

MOVING OBJECT RECOGNITION USING IMPROVED RMI METHOD

¹Wong Chuan Ern, ²Ong Teong Joo

Faculty of Information and Communication Technology
Universiti Tunku Abdul Rahman

No. 9, Jalan Bersatu 13/4, 46200 Petaling Jaya, Selangor, Malaysia
e-mail: ¹wongce1@mail2.utar.edu.my, ²ongtj@mail.utar.edu.my

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. RMI is a specific feature vector that calculates recurrent motion of moving objects, which is denoted as repetitive changes in silhouette of the objects. RMI approaches use this feature vector to differentiate and classify the moving objects into various categories based on different recurrent motion behavior. In addition to single person, group of persons and vehicle, we extended the object classes recognized by the RMI algorithm to include four-legged animals (such as dog and cat). Preprocessing and shadow removal algorithms are also refined to enhance moving object segmentation to achieve better recognition rate. Foreground points obtained from background subtraction in the RGB color space will now go through a shadow removal stage in the HSV color space and several layers of noise filter. Object blobs obtained after the aforementioned processes are then used to generate the corresponding RMI signatures. Finally, the recurrent motion behaviors of the RMI signatures are examined to classify the objects. The new framework with improved object detection and classification algorithm is implemented using Matlab and tested on several real world 320 x 240 resolution color image sequences captured with a low-end digital camera, Olympus FE-280. The RMI of one second duration are generated from the image sequences after the moving objects have completely entered the field of view of a single camera. A recognition rate of approximately 96% (since 22 out of the 23 samples were classified correctly) was achieved from the image sequences indicating the applicability of the new framework in similar task environment. Lastly, the new framework is computationally and memory efficient with great potential for use in outdoor surveillance systems.*

Keywords. *Moving object recognition, classification, recurrent motion.*

1. Introduction

Moving object recognition has been one of the active research areas in computer vision and pattern analysis. It plays a major role in advanced security systems and video surveillance applications. With the aim to recognize moving objects through video monitoring system, it is able to detect and track for moving objects in a surveillance area, and classify these objects into predefined categories. Enhancement to current security systems and surveillance applications can be realized with this recognition function. For instance, intruder recognition function can be incorporated into a security system to classify intruders in order to reduce nuisance alarm and minimize human errors in manned surveillance system. Another example of its application is in traffic monitoring system, where it can be used to estimate traffic flow by making vehicle and pedestrian counts.

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. Moving objects detected from image sequences are classified based on their periodic motion patterns captured in the RMI. In our approach, a new object class has been introduced to the existing RMI framework and various refinements have been done on the preprocessing algorithms, to enhance its recognition accuracy and compatibility for use in general outdoor scenes.

2. Related works

Extensive research efforts have been dedicated to moving object recognition, where many approaches have been presented to tackle this problem. Due to page constraint, instead of discussing all these methods in detail, an overview of the related techniques is presented. For example, a view-based method for recognizing 3D objects from 2D images [2] is exploited to serve this purpose. An aspect graph structure is implemented to generate aspects using

a notion of similarity between views. A viewing sphere is endowed with a metric of dissimilarity for each pair of views. The viewing sphere is sampled at regular intervals, and the views are combined into aspects, each represented by a prototypical view. Unknown views of unknown objects are compared with the prototypical views hierarchically, and results are ordered by similarity. Two similarity metrics for shape were used, one based on curve matching and another based on matching shock graphs. The identification results are 98 percent and 100 percent respectively. The results are promising but this approach is significantly time consuming and memory intensive. With 5 degrees sampling rate, database construction for each 3D object takes up to 11 hours, while a recognition process requires up to 45 minutes.

An approach adopted statistical motion detection and Fourier descriptors for shape-based moving object recognition [3]. A statistical, adaptive, illumination invariant motion detection algorithm is used to identify moving object candidates. Fourier descriptors [4] are computed as feature vectors to describe the shape of object. The classification module utilizes a four-layered feed-forward neural net. Objects are categorized as human, vehicle or background clutter. This system is robust and yields correct classification in more than 90 percent of all tested cases except scenes involving occlusion and shadow handling which generated numerous misclassifications.

There is an example of motion-based recognition method using Generalized Symmetry Operator to extract motion symmetry of moving objects for gait recognition [5]. First, Sobel operator is applied to the object silhouette to obtain edge map. A symmetry operator is then applied to the edge map to produce symmetry map. Fourier transform is applied to the gait signature (obtained by averaging all symmetry maps in an image sequence) and the k-Nearest Neighbor rule [6] is adopted for classification. This approach has a promising recognition rate of over 95 percent, and it is relatively immune to noise and capable of handling occlusion. However, for practical purposes, the recognition rate on a larger database may diminish by selection of fewer Fourier components to improve recognition speed.

A specific feature vector called Recurrent Motion Image (RMI) [1] is 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. 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 at the areas of hands and legs. This approach produces accurate classification results while remaining computationally and space efficient. However, shadow removal suffered an error rate of 30 percent due to poor segmentation.

A hybrid classification system [7] can be used to recognize moving objects based on motion and appearance features simultaneously. At the first layer of the hybrid classifier, appearance data is processed by support vector machine (SVM) classifier [8], the resulting feature vectors are known as shape and appearance features. These features are combined with motion-based features, and used as input to the SVM classifier at the second layer. The hybrid approach saw a 15.5% increase in recognition rate for human, animal and vehicle, as compared to single SVM classifier which uses motion or shape features. However, its framework architecture is of higher complexity, where two classification layers are needed with a SVM for each. It also uses multiple hypothesis approach at which a classification hypothesis is updated every 24 frames, and statistics are accumulated for 3 seconds before a classification decision is made. This limits the recognition speed of the hybrid classifier.

Based on the discussion above, motion-based recognition with RMI is one of the few approaches that produce high recognition rate while remaining computationally and space efficient. However, existing RMI method has only been tested on a small set of object classes (human and vehicle) and performs poorly in shadow removal. In this paper, we propose several refinements to the existing RMI method to increase its recognition classes and recognition rate for use in general outdoor scenes.

3. Methods

In the original RMI motion-based recognition approach, moving objects detected are classified as single person, group of persons or vehicle, based on a feature vector called Recurrent Motion Image (RMI). Our method extends the object classes of the existing framework to include four-legged animal, and we made various refinements to the shadow removal and preprocessing algorithms to improve recognition accuracy of the new framework.

3.1 Object detection and tracking

The moving object detection algorithm in our new framework differs from [1] in that it extends the preprocessing stages to include shadow removal and multiple levels of noise filtering to enhance moving object segmentation. Background subtraction is carried out by computing an L-inf distance image [9] in the Red-Green-Blue (RGB) color space. Subsequently, foreground points obtained by applying a low threshold to the L-inf distance image will go through a morphological opening denoise layer [9]. We locate the shadow points in the Hue-Saturation-Value (HSV) color space [10] and these are removed before connected components labeling is carried out to extract the foreground blobs. Blob level analysis [9] is then performed to filter background noise clutters using a blob size threshold. Lastly, a high threshold 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 while the others are removed as non-salient blobs.

The validated blobs are tracked using region correspondence [11]. Parameters such as centroid, bounding box, size, velocity and change in size of each blob are extracted. 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. Non-corresponded region may possibly involve in object entry or exit occlusions, so such region from previous frame is examined for object exit. If its predicted position exceeds frame boundary, it is determined to have exited the surveillance area. Otherwise, object occlusion is suspected. If its bounding box overlaps the bounding box of another region Q in the current frame then Q is marked as an occluded region while all the non-corresponded regions in previous frame overlapping Q are marked as occluding each other. On the other hand, non-corresponded region in current frame is set to be an object entry.

3.2 Object classification

Each of the moving objects detected and tracked in 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. RMI is computed with the following equations to determine the areas of moving object's silhouette undergoing repetitive changes:

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

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

S_a is a binary silhouette image 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.

The RMI is partitioned into N equal-sized blocks in order to compute the average recurrence for each block, as illustrated in Figure 1. Blocks with average recurrence value greater than a threshold τ_{RMI} are set to 1 (white) and vice versa. If there are white blocks at the middle and bottom sections, the object is classified as a single person or a group of persons. If there are no white blocks, which means no recurrent motion, the object is classified as a vehicle.

There are two cues to differentiate between single person and group of persons – multiple peak points in a silhouette indicates more than one headcount, therefore representing a group of persons, and, the normalized area of recurrence response at the top section of RMI being greater than that of a single person, due to presence of multiple heads. If either one of the aforementioned criteria is satisfied, the object is classified as a group of persons, whereas if none of them are satisfied, the object is classified as a single person.

To include four-legged animal as an additional object class, we studied the recurrent motion behavior of dogs and cats to derive the motion pattern and criteria for classification. Figure 2 shows the RMI generated for a dog and a cat. Notice that the RMI for dog and cat are similar, whereby their legs and tail demonstrated repetitive motion. This causes white blocks to occur in all top, middle and bottom sections. However, for dogs and cats without a tail, only middle and bottom sections will have white blocks. In addition, dogs or cats which are short in height but long in length may cause white blocks to occur only at the middle section. Thus, blocks in the middle section can be taken into consideration for four-legged animal classification. This is similar to the cue for classifying a moving object as single person or group of persons. Therefore the classification criteria are modified specifically to differentiate between human and four-legged animal (dog or cat in this case).

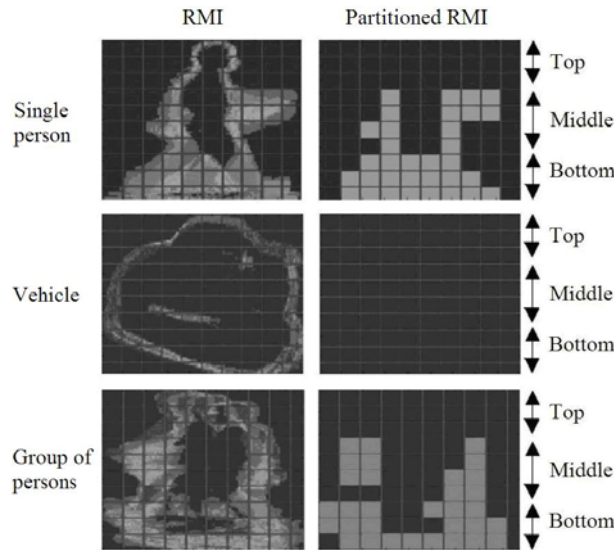


Figure 1. RMI for classification [1]

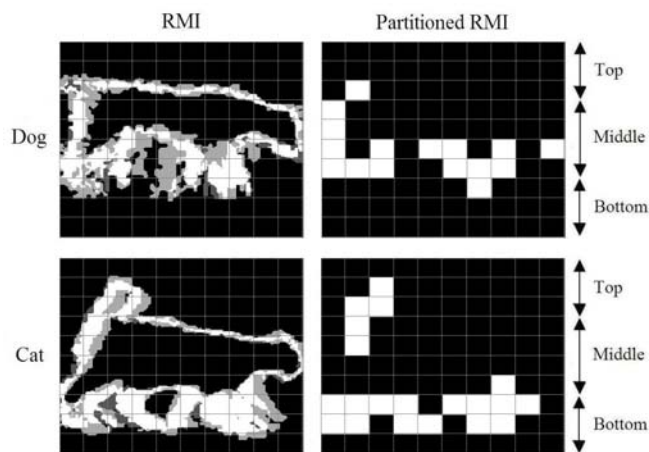


Figure 2. RMI of a dog and a cat

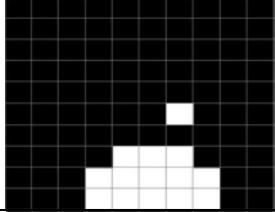

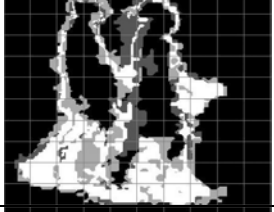
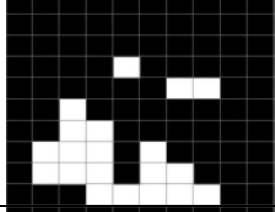
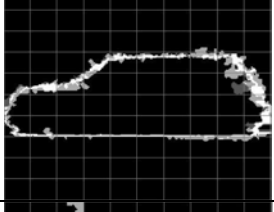

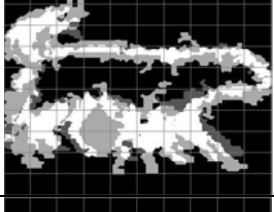
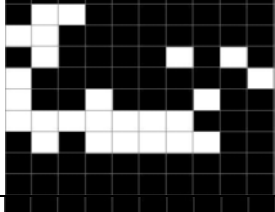

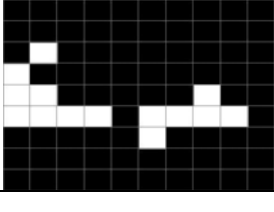
The black area within the RMI of an object corresponds to the area where the object demonstrates no recurrent motion. Human and four-legged animals generally do not show any recurrent motion at the main part of the body where the backbone is located. This can be seen in Figure 1 and Figure 2. Notice that the black area within the RMI of a single person and group of persons (Figure 1) has vertical major axis, whereas the black area within the RMI of a dog and a cat (Figure 2) has horizontal major axis. This acts as the cue to differentiate between human and four-legged animal (dog or cat in this case).

When there are white blocks in the middle section of a partitioned RMI, the black area within the respective RMI is observed. If the black area has a vertical major axis, the corresponding object is classified as human, which will then be further categorized as a single person or group of persons. Otherwise, if the black area has a horizontal major axis, the object is classified as a four-legged animal. If the object does not fall under any of the predefined categories (vehicle, single person, group of persons and four-legged animal), it will be classified as other objects. Note that accurate differentiation between all of the four-legged animals classified by the new algorithm can be carried out using texture, size, color and other relevant information of the object. 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.

4. Results

The moving object detection, tracking and classification algorithm was implemented on several image sequences captured using a low-end digital camera (Olympus FE-280) at housing areas. Those image sequences contain single persons, groups of persons, vehicles, four-legged animals (dogs and cats). After the moving object has completely entered the frame, the RMI was generated for one second duration. 23 image sequences of various time lengths were taken in parallel to the ground plane. The frames were 320 x 240 pixels in size and sampled at a rate of 8 frames per second. The algorithm was implemented in Matlab and executed on a 1.5GHz Core 2 Duo CPU. Table 1 shows several instances of the moving object detected and classified using the framework with a threshold (τ_{RMI}) of 2.

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

There were a few moving objects that the framework failed to detect or classify accurately. They were mainly caused by foreground extraction problem due to color similarity between moving object and the background. For example, the moving object in Figure 3 was having similar color with the background at the pointed location. The background subtraction method used in this framework was unable to completely extract the foreground points in the highlighted locations.



Figure 3. Defective silhouette of a single person

As indicated by the red arrow in Figure 3, one of the shoes was separated from the main silhouette of the person, and later removed as noise clutter when passed through the blob size filter. Although the silhouette was incomplete but the absence of a shoe did not affect the classification, because the recurrent motion was still able to be seen in the RMI generated, and the moving object could still be correctly classified as a single person. Hence, we would say that in this paper, moving object detection and segmentation defects may or may not affect the classification results, depending on the defective location and defect's severity. Little defects at regions which carry unimportant or less important information for classification, will not affect the recognition rate. For instance, the shoe in Figure 3 is less important as compared to the legs which carry more motion information.

5. Conclusion and future work

22 out of 23 moving objects in our experiments were successfully detected, tracked and classified into categories of single person, group of persons, vehicle, and four-legged animal, except objects with areas that exhibit similar color to the background. These objects may pose some difficulties to the detection and segmentation algorithms, but they may or may not cause problem to the classification algorithm, depending on the defective areas of the object silhouettes. Most of the experiments showed that the defects were not remarkable, thus did not affect the classification results. As illustrated in Table 2, 96% of the moving objects were correctly classified thus indicating the backward compatibility and success of the new classification list (human, vehicle and four-legged animal) introduced in the new method.

Table 2. Classification results

Object class	Number of samples tested	Number of samples correctly classified
Single person	5	5
Group of persons	4	4
Vehicle	6	6
Four-legged animal	8	7

The aforementioned defective silhouette problems may be handled via segmentation and texture analysis to combine several disjointed clusters in close proximity as a single object for analysis. Further investigations are needed to ascertain the effectiveness of these analyses in the RMI recognition framework.

6. References

- [1] O. Javed and M. Shah, "Tracking and Object Classification for Automated Surveillance," *Proc. 7th European Conf. Computer Vision-Part IV (ECCV 02)*, 2002, pp. 343-357.
- [2] C.M. Cyr and B.B. Kimia, "A Similarity-Based Aspect-Graph Approach to 3D Object Recognition," *Int'l J. Computer Vision*, vol. 57, no. 1, 2004, pp. 5-22.
- [3] 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 (ICIAP 03)*, 2003, pp. 430-435.

- [4] J.C. Russ, *The Image Processing Handbook*, 5th ed., CRC Press, Boca Raton, 2006, p. 589.
- [5] 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 (ICIP 01)*, 2001, pp. 290-293.
- [6] J.C. Russ, *The Image Processing Handbook*, 5th ed., CRC Press, Boca Raton, 2006, p. 619.
- [7] 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. (BMVC 03)*, 2003, pp. 429-438.
- [8] D.A. Forsyth and J. Ponce, *Computer Vision: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ, 2002, p. 615.
- [9] 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.
- [10] 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 (ITSC 01)*, 2001, pp. 334-339.
- [11] B. Jahne and H. HauBecker, *Computer Vision and Applications: A Guide for Students and Practitioners*, Academic Press, San Diego, USA, 2000, p. 379.