An Improved Recurrent Motion Image Framework for Outdoor Objects Recognition

Chuan Ern Wong and Teong Joo Ong

Abstract—**In this paper, we present an extension to the Recurrent Motion Image (RMI) motion-based object recognition framework for use in development of automated video surveillance systems. We extended the original object classes of RMI to include four-legged animals (such as dog and cat). Various enhancements are made to the object detection and classification algorithms for better object segmentation, error tolerance and wider range of recognition. Under the new framework, object blobs obtained from background subtraction of scenes are tracked using region correspondence. In turn, we calculate the RMI signatures based on the silhouettes of the object blobs for proper classification. This new framework is tested on several real world 320 x 240 resolution color image sequences captured with a low-end digital camera, and also on the PETS 2001 dataset. A recognition rate of approximately 98 percent (39 out of 40 moving objects in the experiments were correctly classified) was achieved, indicating the applicability of the new framework in similar task environment.**

Index Terms—**Moving object recognition, object classification, recurrent motion.**

I. Introduction

Moving object recognition has been an active area of research for computer vision and pattern analysis applications. It plays a major role in advanced security systems and video surveillance applications - to recognize moving objects through a video monitoring system and generate appropriate alert for other parts of the system to respond to the situation. Thus, an object recognition algorithm should be able to detect and track moving objects within a surveillance area in real time, and classify objects of interest into various predefined categories efficiently.

Object recognition function enhances the feature sets and functions of security systems and surveillance applications. For instance, an intruder recognition function can be incorporated into a security system to categorize intruders into various threat levels to reduce nuisance alarm and minimize human errors in manned surveillance system. Analogously, the recognition function may also be used in traffic monitoring system to estimate traffic flow by making vehicle and pedestrian counts.

Manuscript received December 6, 2008.
C. E. Wong is with the Faculty of Information and Communication Technology, Universiti Tunku Abdul Rahman, Petaling Jaya, 46200 Selangor, Malaysia (phone: 603-79551511; fax: 603-79551611; e-mail: chuanern@hotmail.com).
T. J. Ong is with the School of Computer Technology, Sunway University College, Petaling Jaya, 46150 Selangor, Malaysia (e-mail: teongjoo@gmail.com).

This paper presents an improved motion-based recognition approach using a specific feature vector called Recurrent Motion Image (RMI) [1] to classify moving objects into predefined categories, namely single person, group of persons, vehicle and four-legged animal (dog or cat in this case). Moving objects detected from image sequences are classified based on their periodic motion patterns captured with the RMI.

Preprocessing and complex scene handling routines (such as removal and processing of noise and object occlusion) can improve the accuracy of the recognition system. Hence, various refinements are added to the new framework to improve its recognition accuracy and compatibility for use in general outdoor scenes.

II. Background

Extensive research efforts have been dedicated to moving object recognition, where many approaches, such as [1]-[4] have been presented to tackle this problem.

RMI method [1] is one of the few approaches that produce high recognition rate while remaining computationally and space efficient. A specific feature vector called RMI was proposed to estimate repetitive motion behavior of moving objects. Different object has different motion behavior yielding different RMI. Thus, moving objects can be classified as single person, group of persons or vehicle based on their corresponding RMI, as shown in Fig. 1.

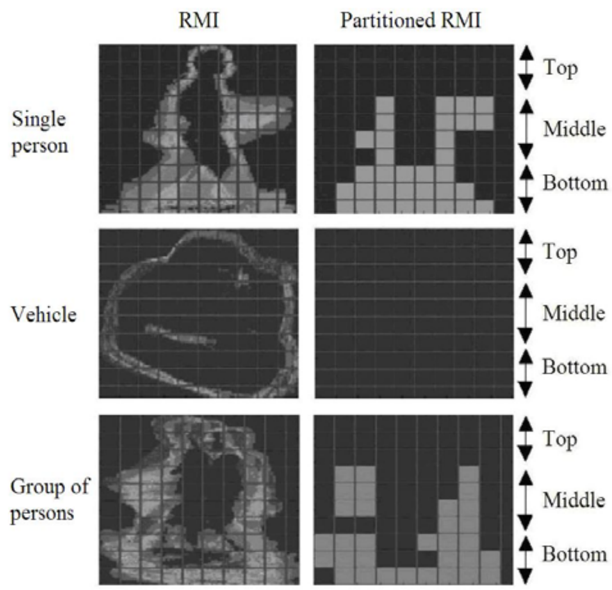


Fig. 1. RMI for classification [1]

This approach starts with background subtraction and shadow removal, followed by region-based tracking to establish motion correspondence. Repetitive changes in the shape of object yield recurrent motion behavior which is used to generate the RMI. The areas of RMI demonstrating high motion recurrence will be used to determine the object's class. For example, the RMI of a walking human has high recurrence near the hands and legs, whereas the RMI of a moving vehicle shows no motion recurrence.

Experiments conducted in [1] indicate that this approach yields correct classification in about 97 percent of all tested samples. However, the shadow removal algorithm in the original framework suffered an error rate of 30 percent due to segmentation failure. The segmentation algorithm failed to divide cast shadows and self shadows in different regions. Besides, the framework has only been tested on a small set of object classes (human and vehicle).

In addition, the error tolerance of the original algorithm is low since it is unable to accommodate slight deviations in the RMI data. For instance, a person who walks with hands in the pockets will not be recognized as a human because the resultant RMI does not exhibit significant hand movements. The person is categorized as other object since the RMI does not match any of the predefined classes. Such limitations, in essence, confine the recognition range and accuracy of the framework.

The approach presented in this paper consists of various refinements we have made to the RMI framework to increase its recognition rate and error tolerance for outdoor settings and complex scenes.

III. Methods

A. Object Detection

Before we can classify moving objects accordingly, we have to extract the object silhouettes from image sequences in order to generate their respective RMI. Previous work [1] uses a mixture of K Gaussian distributions [5] to perform background subtraction, followed by connected components labeling [6] to segment the foreground pixels into regions. A combination of color segmentation using K-means approximation of the EM algorithm and gradient direction [1] is used to identify and remove shadows. However, it was shown in their results that the shadow removal process failed in about 30 percent of the frames that contain significant shadows. The errors were caused by failure to divide cast shadows and self shadows in different regions.

Our framework extends the preprocessing stages from [1] to include a better shadow removal algorithm and multiple levels of noise filtering for better moving object segmentation, as illustrated in Fig. 2. Firstly, background subtraction is carried out by computing an L-inf distance image [7] in the Red-Green-Blue (RGB) color space. Foreground points are obtained by applying a low threshold (0.08 in our experiments) to the L-inf distance image, and these points will go through a morphological opening [8] denoise layer before shadow removal.

Shadow points are located by transforming the image pixel values to the Hue-Saturation-Value (HSV) color space [9] and removed from the foreground points. Foreground blobs are extracted using connected components labeling and subsequently blob analysis is performed to filter noise clutters using a blob size threshold.

Lastly, a high threshold (0.4 in our experiments) is applied to the L-inf distance image to select points with large difference from the background. Blobs consisting of at least one of these salient points are validated whereas the others are removed as non-salient blobs.

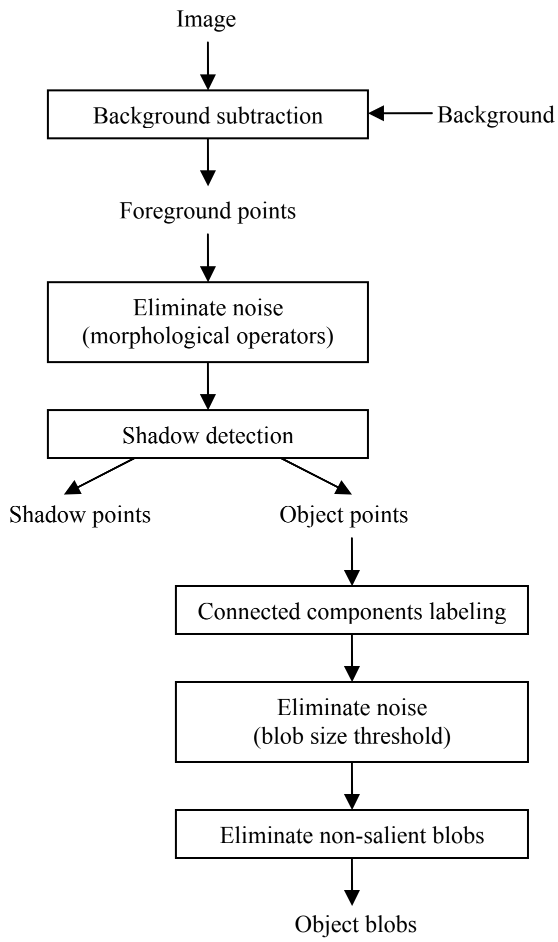


Fig. 2. Object detection algorithm

B. Object Tracking

Blobs obtained from the detection phase are tracked using region correspondence [10]. Various parameters and descriptors (such as centroid, bounding box, size, velocity and change in size of each blob) are extracted from the blobs. Correspondences between regions in previous frame and current frame are established using the minimum cost criteria [1] to update the status of each object over the frames.

As shown in Fig. 3, there might be non-corresponded regions in the previous and current frames. Since object exit or occlusion events may be associated to some of the regions in the previous frame, they must be examined based on the following rules:

1) 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.
2) 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 previous frame overlapping Q are, thus, marked as occluding each other.

3) Lastly, non-corresponded region in the current frame is set to be an object entry.

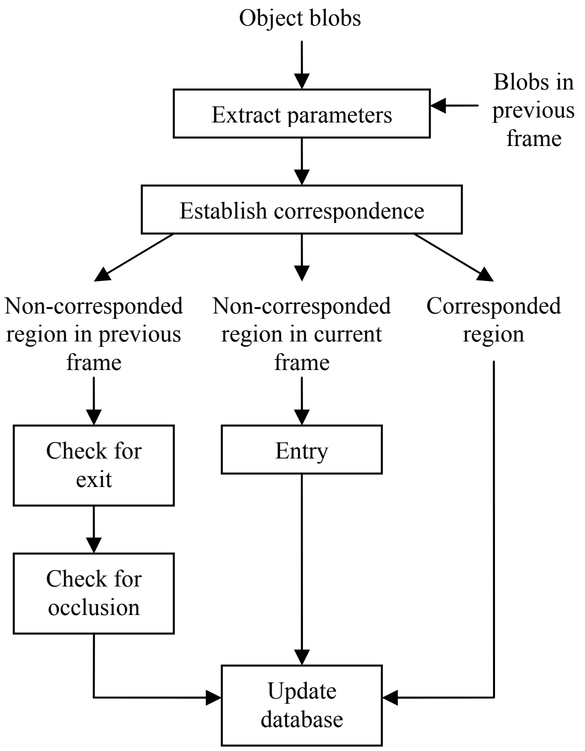


Fig. 3. Object tracking algorithm

C. Object Classification

Each of the moving objects detected and tracked in the image sequences are classified as a single person, a group of persons, a vehicle or a four-legged animal. Recurrent motion which is denoted as repetitive changes in shape of the objects is the main essential feature that differentiates the object classes. RMI will have high values at pixels where motion occurred repetitively and low values at pixels where little or no motion occurred.

$$DS_a(x,y,t) = S_a(x,y,t-1) \oplus S_a(x,y,t) \quad (1)$$

$$RMI_a = \sum_{k=0}^{T} DS_a(x,y,t-k) \quad (2)$$

RMI is computed with (1) and (2) to determine the areas of moving object's silhouette undergoing repetitive changes. S_a is a binary silhouette for object *a* at frame *t*, and DS_a is a binary image indicating areas of motion for object *a* between frame *t* and *t*-1. RMI_a is the RMI for object *a* calculated over *T* frames. Subsequently, the RMI is partitioned into N equal-sized blocks in order to compute the average recurrence for each block. Blocks with average recurrence value greater than a threshold τ_{RMI} are set to 1 (white) and vice versa. Hence, white blocks indicate image regions with high motion recurrence whereas black blocks indicate the areas with insignificant or no motion recurrence.

The white areas extracted from the RMI are matched against templates stored in the knowledge base. In [1], [11], the matching scheme for human search for 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 persons) when significant recurrent motion is present in the corresponding sections.

In our experiments, we discovered that the matching scheme is only true for common cases, when the humans demonstrate periodic motion at both the hands and legs while walking. The matching scheme is insufficient to account for cases when the humans are walking with hands in their pockets, at the back, or lifting or carrying things as they walk. Fig. 4 and Fig. 5 illustrate the absence of white blocks in the middle section of partitioned RMI because no periodic motion is demonstrated by the hands of the person. Since recurrent motion is not always observable from human hands, it is insufficient to rely on hands movement as a cue to classify an object as human.

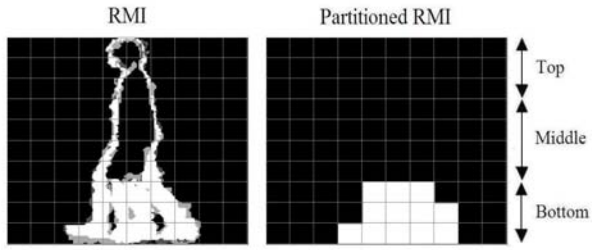


Fig. 4. RMI of a walking human with hands in pockets

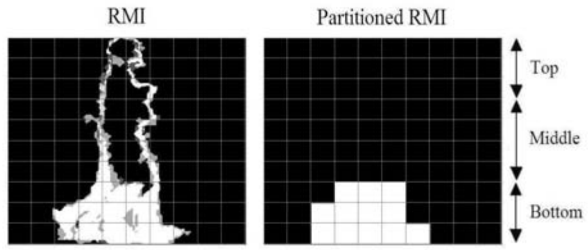


Fig. 5. RMI of a walking human with hands carrying things in front

To overcome such issues, we separated the classification rules of human into two sets - generic case (where a walking human demonstrates high motion recurrence at the hands and legs, as shown in Fig. 6) and special case (where a walking human demonstrates high motion recurrence at the legs only, such as the walking humans in Fig. 4 and Fig. 5).

The generic cases can be handled by the matching scheme proposed in [1], whereas special cases are handled by noting that the RMIs in Fig. 4 and Fig. 5 demonstrated high motion recurrence near the legs region, as evident from high concentration of white blocks at the bottom section of the partitioned RMI. Therefore, to account for the special cases when the matching rules in [1] failed, a second set of matching scheme should search for region with the most significant recurrent motion. If white blocks are detected around the legs (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.

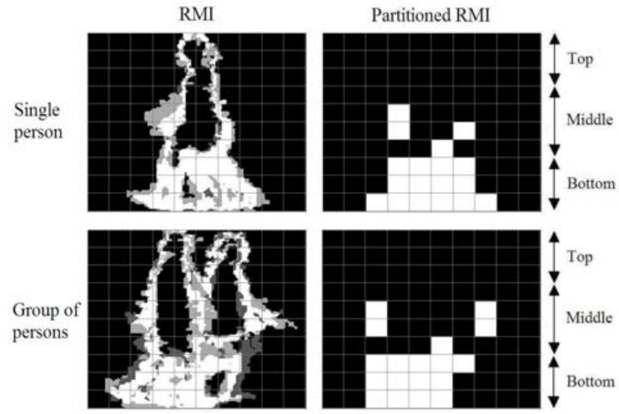


Fig. 6. RMI of a single person and a group of persons

Analogously, derivation of the criteria for classifying a moving object as a four-legged animal is based on the following observations:

1) The RMIs of dog and cat (Fig. 7) reminisces each other, whereby their legs and tail demonstrated repetitive motion. White blocks tend to occur in the top, middle and bottom sections of the resultant partitioned RMIs.
2) As for dogs and cats without tail, the white blocks are observable only in the middle and bottom sections.
3) Lastly, dogs and cats which are short in height but long in length may cause white blocks to occur only at the middle section, whereas dogs and cats that are long in height but short in length may cause white blocks to occur only at the bottom section.

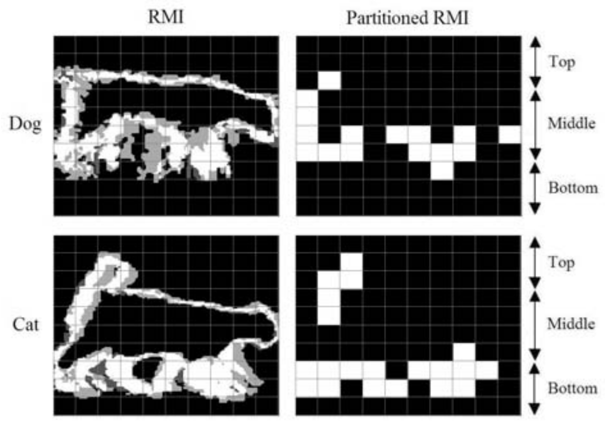


Fig. 7. RMI of a dog and a cat

From the observations, we noted that the location of white blocks in the partitioned RMI for dogs and cats may differ on their size, and whether or not the tail is present. Nonetheless, the partitioned RMI is similar to a human's where the middle or bottom section contains numerous white blocks – indicating that distinct classification criteria should be defined to allow proper differentiation between the RMI of four-legged animal and human.

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. 4, Fig. 5, Fig. 6, and Fig. 7, human and four-legged animals generally do not show any recurrent motion at the main part of the body where the backbone is located. Furthermore, the black area within the RMI of a single person or a group of persons has a vertically aligned major axis, whereas the black area within the RMI of a dog or a cat has a horizontal major axis. Therefore, the alignment of the major axis serves as our cue to differentiate between human and four-legged animal from their RMI.

As a result, the rules used by the classification algorithm for human is refined to be:

1) If there are white blocks in the middle or bottom section of a partitioned RMI, the black area within the respective RMI is extracted. If the black area has a vertical major axis, the corresponding object is classified as a human.
2) Since rule 1 may yield inconclusive results (special case), the rules will also search for high recurrent motion at the bottom section (around the legs region) while utilizing specific matching scheme. If large number of white blocks are found at the bottom section of partitioned RMI, the object is classified as a human.

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

1) Multiple peak points in a silhouette indicate more than one headcount, therefore representing a group of persons, for instance there are 2 peak points in the group of persons in Fig. 6 since there are 2 headcounts.
2) Normalized area of recurrence response at the top section of RMI for a group of persons is greater than that for a single person, due to presence of multiple heads.
3) If rule 1 and rule 2 failed, the object is classified as a single person.

Lastly, if there are no white blocks in a partitioned RMI, which indicates no recurrent motion, the corresponding object is classified as a vehicle, as shown in Fig. 8. An object that does not fall into any of the predefined categories (vehicle, single person, group of persons and four-legged animal) will be classified as other object.

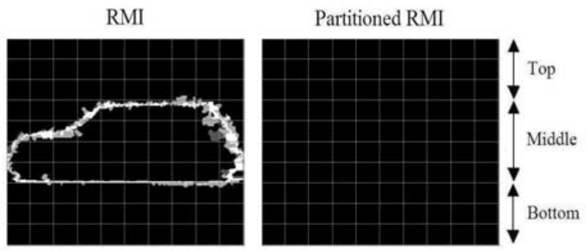
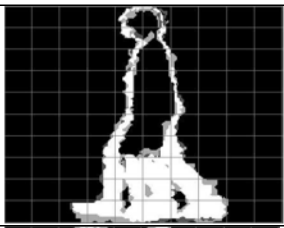
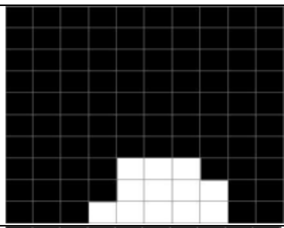
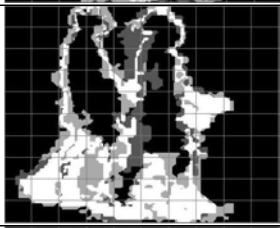
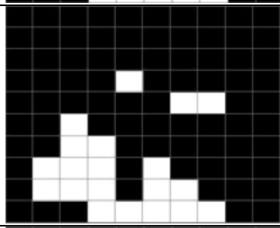
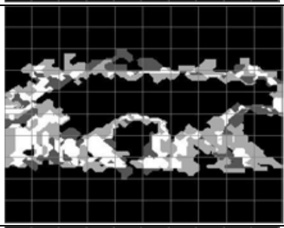
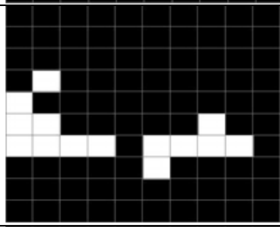
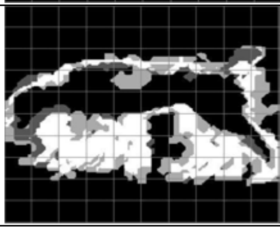
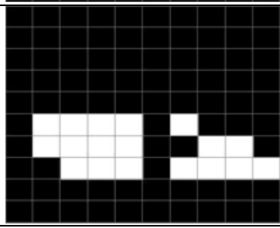


Fig. 8. RMI of a vehicle

IV. Results

The moving object detection, tracking and classification algorithm is implemented in Matlab [12] and executed on a 1.5GHz Core 2 Duo CPU. We captured several image sequences with a low-end digital camera (Olympus FE-280) at various housing areas. The image sequences consist of a variety of single persons, groups of persons, vehicles, and four-legged animals (dogs and cats). The frames are 320 x 240 pixels in size and sampled at a rate of 8 frames per second. Table 1 shows several instances of moving object classified using the framework.

TABLE 1. EXAMPLES OF MOVING OBJECT DETECTED AND CLASSIFIED

Image frame	RMI	Partitioned RMI	Object class
			Single person
			Group of persons
			Vehicle
			Four-legged animal
			Four-legged animal

In our experiments, the RMI of a moving object was generated for one second duration after the object has completely entered the scene, and the partitioned RMI was computed with a threshold (τ_{RMI}) of 2. The framework took an average of 4 seconds to process the image sequence to produce the classification result of a moving object. All of the moving objects tested were correctly classified into the predefined categories. The classification results are summarized in Table 2.

TABLE 2. CLASSIFICATION RESULTS FOR OUR DATASETS

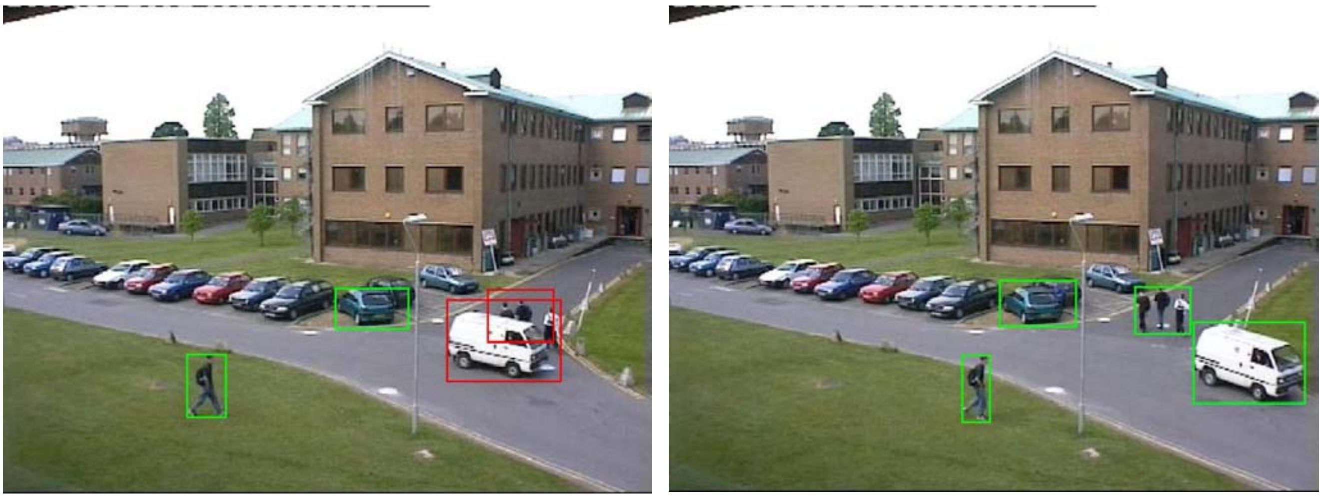
Object class	Number of samples tested	Number of samples correctly classified
Single person	10	10
Group of persons	5	5
Vehicle	6	6
Four-legged animal	9	9

In addition to the image sequences mentioned above, we also applied the framework on PETS 2001 dataset from the Second IEEE International Workshop on Performance Evaluation of Tracking and Surveillance [13]. The dataset consists of several 768 x 576 pixels image sequences of vehicles, single persons, and groups of persons at a wide surveillance area. Screenshots of the objects detected and tracked are shown in Fig. 9 where occlusion between a vehicle and a group of persons was successfully handled by our framework.

As listed in Table 3, all of the moving objects in the surveillance area were properly classified, except for a vehicle that was misclassified as a four-legged animal due to the size of the vehicle silhouette in the JPEG image sequence of the PETS 2001 image database. The silhouette of the misclassified vehicle is rough and poorly segmented by the algorithms - resulting in misclassification of the vehicle due to an inaccurate RMI.

TABLE 3. CLASSIFICATION RESULTS FOR PETS 2001 DATASET

Object class	Number of samples tested	Number of samples correctly classified
Single person	4	4
Group of persons	2	2
Vehicle	4	3



(a) (b)

Fig. 9. Screenshots of PETS 2001 dataset (a) during an occlusion – as shown in the red rectangles (b) after an occlusion

V. Conclusion

A total of 39 out of 40 objects were correctly classified in our experiments, indicating the backward compatibility and successful integration of the new classification list (single person, group of persons, vehicle and four-legged animal) in this framework. The modified preprocessing, object detection and classification algorithms also enhanced the recognition accuracy and practicability of the new framework.

Our proposed RMI classification method can be used as a filter that classifies moving objects into the proper categories for further processing in a recognition engine. For instance, differentiation between different species of four-legged animals after classification can be performed using texture, size, color and other relevant information of the species class.

However, there was one misclassified sample which was likely due to weak segmentation for objects in noisy image sequences. To circumvent such problem, smoothing and image enhancing routines may be applied to the rough silhouettes obtained from noisy images. Lastly, this framework can be further enhanced by incorporating texture and illumination analysis to improve the RMI extraction process for better recognition rate.

References

[1] O. Javed and M. Shah, "Tracking and Object Classification for Automated Surveillance," *Proc. 7th European Conf. Computer Vision-Part IV*, 2002, pp. 343-357.

[2] D. Toth and T. Aach, "Detection and Recognition of Moving Objects Using Statistical Motion Detection and Fourier Descriptors," *Proc. 12th Int'l Conf. Image Analysis and Processing*, 2003, pp. 430-435.

[3] J. B. Hayfron-Acquah, M. S. Nixon, and J. N. Carter, "Recognising Human and Animal Movement by Symmetry," *Proc. IEEE Int'l Conf. Image Processing*, 2001, pp. 290-293.

[4] Y. Bogomolov, G. Dror, S. Lapchev, E. Rivlin, and M. Rudzsky, "Classification of Moving Targets Based on Motion and Appearance," *Proc. British Machine Vision Conf.*, 2003, pp. 429-438.

[5] C. Stauffer and W. E. L. Grimson, "Learning Patterns of Activity Using Real-Time Tracking," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, 2000, pp. 747-757.

[6] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using Matlab*, Prentice Hall, Upper Saddle River, NJ, 2004, p. 359.

[7] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting Moving Objects, Ghosts and Shadows in Video Streams," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, 2003, pp. 1337-1342.

[8] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital Image Processing Using Matlab*, Prentice Hall, Upper Saddle River, NJ, 2004, p. 347.

[9] R. Cucchiara, C. Grana, M. Piccardi, A. Prati, and S. Sirotti, "Improving Shadow Suppression in Moving Object Detection with HSV Color Information," *Proc. IEEE Int'l Conf. Intelligent Transportation Systems*, 2001, pp. 334-339.

[10] B. Jahne and H. HauBecker, *Computer Vision and Applications: A Guide for Students and Practitioners*, Academic Press, San Diego, USA, 2000, p. 379.

[11] C. E. Wong and T. J. Ong, "A New RMI Framework for Outdoor Objects Recognition," *Proc. Int'l Conf. Advanced Computer Control*, to be published.

[12] The MathWorks, Inc., "MATLAB – The Language of Technical Computing." [Online]. Available: http://www.mathworks.com/products/matlab/

[13] The University of Reading, UK, "PETS2001 Datasets." [Online]. Available: http://ftp.pets.rdg.ac.uk/PETS2001/