CLASSIFICATION OF MOVING OBJECTS USING RECURRENT MOTION IMAGE CLASSIFIER

C. J. Kavithapriya
Assistant Professor, Department of Information Technology, Jeppiaar Institute of Technology, Sriperumbudur

Abstract: In advanced security and automated surveillance systems, Object Recognition and Classification plays a vital role. Motion Based Recognition has been improved by using Recurrent Motion Image (RMI) classifier. For efficient Object Recognition the input video frames endures Preprocessing method such as (i) Background Subtraction using L-infinite distant image method, (ii) Shadow Points located by transforming the pixels from RGB (Red Green Blue) color space to HSV (Hue Saturation Value) color space and these points are remove by exploiting Gaussian Filters, (iii) Foreground Blob extraction by incorporating Connected Components Labeling algorithm and Blob Analysis using Blob size threshold to filter noise clutters. Postprocessing method encompasses (i) Blob Tracking by Region Correspondence algorithm, (ii) Blob Classification using RMI based on their periodic motion patterns. Most of the existing classification algorithms deal the Object Classification with the constraint of classifying only Human Being, Vehicles. But, the proposed algorithm has been extended such that it violates this constraint by classifying Four-Legged Animals in addition to Human Being, Vehicles and upgrades the Object Classification methodology. By implementing this proposed algorithm, all moving objects were classified into proper categories as (i) Single Person, (ii) Group of Persons, (iii) Vehicles, (iv) Four-Legged Animals.

Keywords: RMI, Blob Tracking, Recurrent Motion Image.

I. INTRODUCTION

Motion based recognition paves the way for recognizing an object based on its motion. To acquire relative motions of an object, few numbers of frames are insufficient. Due to this constraint most of the algorithms uses more number of frames, with which the motion information is extracted. Typically, motion based recognition involves two basic steps such as extracting features of object and then matching those features in the model. Recovering of object’s parameters even in the motion is the process of object tracking. Eventually, the motion based recognition engrosses object classification which uses the extracted motion information and the tracked information of the moving object.

In the recognition process, the knowledge about the object is utilized for constructing précised model. Motion based recognition approach has been exploited in various fields such as in telephone networks for real time sign language communication over phones, for study of left ventricular motion and for tracking the location of various joints of human body in gait analysis, in surveillance systems for classifying intruders, in automatic monitoring systems for the inference of traffic flow by counting pedestrians and vehicles. For augmenting the approach of motion based object recognition a specific feature vector, the Recurrent Motion Image (RMI) classifier has been proposed. The new improved RMI classifier violates the constraints of previous classifiers of classifying only human being, vehicles. As the improved classifier this classifies not only the human being and vehicles but also the four-legged animal.

II. LITERATURE REVIEW

Most of the Far-reaching research efforts have been dedicated to moving object recognition, where the problems in moving object recognition have been tackled by many approaches that has been discussed in this section.
A view-based method [5] has exploited for recognizing the 3D form of images from the 2D form of images. To generate aspects using a conception of similarity between views, the construction of aspect graph has been used in view-based method. The viewing sphere, with the artistic effects of divergence between each pair of views is sampled at regular intervals, and the resultant views that are combined into aspects are represented by a prototypical view individually. Finally, the unknown views of indefinite objects are compared with prototypical views hierarchically and the results are arrayed by similarity. Curve matching and matching shock graphs were used as the similarity metrics for shapes of the object. The results achieved in this method are of 98 percent. Even though it furnishes promising results, its efficiency is degraded by time and memory consumption.

Statistical motion detection and Fourier descriptors for shape-based moving object recognition [12] is another approach in motion based object recognition method. To describe the shape of object, Fourier [11] are computed as feature vectors. Four-layered feed-forward neural net has been exploited in this classification module. Objects are categorized as human, vehicle or background clutter. In more than 90 percent of all tested cases the system is robust and also yields correct classification, except the scenes entailing occlusion and shadow handling in which numerous misclassifications are produced.

To extract motion symmetry of moving objects for gait recognition Motion-based recognition method using Generalized Symmetry Operator has been incorporated. Initially, Sobel operator is applied to the object silhouette to obtain edge map and then Symmetry Operator is applied to the edge map to produce symmetry map. For the purpose of classifying the objects Fourier transform is applied to the gait signature [10] (obtained by averaging all symmetry maps in an image sequence) and the k-Nearest Neighbor rule [11] This approach has a promising recognition rate of over 95 percent, and it is relatively immune to noise and capable of handling occlusion. Inspite of providing promising recognition rate of over 95 percent and with the advantage of immunity to noise and occlusion handling capability it is not appropriate for practical purposes. On a larger database the recognition rate diminishes by selection of fewer Fourier components to improve recognition speed.

A hybrid classification system [1] 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 [6], 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. Its framework architecture is of higher complexity, where two classification layers are needed with a SVM (Support Vector Machine) 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 above discussions, motion-based recognition with RMI [2] is one of the few approaches that produce high recognition rate of over 98 percent.

III. SUMMARY OF PROPOSED WORK

In order to boost up the efficiency of motion based object recognition, an improved classifier Recurrent Motion Image (RMI) has been proposed in this work. The procedure involved in the classification of moving
objects using RMI are (i) Background Subtraction, (ii) Shadow Removal, (iii) Foreground Blob Extraction, (iv) Blob Tracking, (v) Blob Classification. Initially, the input video frames undergoes the preprocessing steps for extracting the Blob and then the resultant blob passes onto the postprocessing steps for the classification of the moving object in each frame. To detect the motion of the object more number of frames is used. The classified objects are categorized as (i) single person, (ii) group of persons, (iii) vehicles, (iv) four-legged animals based on their periodic motion and the calculated RMI values.

The block diagram that is shown in fig. 1 symbolize the procedure involved in preprocessing and post processing of determining the RMI for the input frames. Initial block in the shown figure receives the input frames as background and foreground frames and generates the output frame. The resultant of the first block has been passed to the next block for removing the shadow points that exists in the previous result.

After the shadow removal process has been refined then the process of foreground blob extraction is initialized with the shadow removed image and generates the foreground blob. As the preprocessing method blob tracking is initialized and the corresponding regions between current and previous frames have been tracked.

By utilizing the tracked regions of the blob, the blobs are classified to categories as human, vehicles, four-legged animals. Finally, the motion pattern of the classified objects has determined by using the equation (7).

With the reference of the motion pattern determined, RMI is calculated for each object using the equation (8) and the calculated RMI is partitioned into blocks. The RMI values for each category will be explained in Section V.

**IV. PREPROCESSING PROCEDURE**

The preprocessing procedure involved in the RMI classifier has been discussed in this Section.

**A. Background Subtraction**

Every segmentation process initiates with the process of background subtraction which performs the extraction of foreground from the background by comparing each frames in the input video sequence. In this work the process of background subtraction has been performed by exploiting L- inf distance image computation in RGB (Red Green Blue) color space [3]. The background model is computed by using the Median function as given in equation (1).
Figure 2: Results of Shadow Removal process. (a) and (b) represents the background and foreground frames respectively, the bounding box indicated in red color denotes the foreground with shadow. The shadow of the foreground lies on the floor has to be removed in this process of shadow removal. (c) Denotes the background subtracted image in RGB color space. (e) Shows the morphological operations exploited on (c). (d) Specifies the shadow points in HSV color space (bounding box in white color characterizes the foreground and the ellipse mark denotes the shadow points). (f) Final result of process involving removal of shadow is shown.

\[ M_{s}^{t=M}(p) = \arg \min_{i=1..k} \sum_{j=1..n} \text{Distance} \left( x_{i}, x_{j} \right) x_{i}, x_{j} \in SS \]

(1)

Where the distance is a L-inf distance in the RGB color space which is given by

\[ \text{Distance} \left( x_{i}, x_{j} \right) = \max \left( |x_{i,c} - x_{j,d}| \right) \]

(2)

The distance parameter given in equation (2) is determined by evaluating the distance between each pixel and its neighboring pixels in RGB color space with the parameter ‘c’ which notifies the R, G, B pixels. The resultant distances between RGB pixels are averaged and the median function is generated in equation (1).

The median function has proven effective than the Gaussian or other complex statistics in the aspect of less computational cost. In order to improve the object detection accuracy, background subtraction is performed by taking into account not only point’s brightness, but also its chromaticity.

\[ DsBg^{t}(p) = \text{Distance( Is}^{t}(p), Bg^{t}(p)) \]

(3)

L-inf is less computationally expensive than other distances in which the selection of initial set of foreground points is carried out by selecting the distance image DsBg\(^{t}\) defined in (3) with an adequately low threshold \( T_{L} \). Among these points some points are discarded as the noisy points by applying morphological operators [6] and the result generated is a smoothened image with the noisy pixels, discarded as shown in fig. 2 (e).

The noise removed image is then passed for shadow removal process. The major problem is how to
distinguish between moving cast shadows and moving object points. In effect, points belonging to both moving objects and shadows are detected by background subtraction by means of (4).

To accomplish these criteria, the pixels in the Hue-Saturation-Value (HSV) color space [4] has been analyzed which is illustrated in fig. 2(d). The main motive of using HSV color space is that it explicitly separates chromaticity and luminosity and differentiate the foreground and shadow points, also has proven trouble-free than the RGB space to set a mathematical formulation for shadow detection.

1) If a shadow is cast on a background, the hue component (5) varies, but within a certain threshold.

2) In addition, the saturation component has also been considered, for proving experimentally that it also varies within a certain threshold.

3) The difference in saturation will be an absolute difference, while the difference in hue will be an angular difference.

We define a shadow mask SM', which is specified in equation (4) for each point ‘p’ resulting from motion segmentation based conditions given in (4). The lower bound ‘a’ defines a maximum value for the darkening effect of shadows on the background and approximately proportional to the light source intensity.

The upper bound ‘b’ precludes the system from identifying as shadows, those points where the background was darkened too little relative to the expected effect of shadows.

Equation (4)

\[
SM'(x,y) = \begin{cases} 
1 & \text{if } a \leq \frac{I_S^v(x,y)}{B_g^v(x,y)} \leq b \\
& \land \left( \left| I_S^s(x,y) - B_g^s(x,y) \right| \right) \leq t_s \\
& \land \left| I_S^h(x,y) - B_g^h(x,y) \right| \leq t_h \\
0 & \text{otherwise}
\end{cases}
\]

\( B_g^v(x,y) \) - Intensity value for the component V of the HSV pixel at coordinates (x, y) in the frame k of the background frame.

\( I_S^v(x,y) \) - Saturation value for the component V of the HSV pixel at coordinates (x, y) in the frame k of the current frame

\( B_g^s(x,y) \) - Saturation value for the component V of the HSV pixel at coordinates (x, y) in the frame k of the background frame.

\( I_S^h(x,y) \) - Hue value for the component V of the HSV pixel at coordinates (x, y) in the frame k of the current frame

\( B_g^h(x,y) \) - Hue value for the component V of the HSV pixel at coordinates (x, y) in the frame k of the background frame.

\( \text{Hue}_p^i = \min \left( \left| I_S^i(p).H - B_g^i(p).H \right| \right) , 360 - \left| I_S^i(p).H - B_g^i(p).H \right| \) (5)

Black pixels obtained are classified as belonging to the background model, dark gray pixels obtained are classified as foreground, light gray ones can be identified as shadows by means of only the luminance information given in (5), while white pixels are detected as shadow points using the specific criteria,
the chrominance information. Removing light gray pixels from the shadow mask improves the accuracy and avoids the misclassification of pixel belonging to the object as shadow.

B. Foreground Blob Extraction

The resultant shadow removed frames undergoes foreground blob extraction process by incorporating connected component labeling algorithm [7]. In order to identify connected pixel regions, that is the regions of adjacent pixels which share the same set of intensity values I connected component labeling employs scanning of an image, pixel-by-pixel (from top to bottom and left to right).

The path between $P_1$ and $P_n$ (pixels) is the sequence of pixels, $P_1, P_2, ..., P_{n-1}, P_n$, such that $P_k$ is adjacent to $P_{k+1}$, with the $k$ values limiting from 1 to $n$. The path can be classified as 4-connected or 8-connected, depending on adjacency used. The extracted foreground blob as described in fig. 3 (a) has been analyzed to segregate the noise clutters by using blob threshold with certain constraints as given.

1) Extracted blob must be larger than the threshold $B_H$, so that the blobs of few pixels like movements of trees can be removed and also selects points with large difference from background and confirm blobs with one of these points.

2) Low threshold $B_L$ set on the difference image $D_{sBg}$ in (3) definitely selects noise pixels with definite foreground points.

V. Postprocessing Procedure

A. Blob Tracking

The blobs extracted from the previous process are tracked by exploiting Region Correspondence algorithm. In each blob the corresponding parameters such as bounding box, centroid, velocity, and change in size of the blobs are extracted.

To update the status of each object over the frames the minimum cost criteria [9] has been used for determining the correspondences between regions in previous frame and current frame. In previous frames, the exit of object or occlusion has been examined by detecting the non corresponding region. The cost function of the minimum cost criteria between two regions is given by,

$$MC_{Li} = w \left\| C_t^{LR}_i - 1 + V_t^{LR}_i - 1 - C_t^{LR}_i \right\| + (1 - w) \left\| S_{t}^{LR}_i - 1 + S_{t}^{LR}_i - 1 - S_t^{LR}_i \right\|^2$$

Where the coordinates of the centroid is given by $C$, blob size represented as $S$.

1) Process Involved In Region Correspondence Algorithm: The corresponding regions that are established has been associated with the velocity $V$, and variation in size $\nabla S$. In the frame $t$ there occurs $N$ regions with the centroids $C_t^{LR}_i$, with no correspondence to the previous frame. In frame $t-1$ there occurs $M$ regions with the centroids $C_{t-1}^{LR}_i$, with established correspondence with the previous frame. The major goal of the region correspondence algorithm is to recognize the corresponding regions between the frames $t$ and $t-1$ with the minimum cost criteria function in (6).

C. Blob Classification

Blobs that are detected and tracked endure the process of classification. Based on the repetitive changes in the

© 2014 PISER Journal

http://piserjournal.org/
shape of the objects are classified as (i) single-person,(ii) group of persons, (iii) vehicles, (iv) four-legged animals. Repetitive motion and non-repetitive motion can be determined by the RMI values. Higher values of RMI represent the recurrent motion and lower values of RMI represent the non-recurrent motion or the motionless region. In order to determine the areas of moving object’s silhouette that experience repetitive changes, RMI is computed by the following equation

$$Mo(x, y, t) = B_S(x, y, t - 1) \oplus B_S(x, y, t)$$

(7)

$$RMI = \sum_{k=0}^{F} Mo(x, y, t-k)$$

(8)

$B_s$ is the binary silhouette image for object ‘a’ at frame $t$, and $Mo$ is the binary image indicating areas of motion for object $a$ between frame $t$ and $t-1$. For computing RMI of an object, initially the recurrent motion areas of object are determined by (7). Then the RMI for object is calculated over $F$, the total number of frames by the equation (8) and partitioned into $X$ number of blocks in order to figure out the average recurrence for each block. The threshold $\alpha_{RMI}$ is used to set the higher and lower recurrence values of each block with average recurrence value. If $\alpha_{RMI}$ is lesser than the average recurrence value then blocks are set to 1 i.e. blocks contains white pixels and if $\alpha_{RMI}$ is greater than the average value the blocks are set to 0 i.e. blocks contains black pixels.

1) **Classification of Person:** The incidence of white pixels at the middle and bottom sections as shown in Table 1. is classified as person. The category person is classified as single person or group of persons. If a silhouette includes multiple peak points then it indicates the existence of more than one headcount, hence it represents a group of persons otherwise it represents a single person. For group of persons the response of the normalized area of recurrence at top section of RMI will be greater as shown in fig. 4 (b) than the response of the single person. In addition to these types, another type is person with hands inside pocket, for this type the response will be over top, middle and bottom sections with vertical major axis to differentiate from four-legged animals.

![PARTITIONED RMI](image1)

Figure 4: (a) shows the RMI partition for person with hands inside the pocket with RMI responses at bottom sections, (b) shows Partitioned RMI for the category group of persons with responses at top, middle, bottom sections

2) **Classification of Four-Legged Animals:** As an additional object class, four-legged animal has been included. In the classification of four-legged animals, the repetitive motion will be higher at the regions of its legs and tails. Since the tail region of the four-legged animals have repetitive motion it occupies the top section of partitioned RMI and the repetitive motion of leg region occupies the middle, bottom section as shown in fig. 5(a).

![RMI & PARTITIONED](image2)

Figure 5: Above figure illustrates the difference between the RMI responses for the animal (dog) with tail (a) and without tail (b)
For tailed animals the above criteria has been satisfied, also additionally tail less animals also can be classified with its recurrence response at only bottom and middle sections as in fig. 5(b). Exclusion of top section is due to the absence of motion pattern of the tail.

3) **Classification of Vehicles:** Compared to person and four-legged animals the repetitive motion pattern of vehicles is dissimilar. For person and four-legged animals the motion pattern exists at the regions of hands, legs and tail respectively. But for vehicles the motion pattern exists at the region of wheels while the vehicle at motion. At movement the category of vehicles are classified by the reference of white blocks at only bottom section and for vehicles with no movement no sections will be selected and RMI holds only black blocks.

4) **Remedy for Confliction between Classification of Person and (Four-Legged Animals, vehicles):** The confliction between the classification of person and four-legged animals results due to the same response of recurrent motion at middle and bottom sections. This can be differentiated by considering the criteria that for human, the black area i.e. the black blocks will be along vertical major axis and for four-legged animals the black blocks will be along horizontal major axis. Even in the case of classification of single person with hands inside the pockets provides the responses of RMI at the bottom sections as shown in fig. 4(a) which conceives a conflict with the responses of the vehicles. To solve this conflict, the considered criteria are vehicles have horizontal major axis and apparently human has vertical major axis.

**TABLE 1: INTERPRETATION OF RMI CLASSIFIER RESULTS**

<table>
<thead>
<tr>
<th>INPUT FRAME</th>
<th>RMI</th>
<th>PARTITIONED RMI</th>
<th>OBJECT CATEGORY</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Person" /></td>
<td><img src="image2.png" alt="RMI Person" /></td>
<td><img src="image3.png" alt="Partitioned RMI Person" /></td>
<td>SINGLE PERSON</td>
</tr>
<tr>
<td><img src="image4.png" alt="Group" /></td>
<td><img src="image5.png" alt="RMI Group" /></td>
<td><img src="image6.png" alt="Partitioned RMI Group" /></td>
<td>GROUP OF PERSONS</td>
</tr>
<tr>
<td><img src="image7.png" alt="Four-Legged Animal" /></td>
<td><img src="image8.png" alt="RMI Animal" /></td>
<td><img src="image9.png" alt="Partitioned RMI Animal" /></td>
<td>FOUR-LEGGED ANIMAL</td>
</tr>
<tr>
<td><img src="image10.png" alt="Vehicle" /></td>
<td><img src="image11.png" alt="RMI Vehicle" /></td>
<td><img src="image12.png" alt="Partitioned RMI Vehicle" /></td>
<td>VEHICLE</td>
</tr>
</tbody>
</table>
VI. RESULTS

The algorithms of motion detection, object tracking, and object classification has been implemented in MATLAB\textsuperscript{8}. Various video frames has taken as the input and classified as categories of single person, group of persons, vehicles, four-legged animals. RMI signature has been generated for various frames with the size of 640×480 pixels and sampled at rate of 3 frames per second. Table 1. interprets the classification of objects using RMI classifier at threshold of 0.5

VII. CONCLUSION AND IMMINENT WORK

In this work the input datasets taken has been successfully classified as single person, group of persons, four-legged animals and vehicles. Minor problems have been occurred at the preprocessing methods such as in the noise clutter removal and shadow removal but that has been recovered by using efficient filters.

As an imminent work this paper could be extended by exploiting new algorithms to classify the four-legged animals as specified animal by studying the recurrent motion behavior.

REFERENCES


