



# People/vehicle classification by recurrent motion of skeleton features

B. Yogameena S.Md. Mansoor Roomi R. Jyothi Priya S. Raju V. Abhaikumar

Thiagarajar College of Engineering, Thiruparankundram, Madurai 625015, India  
 E-mail: ymece@tce.edu

**Abstract:** Object classification is a major application in video surveillance such as automatic vehicle detection and pedestrian detection, which is to monitor thousands of vehicles and people. In this study, an object classification algorithm is proposed to classify the objects into persons and vehicles despite the presence of shadow and partial occlusion in mid-field video using recurrent motion image (RMI) of skeleton features. In this framework, the background subtraction using a Gaussian mixture model is followed by Gabor filter based shadow removal in order to remove the shadow in the image. The star skeletonisation algorithm is performed on the segmented objects to obtain skeleton features. Then the RMI is computed and it is partitioned into two sections such as top and bottom. Based on the signatures derived from the bottom section of the partitioned RMI using skeleton features, the object is classified into people and vehicles.

## 1 Introduction

Video surveillance is currently an active area of research in computer vision field, with the goal of developing visual sensing, processing algorithms and hardware that can see and understand the world around them. Recent research in video surveillance systems has focused on foreground object detection, moving object classification, tracking and abnormal activity analysis [1]. A more interesting part of computer vision is the classification of segmented foreground regions in terms of the meaningful objects it comprises such as people, cars and animals. The classification of a group of image pixels into meaningful object regions probably requires an intelligent combination of multiple visual cues: colour, shape, which is defined by a silhouette, local features such as internal edges, corners and spatial continuity [2]. The common architecture of classification systems consists of the following three main steps: motion segmentation, object tracking and object classification [3]. Motion segmentation provides essentially a collection of foreground regions and in each frame that might correspond to moving objects, hence it acts as a filter that focuses attention on only certain regions of an image [4]. Stauffer and Grimson [5] used the mixture of Gaussians to perform background subtraction and applied a two-pass grouping algorithm to segment foreground pixels. Other simple and common techniques are based on frame differencing or using a median filter [6]. The difficulties associated with foreground detection deal with illumination change, camouflage and shadow. Shadow is considered to be an important issue, which leads to misclassification in moving object classification. Although there are many approaches available to deal with shadows, Cucchiara *et al.* [7] in their work, characterised and classified many types of

shadows. The major issue in the algorithms proposed in the literature for dynamic shadow elimination is their computational cost and usage of empirically determined threshold values for shadow removal. The use of complex geometrical approaches for detecting rough shadow regions makes the system computationally multifarious. In model based shadow elimination methods, various parameters remain to be tuned during the modelling stage, which requires user intervention. Therefore an effective shadow elimination algorithm has to be developed.

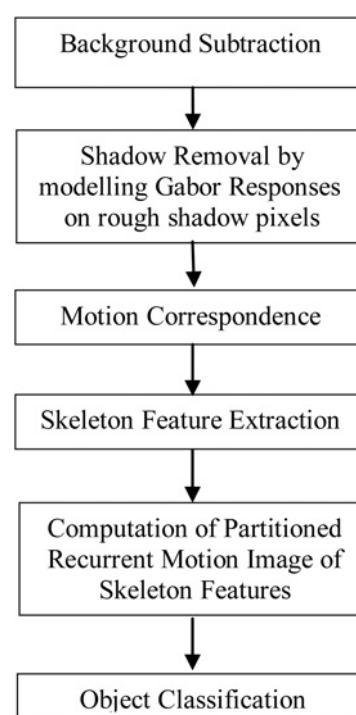
For a static camera environment, two main categories of approaches for classifying moving objects are reported in the literature, namely shape based classification and motion based classification. Shape based classification depends on the descriptions of shape information of motion regions such as points, boxes and silhouettes for classifying moving objects. The visual surveillance and monitoring system takes object's dispersedness, area, apparent aspect ratio of bounding box as key features and classifies moving-objects into four classes: single human, vehicle, group of people and clutter [8]. Various types of learning algorithms have also been used for object classification problems. A simple yet effective classifier is based on modelling class-conditional probability densities as multivariate Gaussians [8, 9]. Other types of classification algorithms include support vector machines (SVMs) [10, 11], boosting [12], logistic linear classifiers [13], neural networks [14] and Bayesian using mixtures of Gaussians. The other factor affecting the performance of all these types of classification algorithms is the choice of features used for representing objects. Many possible features exist, including entire image [15], wavelet/rectangular filter outputs [16], shape and size [17], morphological features [18] and recurrent motion [19]. However, the features used for classification

purposes should be reliable, descriptive and representative for both static and dynamic properties of moving objects. In general, non-rigid articulated human motion shows a periodic property, and has been used as a strong cue for classification of moving objects. Based on this useful cue, human motion is distinguished from motion of other objects, such as vehicles. Cutler and Davis [19] described a similarity based technique to detect and analyse periodic motion. By tracking the moving object, its self-similarity is computed as it evolves over a period of time, which is the measure of periodicity. Therefore time–frequency analysis is applied to detect and characterise the periodic motion. Fujiyoshi *et al.* [20] proposed the use of a star skeletonisation procedure for analysing the motion of human targets however not for classifying the objects. Miller *et al.* [21] improved the SVM based person vehicle classification with automated parameter tuning and illumination change compensation. Later, Yuan and Yang [22] presented a star skeleton feature based SVM classifier to classify actions. Yu and Agarwal [23] developed variable star skeleton (VSS) to obtain the accurate extremities in star skeleton, which also classifies action. These techniques need to be trained via test sequences. Javed and Shah [17] produced an algorithm based on recurrent motion image (RMI) that does not need training sequences. The RMI method is one of the few approaches that produce a high recognition rate while remaining computational and space efficient. Wong and Ong [24, 25] have extended the RMI analysis based object classification work to special cases like, people walking with hands in their pockets, at the back or lifting or carrying things as they walk. Since recurrent motion is not always observable from human hands, it relies on hands movement as a cue to classify an object as human. Johnsen and Tews [4] later modified the RMI based classification further by including boundary features, it suffers under occlusion. Ryoo *et al.* [26] presented a methodology for analysing a sequence of scene states from videos of human–vehicle interactions; however it focuses on events occurring in a video to contextually coincide with scene states of humans and vehicles. HouLR-based method for people and vehicles classification [27] far distance scene surveillance with static camera has recently been cited. A feature driven method based on the difference of histogram of oriented gradients [28], object size, object velocity and location is also available to classify people and vehicles. However, their algorithm fails when part of the moving object just enters the region of interest (ROI)s, which makes the line character of objects unstable.

The challenging issues, towards the object classification still prevalent are, lack of performance of object classification techniques because of the presence of shadow and lack of invariance of the existing methods to scene-dependent parameters, such as position, orientation, illumination, scale or visibility, which prevent the widespread use of vision systems. Hence, the development of a practical object classification algorithm becomes a challenge. Although the existing RMI analysis solves the problem of object classification, it fails in the case of partial occlusion. In this paper, the objective is to develop an algorithm that could overcome the challenges such as shadow, scene independence and occlusion.

## 2 Proposed method

In this paper, a reliable system is proposed for person–vehicle classification, which works well in challenging real-world



**Fig. 1** Block diagram of the proposed object classification algorithm

conditions, including the presence of shadows and partial occlusion between objects. An overview of the proposed object classification method is shown in Fig. 1. In the first stage, the background is modelled using a Gaussian mixture model (GMM) and the foreground pixels are detected from the background model. Then, the foreground pixels are separated as object and rough shadow regions, using a projection histogram approach [28]. The maximal response for rough shadow pixels are extracted using a Gabor filter for different ranges of spatial frequency and orientation for some frames. In the next stage, these values of the shadow pixels are further refined through Gaussian shadow modelling. With these features, the proposed shadow modelling method can precisely model the shadow. After modelling a few initial frames, rough shadow pixels are removed using Gabor responses. Moving objects obtained from the detection phase are tracked using region correspondence. The descriptors such as centroid, bounding box and size are extracted from the blobs to achieve the correspondence of the blobs between frames. Correspondence between regions in current frame and previous frame is established using the minimum cost criteria to update the status of each object over the frames. Then, the star skeleton features are extracted from the silhouettes. The proposed skeleton feature based RMI is partitioned to the top and bottom sections. Based on the analysis of RMI on skeleton features, the moving objects detected and tracked in the image sequences are classified as ‘single person’, ‘group of people’ or ‘vehicle’. The classification rules are assigned only to the bottom section of the partitioned RMI which classifies the objects under partial occlusion.

### 2.1 Background subtraction and shadow removal

In this work, Gaussian modelling of shadow is developed by extracting maximal Gabor responses over rough shadow regions where the camera is assumed to be static. In the

first stage, the background is modelled using GMM and the foreground pixels are separated as object regions and rough shadow regions. Once the rough shadow region has been determined, the Gabor filter kernels are used to build a shadow model for the pixels. Since Gabor possesses optimal localisation properties in both spatial and frequency domains it is ideal for accurate texture identification of the surface which is related with shadow. Also, Gabor filter response can be represented as a sinusoidal plane of a particular frequency and orientation, modulated by a Gaussian envelope [29]. Considering the advantages of Gabor filters, which include the robustness to illumination change, multi-scale and multi-orientation nature, the following form of a normalised 2D Gabor filter function in the frequency domain is employed for the extraction of maximal response of spatial frequency and orientation for the shadow pixels as in (1) and (2)

$$\psi(x, y, f, \theta) = \exp\left[-\left(\frac{x^2 + y^2}{\sigma^2}\right)\right] \exp(2\pi f i x') \quad (1)$$

$$x' = x \cos \theta + y \sin \theta \quad (2)$$

where  $\psi$  is the Gabor kernel,  $f$  is the spatial frequency,  $\theta$  is the orientation and  $\sigma$  is the standard deviation of the Gaussian kernel and it depends upon the spatial frequency. The responses of convolutions in a shadow region at location  $(x, y)$  are given in spatial domain as in (3) and it represents important features of the shadow pixels

$$r_{\xi}(x, y, f, \theta) = \psi(x, y, f, \theta) \xi(x, y) \\ = \int_{-\alpha}^{\alpha} \int_{-\alpha}^{\alpha} \psi(x - x_x, y - y_x; f, \theta) \xi(x_x, y_x) dx_x dy_x \quad (3)$$

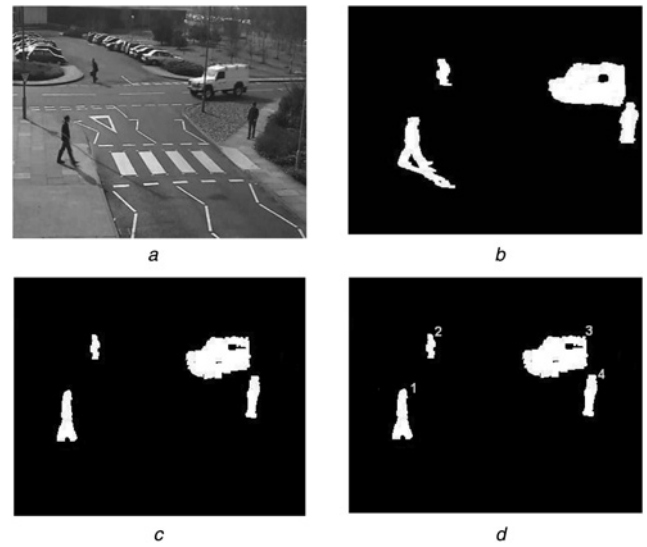
where  $\xi(x, y)$  is the shadow pixel in the rough shadow region and  $r_{\xi}(x, y, f, \theta)$  represents the Gabor response on that pixel. For each shadow pixel, Gabor filters with multi-scale and multi-orientation are first used to extract Gabor features. Here, the magnitude of complex outputs of Gabor convolutions are used as features.

The maximal response of the pixels in the shadow region is obtained by multiplying it with the Gabor kernel in frequency domain. These maximal Gabor responses are modelled as Gaussian since the original shadow data are random variables and subsequently these responses are also random and large in size. Once these responses are assumed as Gaussian, the likelihood is derived as in (4)

$$L_1(i, j) = \exp(-0.5(X - m_f)^T C^{-1}(Y - m_{\theta})) \quad (4)$$

where  $L_1(i, j)$  is the proposed Gaussian shadow model for the Gabor response of the shadow pixels,  $X$  and  $Y$  are the Gabor response obtained by varying the spatial frequency, ' $f$ ' and orientation  $\theta$ , respectively,  $m_f$  is the mean value of the Gabor filter response when  $\theta$  is kept constant,  $m_{\theta}$  is the mean value of the Gabor filter response when  $f$  is kept constant and  $C$  is the covariance of both responses.

**2.1.1 Shadow elimination by Gabor based shadow modelling:** To eliminate the shadows completely, the Gaussian model fitted earlier is used as a tool on the currently available pixels representing the moving object along with their shadow. These pixels are subjected to a Gabor filtering action as developed earlier in modelling



**Fig. 2** Background subtraction, shadow removal and motion correspondence

a Original image  
b Extracted foreground blobs  
c Shadow removed foreground blobs  
d Motion correspondence

with the filter function and the Gabor response is represented as  $F_g$ . The mean spatial frequency  $m_f$  and the mean orientation  $m_{\theta}$ , which were responsible for producing maxima during modelling are used to produce Gabor responses. The likelihood function of the response is given by (5)

$$L_2(i, j) = \exp(-0.5(F_g(i, j) - m_f)^T C^{-1}(F_g(i, j) - m_{\theta})) \quad (5)$$

where  $L_2(i, j)$  is the shadow likelihood. This likelihood is thresholded '(th<sub>s</sub>)' by the parameter mean '(μ)' of the Gaussian modelled Gabor response to ascertain whether the pixel belongs to shadow and hence to remove. The result of shadow removal by the likelihood estimation is shown in Figs. 2a–c.

## 2.2 Motion correspondence and object tracking

Correspondence between regions in current frame and previous frame is established using the minimum cost criteria [25] to update the status of each object over the frames. If a region's predicted position exceeds the frame boundary, the corresponding object is determined to have exited the surveillance area; otherwise, object occlusion may have happened. If an object's bounding box overlaps the bounding box of another region  $Q$  in the current frame,  $Q$  is marked as an occluded region, and all of the non-corresponded regions in the previous frame overlapping  $Q$  are, thus, marked as occluding each other. Lastly, a non-corresponding region in the current frame is set to be an object entry. Such motion correspondence of the objects is shown in Fig. 2d.

## 2.3 Skeleton feature extraction

A straightforward way of detecting only the gross extremities of the target to produce a star skeleton is the simple form of skeletonisation that extracts the broad internal motion features of a target to analyse the target's motion [20]. The

contour of a human blob is extracted and the centroid  $(x_c, y_c)$  of the human blob is determined by using the following (6) and (7)

$$x_c = \frac{1}{N} \sum_{j=1}^N x_i \quad (6)$$

$$y_c = \frac{1}{N} \sum_{j=1}^N y_i \quad (7)$$

where  $(x_c, y_c)$  represent the average contour pixel position,  $x_i$  and  $y_i$  represent the points on the human blob contour and  $N$  is the total number of contour points. The distance  $d_i$  from the centroid to contour points is given by (8)

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad (8)$$

The distance computed is shown in Fig. 3a, from which the local maxima (indicated in ‘\*’) are found to represent the extremities of the object under consideration.

The line joining the centroid and extreme points provides a star skeleton and is shown in Fig. 3b. The star skeleton provides a unique signature for a class of objects even for vehicles which are non-star shaped objects. It also reduces the noise for the splotchy motion blobs by smoothing the plot of distance from the centroid against contour points plot by moving average.

## 2.4 Determination of recurrent motion image

An improved recurrent motion-based object classification approach using star skeleton features is proposed to classify moving objects. Recurrent motion which is denoted as repetitive change in the shape of the objects is one of the essential features that differentiate the object classes. Recurrent motion will have high values at pixels where motion occurs repetitively and low values at pixels where little or no motion occurs. In most cases the whole object is moving in addition to local changes in shape, for example, when a person is walking. Hence, the compensation for the translation and scale of the object over time to detect the local change is essential. Translation is compensated by aligning the object in subsequent frames along its centroid. For compensation of scale, the object is scaled equally in horizontal and vertical directions such that its vertical

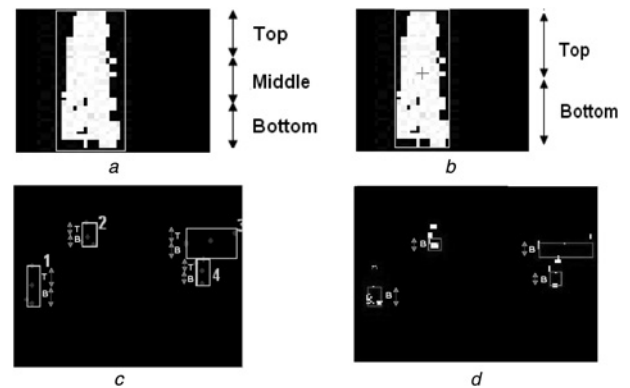
length, that is, projected height, is equal to its vertical length at the first observation. Here, the assumption that the only cause of change in the projected height of an object is the variation in the object’s distance from the camera. Once the objects are aligned the recurrent motion is used to determine the areas of silhouette undergoing repeated change.

Recurrent motion is computed using (9) and (10) to determine the areas of moving object’s silhouette undergoing repetitive changes.  $S_{a_0}$  is a binary silhouette for an object ‘ $a_0$ ’ at frame  $t$ , and  $DS_{a_0}$  is a binary image indicating areas of motion for object ‘ $a_0$ ’ between frame  $t - 1$  and  $t$ .  $RMI_{a_0}$  is the RMI for object ‘ $a_0$ ’, calculated over  $T_f$  frames. Subsequently, the RMI is partitioned into  $N_p$  equal-sized blocks in order to compute the average recurrence for each block. Each block belongs to top, middle or bottom section of the silhouette as shown in Fig. 4a. Blocks with average recurrence value greater than a threshold  $\tau_{RMI}$  are set to 1 (white) and vice versa. Hence, white blocks indicate image regions with high motion recurrence whereas black blocks indicate areas with insignificant or no motion recurrence.

$$DS_{a_0}(x, y, t) = S_{a_0}(x, y, t - 1) \oplus S_{a_0}(x, y, t) \quad (9)$$

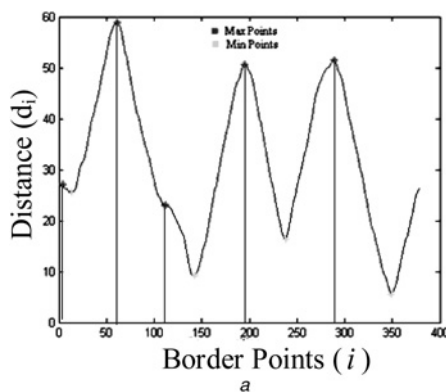
$$RMI_{a_0} = \sum_{k=0}^{T_f} DS_{a_0}(x, y, t - k) \quad (10)$$

where  $\oplus$  is the exclusive-OR operator.



**Fig. 4** Partitioned RMI analysis on star skeleton features

- a Partitioned RMI based on boundary features (top, middle, bottom)
- b Proposed partitioned RMI based on boundary features (top, bottom)
- c Proposed partitioned RMI analysis on star features (top, bottom)
- d Proposed partitioned RMI analysis on star skeleton features (bottom)



**Fig. 3** Skeleton feature extraction

- a Contour points against distance
- b Image with contour and skeleton feature



The white areas extracted from the RMI are matched against the templates stored in the knowledge base. Javed and Shah [17] presented the matching scheme for human search using white blocks (or recurrent motion) in the middle and bottom sections of the partitioned RMI. An object is classified as human (single person or group of people) when significant recurrent motion is present in the corresponding sections.

**2.4.1 RMI analysis on star skeleton features for object classification:** The earlier work on RMI based classification partitions the RMI into three regions and analyses for the recurrences of boundary features to classify actions, whereas the current work aims to analyse the recurrences of skeleton features by partitioning RMI into two with the focus on the bottom region alone. The intermediate result of Wong and Ong [25] and the proposed star skeleton RMI are shown in Figs. 4a–d. The matching scheme for classification suggested earlier works for common cases like human demonstrating periodic motion, with both hands and legs while walking. This may be insufficient when human hand motion is restricted while the hands are in their pockets, at the back or lifting/carrying some objects as they walk where it leads to the absence of recurrent motion in the middle section.

To overcome such issues, the classification rules of humans have been separated into two sets – generic case (where a walking human demonstrates high motion recurrence at the hands and legs) and special case (where a walking human demonstrates high motion recurrence at the legs only, such as the walking humans with hands in the pockets). The special cases handled by the matching scheme proposed by Javed and Shah [17] failed and another matching scheme is used by Wong and Ong [25]. This scheme searches for particular parts of the human body with the most significant recurrent motion in the partitioned RMI. If white blocks are detected in the bottom section of the partitioned RMI of the object of interest, the object is classified as human since recurrent motion at the legs is always clearly seen from all human RMIs. However, the RMI using boundary features, which are described above and the matching scheme for both generic and special cases fail when objects are partially occluded.

The proposed algorithm classifies objects into vehicle and persons. Moving objects detected from image sequences are classified by analysing their periodic motion patterns, in the bottom section of the partitioned RMI based on skeleton features. The classification rule can be derived from the black area within the RMI of an object, which corresponds to the area where the object demonstrates no recurrent motion. As seen in the RMI from Fig. 3a humans generally do not show any recurrent motion in the main part of the body where the backbone is located. As a result, the rules for the proposed classification algorithm for humans are refined as

1. If there are white blocks in the bottom section of a partitioned RMI, it is considered to be as a ‘human’ (recurrent motion should be there if the person is moving).
2. The rule will also search for high recurrent motion at the bottom section (around the legs region), if a large number of white blocks are found at the bottom section of the partitioned RMI analysis of the star skeleton features, the object is classified as ‘group’.

Subsequently, when a moving object is classified as ‘human’, it will be further categorised as a single person or a group of persons based on the following rules:

1. It counts the number of centroids present in the partitioned RMI of star skeleton features.
2. According to the count of centroids the number of persons can be estimated.

Finally, if there are no white blocks in a partitioned RMI analysis of star skeleton features in the bottom side, which indicates no recurrent motion, the corresponding object is classified as vehicle. In order to evaluate the proposed method, the precision and recall rate are calculated using (11) and (12).

$$P = \frac{TP}{TP + FP} \times 100 \quad (11)$$

$$R = \frac{TP}{TP + FN} \times 100 \quad (12)$$

where TP is the true positive, FN is the false negative and FP is the false positive. Also the *F*-measure based on precision and recall is also used to evaluate the accuracy. The *F*-measure is given by (13)

$$F\_Measure = 2 \left( \frac{(Precision \times Recall)}{(Precision + Recall)} \right) \quad (13)$$

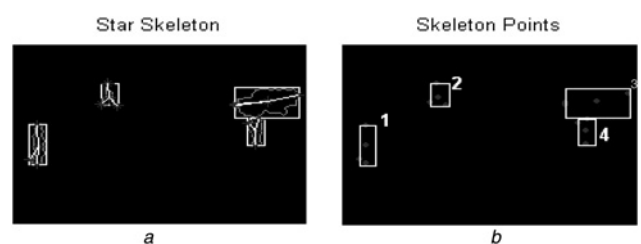
### 3 Results and discussion

To demonstrate the effectiveness of the proposed framework, the object classification experiment has been carried out on a variety of benchmark video datasets such as PETS 2001 [http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html], Challenge [http://gps-tsc.upc.es/imatge/\_j/Tracking/



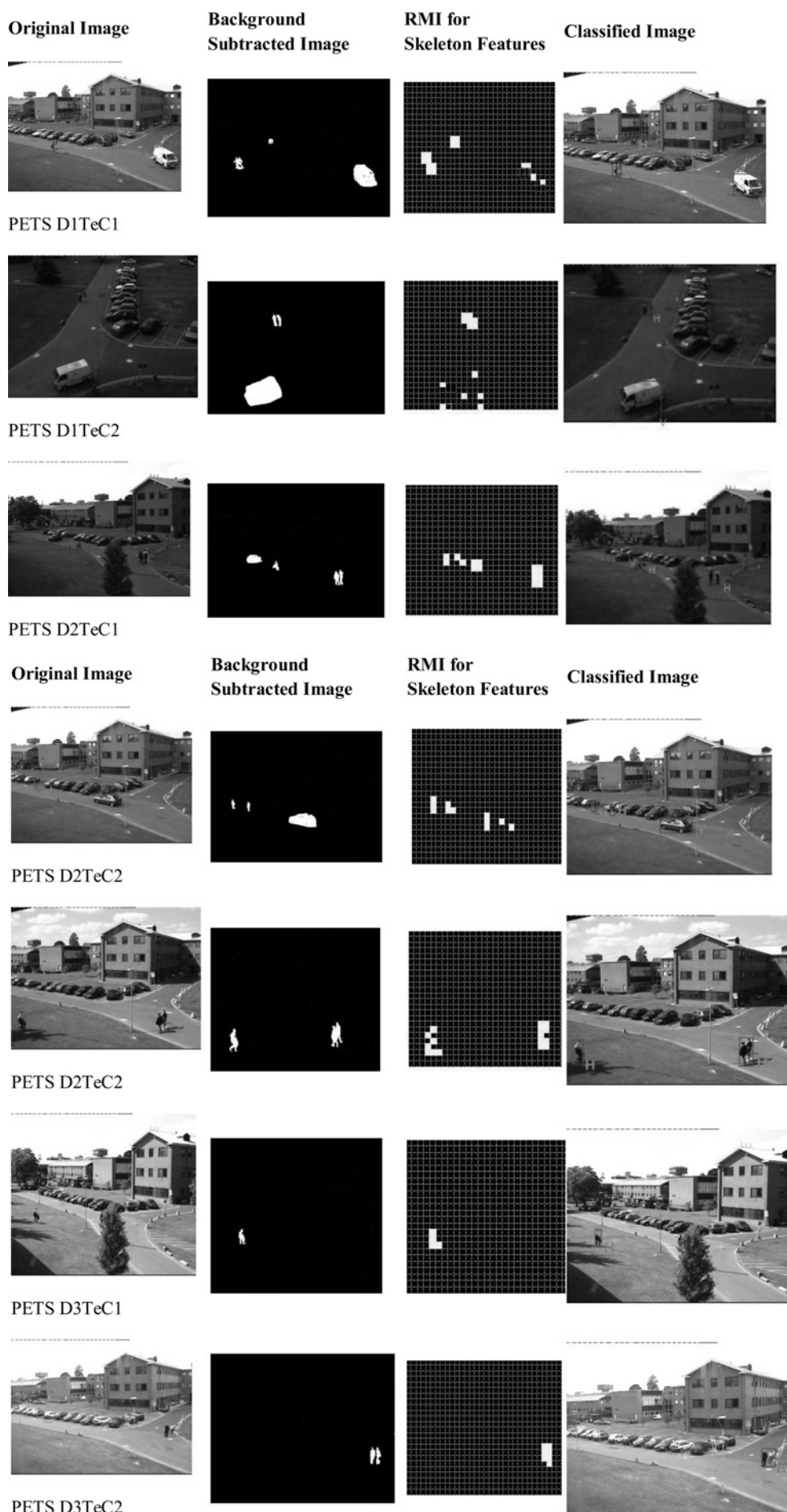
**Fig. 5** Background subtraction and shadow removal

- a Extracted foreground blobs
- b Shadow removed foreground blobs with motion correspondence



**Fig. 6** Partitioned RMI computation using skeleton features

- a Object boundary with skeleton feature extraction
- b Translation and scale compensated RMI



**Fig. 7** Benchmark results for object classification

**Table 1** Classification results for benchmark datasets

Videos	Ground truth		Classification based on partitioned RMI using boundary		Classification based on RMI using skeleton	
	Number of people	Number of vehicles	Number of people	Number of vehicles	Number of people	Number of vehicles
PETS D1TeC1	6	2	5	3	6	2
PETS D1TeC2	5	2	6	1	5	2
PETS D2TeC1	6	3	8	1	7	2
PETS D2TeC2	7	2	8	1	8	1
PETS D3TeC1	16	0	16	0	16	0
PETS D3TeC2	13	0	13	0	13	0
CHALLENGEDATASET	4	3	6	1	4	3

[Challenge.avi](#)], comprising the presence of vehicle, human in differing views covering a wide surveillance area. The PETS 2001 dataset is of 15 frames/s and each frame contains  $768 \times 576$  pixels and has more number of people than vehicles. The ‘Challenge’ dataset contains more vehicles since they are mostly recorded from street scenes. The testing videos are of the length taken with range starting from 5 to 15 min long. They are resized to the image dimension of  $320 \times 240$ . The image sequences consist of a variety of objects like single person, group of persons and vehicles.

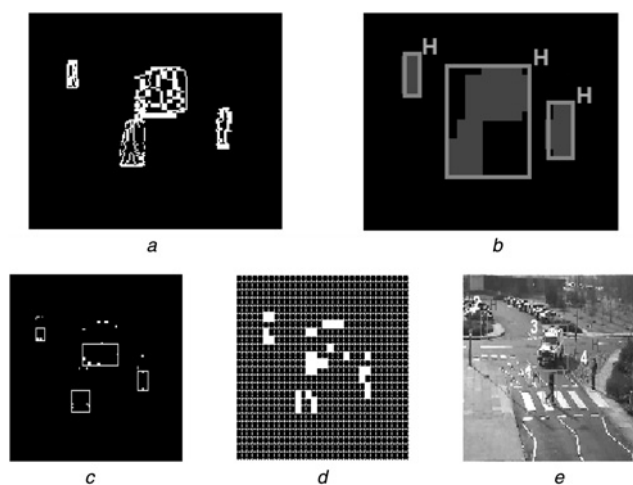
In the first step of the object classification algorithm, the Gabor response based Gaussian shadow modelling algorithm is used to obtain the foreground. First, the influence of Gabor filter response has been analysed with the best values of frequencies and orientations for shadow modelling. Gabor kernels having seven different frequencies and eight different orientations are used for the analysis. Experiments are carried out to show independent contributions of each frequency and orientation on the shadow modelling problem. It could be observed that the best frequency and orientation values for the particular shadow models are obtained using the responses obtained by varying  $\theta$  and  $f$ . Maximal response for different frequencies ranges 2, 4, 8, 16, 32, 64, 128 and for different

orientations 0,  $\pi/8$ ,  $2\pi/8$ ,  $3\pi/8$ ,  $4\pi/8$ ,  $5\pi/8$ ,  $6\pi/8$ ,  $7\pi/8$  with  $\sigma$  is  $0.65f$  are observed for each shadow pixel. Thus, by using the most important subset of the frequency and orientation parameters, one can speed up the shadow modelling module. In the proposed work, the initial frames considered for shadow modelling are 20. For each shadow pixel, maximal Gabor filter response is obtained by varying spatial frequencies at constant orientation and also by varying orientation at constant spatial frequency. The mean spatial frequency and orientation which give maximal response are  $m_f$ ,  $m_\theta$ , and the values for the example dataset are 2 and 1.462. The mean spatial frequency  $m_f$  and the mean orientation  $m_\theta$  which were responsible for producing maxima during modelling are used to produce Gabor responses and the values are 22.5265 and 33.9176. The covariance matrix  $C$  of the Gaussian likelihood is

$$\begin{bmatrix} 178.66 & 106.25 \\ 106.25 & 198.41 \end{bmatrix}$$

After this, the foreground pixel in the current frame is convolved with Gabor and the response is subjected to a likelihood estimation resulting in a likelihood value. This shadow likelihood is thresholded to decide whether the current pixel is shadow or not by the threshold  $th_s = 0.039$ , which is obtained from the mean value of the Gabor response. The simulation results on ‘Challenge’ dataset have shown that the chosen shadow model can model the shadows and can be eliminated and are illustrated in [Figs. 5a](#) and [b](#). The motion correspondence of the objects is also shown in [Fig. 5b](#). Then, the contour of the objects is extracted. The centroid  $(x_c, y_c)$  of the objects is determined by using (6) and (7). From the  $d_i$  plot, the local maximum points are found using (8) and the star skeletonisation is formed as shown in [Fig. 6a](#).

For the experimentation, the computation of the proposed partitioned RMI of skeleton features for a moving object is carried out for every 30 consecutive frames  $T_f$  after the object has completely entered into the scene. They are compensated both in translation and scale as shown in [Fig. 6b](#) [25]. Then, the partitioned RMI is computed. For computation of recurrent motion, the camera views are equally partitioned into grids of blocks with a size of  $5 \times 5$ , and the distributions are established for each block. The blocks having the threshold ( $\tau_{RMI}$ ), that is having a number of white pixels greater than ten are considered for partitioning the RMI. Those blocks are padded with extra white pixels for better clarity; however, care must be taken


**Fig. 8** Classified moving objects under occlusion

- a Translation and scale compensated RMI on boundary features boundary
- b Partitioned RMI on boundary features boundary features (enlarged version)
- c Partitioned RMI analysis on star skeleton features
- d Enlarged version of RMI blocks
- e Object classification

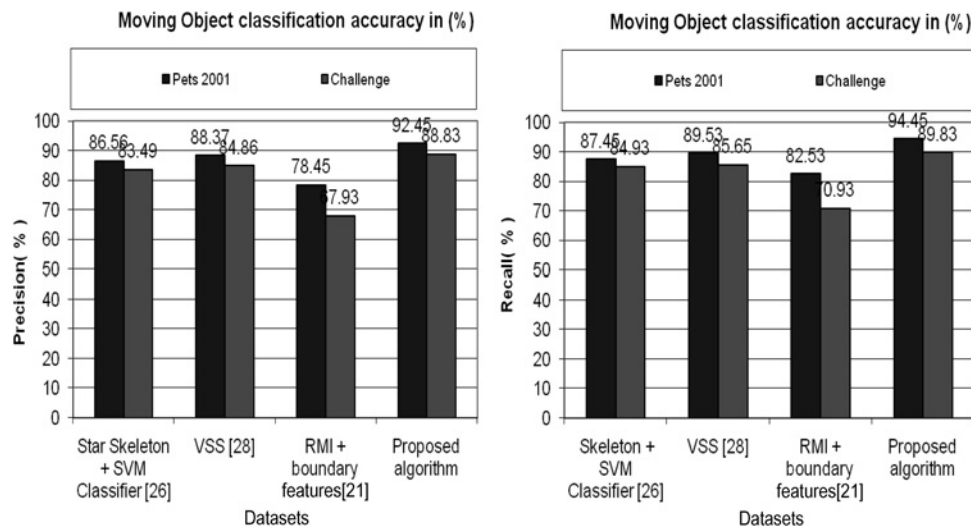


Fig. 9 Performance comparison of object classification methods

Table 2 Comparison of CPU time

Datasets	CPU elapsed time, s/frame			
	Star skeleton + SVM classifier [21]	VSS [22]	RMI analysis using boundary features [29]	Proposed RMI analysis using skeleton features
D1TeC1	2.1	2.3	1.9	1.2
D1TeC2	1.9	2.48	1.87	0.93
D2TeC1	1.93	2.54	1.89	0.98
D2TeC2	2.3	2.434	1.863	1.12
D3TeC1	2.53	2.65	1.83	0.97
D3TeC2	2.67	2.73	1.94	1.43
Challenge	2.63	2.654	1.893	0.98

that this padding should not increase the percentage error in the calculation of recurrent motion. Fig. 7 illustrates the object classification results on PETS 2001 datasets.

The challenging scenario shown in Fig. 7, such as a person with long shadow (PETS D2TeC2) and a group of people (PETS D1TeC1), are correctly classified as 'human' and 'group of people'. Here, it is assumed that the tracking results are perfect before performing object classification. Thus, if the background subtraction and/or tracking module fail to produce reliable results, the object classifier will also fail because of incorrect input data. Table 1 shows the classification performance of the proposed skeleton RMI method and boundary RMI method on standard datasets. They are validated against the ground truth in the video sample. In all the cases, the proposed skeleton RMI is found to correctly classify the people and vehicles although each video sample contains close to 1100 motions established by the objects present in the video.

In addition to that, the framework is applied to the image sequences where the objects are occluded from each other. Such foreground regions and the corresponding object's boundaries are shown in Figs. 8a and b. The translation and scale compensated RMI using boundary features are shown in Figs. 8a and b.

The occlusion between a vehicle and a group of persons was successfully handled by the proposed framework where the RMI based on boundary fails to classify. Since, in partitioned RMI analysis on boundary features, the

recurrence of boundary of the moving objects (here vehicle and the man) is merged and is shown in Fig. 8b. Hence, it is considered as a single object and misclassified as 'human'. However, in the proposed work, the partitioned RMI analysis on skeleton features does not merge with each other and classified the objects as 'human' and 'vehicle'. They are shown in Figs. 8c–e. Hence, in the proposed work, the classification rules using skeleton features are appropriate for the classification of human and vehicle even when the moving objects are occluded.

The performance measures precision and recall are used to validate the proposed work on PETS 2001 and challenge datasets and are shown in Fig. 9. The accuracy of the proposed partitioned RMI analysis on skeleton features for object classification is higher than the others such as star skeleton with SVM classifier [21], VSS [23] and RMI on boundary features [25]. Based on precision and recall of the proposed work the *F*-measure for PETS 2001 and Challenge datasets are 0.94 and 0.90, which is closer to one. Also, it indicates the compatibility and successful integration of the classification list (single person, group of persons) in this frame work. The modified object detection using the proposed shadow removal technique based on Gabor filter response on shadow pixels and classification algorithms enhance the recognition accuracy and practicability of the proposed framework. However, there was one misclassified sample in a cycling man classified as human. Owing to cycling, the recurrent motion is high but



not like a vehicle. Hence, the high recurrent motion in the bottom side of the partitioned RMI misclassified the cycling man as a 'human'. The skeleton feature based RMI algorithm is faster than its counterparts and takes an average of 1.08 s to classify the moving objects.

The comparison of execution time for various algorithms for object classification is employed and shown in Table 2. It depicts that the proposed RMI analysis of skeleton features has the capacity of working in real-time environment.

## 4 Conclusion

Real-world automatic video-surveillance presents the conditions that make the design of an efficient multi-class object classification module a demanding challenge. However, most algorithms fail to satisfy the requirements of real-time operation, flexibility, cost effectiveness and robustness to varying environmental conditions. In this proposed work, a novel approach for eliminating shadows from a static background using a Gabor filter based Gaussian shadow modelling technique is proposed. After modelling a few initial frames rough shadow responses with Gabor, the shadows thereafter are removed. Blobs obtained from the detection phase are tracked using region correspondence. The descriptors such as centroid, bounding box, size and velocity are extracted from the blobs to achieve the correspondence of the blobs between frames. Correspondence between regions in previous frame and current frame is established using the minimum cost criteria to update the status of each object over the frames. Then, the star skeleton features are extracted from the silhouettes. Based on the features, each of the moving objects detected and tracked in the image sequences are classified as a single person, group of persons or a vehicle using RMI. Different types of objects yield very different RMIs and therefore can easily be classified into different categories on the basis of their RMI. The partitioned RMI based on star skeleton features classifies the objects even during occlusion. The rules for the classification of objects using partitioned RMI have been proposed based on skeleton features. They are suitable even when the moving objects are occluded. In future, the classification frame work can be extended for sub-classes (e.g. cars, vans and trucks). Classification will certainly benefit from improvement in detection and tracking. Best possible results will probably be obtained by integrating object tracking and classification, whereby the knowledge of object class is fed back into the tracking system to help locate the object in the next frame.

## 5 References

- Hu, W., Tan, T., Wang, L., Maybank, S.: 'A survey on visual surveillance of object motion and behaviors', *IEEE Trans. Syst. Man Cybern. C, Appl. Rev.*, 2004, **34**, (3), pp. 334–352
- Wang, L., Hu, W., Tan, T.: 'Recent developments in human motion analysis', *Pattern Recognit.*, 2003, **36**, (3), pp. 585–601
- Zhang, L., Li, S.Z., Yuan, X., Xiang, S.: 'Real-time object classification in video surveillance based on appearance learning'. Proc. IEEE Conf. CVPR, 2007, pp. 1–8
- Johnsen, S., Tews, A.: 'Real-time object tracking and classification using a static camera'. Proc. IEEE ICRA 2009 Workshop on People Detection and Tracking, Kobe, Japan, May 2009
- Stauffer, C., Grimson, W.E.L.: 'Learning patterns of activity using real-time tracking', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (8), pp. 747–757
- Yang, T., Li, S.Z., Pan, Q., Li, J.: 'Real-time multiple objects tracking with occlusion handling in dynamic scenes', *Proc. IEEE Comput. Soc. Comput. Vis. Pattern Recognit.*, 2005, **1**, pp. 970–975
- Cucchiara, R., Grana, C., Piccardi, M., Prati, A.: 'Detecting moving objects, ghosts and shadows in video streams', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2003, **10**, pp. 1337–1342
- Kanade, T., Collins, R., Lipton, A., Burt, P., Wixson, L.: 'Advances in cooperative multi-sensor video surveillance'. Darpa Image Understanding Workshop, November 1998, vol. 1, pp. 3–24
- Lipton, A.J., Fujiyoshi, H., Patil, R.S.: 'Moving target classification and tracking from real-time video'. Proc. IEEE Workshop on Applications of Computer Vision, 1998, pp. 8–14
- Rivlin, E., Rudzsky, M., Goldenberg, R., Bogomolov, U., Lepchev, S.: 'A real-time system for classification moving objects'. 16th Int. Conf. on Pattern Recognition, Quebec City, August 2002, vol. 3, pp. 688–691
- Song, Z., Chen, Q., Huang, Z., Hua, Y., Yan, S.: 'Contextualizing object detection and classification'. IEEE Int. Conf. on Computer Vision and Pattern Recognition, 2011
- Levin, A., Viola, P., Freund, Y.: 'Unsupervised improvement of visual detectors using co-training'. Proc. Int. Conf. on Computer Vision, 2003, vol. 1, pp. 626–633
- Diehl, C.P.: 'Towards efficient collaborative classification for distributed video surveillance'. PhD thesis, Carnegie Mellon University, 2000
- Hollis, J.E.L., Brown, D.J., Luckraft, I.C., Gent, C.R.: 'Feature vectors for road vehicle scene classification', *Neural Netw.*, 1996, **9**, (2), pp. 337–344
- Pontil, M., Verri, A.: 'Support vector machines for 3D object recognition', *Proc. IEEE Trans. Pattern Anal. Mach. Intell.*, 1998, **20**, (6), pp. 637–646
- Viola, P., Jones, M., Snow, D.: 'Detecting pedestrians using patterns of motion and appearance'. Proc. Int. Conf. on Computer Vision, vol. 2, pp. 734–741
- Javed, O., Shah, M.: 'Tracking and object classification for automated surveillance'. Proc. Seventh European Conf. on Computer Vision – Part IV, 2002, pp. 343–357
- Foresti, G.L.: 'Object recognition and tracking for remote video surveillance', *IEEE Trans. Pattern Anal. Mach. Intell.*, 1999, **9**, (7), pp. 1045–1062
- Cutler, R., Davis, L.: 'Robust real-time periodic motion detection, analysis, and applications', *IEEE Trans. Pattern Anal. Mach. Intell.*, 2000, **22**, (8), pp. 781–796
- Fujiyoshi, H., Lipton, A.J., Kanade, T.: 'Real-time human motion analysis by image skeletonization', *IEICE Trans. Inf. Sys.*, 2004, **E87-D**, (1), pp. 113–120
- Miller, A., Basharat, A., White, B., Liu, J., Shah, M.: 'Person and vehicle tracking in surveillance video'. *Multimodal Technologies for Perception of Humans*, 2008, **4625**, pp. 174–178
- Yuan, X., Yang, X.: 'A robust human action recognition system using single camera'. Proc. Int. Conf. Computational Intelligence and Software Engineering, 11–13 December 2009, pp. 1–4
- Yu, E., Agarwal, J.K.: 'Human action recognition with extremities as semantic posture representation'. IEEE Computer Society Conf. on Computer Vision and Pattern Recognition Workshops, June 2009, pp. 1–8
- Wong, C.E., Ong, T.J.: 'A new RMI Frame work for outdoor objects recognition'. Proc. Int. Conf. on Advanced Computer Control, 2009, pp. 555–559
- Wong, C.E., Ong, T.J.: 'An improved recurrent motion image frame work for outdoor objects recognition'. Proc. Int. MultiConf. on Engineers and Computer Scientists, 2009, vol. 1
- Ryoo, M.S., Lee, J.T., Aggarwal, J.K.: 'Video scene analysis of interactions between humans and vehicles using event context'. ACM Int. Conf. on Image and Video Retrieval, July 2010
- Xu, T., Liu, H., Qian, Y., Zhang, H.: 'A novel method for people and vehicle classification based on hough line feature'. Int. Conf. on Information Science and Technology, Nanjing, Jiangsu, China, 26–28 March 2011
- Ibrahim, M.M., Anupama, R.: 'Scene adaptive shadow detection algorithm'. World academy of science, Engineering and Technology, 2005, vol. 42, pp. 88–91
- Caleanu, C., Huang, D., Gui, V., Taponut, V., Maranescu, V.: 'Interest operator versus Gabor filtering for facial imagery classification', *Pattern Recognit. Lett.*, 2007, **28**, pp. 950–956