Applied Soft Computing xxx (2012) xxx–xxx

Granulation, rough entropy and spatiotemporal moving object detection

Debarati Chakraborty[a,*], B. Uma Shankar[b], Sankar K. Pal[a,b]

[a] Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700 108, India
[b] Machine Intelligence Unit, Indian Statistical Institute, Kolkata 700 108, India

ABSTRACT

A new spatio-temporal segmentation approach for moving object(s) detection and tracking from a video sequence is described. Spatial segmentation is carried out using rough entropy maximization, where we use the quad-tree decomposition, resulting in unequal image granulation which is closer to natural granulation. A three point estimation based on *Beta Distribution* is formulated for background estimation during temporal segmentation. Reconstruction and tracking of the object in the target frame is performed after combining the two segmentation outputs using its color and shift information. The algorithm is more robust to noise and gradual illumination change, because their presence is less likely to affect both its spatial and temporal segments inside the search window. The proposed methods for spatial and temporal segmentation are seen to be superior to several related methods. The accuracy of reconstruction has been significantly high.

Keywords:
Spatial segmentation
Temporal segmentation
Granular computing
Rough entropy
Three point estimation
Video tracking

1. Introduction

In computer vision object detection and tracking is an important task. The application of object tracking in video sequences has been studied over the years [8,21]. In this task there are different types of uncertainties and ambiguities which is making this task a difficult problem. The first step towards detecting an object from a video sequence is to find the object as a separate segment in the frame. The object can be separated out from its background according to spatial homogeneity and/or based on temporal homogeneity. If the object and background can be separated properly in any feature space, then only spatial segmentation can give satisfactory result. But, in most of the cases the total object cannot be solely extracted from the background. On the other hand temporal information plays an important role in detecting object. But, without huge change from frame to frame or without having predefined background, temporal segmentation technique is also unable to extract the total object out in the video sequence [21]. In the present article we have defined a spatio-temporal segmentation technique, where the spatial and temporal segmentation outputs are merged together to construct the target object for detection and tracking of it efficiently and accurately.

Understanding of an image depends on its proper partitioning. Granulation can be useful to find segments/regions in an image. Pedrycz et al. showed how granular computing plays an effective role to define vagueness [14,1] in several sets with soft boundary. Depending on the application, granulation could be of different types with granules of equal or unequal size, although unequal granules are more natural for real life problems. Here we use rough entropy [12], with unequal image granulation, incorporating some modifications for video image segmentation. The basic idea underlying the rough sets approach to information granulation is to discover the extent to which a given set of objects (e.g., pixels in windows of an image) approximates another object of interest (e.g., image region). Rough sets along with granular computing have been applied to several areas of image processing [12,10] but hardly on video processing. Here we use rough set and granular computing to handle the uncertainties and ambiguities in video images and find these to be effective.

Detection of an object can be viewed in another way as removal of background. In case of video sequences the proper estimation of a background plays an important role. There are several methods for background construction [7,18]. But, the time complexity or memory requirements of most of the techniques are very high. We know that, when a random variable follows *Beta Distribution* [6], the mean and standard deviation of the variable can be estimated by three point estimation. It is normally used in time estimation in management issues [9]. We found this technique useful to estimate the background and foreground deviation from the background of a video sequence and applied it for temporal segmentation.

Temporal and spatial information need to be used in such a way that the proper object can be reconstructed. We have considered the color distribution combining the spatial and temporal segmentation outputs and the object locations in the previous two frames

[*] Corresponding author.
E-mail address: debarati.earth@gmail.com (D. Chakraborty).

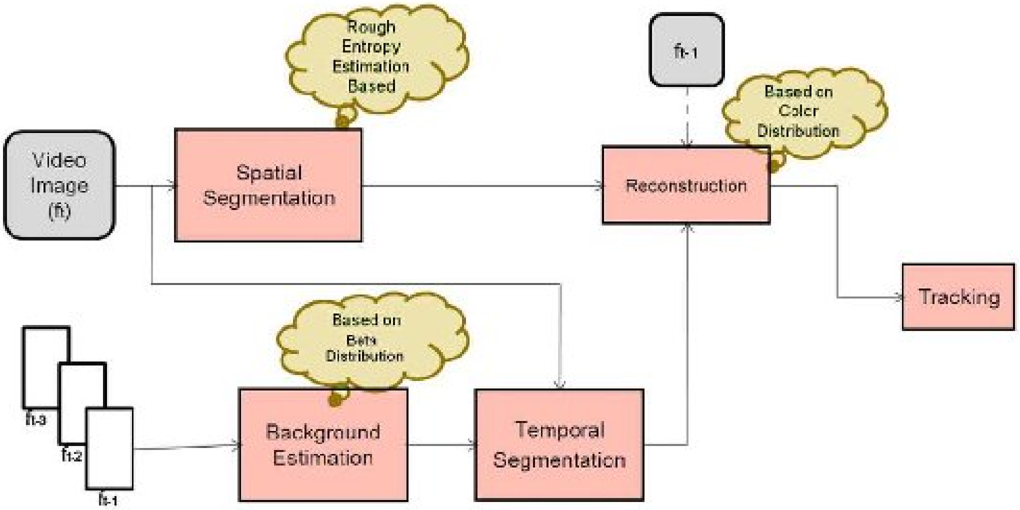

Fig. 1. Block diagram of the proposed methodology.

to optimize the search region. There exist several methodologies dealing with object location estimation from previous frames, but most of them consume large memory. Whereas, in our case, less memory is required as information only from the previous two frames is to be stored. The block diagram of the proposed methodology for tracking is given in Fig. 1, where the bubbles show defined methodologies used for the steps in the rectangular blocks.

The objective of the investigation is to develop a spatiotemporal approach to detect and track the moving objects using still cameras efficiently and accurately. We propose here the rough entropy based image segmentation, background estimation using three point estimation for single and double objects under various illumination conditions and a method of combining them judiciously. We also perform the comparison of segmentation to evaluate with similar methods. Accuracy is evaluated with ground truth for detection and tracking. The novelty of the contribution mainly lies with

(a) formation of granules of unequal size which is more natural,
(b) formulating three point approximation method for background estimation, and
(c) combining spatial and temporal segmentations using both color and shift information of the object.

This paper is organized as follows. In Section 2 we have discussed the basic concepts of rough sets in brief and a spatial bi-level segmentation technique according to rough entropy maximization [12,19]. In Section 3, we have proposed a new technique of background estimation based on a three point estimation. Section 4 describes combination of spatial and temporal segmentation results. In Section 5, we have presented the results and comparative performance on several types of video images. The spatial segmentation technique is compared with Otsu's thresholding, rough entropy maximization with uniform granules [12] and the recently reported rough fuzzy entropy based segmentation [17] both visually and quantitatively. Temporal segmentation is compared with a popular and widely used technique: mixture of Gaussian (MoG), and a change detection technique: linear change detection (LDD) [5]. We show how our proposed simpler technique results in less noisy foreground. The reconstruction results have been validated with ground truth. The object(s) in the sequences has (have) been found to be successfully tracked.

2. Rough entropy based spatial segmentation

Incompleteness of knowledge within an universe leads to granulation. A measure of the uncertainty in granulation quantifies this incompleteness of knowledge. Theory of Rough sets [13] has recently become a popular mathematical framework for granular computing, as the effect of the incompleteness of knowledge about a universe becomes evident only when an attempt is made to define a set in it. The focus of rough set theory is on the ambiguity caused by limited discernibility of objects in the domain of discourse. In case of gray level images, gray levels have limited discernibility due to inadequacy of contrast. That is why rough set and granular computing comes to be an effective tool for gray level image analysis.

2.1. Concept of rough set and image definition

Let $\mathcal{A} =< U, A >$ be an information system, where, U represents the universe and A represents the set of attributes. Let $B \subseteq A$ and $X \subseteq U$. The set X (where $X \subseteq U$) can be approximated using the information contained in B only and this could be done by constructing the lower and upper approximations of X. If $X \subseteq U$, the set $\{\mathbf{x} \in U : [\mathbf{x}]_B \subseteq X\}$ is known as *B-lower approximation* of X ($\underline{B}X$), i.e., this set will always be a subset of X. Similarly, the set $\{\mathbf{x} \in U : [\mathbf{x}]_B \cap X \neq \emptyset\}$ represents the *B-upper approximations* of X in U ($\overline{B}X$) which will always have a nonzero intersection with X. Here $[\mathbf{x}]_B$ denotes the equivalence class of the object $\mathbf{x} \in U$ relative to I_B (the equivalence relation). These are illustrated in Fig. 2, where the set of granules within red lined area represent the lower approximation of the object of interest, whereas the granules within the green lined area represent its upper approximation (For interpretation of the references to color in this sentence, the reader is referred to the web version of the article.). Therefore, a rough set is nothing but a crisp set with rough representation.

The roughness of a set X with respect to B is characterized numerically [13] as $R_I = 1 - (|\underline{B}X|/|\overline{B}X|)$.

Let the universe U be an image (I of size $M \times N$) consisting of a collection of pixels. Then if we partition U into a collection of non-overlapping windows of unequal size (of size $m_i \times n_i$, where i reflects the ith granule say), each window can be considered as a granule G_i. Object regions in the image can be approximated by rough sets depending on the granulation. Let it needs to be classified into object and background classes (O and B) and the gray level

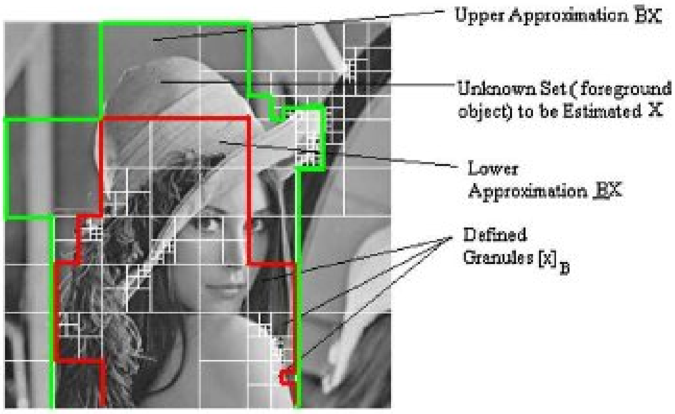
Upper Approximation $\overline{B}X$
Unknown Set (foreground object) to be Estimated X
Lower Approximation $\underline{B}X$
Defined Granules $[x]_B$

Fig. 2. Rough sets representation (with unequal granules) of a set (foreground object) with upper and lower approximations.

threshold be T. The object lower ($\underline{O}_T$) and upper ($\overline{O}_T$) approximations are constructed given the granulation as:

$\underline{O}_T$: *The set of the granules with all the pixel values grater than T*

$\overline{O}_T$: *The set of the granules with at least one pixel value grater than T*

The background lower ($\underline{B}_T$) and background upper ($\overline{B}_T$) approximations are constructed in similar manner. An example of this kind of approximation in a gray level image is shown in Fig. 2.

The rough set representation of the image (i.e., object O_T and background B_T) for a given granulation depends on the value of T.

The roughness of object O_T and background B_T are defined as [12]

$$\begin{aligned} R_{O_T} &= 1-\frac{|\underline{O}_T|}{|\overline{O}_T|}=\frac{|\overline{O}_T|-|\underline{O}_T|}{|\overline{O}_T|} \\ R_{B_T} &= 1-\frac{|\underline{B}_T|}{|\overline{B}_T|}=\frac{|\overline{B}_T|-|\underline{B}_T|}{|\overline{B}_T|} \end{aligned} \quad (1)$$

where | . | reflects the cardinality of a set. In the aforesaid discussion, we have considered the unequal sized granules instead of equal size as it was in [12]. The details about the formation of the granules have been discussed in Section 2.2.

2.1.1. Rough entropy measure and object extraction

Rough entropy (RE) of an image is defined as

$$RE_T=-\frac{BASE}{2}[R_{O_T}\ \ log_{BASE}(R_{O_T})+R_{B_T}log_{BASE}(R_{B_T})]. \quad (2)$$

This is a modified version of the one defined by Pal et al. [12]. We choose different values of base (*BASE* in Eq. (2)) instead of considering only "*e*". This equation provides the flexibility to choose the right value of the *BASE* for the image under consideration with the amount of noise present in it. Through experimentation, we found that for most of the images a *BASE* value of "10" is quite suitable. However one can choose the value "e" or "2" as the *BASE*. In this definition of rough entropy the maximum value of "1" will be attained at 1/(*BASE*), which is a function of *BASE*, providing the necessary flexibility (as all the entropy values for roughness <1/*BASE* is set as 1 to eliminate the noise) to compensate for the noise, while computing the rough entropy.

To detect the object, we have used the method of object enhancement/extraction based on the principle of minimizing the roughness of both object and background regions, for maximizing RE_T, with unequal granules. The RE_T, for every T, of the image is computed, representing the entropy of background and object regions $(0, \cdots, T)$ and $(T+1, \cdots, L-1)$, respectively. The value of T for which RE_T is maximum is selected as the optimum threshold to provide the object background segmentation. Note that maximizing the rough entropy to get the required threshold basically implies minimizing both the object *roughness* and background *roughness*.

2.2. Formation of granules

A granule is a clump of objects (points), in the universe of discourse, drawn together by indistinguishability, similarity, proximity, or functionality [22]. There exists several methods about the formation of a granule for measuring the ambiguity in images [17,20]. In the present article the unequal granules are formed by drawing the pixels of an image together based on their spatial adjacency [19], as well as gray level similarity. The granulation is based on quad-tree decomposition of each frame of the video sequence. The quad-tree decomposition of the image is performed based on the gray level difference among the pixels, within a window of the image. That is, the image will be divided into granules as long as the difference between the maximum and minimum gray values in a granule is greater than a certain threshold, in this way both the gray level and spatial ambiguities are taken care of. The first and the third quartiles of the image gray level distribution (denoted as Q_1 and Q_3) are considered to calculate the threshold (*GrTh*) for granule detection with

$$GrTh=\frac{Q_3-Q_1}{2}. \quad (3)$$

The resulting granules, thus formed automatically using image statistics, are unequal in size which is more appropriate and natural for real life problems.

Note: In the proposed method the quad-tree decomposition is done only once and the same decomposition is considered for the consecutive frames. Moreover, the threshold detection problem is carried out first time on all possible values of gray-level, and in the consecutive frames the search for the threshold is limited only around that obtained in the previous frame. This two implementation strategies improve the speed of detection of object without deterioration in accuracy.

3. Background estimation based temporal segmentation

The aim of background estimation approach is to construct a background model from the available information of the video sequence so that the object can easily be separated. This process is also known as temporal segmentation, that is extracting the moving pixels from a video as a segment.

There are several methods to estimate the background of a video sequence [2]. Frame difference, median filter, linear predictive filter, approximate median filter, Kalman filter, mixture of Gaussian are the commonly used approaches to construct a background of a video sequence taken from a still camera. Here we are proposing a three point approximation based background estimation method.

3.1. Background estimation technique

In our method of temporal segmentation we have dealt with the statistical distribution of gray levels for one pixel in previous N frames. We can say that the video sequence is a discrete probability distribution of gray levels for every pixel. According to [6], when a random variable follows *Beta Distribution*, the mean and standard deviation of the variable can be estimated by three point estimation. We have applied this technique to estimate the background. The standard deviation of distribution for possible error due to randomness is considered. These two (i.e., mean and standard deviation) together helps in characterizing the background. The estimated background is subtracted from target frame to delineate the object.

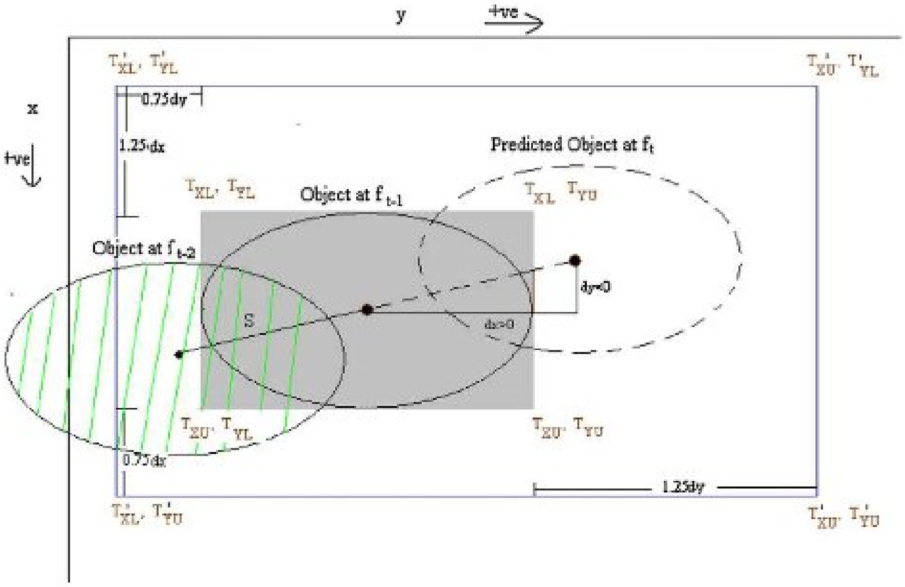

Fig. 3. Pictorial representation of the defined tracker with labeled vertices.

3.1.1. Three point estimation techniques

The beta distribution along with triangular distribution [6] can also be used to model events which are constrained to take place within an interval defined by a optimistic and pessimistic value. In PERT (Program Evaluation and Review Technique) this property of beta distribution was extensively used. The technique evolved around the approximations for the mean and standard deviation of a Beta distribution. Here the three points used to estimate the mean and the standard deviation are named as optimistic point, most likely point and pessimistic point.

In our proposed approach of temporal segmentation we have used previous three frames to estimate the background and standard deviation. Let the previous three frames of the tth frame be denoted by f_{t-1}, f_{t-2} and f_{t-3} respectively. Our estimation is based on the aforesaid three point distribution, where,

$$a = max(f_{t-1}, f_{t-2}, f_{t-3}) \tag{4}$$

$$b = median(f_{t-1}, f_{t-2}, f_{t-3}) \tag{5}$$

$$c = min(f_{t-1}, f_{t-2}, f_{t-3}) \tag{6}$$

a, b and c denote the optimistic, most likely and pessimistic points respectively. The mean and standard deviation are defined accordingly as:

$$\widehat{f_t} = \frac{a + 4b + c}{6} \tag{7}$$

$$\widehat{\sigma}_t = \frac{a - c}{6} \tag{8}$$

In Eq. (7), $\widehat{f_t}$ is the estimated background for frame f_t. A pixel $f_t(x, y)$ will be detected as a foreground pixel if its difference from the estimated value is greater than E (normally chosen as 3, as it is found experimentally suitable) times of variance. Therefore, (x, y) will be a foreground pixel if

$$|f_t(x, y) - \widehat{f_t}(x, y)| > E\widehat{\sigma}_t \tag{9}$$

Otherwise, the pixel will be treated as a background pixel. The choice of a, b and c are relevant, because in a video sequence, it is natural to assume that the median values will be the most likely one and the fraction of the difference between maximum and minimum values of a distribution will be helpful to find the variance. Therefore such an approximation can give a better estimate of the background. It is also fast to approximate, because it is based on only three previous frames.

Note: Three points have been derived here from the previous three frames. However, one can also use more frames.

4. Target localization and tracking

In this section we show how the system combines both the spatial and temporal segmentation techniques and apply it to reconstruct the moving object in the target frame and track it. At first, the reference model from the first frame is marked. After having the two (spatial and temporal) segmented images of the target frame, we have taken the *intersection* of the two segments to get the object region which is common to both. The statistics of this common region (as this region will definitely belong to the object and likely to have the color variation of several object regions due to containing the boundary parts) is estimated in the RGB feature space. Here, we used the mean (mn) and maximum deviation (max_dev) of this intersecting region in the RGB feature space. After having these information, we place a tracker on the current frame within which the pixels will be scanned. The size of the tracker depends on the spatial shift of the object centroid and its direction of movement between the previous two frames with the location of center the same as that was in the previous frame. Let, the locations of the centroid in $(t-1)$th and $(t-2)$th frames be (x_c, y_c) and (x'_c, y'_c) respectively. The Euclidian distance (the distance s) between two of them will determine the size of the tracker (T_x, T_y) and the signs of $d_x = (x_c - x'_c)$ and $d_y = (y_c - y'_c)$ will determine the direction of the object movement. That is, we assume that after changing the location from the previous frame to the target frame, the whole object should be within the area considered by the tracker. This is done by giving more importance to the direction in which change occurs. The details of it is given in Section 4.1. Then we start to scan each pixel within the tracker, which belongs to either spatial or temporal segment (i.e., *union* of both segments within the tracker). The feature-wise difference between the value of the pixel and the mean (mn) computed for each pixel are checked. If the difference is less than the maximum deviation (i.e., max_dev) then that pixel is considered as a part of the object, otherwise of background.

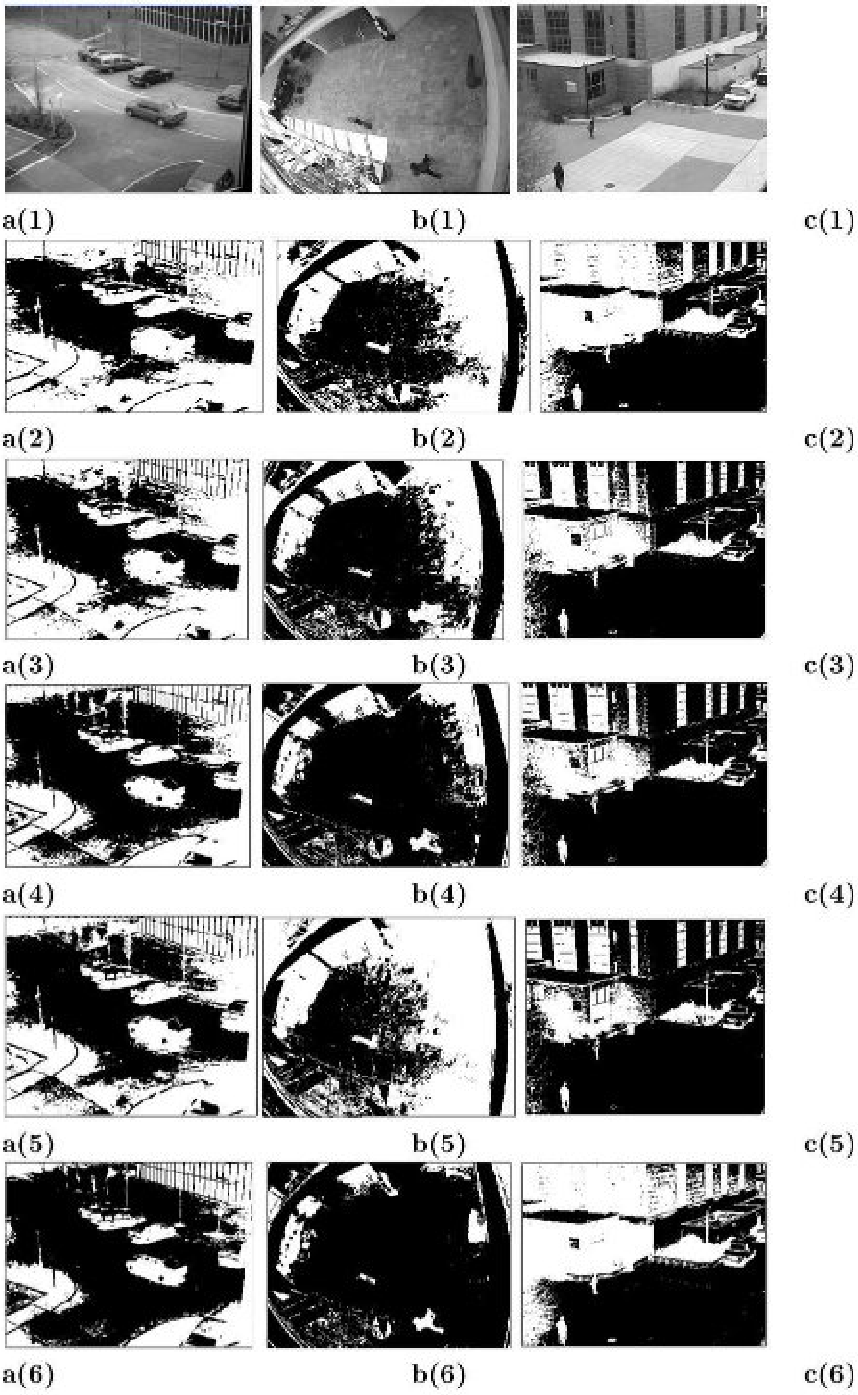

Fig. 4. Spatial segmentation results on (a) frame no. 135 of the *Surveillance Scenario Sequence* from PETS-2000, (b) frame no. 56 of the *Walk3 sequence* from PETS-2004 and (c) frame no. 370 of *Dataset03* from OTCBVS-2007 (1) original, (2) Otsu's thresholding, (3) RE with 4 × 4 granule, (4) RE with 6 × 6 granule, (5) RFE and (6) proposed method.

4.1. Algorithm for reconstruction and tracking

The object detection and tracking starts with selection of a tracker. Here we have considered a rectangular tracker, which completely covers the object that is to be tracked. The initial tracker is defined as a rectangular box, whose vertices are given as ((T_{XL}, T_{YL}), (T_{XU}, T_{YL}), (T_{XU}, T_{YU}) and (T_{XL}, T_{YU})), shown in Fig. 3. Then the following are the steps to detect the object and to track it. In the beginning of the algorithm the first frame is considered as the reference frame. Then it is segmented into regions and the object of interest is marked.

Step 1: Convert the input color (RGB) image (I) to gray level image (Y) by using the equation: $Y = 0.3R + 0.59G + 0.11B$.

Step 2: Apply the rough entropy based image segmentation. (*Note*: Here, the advantage of considering a window around the threshold in the previous frame is exploited for segmentation of video images in the current frame.)

Step 3: Do a temporal segmentation according to three point estimation (Section 3.1).

Step 4: Design the tracker according to the following:$IF\, d_x > 0$ then, $T_{xu} = T_{xu} + 1.25s$ and $T_{xl} = T_{xl} - 0.75s\, ELSE\, T_{xu} = T_{xu} + 0.75s$ and $T_{xl} = T_{xl} - 1.25s IF\, d_y > 0$ then, $T_{yu} = T_{yu} + 1.25s$

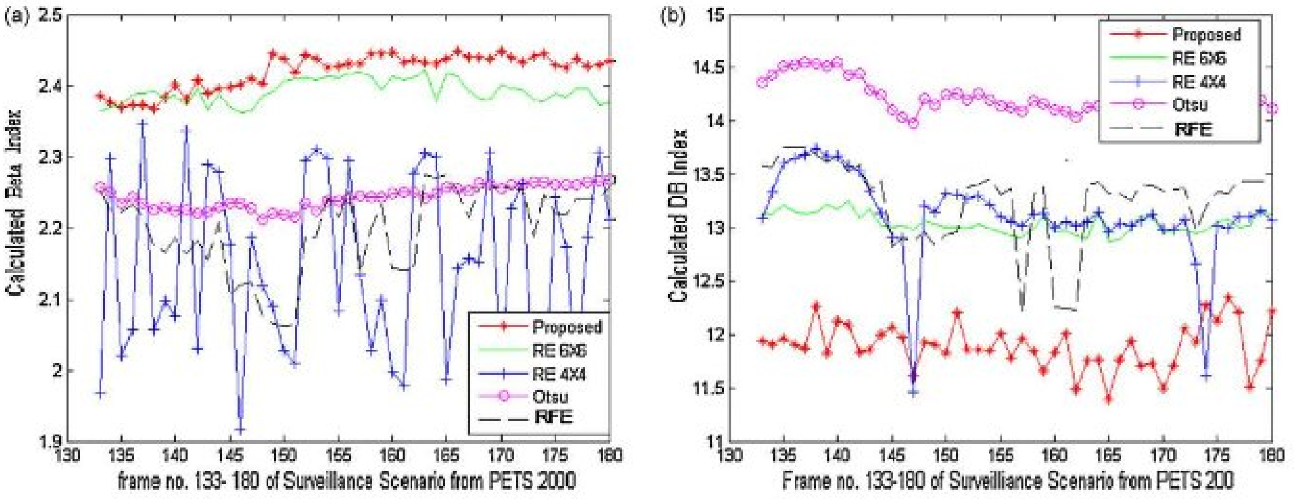

Fig. 5. Plot of (a) beta index (β) and (b) DB index with frames for Otsu's thresholding, RE with 4 × 4 and 6 × 6 granules, RFE and proposed method.

and $T_{yl} = T_{yl} - 0.75s$ *ELSE* $T_{yu} = T_{yu} + 0.75s$ and $T_{yl} = T_{yl} - 1.25s$.Here, $T_x = T_{xu} + T_{xl}$ and $T_y = T_{yu} + T_{yl}$.

Step 5: The intersection of the two segmented images within the tracker is considered in RGB feature space and the mean (mn) and maximum deviation (max_dev) of those points are calculated in the same feature space.

Step 6: For every pixel (x, y) within the tracker, which belongs to either spatial or temporal segmented region, do the following:*IF* $|f_R(x, y) - mn| < max\_dev_R$ and $|f_G(x, y) - mn| < max\_dev_G$ and $|f_B(x, y) - mn| < max\_dev_B$ then $F(x, y) = I(x, y)$ *ELSE* $F(x, y) = 0$ (here, F is the detected object image).

Step 7: Redefine the tracker around the detected object for tracking.

Step 8: Repeat Steps 1–8, for all the frames in the video sequence for tracking of the moving object. ♠

5. Results and discussions

Experiments along with comparisons were conducted to evaluate the effectiveness of the proposed algorithm in (a) spatial segmentation, (b) temporal segmentation, (c) reconstruction, and (d) tracking. We have performed our experiments with different types of data sets: (1) *Surveillance Scenario* from PETS-2000 [15], (2) *Walk3* from PETS-2004 [16], (3) *Dataset03* from OTCBVS-2007 [4]. However, to limit the size of the paper, we have shown the results on some of the frames of these video sequences, only.

At first we present the results of spatial segmentation and temporal segmentation, and demonstrate their comparative performance with other existing methods. This is followed by the results of reconstruction and tracking.

5.1. Results of spatial segmentation

In Fig. 4, we have shown some results of bi-level segmentation performed on three frames of the aforesaid video sequences. Here one individual frame of a video sequence is considered as a still image. For the sake of comparison we have performed Otsu's thresholding, rough entropy maximization based thresholding (RE) with uniform granules [12] of sizes 4 × 4 and 6 × 6 and rough fuzzy entropy based segmentation (RFE) [17] with crisp granule, crisp set and 6 × 6 granule size. Fig. 4 shows the comparative segmented outputs. Since in our proposed method with unequal granules, the smallest size was found to be 5 × 5, we showed the results corresponding to equal sized granules of 4 × 4 and 6 × 6 only (as they are closest to 5 × 5) when comparing with rough set theoretic segmentation.

In Fig. 4, it can be seen visually that the proposed segmentation method can separate out the object of interest more clearly than the other methods. Though, rough entropy method with granules of size 6 × 6 gives equally good results in some of the cases, the size of the granule is to be chosen experimentally [12], which is not much practically feasible. It can be seen that thresholding by Otsu's method is the poorest in comparison to other three results.

To compare the performance of the techniques quantitatively we compute the β index [11] and DB index [3] for all frames (frame nos. 133–180) of *Surveillance Scenario* [15] video sequence. We know that, for a given no. of classes the higher value of β and lower the value of DB are desirable for good segmentation.

In our example the no. of classes was taken as two. The variation of β values and DB values is shown in Fig. 5 (a) and (b) respectively.

In Fig. 5 the '*-' line denotes β and DB index obtained by the proposed method, the '-' and '+-' lines denote the β values and DB values obtained by RE method with 6 × 6 and 4 × 4 granules respectively. The '–' line denotes the values of the two indexes according to RFE. The 'o-'denotes the β and DB values obtained after segmentation by Otsu's thresholding. As β-values are higher and DB values are lower in case of our method, we can say that, the proposed algorithm is also superior quantitatively in the current scenario.

5.2. Results of temporal segmentation

Fig. 6 shows comparative performance of our method with one of the most popular techniques: (MoG) [18], and a change detection technique: linear change detection (LDD) [5]. LDD has been implemented with 5 × 5 window, 0.05 threshold and the frame before the previous frame as the reference frame. In case of our technique the noise present in the segmented images is much less compared to other two methods and the object in the current frame can be detected properly as seen from Fig. 6. (One may notice that, if an ideal reference frame was available in case of LDD, the results could be better, but, it is not available in the current scenario.) The boundaries of the objects are seen to be clearly extracted with our method almost covering the region of interest. This shows that the proposed technique is more efficient, accurate and simpler.

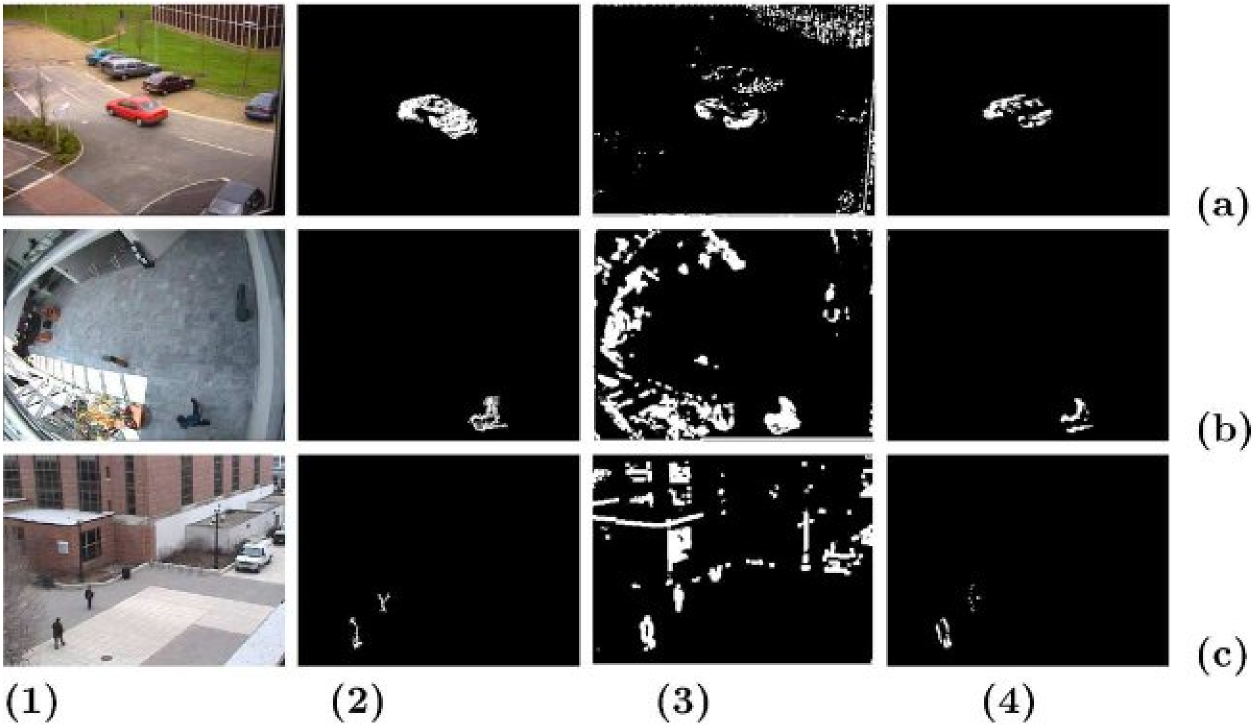

Fig. 6. Results of temporal segmentation by (2) MoG, (3) LDD and (4) proposed method on (a) frame no. 145 of the *Surveillance Scenario Sequence* from PETS-2000, (b) frame no. 52 of *Walk3 sequence* from PETS-2004, and (c) frame no. 448 of *Dataset03* from OTCBVS-2007.

5.3. Results of reconstruction

We have implemented the algorithm on several video sequences and got satisfactory result, i.e., the object/objects of interest can be reconstructed properly (very close to the ground truth). In Fig. 7, we have shown only three of such video sequences, one frame from each sequence along with its ground truth. We have also evaluated the performance of the reconstruction results by calculating the values of *Precision*, *Recall* and F_{score} for the sequences.

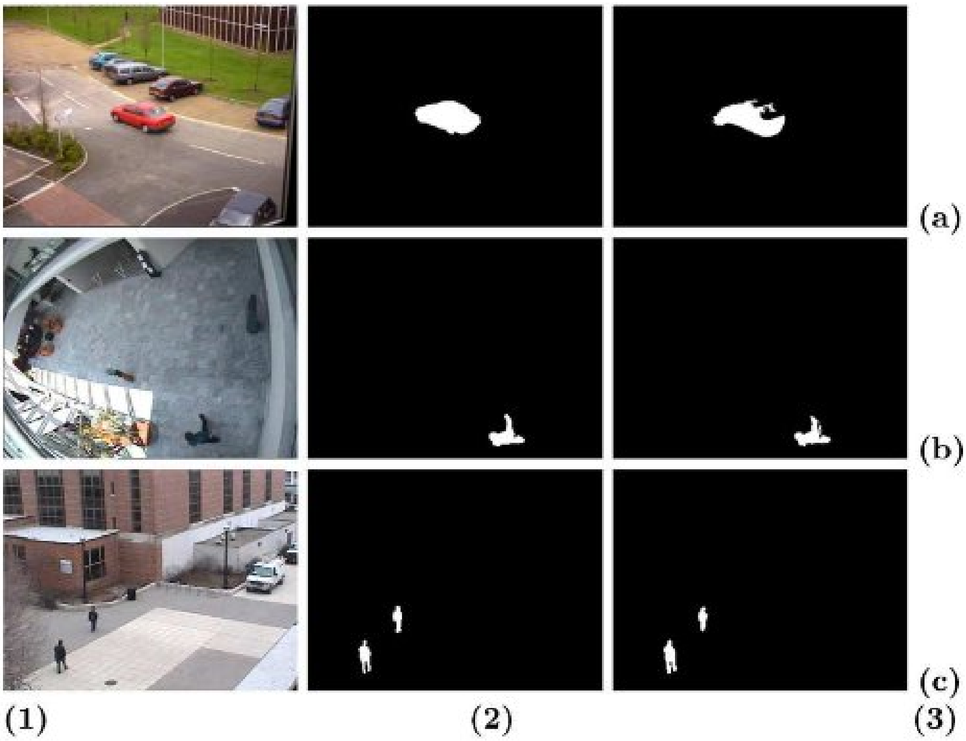

Fig. 7. Results of validation (1) original frame, (2) ground truth, and (3) reconstruction by proposed method of (a) frame no. 145 of the *Surveillance Scenario Sequence* from PETS-2000, (b) frame no. 52 of the *Walk3 sequence* from PETS-2004, and (c) frame no. 448 of *Dataset03* from OTCBVS-2007.

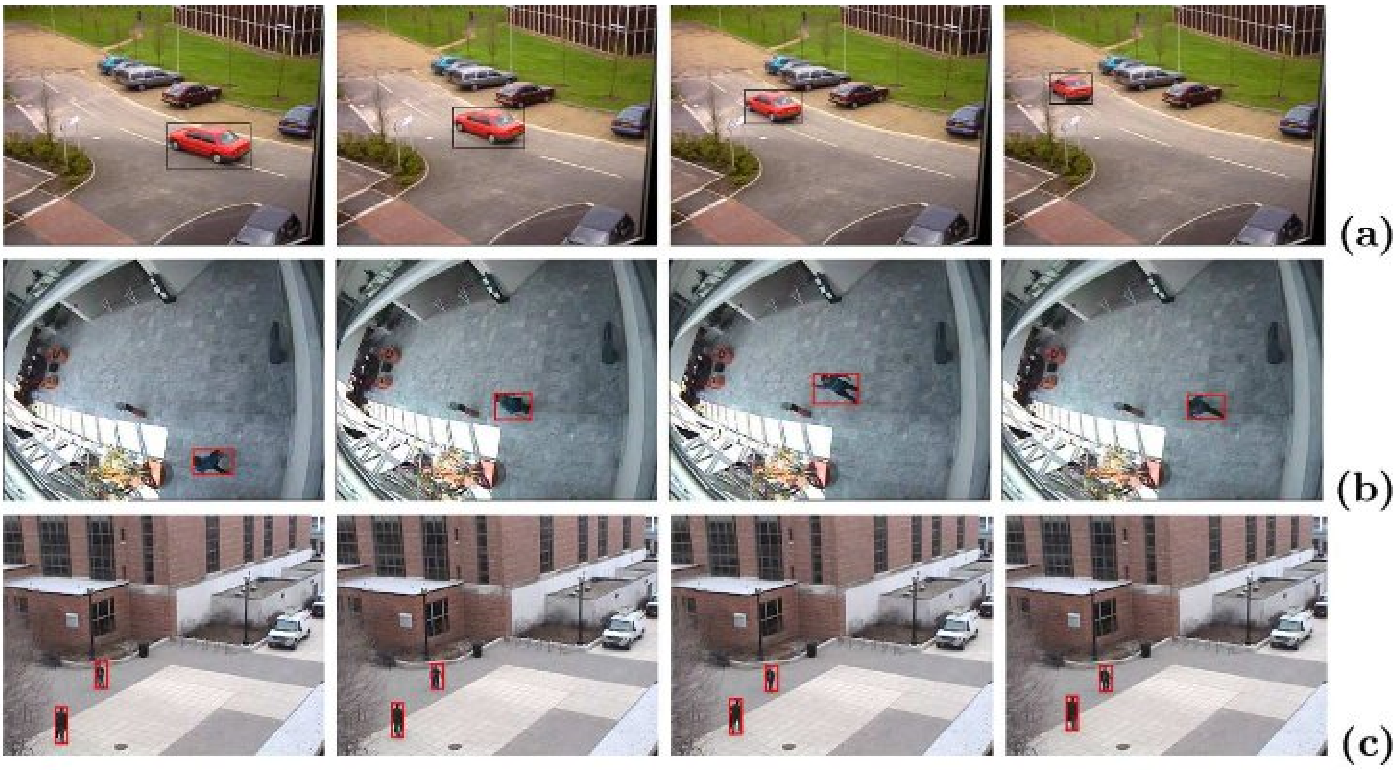

Fig. 8. Results of tracking (a) frame nos. 133, 145, 161, 180 of the *Surveillance Scenario Sequence* from PETS-2000, (b) frame nos. 58, 87, 101, 159 of *Walk3 sequence* from PETS-2004, and (c) frame nos. 416, 436, 462, 478 of *Dataset03* from OTCBVS-2007.

From Table 1, it can be said that the proposed algorithm is very effective for reconstruction of the object/objects of interest. All the values of *Precision*, *Recall* and F_{score} obtained for the different kinds of sequences with single or multiple moving objects are grater than 0.85; signifying high accuracy of the algorithm.

Note that here we have taken four frames from each video sequence. *Recall*, *Precision* and F_{score} are computed for each frame. An average of these four frames is reported in Table 1. However the results will be similar for the other frames also.

5.4. Results of tracking

Here, the results of tracking by implementing the algorithm on four frames from each of the aforesaid sequences are shown. Each of the sequences has different types of object with different kinds of movement. In the first sequence (Fig. 8(a)), i.e., the *Surveillance Scenario* from PETS-2000 [15], a car is moving through out the frame and its shape and size are changing gradually. In the second sequence (Fig. 8(b)) *Walk3* from PETS-2004 [16], a man is walking through the h-line, getting stopped, waving his hand and then returning back. There are two objects moving in different directions in the third sequence (Fig. 8(c)) *Dataset03* from OTCBVS-2007 [4]. In all the sequences the object/objects are being tracked properly which shows the effectiveness of the algorithm on videos of different characteristics.

Note that, the algorithm of tracking is more robust to gradual illumination change and noise. Because, even if those occur in a certain frame it could not affect both the segments of that frame within the search window.

Table 1
Reconstruction accuracy on *Surveillance Scenario*, *Walk3* and *Dataset03* sequences.

Sequences	*Recall*	*Precision*	F_{score}
Surveillance Scenario	0.85	0.86	0.85
Walk3	0.89	0.94	0.91
Dataset03	0.86	0.95	0.90

6. Conclusions

A spatio-temporal video segmentation approach for object(s) detection and tracking is described. The method uses granular computing and rough entropy for spatial segmentation and a three point estimation for temporal segmentation. The combination of these two segmentation results is used for the detection, reconstruction and tracking of the object(s) in video sequences. Image granulation is performed using quad-tree decomposition resulting in unequal granules which is closer to natural granulation. Here, we have used gray level as a feature for formation of granules. One may use any other feature (color, texture, etc. or combination of them) depending on the application.

The proposed spatial segmentation technique is seen to perform well on several kinds of video images. The method is faster as the gray levels for optimal threshold are searched only within a window of the threshold value obtained in the previous frame. Results with unequal granules are found to be superior both visually and quantitatively than those of three other techniques. β index and DB index also reflect it well. The temporal segmentation results are superior to MoG and LDD as less amount of noise is present there. The reconstruction results are validated with the ground truths and it has an accuracy of more than 85% for all the sequences under consideration, particularly in some cases it is as high as 91%. The method is robust to noise and gradual illumination change as it judiciously combines both spatial and temporal segments for detection.

The overall methodology is less complex and results in an accurate and efficient algorithm for tracking. The novelty of the investigation mainly lies with formation of granules, background estimation with three point approximation, and using both color and shift information of object(s) during reconstruction.

Acknowledgement

S.K. Pal acknowledges the J.C. Bose National Fellowship, Government of India.

References

[1] O. Castillo, P. Melin, W. Pedrycz, Design of interval type-2 fuzzy models through optimal granularity allocation, Applied Soft Computing 11 (8) (2011) 5590–5601.

[2] S.-C.S. Cheung, C. Kamath, Robust background subtraction with foreground validation for urban traffic video, EURASIP Journal on Applied Signal Processing 2005 (2005) 2330–2340.

[3] D.L. Davies, D.W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence 1 (1979) 224–227.

[4] E.O.W.S.B.J. Davis, V. Sharma, Background-subtraction using contour-based fusion of thermal and visible imagery, Computer Vision and Image Understanding 106 (2007) 162–182.

[5] E. Durucan, T. Ebrahimi, Change detection and background extraction by linear algebra, Proceedings of IEEE 89 (10) (2001) 1368–1381.

[6] D.L. Keefer, S.E. Bodily, Three-point approximations for continuous random variables, Management Science 29 (1983) 595–609.

[7] Madalena, A. Petrosino, A self organised approach for background subtraction for visual surveillance applications, IEEE Transactions on Image Processing 17 (2008) 1070–1077.

[8] E. Maggio, A. Cavallaro, Video Tracking – Theory and Practice, Wiley, West Sussex, UK, 2010.

[9] D.G. Malcolm, J.H. Roseboom, C.E. Clark, W. Fazar, Application of a technique for research and development program evaluation, Operations Research 7 (1959) 646–669.

[10] S.K. Meher, S.K. Pal, Rough-wavelet granular space and classification of multispectral remote sensing image, Applied Soft Computing 11 (2011) 5662–5673.

[11] S.K. Pal, A. Ghosh, B. Uma Shankar, Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation, International Journal of Remote Sensing 21 (2000) 2269–2300.

[12] S.K. Pal, B. Uma Shankar, P. Mitra, Granular computing, rough entropy and object extraction, Pattern Recognition Letters 26 (2005) 2509–2517.

[13] Z. Pawlak, Rough Sets: Theoretical Aspects of Reasoning about Data, Kluwer Academic Publishers, Norwell, MA, USA, 1992.

[14] W. Pedrycz, Granular Computing: An Emerging Paradigm, Physica-Verlag, Heidelberg, 2001.

[15] PETS-2000, IEEE Int. WS Perfor. Evaluation of Tracking and Surveillance, 2000.

[16] PETS-2004, IEEE Int. WS Perfor. Evaluation of Tracking and Surveillance and EC Funded CAVIAR project/IST 2001, 2004.

[17] D. Sen, S.K. Pal, Generalized rough sets, entropy, and image ambiguity measures, IEEE Transactions on Systems, Man, and Cybernetics, Part B 39 (2009) 117–128.

[18] C. Stauffer, W.E.L. Grimson, Adaptive background mixture models for real-time tracking, in: Proc. Computer Vision and Pattern Recognition, IEEE Computer Society, 1999, pp. 246–252.

[19] B. Uma Shankar, D. Chakraborty, Spatiotemporal approach for tracking using rough entropy and frame subtraction, in: Proceedings of the 4th International Conference on Pattern Recognition and Machine Intelligence, PReMI'11, Springer, 2011, pp. 193–199.

[20] Y.Y. Yao, Granular computing: basic issues and possible solutions, in: Proceedings of the 5th Joint Conference on Information Sciences, 2000, pp. 186–189.

[21] A. Yilmaz, O. Javed, M. Shah, Object tracking: a survey, ACM Computing Surveys 38 (4) (2006) 1264–1291.

[22] L.A. Zadeh, Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic, Fuzzy Sets and Systems 90 (September) (1997) 111–127.