@live.com)

Shadow removal is one of the most employed image refining tasks performed upon foreground detection. Defining a threshold for HSV colour values is employed by Qin *et al*. (2013) to remove shadowing effects. But this method requires manual tuning in the case of a dynamic scene. Thresholding grey-scaled foreground objects to detect shadows has been proposed by Chen *et al*. (2013). However, such ad-hoc thresholding could result in considering dark trousers, shoes and accessories erroneously as shadow components. The paper utilises the object tracking methodology proposed by Herath *et al.* (2014), which is capable of handling inter-object interactions such as occlusions and branching.

Distance metric learning extends from traditional Mahalanobis (Köstinger *et al*., 2012) metric learning to recent deep-net solutions (Ahmed *et al*., 2015; Cheng *et al*., 2016). However, most of them do not comply with spatio-temporal data. Although shape oriented comparison methods such as dynamic time warping (DTW) (Muller, 2007; Salvador & Chan, 2007) and Hausdorff distance (Williem *et al*., 2008) possess accurate results, they are less suited for online systems as they require almost completed traces for comparison. Many unsupervised clustering algorithms proposed by Jain (2010), Huang *et al*. (2014), and Anjum and Cavallaro (2008) have been attempted to use for event clustering. However, spectral clustering (SC) (Ng *et al*., 2001; Porikli & Haga, 2004; Perera, 2015) outperforms other methods due to its ability to capture non-linear connectivity. Recent work of Langone *et al*. (2016) proposes a Kernelised SC solution. Despite this, the proposed ad-hoc parameter selection methods suggested in these works do not guarantee good clustering in general.

The major contributions of the present study are as follows. First a behavioural analysis of the adaptive Gaussian mixture model (AGMM) was carried out to interpret the model qualitatively. Based on the inferences made, methods of deriving initial AGMM parameters are provided. The structure of the AGMM was modified to obtain a more consistent foreground by removing the effects of shadowing, visual impulses and visual distortions. Usability of the AGMM for event detection was made possible by introducing a still-object handling mechanism. Hence, the proposed methodology utilises the key features of AGMM, while the suggested modifications address its limitation in its application for surveillance based event detection applications.

SC principles were employed in identifying event clusters in the event space, and the methods to determine initial parameters are proposed. Time series modelling is proposed to model events and a cross-substitution based mechanism is suggested to estimate the disparity between two events in constructing the required affinity matrix.

METHODOLOGY

As in Figure 1 a raw video signal obtained was subjected to foreground estimation where foreground objects are identified. These objects were tracked to generate a feature space that describes the behaviour of the object while in frame; this feature space is what defines object events. Variation of features over time is modelled, compared and subjected to clustering through SC.

Foreground estimation

In the proposed work object based features are continuously tracked to infer the characteristics of each object event. Hence, proper isolation of each object as a foreground element is essential. Any distortion in extracted features (dynamics, size etc.) of an object will

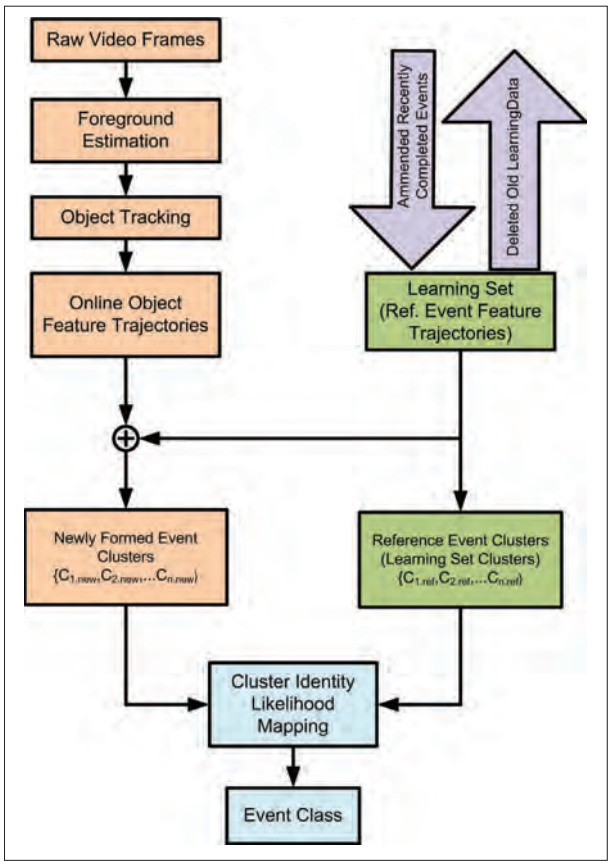

Figure 1: Overview of the overall event clustering process

result in erroneous conclusions about its nature. The proposed modifications to AGMM based foreground estimation mitigates the incorrect estimation of foreground pixels due to shadowing, camera and noise effects, visual bursts and stillness of tracked objects. Many studies have extended the core concept of AGMM to generate more sophisticated algorithms (Katsarakis *et al.*, 2016). AGMM has been fused with optical flow by Fradi and Dugelay (2012) and Rasheed *et al*. (2014) to generate more robust results. The EM algorithm has been used to automatically learn Gaussian parameters by Liu *et al*. (2012). However, since all these methods are built on the foundation of foreground estimation provided by the AGMM, they all suffer when there are erroneous estimates. Unlike in most event recognition studies where foreground estimation is of minor focus and only seen as an input received from a different system, this study attempted to scrutinise this stage as well.

The underlying logic behind the distinction of a foreground and background pixel lies in the sustainability of the temporal fluctuation of pixel intensity in background pixels as opposed to foreground pixels, which display a more sporadic and abrupt intensity fluctuation. For instance, through a human perspective, objects that exhibit a certain sustained dynamism over time are classified as background objects (still objects, fluttering leaves, ripples in a body of water etc.). Moreover, the perception of background would stay constant regardless of the gradual changes they undergo such as illumination changes. For example, a robust foreground estimation technique should isolate a man walking in the backdrop of fluttering leaves as a foreground object, while ensuring that the dynamic background in spite of its dynamism is still categorised as background.

Adaptive Gaussian mixture model (AGMM)

The AGMM based foreground estimation is capable of catering to the above mentioned requirements of a robust foreground estimation technique. The underlying argument in AGMM based techniques is that, a pixel's intensity fluctuation can be modelled based on a collection of Gaussian models where the more sustained and low variance models are isolated as background behaviour. For example a still background element will manifest itself as a near zero variance Gaussian and an oscillatory flutter as a bi-modal low variance Gaussian. The foreground components, which are temporally less sustained are modelled as high variance Gaussians with a low weighting, where the weighting reflects the temporal sustainability of the Gaussian. AGMM proposes a set of possible background models for each pixel in a video frame. If the currently observed intensity value of a particular pixel falls within the background model it is classified as background and *vice-versa*.

This process is implemented mathematically by modelling each pixel as a set of *K* Gaussian distributions. Each distribution has a mean (μ), variance (σ^2) and weight (*w*), where weights across them of a particular pixel are normalised. Initial mean values were selected such that they are placed equidistantly in the RGB colour space, and variances are given a high initial value. An observed pixel intensity becomes a match for a particular Gaussian if,

$$match = \begin{cases} True & \|X_t - \mu_t\| < W \\ False & Otherwise \end{cases}, \quad ...(1)$$

where X_t is the observed pixel value, and μ_t and σ_t are mean and variance of the distribution under consideration, respectively.

Parameter W is given by,

$$W = 2.5 \times \sigma_t . \quad ...(2)$$

When a distribution is matched, its distribution parameters are updated as,

$$w_{k,t} = (1 - \alpha)\, w_{k,t-1} + \alpha , \quad ...(3)$$

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

$$\sigma_t{}^2 = (1 - \rho)\, \sigma^2{}_{t-1} + \rho\, (X_t - \mu_t)^T (X_t - \mu_t), \quad ...(5)$$

$$\rho = \alpha \eta(X_t | \mu_k, \sigma_k), \quad ...(6)$$

The weight is increased using a universal incremental parameter α through the linear learning formula in equation (3). The mean and variance are both subjected to linear learning processes as in equations (4) and (5), respectively based on the likelihood match of the intensity to the Gaussian given in equation (6). If none of the Gaussians match the current intensity measurement, a new distribution is created around the observed value and the distribution, while dropping the distribution with the least *w*. The new Gaussian distribution added to the model will have the earlier stated high initial variance and a low weight.

Effectively, the weight increases and the variance decreases while the mean converges to a static value when the same Gaussian matches the observed intensity value continuously. This would be the case for a static background pixel. Since the weight and variance of background intensity fluctuations tend to be high and

low, respectively, it is expected that the parameter w/σ would be high for Gaussian's that model the background. In other words this parameter quantifies the likeliness of each Gaussian to be a background element. Hence, all *K* distributions are sorted based on their respective w/σ values and it is estimated as to how many Gaussian (b) distributions it takes to construct the background model of a pixel. The number of Gaussian distributions that make up the background model can be estimated as,

$$b = argmin_n \sum_{k=1}^{n} w_k > T, \quad ...(7)$$

where T is the user's belief as to how likely a pixel is to belong to a background in a given video. As pixel intensity distributions are assumed to be independent in AGMM, it stands to reason that the parameter in a given video is the same as the percentage contribution of background pixels in the selected surveillance application. This proposed logic enables the user to determine *T* more conveniently according to the context of the application. The Gaussians selected in this manner using equation (7) constructs the current background model for a particular pixel. Thereafter, pixels are classified as background or foreground pixels in the current frame according to equation (1).

Although AGMM is capable of estimating foregrounds in videos under slight background disturbances, this inherits a number of deficiencies that makes it unsuitable for object-event tracking applications. The deficiencies are as follows;

(a) Recognition of object shadows as part of the foreground

In addition to objects, shadows of objects satisfies the foreground conditions stated in AGMM. The inclusion of shadows overstates the dimensions of objects in feature space construction. Moreover, shadows that connect (merge) two or more objects on ground would cause false merges between objects, which are in fact separated in reality.

(b) Vulnerability to camera effects and noises

Cameras have a built-in automated colour histogram normalising feature. When a larger object suddenly appears in the video frame, the camera attempts to balance the overall intensity histogram. This would result in a sudden change in observed pixel intensity values for a large portion of the frame. This generates a huge error in the foreground estimate as the sudden burst manifests itself as a foreground behaviour that appears in a large portion of the video frame.

(c) Vulnerability to visual bursts in transients

Even though AGMM is capable of absorbing gradual background changes into the background model, it is unable to capture sudden changes in the entire frame resulting due to lightning, sudden cover of the sun due to clouds, camera shake, etc. As a result huge erroneous foreground blobs appear for a considerable amount of time, which would generate complications in object tracking.

(d) Absorption of still objects into the background

AGMM is based on the premise that foreground object mobility is higher than the background. On that logic, a human that walks into a frame and stops moving for a given time will get absorbed into the background due to lack of mobility. This would result in a disappearance of the object while within the frame. This is highly undesirable in event detection applications as it might even be falsely concluded as an abnormal event.

As for them, AGMM in its original form fails to deliver the desired estimation required for event detection. In addition to the behavioural analysis conducted for AGMM based foreground estimation and introduction of reliable tools to estimate and interpret parameters in the algorithm, the work presented in the paper proposes necessary modifications to the original AGMM to address these issues.

Shadow and noise removal

By utilising the understanding about the AGMM method it was concluded that a modification is required to refine the foreground estimation process to remove the impact due to shadow objects from the foreground. Hence, an analysis was conducted to understand the impact of shadows in the RGB colour space for practical real world measurements and thereby utilise these measured observations to remove shadow objects from the foreground estimated using the AGMM technique.

First, the per-pixel behaviour as it transits from background to foreground and background to shadow were visualised in the RGB space to distinguish a pattern to separate foreground from shadow objects. Figure 2 illustrates the results of this analysis. It was observed that the relative colour motion vector from background to shadow transition lies within a cylindrical space with the reference background as the axis of the cylinder, while the motion vector for background to foreground transition is more sporadic. This is intuitive as the foreground actually occludes the background for a given pixel, while the shadow simply casts a shadowy-veil

upon the background, resulting in a controlled colour distortion that is symbolised by the motion vector restricting itself to the cylinder mentioned above.

It should also be noted that this proposed method utilises the relative motion in colour space with respect to each pixel's background model when a shadow is cast upon it. As the classification is always conducted with respect to each pixel's reference background it embodies the fact that the colour distortion caused by a shadow is dependent upon the background it casts itself upon. These inferences were made by analysing real video streams under conditions that are common for applications of this nature.

The constrained colour distortion explained above due to background to shadow transition can be identified by constructing a cylindrical boundary as shown in Figure 2, to contain the colour motion for this transition. The array consisting of the RGB values of a pixel is denoted as the colour vector. The motion vector in the colour space is obtained by differencing the reference background colour vector from the currently observed colour vector for the pixel in concern. Thereafter, the differential colour vector is resolved along the radial and axis direction of the cylinder and checked whether both projected values are less than a selected threshold. If so it is classified as a shadow pixel, otherwise it becomes part of the foreground as for,

$$pixel = \begin{cases} Shadow\ ; axial\ projection < h_o\ AND\ radial\ projection < r_o \\ Foreground\ ; otherwise, \end{cases} \quad ...(8)$$

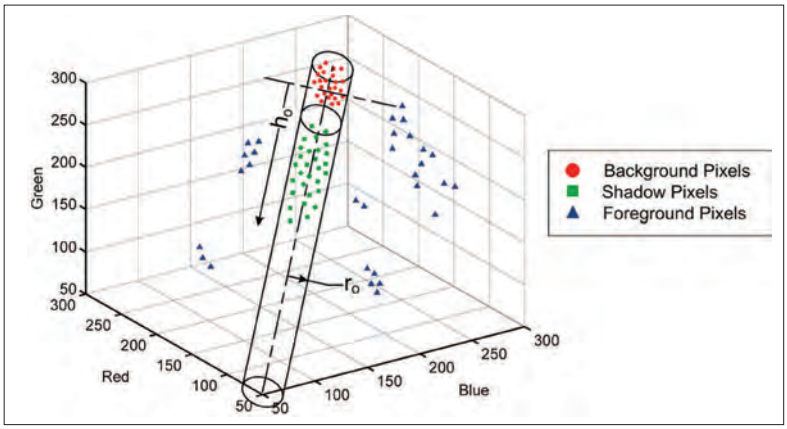

Figure 2: Spatial clustering of foreground, background and shadow pixels in RGB domain

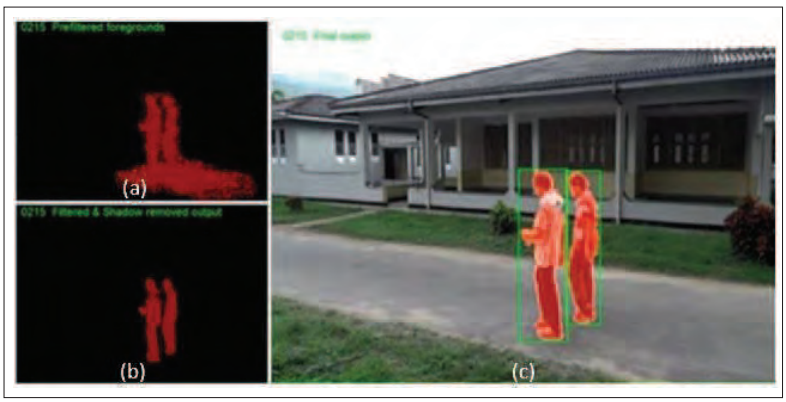

Figure 3: Visual coalition of objects eliminated through shadow removal: (a) visually merged silhouettes; (b) filtered silhouettes; and (c) filtered result

Where, the radial threshold $(r_0) = 3$ and the axial threshold $(h_0) = 3.5$ are selected on tuning for defining the shadow bounds. Illustrated in Figure 3 is a frame where the above shadow removal is incorporated. AGMM is a method that concludes whether a pixel belongs to foreground or background. The proposed RGB domain motion vector classifier concludes on the status of a pixel as a foreground or a shadow element and is constructed upon the reference background value provided by the AGMM.

Since both the above methods conclude on a pixel behaviour, to improve the reliability, a decision level fusion of the above two methods is proposed. However, neither the AGMM nor RGB gives a perfect decision on pixel status as to whether it is beyond any doubt a foreground pixel. A foreground element with a high colour similarity with the background would create a colour motion vector that remains inside the cylindrical boundary for shadow elements. Hence, a foreground pixel could be falsely identified as a shadow element from the RGB classifier. This is due to the fact that while colour distortion due to background to shadow transition is limited to the RGB colour cylinder, the colour distortion due to background to foreground transition can be in or outside the cylinder. Hence, the decision in equation (8) might cause an error because the shadow space is a subset of the foreground space. Therefore, a secondary validation is required for such cases based on the strength of the conclusions made in the AGMM method.

In such a situation, using a crisp decision fusion method such as the 'AND' operation would not be proper. Illustrated in Figure 4(a) is a foreground estimate obtained through the crisp 'AND' operation based decision fusion. Despite the identification of shadows, the inherent loss of foreground information through fusion has caused the object to fragment in this case.

A fusion method considering the level of confidence each method imposes on its decision, would be appropriate. In order to bring the level of confidence that each of the method produces on its decision, a probabilistic approach is considered. The decision level fusion is conducted through a Bayesian framework by incorporating both methods mentioned above, to obtain the probability for a considered pixel to be foreground. Hereafter, an experimentally decided threshold is imposed on the fused probability in deciding the status of the pixel. The evaluated foreground through fusion results in a better approximation as illustrated in Figure 4(b). For example, as shown in Figures 4(a) and (b) the head, which was incorrectly regarded as a shadow by the RGB cylinder method is reassessed as a foreground by the Bayesian fusion algorithm, due to the strength of the probability that it is foreground based on the AGMM method, resulting in a more desirable less fragmented outcome as seen in the results.

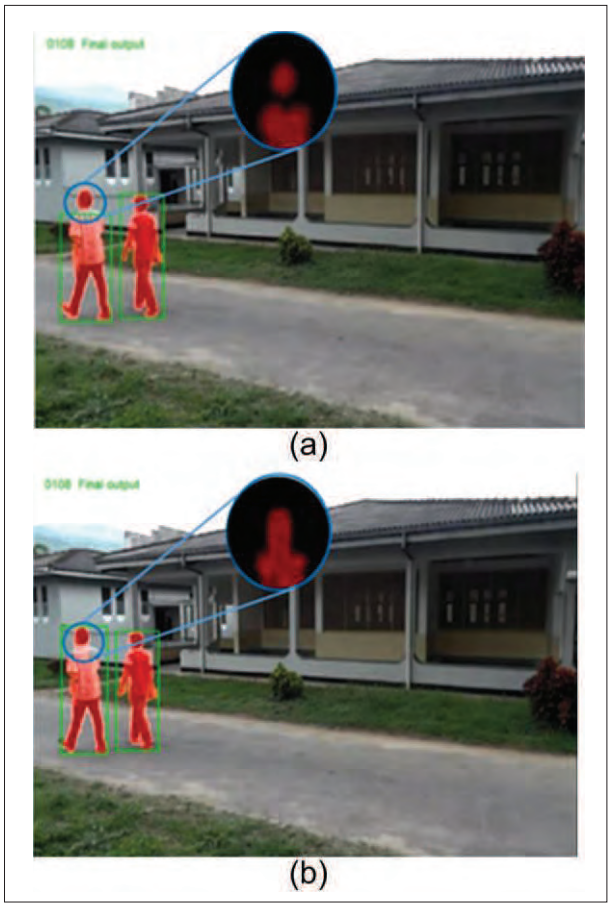

Figure 4: (a) Foreground estimate obtained through crisp decision fusion through AND operation; (b) Foreground estimate obtained through Bayesian fusion.

Visual impulse filtering

Visual bursts can be defined as situations where the estimator associates a large number of background pixels as a part of foreground. AGMM (Mukherjee & Das, 2013; Santosh *et al*., 2013) method matches the observation to a distribution if and only if the condition in equations (1) and (2) are satisfied. Parameter W in equation (2) will directly affect the performance of the estimator. Selecting a very large W value will result in matching an observed pixel into undesired distributions, whereas using a very small W will make the estimator vulnerable to noise.

It was observed that when a visual impulse occurs, the affected pixel values change rapidly even if they belong to the background. This work proposes to modify equation (2) of AGMM to obtain parameter W as in,

$$W = \alpha\sigma + \beta e^{-\gamma n}, \quad ...(9)$$

to overcome this issue. Here, α, β, γ are constants and n is the duration starting from the occurrence of visual impulse. The exponential function has been introduced to ensure rapid absorption of visual impulses into the background. As illustrated in Figure 5, the proposed modified method with the exponentially decaying matching parameter is capable of absorbing visual bursts much faster compared to the original method.

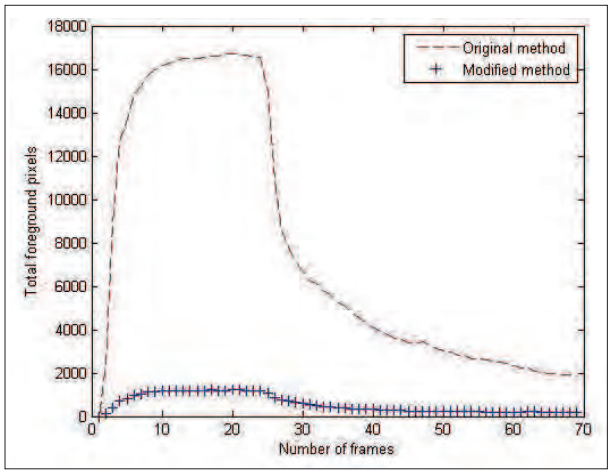

Figure 5: Absorption of a visual burst under the proposed modification for learning *w* parameter

Removing effect of still object absorption

In equation (3), a universal parameter α is used to increase the weight of a matched distribution. Weight is one of the contributory factors that later on decide whether a given distribution describes a background or not. Therefore, if a pixel wise α is defined in place of universal α, and if the α value corresponding to object pixels are set to a lower value, the time taken for the object to get absorbed into the background can be increased. This can be carried out using a closed loop feedback mechanism where pixel wise α is altered as,

$$\alpha_t = \begin{cases} \alpha & ; If\ not\ foreground\ at\ t-1 \\ \alpha' & ;\quad If\ foreground\ at\ t-1 \end{cases}, \quad ...(10)$$

where $\alpha > \alpha'$.

The feedback mechanism alters the α value of the present foreground pixels and slowdowns or prevents (depending on the value of α') the learning of those pixels in the next time instant. As a result, although the same pixel distribution gets matched over a long period of time due to a still foreground object, it will not be adapted as a part of the background.

Object tracking and feature space construction

Avidan (2007) addresses the tracking problem as an object and immediate background pixel classification problem. However, this is limited in its applications when it needs to recognise multiple interacting objects such as merges and splits. Event detection requires a consistent feature space of tracked objects. Tracked object location, size and speed are some informative features used in event detection. Region tracking approaches (Matthews *et al*., 2004) might be less useful in this situation where specific object features need to be recorded for analysing behaviour patterns. Considering the above, here the object tracking methodology utilises the tracking principle proposed by Yang *et al*. (2005) together with the improvements in Herath *et al.* (2014), which is capable of detecting object-events such as merges and splits. The colour histogram based object identity recall mechanism included has enabled to keep track on objects even under temporary losses from the scene. The recorded temporal variation of any feature is considered as a feature trajectory.

EVENT CLUSTERING

The objective of event clustering is to produce an automated system that groups similar events based on event feature trajectories in such a way that the human intuition is closely reflected. Algorithms, such as *K–means* (An *et al*., 2008) are unable to identify similarities between two feature trajectories due to their incapability of mapping all trajectories into a constant dimensional vector space (Porikli *et al*., 2004). Since events are defined based on their feature trajectories, these existing methods are less suitable for event clustering applications. SC is an alternative approach that tackles the issue of event trajectory comparison by modelling the event set as a closed network (Porikli & Haga, 2004). Event clusters obtained through such a mechanism could be subjected to classification schemes (Tuzel *et al*., 2008; Tosato *et al*., 2010) based on their topology in further processing. However, all these algorithms are less appropriate to be applied for real life unsupervised grouping mechanisms as they require a number of clusters as a priori.

Spectral clustering (SC)

SC is capable of mapping events into a cluster space where similar events are clustered together (Porikli & Haga, 2004). The output of this process would be a list of clusters with information on constituent events in each cluster. SC algorithm proposed by Ng *et al*. (2001) is used in this work. Here, an affinity matrix $A \in R^{n \times n}$ is formed by mapping the entries according to,

$$\mathrm{D}(i,j) = e^{\frac{-D_{i,j}{}^2}{2\sigma^2}}. \quad ...(11)$$

where $D_{i,j}$ is a disparity between the events i and j while σ is a tunable initial parameter.

Parameter and variable determination

The performance of SC is highly dependent upon the structure of the affinity matrix and pre-defined parameter values. Affinity matrix entries should closely reflect inter event similarity where parameters K and σ are input variables to the algorithm.

Affinity entry determination

In event comparison a crucial factor that decides its performance would be the distance metric. Real life events do not exist for the exact same duration; hence their feature trajectories are not of the same length. Moreover, distance evaluation methods employed to tackle this phenomenon such as DTW and Hausdorff distance require complete event trajectories. Therefore they are less suitable for online event disparity evaluation.

Proposed here is a method where an event trajectory is considered as a time-series and modelled through multivariate auto regressive (AR) modelling to capture the dynamic behaviour, which governs object motion. AR modelling boils down the dynamic properties of a temporal feature trajectory into a comparable measure. However, direct model structure comparison methods fail in event comparison due to variable model orders across different trajectories. To handle this, a comparison method based on cross-substitution is proposed.

Consider the comparison of two feature trajectories, $a(n)$ of length N and $b(m)$ of length M describing distinct events A and B, respectively. If $M_A(a(n-1), \ldots a(n-k))$ describes the prediction of a k^{th} order dynamic model for $a(n)$, then modelling error $E_A(a)$ is evaluated according to,

$$E_A(a) = \sqrt{\frac{\sum_{k+1}^{N}(a(n) - M_A(a(n-1)\ldots a(n-k)))^2}{N-k-1}}. \quad ...(12)$$

Similarly the error of cross-substitution would be given by the terms $E_A(b)$ and $E_B(a)$. Incorporating the above substitution errors, a scalar mutual distance d_{AB} is defined as in,

$$d_{AB} = \left|(1-\alpha)\left(E_A(b) - E_A(a)\right) + \alpha\left(E_B(a) - E_B(b)\right)\right|. \quad ...(13)$$

Where α is defined according to,

$$\alpha = \frac{E_A(a)}{E_A(a) + E_B(b)} \quad ...(14)$$

as a weigt to encounter the modelling error of an event in d_{AB} evaluation process.

Distance d_{AB} is structured such that the disparity matrix arrangement is substantiated. The used weighting of the distance components from modelling error (α) compensates for the linearity and weak stationarity assumed in time-series modelling.

Illustrated in Figure 6(a) is a situation where the location trajectory of an emerging event is compared with an event of the same class. The variation of mutual

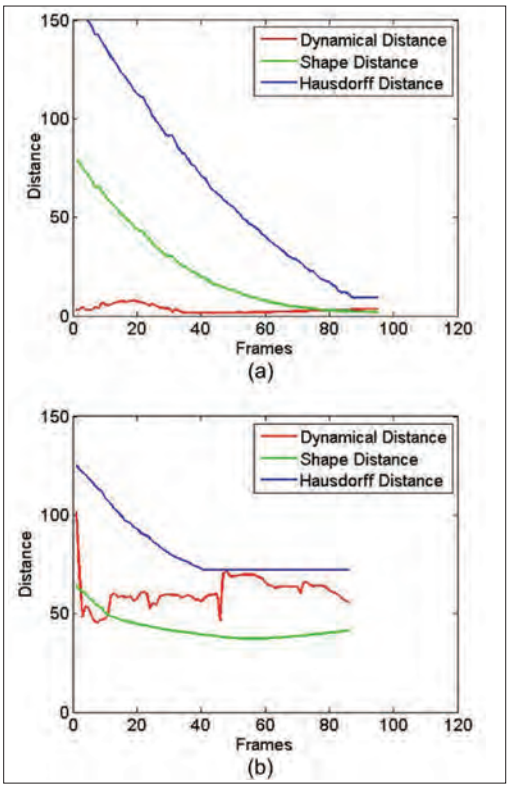

Figure 6: (a) Trajectory distance for an emerging event with a similar class (b) Trajectory distance for an emerging event with dissimilar class

event distance (d_{AB}) with the available emerging event is compared with the response of conventional shape comparison methods, DTW and Hausdorff distance. It is evident from Figure 6(a) that for a given event, the dynamic mutual distance drops much earlier compared to the distance evaluated using other comparison techniques. Hence, emerging event trajectories yield better correspondence to their classes from the onset, which facilitates reliable online event clustering.

Unfolding the distance of an incoming trajectory with a signature trajectory from a different class for various distance metrics is illustrated in Figure 6(b). As for them it is evident that the dynamic modelling based distance has an adequate intra class separation comparable to the shape comparison methods.

Initial parameter determination

In a perfect clustering scenario, which includes distinct events, the affintity matrix would take the form of a block diagonal matrix and there exists a number of dominant Eigen values (Ng *et al*., 2001). This result cannot be generalised for a generic clustering scenario. However, according to the matrix perturbation theory, diviation from block diagonal structure will not affect its Eigen vectors given the Eigen gap is sufficiently large. This paper proposes the selection of parameters such that it maximises the Eigen spread of the affinity matrix. In determining the cluster order, first the difference plot of ordered absolute Eigen values is constructed. The index that relates to the extrema in the above plot is equal to the number of clusters in the dataset.

RESULTS

The effectiveness of the proposed mechanism was tested in two phases independently where the performance of foreground estimation and event clustering were evaluated.

Foreground estimation

In the first phase, videos that contain noise, visual bursts and background dynamics were considered to evaluate the proposed method. Bayesian fusion based noise removal process improved the quality of foreground estimation as shown in Figure 4. The proposed visual impulse filtering method drastically reduced the transient time of sudden foreground bursts occurring due to external phenomenon as shown in Figure 5. With collective effort of all these amendments AGMM was able to produce more consistent object silhouettes for tracking. Compared in

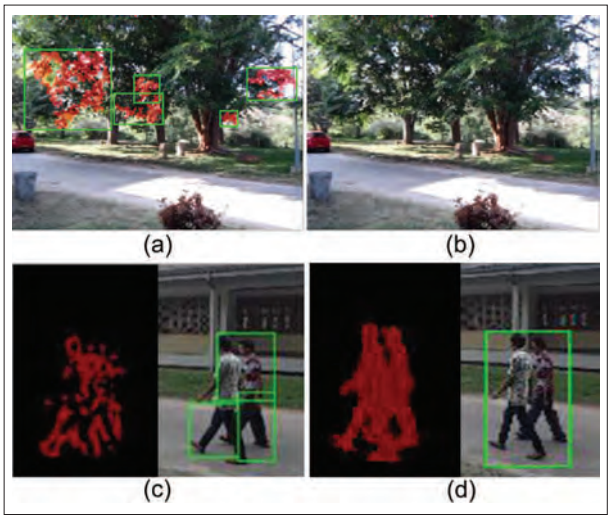

Figure 7: (a) Erroneous foreground detection when optical flow is used on dynamic backgrounds; (b) dynamic background handled through the suggested mechanism; (c) fragmentation of foreground estimate through optical flow; (d) Consistent foreground estimate from the suggested mechanism

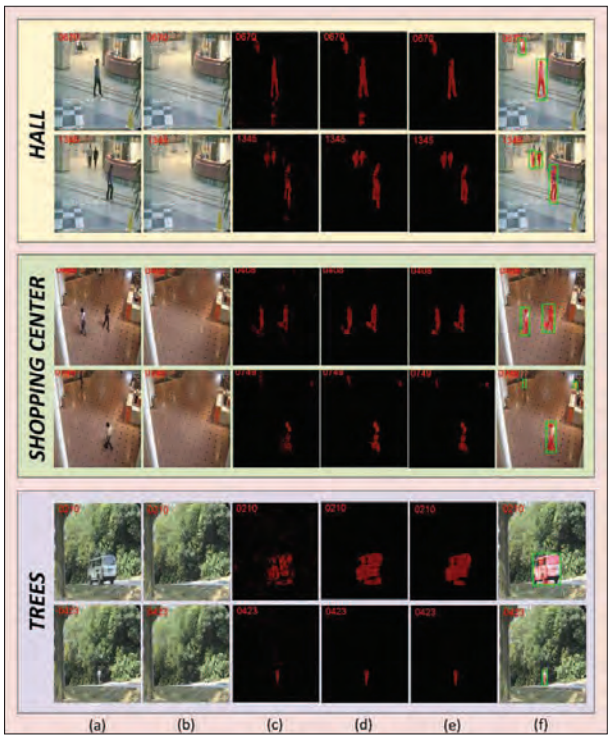

Figure 8: Evaluated results on the i2r dataset: (a) original frame (input); (b) adapted background frame; (c) AGMM output; (d) filtered output; (e) shadow removed; (f) detected object (final output)

Table 1: Comparison of recall (R) and precision (P) (Fradi & Dugelay, 2012) measures with existing methods

Video Sequences	Improved adaptive GMM (Zivkovic., 2004) (R/P)	Foreground object detection method (Li *et al*., 2003) (R/P)	Proposed approach (R/P)
Shopping center	52.18/ 99.69	59.50/ 98.33	67.34/ 99.73
Hall	39.10/ 99.71	47.37/ 99.08	65.23/ 99.65
Trees	73.99/ 97.45	59.50/ 98.33	78.26/ 98.21

Figure 7 are the foreground extracts from the optical flow technique with the proposed AGMM based mechanism. It is evident from Figure 7(a) that the optical flow method is not suited for dynamic background conditions. If it is to be utilised in such circumstances it has to be combined with a mechanism (Fradi & Dugelay, 2012; Rasheed *et al*., 2014) such as AGMM, which effectively handles dynamic backgrounds. However, such a combination would be computationally exhaustive in its execution. Furthermore, it was observed that optical-flow is more vulnerable to lose foreground sections where the object colour spread is uniform. This phenomenon is illustrated in Figure 7(c).

For further verification, the proposed foreground estimation methodology was tested on the publically available I2R dataset, which contains both indoor and outdoor scenes with human subjects (Fradi & Dugelay, 2012). Illustrated in Figure 8 are the outputs obtained at each step of the performed foreground estimation methodology. Listed in Table 1 are the compared measures recall (R) and precision (P) introduced in Fradi and Dugelay (2012), with corresponding figures from existing methodologies (Li *et al*., 2003; Zivkovic, 2004). The measures are calculated on the detected foreground with respect to the provided ground-truth pixel masks according to,

$$Recall\ (R) = {}^{TP\ Foreground}\Big/_{(TP\ Foreground\ +\ FN\ Foreground)} \quad ...(15)$$

$$Precision\ (R) = {}^{TN\ Foreground}\Big/_{(FP\ Foreground\ +\ TN\ Foreground)} \quad ...(16)$$

Event clustering

In evaluating the performance of event clustering, emphasis was given to test the ability of the algorithm to identify clusters based on motion patterns. Event clustering tests were done in an outdoor environment with varying illumination focusing on human motion. The experiment was conducted over 65 object motion

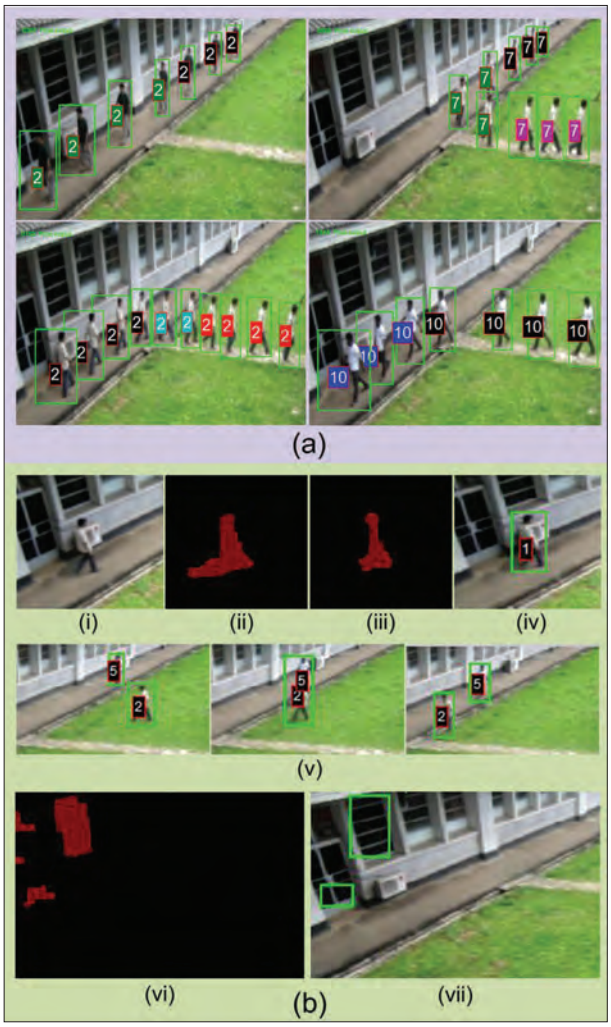

Figure 9: (a) Several detected events of different event classes represented by a corresponding colour with the time of detection; (b) Foreground complications present on the dataset used for event clustering, (i) shadowy input, (ii) foreground estimate without shadow removal, (iii) shadow removed foreground estimate, (iv) detected object, (v) visual occlusions, (vi) noisy foreground visual burst detected at the initialisation with a constant W in place of the proposed W on equation 9, (vii) noises detected as objects

trajectories and 40 of the trajectories were imposed as the initial learning set, which consisted of 5 event classes with multiple occurrences. The execution of the process was conducted as explained in Figure 1. Figure 9(a) illustrates the frame sequences where each of the event is identified and indicated with a respective colour on its identity number box. Some of the foreground complications present in the considered scenes are illustrated in Figure 9(b).

Initially an emerging object undergoes an observation time, during which the object is not subjected to classification and is indicated in black colour. At the completion of the observation time, the events are classified based on the dynamic distances evaluated with the learning set and is indicated in the relevant event colour. The events continue to adapt according to the prevailing conditions due to unsupervised progression of SC.

Table 2: Number of event identification through shape and dynamic comparison

Comparison method	Minimum required trajectory percentile		
	50 %	65%	75 %
Dynamical dist.	11	9	5
DTW	0	15	10

Given in Table 2 is a quantitative comparison of event identification from dynamic distance and DTW shape comparison methods. Experimentally it was found that on average, dynamic mutual distance yields correct predictions at the completion of 60 % of a trajectory whereas methods such as DTW requires on average 70 % of the trajectory for the considered cases. It is evident from the above experimental figures that comparison of the the dynamic modelling has enabled to identify events with a less available trajectory percentage than the existing shape comparison methods. As a result, online event clustering has become a possibility.

CONCLUSION

AGMM in its original form is unsuited to be applied for event detection applications. These shortcomings were mitigated using the introduced stillness handling method, visual impulse filtering and the Bayesian fused shadow removal to produce more consistent event traces. Event classification was made possible preceded on tracked objects by the above due to their imposed reliability and consistency.

The proposed dynamic modelling based event comparison method enabled different length trajectories. The paper utilises the usage of SC for event grouping and states a methodology to tune selected clustering parameters. The developed methodology enabled object-event clustering, which can be further extended in use for classification purposes.

REFERENCES

1. Ahmed E., Jones M. & Marks T.K. (2015). An improved deep learning architecture for person re-identification. *Proceedings of the 2015 IEEE Conference on Computer Vision and Pattern Recognition,* Boston, MA, USA, 7 – 12 June, pp. 3908 − 3916.
   DOI: https://doi.org/10.1109/CVPR.2015.7299016
2. An Y., Baek J., Shin S., Chang M. & Park J. (2008). Classification of feature set using k-means clustering from histogram refinement method. *Proceedings of the 4th International Conference on Networked Computing and Advanced Information Management*, volume 2, 2 − 4 September, pp. 320 − 324.
   DOI: https://doi.org/10.1109/ncm.2008.112
3. Anjum N. & Cavallaro A. (2008). Multifeature object trajectory clustering for video analysis. *IEEE Transactions on Circuits and Systems for Video Technology* **18**(11): 1555 − 1564.
   DOI: https://doi.org/10.1109/TCSVT.2008.2005603
4. Aqel S., Aarab A. & Sabri A. (2016). Traffic video surveillance: background modeling and shadow elimination. *Proceedings of the International Conference on Information Technology for Organizations Development* (IT4OD), Fez, Morocco, 30 March – 01 April, pp. 1 – 6.
   DOI: https://doi.org/10.1109/it4od.2016.7479290
5. Archetti F., Manfredotti C.E., Matteucci M., Messina V. & Sorrenti D.G. (2006). Parallel first-order Markov chain for on-line anomaly detection in traffic video surveillance. *The Institution of Engineering and Technology Conference on Crime and Security*, 13 – 14 June, London, UK, pp. 3251 – 3256.
   DOI: https://doi.org/10.1049/ic:20060365
6. Avidan S. (2007). Ensemble tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **29**(2): 261 − 271.
   DOI: https://doi.org/10.1109/TPAMI.2007.35
7. Chen Y., Yang H., Li C., Pu S., Zhou J. & Zheng L. (2013). Robust pedestrian detection and tracking with shadow removal in indoor environments. *Proceedings of the International Joint Conference on Awareness Science and Technology and Ubi-Media Computing (iCAST-UMEDIA)*, Aizu-Wakamatsu, Japan, 2 − 4 November, pp. 590 − 596.
   DOI: https://doi.org/10.1109/icawst.2013.6765508

8. Cheng D., Gong Y., Zhou S., Wang J. & Zheng N. (2016). Person re-identification by multi-channel parts-based CNN with improved triplet loss function. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Washington Convention Center, Seattle, WA, USA, 27 – 30 June, pp. 1335 − 1344.
9. Fradi H. & Dugelay J.L. (2012). Robust foreground segmentation using improved gaussian mixture model and optical flow. *Proceedings of the International Conference on Informatics, Electronics and Vision (ICIEV)*, Dhaka, Bangladesh, 18 – 19 May, pp. 248 − 253.
DOI: https://doi.org/10.1109/iciev.2012.6317376
10. Gowsikhaa D., Abirami S. & Baskaran R. (2014). Automated human behaviour analysis from surveillance videos: a survey. *Artificial Intelligence Review* **42**(4): 747 − 765.
DOI: https://doi.org/10.1007/s10462-012-9341-3
11. Guo H., Wu X., Li N., Fu R., Liang G. & Feng W. (2013). Anomaly detection and localization in crowded scenes using short-term trajectories. *Proceedings of the IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Shenzhen, China, 12 − 14 December, pp. 245 − 249.
DOI: https://doi.org/10.1109/robio.2013.6739466
12. Herath H.M.S.P.B., Perera P.H., Fernando W.S.K., Ekanayake M.P.B., Godaliyadda G.M.R.I. & Wijayakulasooriya J.V. (2014). Human motion tracking under dynamic background conditions. *Proceedings of the 9th International Conference on Industrial and Information Systems (ICIIS)*, Gwalior, India, 15 – 17 December, pp. 1 − 6.
DOI: https://doi.org/10.1109/iciinfs.2014.7036523
13. Huang X., Ye Y. & Zhang H. (2014). Extensions of kmeans-type algorithms: a new clustering framework by integrating intracluster compactness and intercluster separation. *IEEE Transactions on Neural Networks and Learning Systems* **25**(8): 1433 − 1446.
DOI: https://doi.org/10.1109/TNNLS.2013.2293795
14. Jain A.K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters* **31**(8): 651 − 666.
DOI: https://doi.org/10.1016/j.patrec.2009.09.011
15. Jiang X., Dawood M., Gigengack F., Risse B., Schmid S., Tenbrinck D. & Schäfers K. (2013). Biomedical imaging: a computer vision perspective. *Computer Analysis of Images and Patterns: Lecture Notes in Computer Science* **8047**: 1 − 19.
DOI: https://doi.org/10.1007/978-3-642-40261-6_1
16. Katsarakis N., Pnevmatikakis A., Tan Z. & Prasad R. (2016). Improved Gaussian mixture models for adaptive foreground segmentation. *Wireless Personal Communications* **87**(3): 629 – 643.
DOI: https://doi.org/10.1007/s11277-015-2628-3
17. Köstinger M., Hirzer M., Wohlhart P., Roth P.M. & Bischof H. (2012). Large scale metric learning from equivalence constraints. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Rhode Island, USA, 16 – 21 June, pp. 2288 − 2295.
DOI: https://doi.org/10.1109/cvpr.2012.6247939
18. Langone R., Mall R., Alzate C. & Suykens J.A. (2016). Kernel spectral clustering and applications. *Unsupervised Learning Algorithms*, pp. 135 – 161. Springer International Publishing, Switzerland.
DOI: https://doi.org/10.1007/978-3-319-24211-8_6
19. Li L., Huang W., Gu I.Y. & Tian Q. (2003). Foreground object detection from videos containing complex background. *Proceedings of the Eleventh ACM International Conference on Multimedia*, Berkeley, CA, USA, 02 – 08 November, pp. 2 − 10.
DOI: https://doi.org/10.1145/957013.957017
20. Liu Z., Huang K. & Tan T. (2012). Foreground object detection using top-down information based on EM framework. *IEEE Transactions on Image Processing* **21**(9): 4204 − 4217.
DOI: https://doi.org/10.1109/TIP.2012.2200492
21. Matthews I., Ishikawa T. & Baker S. (2004). The template update problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **26**(6): 810 − 815.
DOI: https://doi.org/10.1109/TPAMI.2004.16
22. Mukherjee S. & Das K. (2013). An AGMM approach to background subtraction for application in real time surveillance. *International Journal of Research in Engineering and Technology* **2**(1): 25 − 29.
DOI: https://doi.org/10.15623/ijret.2013.0201005
23. Müller M. (2007). Dynamic time warping. *Information Retrieval for Music and Motion*, pp. 69 − 84. Springer, Berlin, Heidelberg, Germany.
DOI: https://doi.org/10.1007/978-3-540-74048-3_4
24. Ng A.Y., Jordan M.I. & Weiss Y. (2002). On spectral clustering: analysis and an algorithm. *Advances in Neural Information Processing Systems* **2**: 849 − 856.
25. Perera S. (2015). Rigid body motion segmentation with an RGB-D camera. *PhD thesis,* pp. 199 – 203. The Australian National University, Australia.
26. Porikli F. & Haga T. (2004). Event detection by eigenvector decomposition using object and frame features. *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'04)*, volume 7, 27 June – 02 July, p. 114.
DOI: https://doi.org/10.1109/cvpr.2004.335
27. Qin Y.S., Sun S.F., Ma X.B., Hu S. & Lei B.J. (2013). A shadow removal algorithm for ViBe in HSV colour space. *Proceedings of the 3rd International Conference on Multimedia Technology*, Guangzhou, China, 29 November 01 – December, pp. 966 − 973.
DOI: https://doi.org/10.2991/icmt-13.2013.119
28. Rahman F.Y.A., Hussain A., Tahir N.M., Zaki W.M.D.W. & Mustafa M.M. (2010). Modelling of initial reference frame for background subtraction. *Proceedings of the 6th International Colloquium on Signal Processing and its Applications (CSPA)*, Malacca City, Malaysia, 21 – 23 May, pp. 1 − 4.
DOI: https://doi.org/10.1109/cspa.2010.5545327
29. Ranjith R., Athanesious J.J. & Vaidehi V. (2015). Anomaly detection using DBSCAN clustering technique for traffic video surveillance. *Seventh International Conference on Advanced Computing* (ICoAC), Tamilnadu, India, 15 – 17 December, pp. 1 − 6.

DOI: https://doi.org/10.1109/ICoAC.2015.7562795

30. Rasheed N., Khan S.A. & Khalid A. (2014). Tracking and abnormal behavior detection in video surveillance using optical flow and neural networks. *Proceedings of the 28th International Conference on Advanced Information Networking and Applications (WAINA)*, Victoria, Canada, 13 – 16 May, pp. 61 − 66.
DOI: https://doi.org/10.1109/waina.2014.18
31. Rowe E. (1991). The Los-Angeles automated traffic surveillance and control (ATSAC) system. *IEEE Transactions on Vehicular Technology* **40**(1): 16 − 20.
DOI: https://doi.org/10.1109/25.69967
32. Salvador S. & Chan P. (2007). Toward accurate dynamic time warping in linear time and space. *Intelligent Data Analysis* **11**(5): 561 − 580.
33. Santosh D.H.H., Venkatesh P., Poornesh P., Rao L.N. & Kumar N.A. (2013). Tracking multiple moving objects using Gaussian mixture model. *International Journal of Soft Computing and Engineering* **3**(2): 114 – 119.
34. Stauffer C. & Grimson W.E.L. (2000). Learning patterns of activity using real-time tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(8): 747 − 757.
DOI: https://doi.org/10.1109/34.868677
35. Tosato D., Farenzena M., Spera M., Murino V. & Cristani M. (2010). Multi-class classification on riemannian manifolds for video surveillance. *Proceedings of the 11th European Conference on Computer Vision – Part II*, Heraklion, Crete, Greece, 05 – 11 September, pp. 378 − 391.
DOI: https://doi.org/10.1007/978-3-642-15552-9_28
36. Tuzel O., Porikli F. & Meer P. (2008). Pedestrian detection *via* classification on riemannian manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **30**(10): 1713 − 1727.
DOI: https://doi.org/10.1109/TPAMI.2008.75
37. Utsumi Y., Katte M., Iwamura M. & Kise K. (2013). Event detection based on noisy object information. *Proceedings of the 2nd IAPR Asian Conference on Pattern Recognition (ACPR)*, Okinawa, Japan, 05 – 08 November, pp. 572 – 575.
DOI: https://doi.org/10.1109/acpr.2013.85
38. Wiliem A., Madasu V., Boles W. & Yarlagadda P. (2008). Detecting uncommon trajectories. *Digital Image Computing: Techniques and Applications (DICTA)*, Canberra, Australia, 01 – 03 December, pp. 398 − 404.
DOI: https://doi.org/10.1109/dicta.2008.45
39. Yang T., Pan Q., Li J. & Li S.Z. (2005). Real-time multiple objects tracking with occlusion handling in dynamic scenes. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Diego, CA, USA, 20 – 25 June, pp. 970 − 975.
40. Zang Q. & Klette R. (2004). Robust background subtraction and maintenance. *Proceedings of the 17th International Conference on Pattern Recognition*, volume 3, Cambridge, UK, 23 – 26 August, pp.
DOI: https://doi.org/10.1109/ICPR.2004.1334047
41. Zivkovic Z. (2004). Improved adaptive Gaussian mixture model for background subtraction. *Proceedings of the 17th International Conference on Pattern Recognition*, volume 2, Cambridge, UK, 23 − 26 August, pp. 28 − 31.
DOI: https://doi.org/10.1109/ICPR.2004.1333992