Neighborhood Rough Filter and Intuitionistic Entropy in Unsupervised Tracking

Debarati Bhunia Chakraborty and Sankar K. Pal, *Life Fellow, IEEE*

***Abstract*—This paper aims at developing a novel methodology for unsupervised video tracking by exploring the merits of neighborhood rough sets. A neighborhood rough filter is designed in this process for initial labeling of continuous moving object(s) even in the presence of several variations in different feature spaces. The locations and color models of the object(s) are estimated using their lower–upper approximations in spatio-color neighborhood granular space. Velocity neighborhood granules and acceleration neighborhood granules are then defined over this estimation to predict the object location in the next frame and to speed up the tracking process. A novel concept, namely, intuitionsistic entropy is introduced here, which consists of two new measures: neighborhood rough entropy and neighborhood probabilistic entropy to deal with the ambiguities that arise due to occurrence of overlapping/ occlusion in a video sequence. The unsupervised method of tracking is equally good even when compared with some of the state-of-the art partially supervised methods while showing superior performance during total occlusion.**

***Index Terms*—Entropy, granular computing, Internet-of-Things (IoT), neighborhood granules, object estimation, occlusion handling, rough filter, rough sets, unsupervised video tracking.**

I. Introduction

VIDEO tracking, i.e., tracking of moving object(s) from video sequences is one of the basic steps in computer vision. Object tracking is required in several fields of computer vision, e.g., surveillance, gesture recognition, event detection, and behavior recognition. The tracking problem has been studied over decades [1]–[4] and there exist several literatures for supervised, partially supervised (with initial manual interactions, i.e., tracking by detection) [5]–[11], and unsupervised (without manual interactions, mostly based on background estimation) [12]–[15] tracking. One may note that in video tracking the complete information on object-background is not always available. This makes the prediction difficult as there exist several uncertainties arising from, e.g., changes in shapes/sizes of moving object(s), changes in motion of the object(s), and changes in the number of object(s), occurrence of object to object overlapping, and occurrence of total/partial occlusion to background. These problems have been recently addressed and several approaches proposed using statistical modeling, local features extraction, etc., [6], [8], [16]–[18]. But, the initial object(s) need to be labeled manually in all those approaches. Besides, these methods may fail to estimate the locations of the object(s) during their total occlusion to background. The method described in this paper does not need any initial manual interaction and proves to be effective in handling total occlusion. The merits of neighborhood rough set (NRS) theories are exploited here to model the uncertainties, which are present in the tasks of unsupervised initial labeling and tracking.

Theory of NRS is an extension of classical rough sets, which was defined by Pawlak, [19] to deal with the uncertainties or incompleteness of knowledge arising from the limited discernibility of objects in the domain of discourse. The key concepts of Pawlak's rough set (PaRS) are object "indiscernibility" and "set approximation." These characteristics made the theory useful in several areas of pattern recognition and machine learning, e.g., feature reduction and selection [20], image processing [21]–[23], and data mining and knowledge discovery [24]–[27]. The classical rough set is enriched in several ways by addition of some more features to it. The theories, such as rough-fuzzy sets (RFS) [21], [23], fuzzy-rough sets (FRS)[28], and NRS [29], [30], are introduced in that line to make this technology more robust in handling uncertainties. All of them can efficiently model the ambiguities that arise due to the overlapping nature of the data. Although the nature of overlapping needs to be known in RFS or FRS, it is not required in NRS. This phenomenon of NRS makes it more applicable to unsupervised tasks. NRS has recently been used in image processing but its merits in video processing have not yet been explored. In this paper, the moving object(s) are modeled using the concept of NRS to take care of the overlapping nature of datasets. The new uncertainty measures: neighborhood rough entropy (NRE), neighborhood probabilistic entropy (NPE), and intuitionists entropy, defined over NRS, enable in quantifying the ambiguities.

While modeling the overlapping nature of the data, NRS uses granules (clump of objects) formed over its every point. It may be mentioned here that the concept of granulation was first introduced by Zadeh [31] and it is the primary step in granular computing. Granules so formed over every point take into account their neighborhood as well as the overlapping characteristics of the data. While doing so, their sizes or shapes need to be defined. But, natural granulation is arbitrary and it does not

Manuscript received January 11, 2017; revised June 7, 2017 and September 4, 2017; accepted October 25, 2017. Date of publication November 2, 2017; date of current version August 2, 2018. The work of S. K. Pal was supported in part by J. C. Bose National Fellowship (Grant-in-Aid) and in part by DAE-BRNS Senior Scientist Scheme (RRF) of Government of India. *(Corresponding author: Debarati Bhunia Chakraborty.)*

The authors are with the Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, India (e-mail: debarati.earth@gmail.com; sankar@isical.ac.in).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TFUZZ.2017.2768322

have any fixed shape or size. In this paper, an attempt of using NRS in forming such natural granules in the frames of video sequences is made. The similarities in both color and spatial feature spaces are considered during this formulation, and the granules thus formed make the processing faster. An NRS filter is developed in this granulated space for initial labeling of the objects. Two types of temporal neighborhood granules, namely, velocity and acceleration granules are defined using the output of the filter to model the nature of motion of object(s).

The proposed method works as follows. Given an input video sequence, spatio-color neighborhood granules are formed over all the frames. Its initial P number of frames are used for initial object labeling with the proposed NRS filter. The color model and velocity profiles of the objects are the outputs of this filter from which the velocity and acceleration granules are generated. The object location in the next frame is estimated based on these information and the roughness in estimation is checked. If it is high then the intuitionistic entropy is computed for the boundary granules and the ones that are found ambiguous are eliminated in order to track the object in the frame accurately. The tracking results, thus produced, are comparable to those of some of the state-of-the-art algorithms, while superior in case of total occlusion.

The novelty of the investigation mainly lies in as follows:

1) forming spatio-color neighborhood granules, and velocity and acceleration granules over video frames;
2) developing NRS filter for unsupervised object estimation;
3) defining NRE by incorporating the overlapping phenomenon of datasets;
4) introducing the concept of intuitionistic entropy for occluded object estimation.

The significance of these features is demonstrated extensively in the task of tracking continuous moving object(s) in static background under different kinds of changes in shapes/sizes of moving object(s), changes in speed, multiple moving objects moving in different directions, occurrence of object to object overlapping, and object to background occlusion.

This paper is organized as follows. In Section II, the basic concepts of rough sets (PaRS and NRS) are described. The proposed work is explained in details in Section III. This includes the underlying concepts and theories of spatio-color granules, neighborhood rough filter, temporal granules, and intuitionistic entropy. Section IV contains the proposed algorithms. In Section V, all these characteristics and their effectiveness are experimentally demonstrated along with several comparisons. The overall conclusion of this paper is drawn in Section VI.

A. Related Work

The problems of multiple object tracking and handling their overlapping have recently become a challenging task in video tracking. This problem was primarily addressed by Reid [32], where the current state is estimated from previous frames using Kalman filter. Later, particle filtering (also known as sequential Monte Carlo) was introduced, where a set of weighted particles sampled from a proposal distribution was maintained to represent the current and hidden states [33], [34]. This allows handling nonlinear multimodal distributions.

Kernel-based tracking [35] is another popular method in visual tracking of nonrigid objects. Here the feature "histogram-based target representation" is regularized by spatial masking with an isotropic kernel. Recently a new kernelized correlation filter [13] is derived to handle the problems of multitarget tracking and occlusion. Unlike other kernel algorithms, it has the same complexity as its linear counterpart.

Sparse-based representation of object templates has become quite a popular approach to handle the ambiguities in videos. A sparse weight constraint is introduced in [16] to dynamically select the relevant templates from the full set of templates. An algorithm for tracking multiobjects with occlusion based on the multifeature joint sparse reconstruction is described over these templates. In [18], an object is represented by the sparse coefficients of local patches based on an over-complete dictionary, and a classifier is learned to discriminate the target object from the background. In [9], the sparse representation and online dictionary learning are unified by defining a sparsity consistency constraint that facilitates the generative and discriminative capabilities of the appearance model. An elastic-net constraint is enforced during the dictionary learning stage to capture the characteristics of the local appearances that are insensitive to partial occlusions.

There are several other approaches aimed to solve the aforesaid problems. The methodology developed in [17] is in the form of a binary integer programming problem where a polynomial time solution approach is proposed that can obtain a good relaxation solution of the binary integer programming. This is then applied for multitarget tracking. In [8], an object representation based on the weakly aligned multi-instance local features is defined, which demonstrates that this representation improves on the inherent limit of local features invariance under occlusion. An energy function is developed in [6] for tracking considering some physical constraints, such as target dynamics, mutual exclusion, and track persistence. In addition, partial image evidence is handled with explicit occlusion reasoning, and different targets are disambiguated with an appearance model. A few deep learning based methods have recently come up to handle the aforesaid challenges. A single convolutional neural network (CNN) is developed in [14] for online tracking by learning the effective feature representations of the target object. A two-layer CNN is introduced for object representation in [15].

The limitation of all these methods is that most of them cannot work without initial labeling of the object of interest. Besides, these methods may fail in detecting the moving object(s) when these are totally occluded to background. The proposed method for tracking the continuous moving object(s) in a sequence does not need any initial manual labeling. This method is proved to be effective in handling even the cases like total occlusion or overlapping.

II. Preliminaries

Here we describe, in brief, some of the characteristics of rough sets and NRS, that are relevant to our investigation.

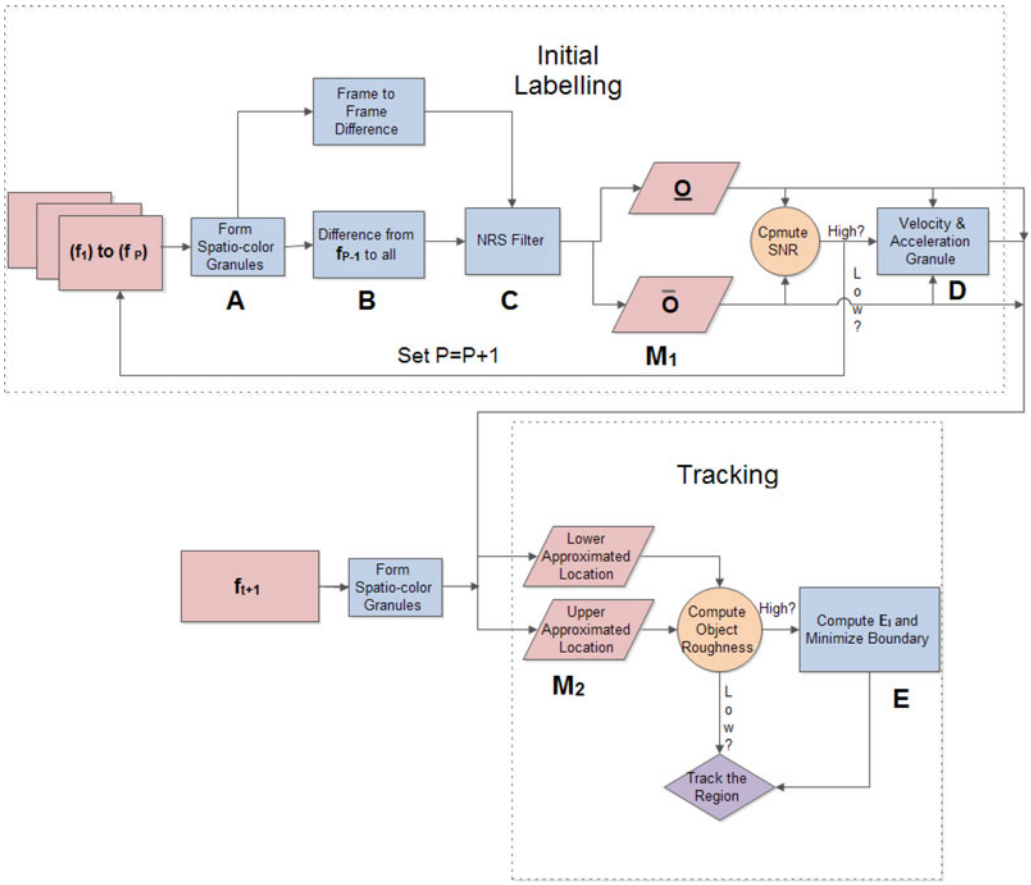

Fig. 1. Block diagram of the proposed method.

A. Rough Set

Rough set, as introduced by Pawlak (PaRS) in 1981, shows the way to divide the universe into several non overlapping segments in a nominal feature space. Let $\mathcal{A} =< U, A >$ be an information system, and let $R \subseteq A$. The set X (where $X \subseteq U$) can be approximated using only the information contained in R by constructing the lower and upper approximations of X. If $X \subseteq U$, the sets $\{x \in U : [x]_R \subseteq X\}$ and $\{x \in U : [x]_R \cap X \neq \emptyset\}$, where $[x]_R$ denotes the equivalence class of the object $x \in U$ relative to I_R (the equivalence relation), are called the *R-lower* and *R-upper approximations* of X in U. They are denoted by $\underline{R}X$ and $\overline{R}X$, respectively.

B. Neighborhood Rough Set

In case of real-life data there are nominal as well as numerical attributes. In numerical feature space, the concept of neighborhood plays an important role [29], [30]. Overlapping is also a usual property in real-life numerical feature space. The concept of NRS [29] comes from these properties of data in heterogeneous feature space. The neighborhood of a point $x_i \in U$, where U is the universe, is expressed as

$$\aleph(x_i) = \{x_j \in U : \triangle(x_i, x_j) \leq \delta\}. \tag{1}$$

In (1), $\triangle$ is a distance function and δ is a threshold dependent on the dataset that may be the same for all points or may vary. So, the granules, defined here around points, are overlapping. Therefore, unlike [19] PaRS, a data point may belong to more than one granule in NRS. In other words, the constituting points of a granule may be shared by other granules too.

There are two types of approximations: point based (type 1) and granule based (type 2) denoted as R_1X and R_2X, respectively. These are constructed as follows:

$$\underline{R_1}X = \{x \in U : \aleph(x) \subseteq X\} \tag{2a}$$

$$\overline{R_1}X = \{x \in U : \aleph(x) \cap X \neq \emptyset\} \tag{2b}$$

$$\underline{R_2}X = \{\aleph(x) \in U : \aleph(x) \subseteq X\} \tag{2c}$$

$$\overline{R_2}X = \{\aleph(x) \in U : \aleph(x) \cap X \neq \emptyset\}. \tag{2d}$$

One may note that R_1 is defined in terms of belonging of points (x) whereas R_2 is defined in terms of belonging of the granules defined around x. If $\underline{R_1}X = \underline{R_2}X$ and $\overline{R_1}X = \overline{R_2}X$, then there is no overlapping neighborhood granules, i.e., granules are crisp, and NRS boils down to PaRS.

III. Proposed Work

A. Overview of the Method

A brief overview of the proposed methodology for unsupervised tracking is shown in Fig. 1. The method has two parts, namely, initial labeling and tracking. Initial P number of frames

of a sequence are the input for initial labeling. Spatio-color granules are formed over these frames as shown in block A. Two types of frame-to-frame difference (between the consecutive, and from the Pth to all its previous frames) are computed in the granulated space as shown in block B. NRS-filter-based object estimation is then performed (block C), and the lower–upper sets of objects are formed (block M1). The temporal granules are generated using these estimated object sets (block D). These information are utilized during the task of tracking in the next frame (f_{t+1}) onward. During tracking, as shown in the figure, the input frame is first granulated and the locations are estimated on the frame (block M2) and the roughness in estimation is checked. If the roughness is high, then the intuitionistic entropy is computed for the boundary granules (block E) to ensure their presence in the object set, after which the object set is tracked.

These blocks are described in details in the following sections.

B. Formation of Spatio-Color Neighborhood Granules

The spatial and color nearness are incorporated to form spatio-color granules in each input video frame (block A). Let f_p be an input video frame and x_i be a pixel in this frame. If x_i does not belong to any other granule then a granule $\aleph(x_i)$ over this point is formed as

$$\aleph_{\text{sp-clr}}(x_i) = \left\{ \bigcup x_j \in U \right\} \tag{3}$$

where x_i and x_j are binary connected over the condition: $|\text{color}(x_j) - \text{color}(x_i)| < \text{Thr}$. That is, x_i and x_j are spatially adjacent.

It means, the granule around x_i contains those pixels for which the color difference from x_i is less than Thr and which are binary connected over this condition. Thr is the color nearness threshold. That is, the granules are formed considering color nearness over spatial feature space, and therefore are of arbitrary shape and size, unlike the existing approaches [29], [30] based on (1). The neighborhood points, as defined using (3), with respect to both the spatial and color similarities, would fall into the same granules. Note that the center point x_i over which a granule is formed cannot belong to any other granule. In this way, the formation of unnecessary overlapping granules as in [29], [30] can be avoided; thereby reducing the complexity.

C. NRS Filter for Unsupervised Object Estimation

Here, the objective is to develop an unsupervised method that will be able to detect continuous moving object(s) moving with different velocities in different directions where the background information is not always available. That is, the decision making is to be done from an incomplete knowledge base. Under this ambiguous situation, the consideration of neighborhood information of data points appears to be helpful in the process of estimation. Hence, an NRS-based filter is designed to handle the uncertainties present in this task where the concepts of lower approximation and upper approximation of object(s) are used to model the estimation approximately.

The NRS filter is one type of spatial bandpass filter that works in temporal feature space over neighborhood granular level. NRS-based lower and upper approximations of each object in the respective frames are performed here. The convolution between the change information is executed to label which temporal segment will probably belong to which object in a certain frame. The different velocity and color models for different objects in a sequence are the outputs of this filter. The entire operation is carried out in the neighborhood granular domain. Therefore, the filter is named as "neighborhood rough filter."

The input to this filter is two types of temporal information corresponding to each frame. Let t denote the current frame and P be the total number of previous frames to be considered. The extraction of temporal information is done as described below.

1) *Difference between the consecutive frames* (δ_i) is extracted in granular level, i.e.,

$$\delta_i = \{\aleph_{\text{sp-clr}}(x) : |f_i(\aleph_{\text{sp-clr}}(x)) - f_{i-1}(\aleph_{\text{sp-clr}}(x))| > \mathcal{T}_1, \; i = t, \ldots, t-P\}. \tag{4}$$

In (4), f_i and f_{i-1} represent two consecutive frames, $f_i(\aleph_{\text{sp-clr}}(x))$ and $f_{i-1}(\aleph_{\text{sp-clr}}(x))$ represent the spatio-temporal granules formed around the point x in the aforementioned two frames, and $\mathcal{T}_1$ is a temporal threshold to detect the changed regions between frames.

2) *Differences from the current frame to all its previous frames* (δ_p) are computed in granular level as

$$\delta_p = \{\aleph_{\text{sp-clr}}(x) : |f_t(\aleph_{\text{sp-clr}}(x)) - f_i(\aleph_{\text{sp-clr}}(x))| > \mathcal{T}_2, \; p = t-1, \ldots, t-P\}. \tag{5}$$

In (5), $\mathcal{T}_2$ is another temporal threshold to roughly detect the changed regions from current to all its previous frames.

These two information in the filter get convolved resulting in the color models and velocity profiles of the objects in a frame. Based on these, locations of the objects are estimated in the next frame.

Let P previous frames be considered for approximation. The inputs to the filter are: 1) the union of the changed regions among current to all its previous frames ($\delta_P = \bigcup \delta_p : p = t-1, \ldots, t-P$) and 2) the changed regions between the consecutive frames (δ_i). That is, a "one to P points convolution" will take place in the filter which is expected to result in a P point matrix. Since two approximated decision spaces (lower and upper) over the filter are defined here to deal with the uncertainties, there will be two P point output matrices. The design of the filter is described as follows:

$$\{\delta_i : i = 1, \ldots, P\} * \delta_P = \{\underline{O}_c : c = 1, \ldots, P\} \; \{\overline{O}_c : c = 1, \ldots, P\} \tag{6}$$

where the left-hand side and the right-hand side represent the input and output of the NRS filter and $*$ denotes the convolution operation. The two convolution parameters that are used during

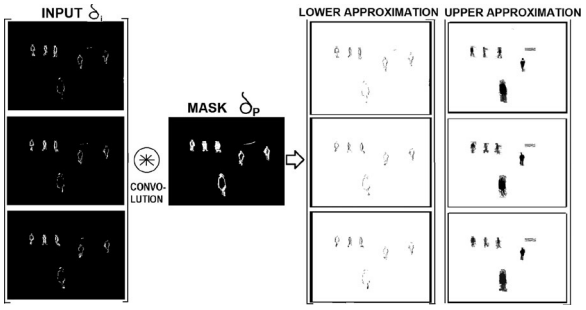

Fig. 2. Example of input and output of NRS filter.

the approximation are as follows:

$$U_c = \delta_P \cup \{\delta_i : i = 1, \ldots, P\} \quad \text{and} \tag{7}$$

$$I_c = \delta_P \cap \{\delta_i : i = 1, \ldots, P\}. \tag{8}$$

$\underline{O}_c$ and $\overline{O}_c$ in (6) are defined as

$$\underline{O}_c = \{\aleph(x) \in U : \aleph(x) \in I_c\} \quad \text{and} \tag{9a}$$

$$\overline{O}_c = \{\aleph(x) \in U : \aleph(x) \in U_c \quad \& \quad \aleph(x) \cap I_c \neq \emptyset\}. \tag{9b}$$

The neighborhood granules $\{\aleph(x)\}$, with which the approximations are performed, are the spatio-color granules (as described in Section III-B) belonging to the regions δ_P and δ_i. An example of input and output of NRS filter is shown in Fig. 2 where $P = 3$. The convolution results in two 3×1 matrices representing two types of approximated regions of the objects as the output of NRS filter.

D. Performance Measure of NRS Filter

The performance of NRS filter is measured with the ratio of the cardinality of the obvious object sets (signal) to that of the boundary sets (probable noise). The cardinality of the object/boundary sets is viewed as the amplitude of the signal (noise) sets therein.

The signal-to-noise ratio (SNR) [36] of the NRS filter is expressed as

$$\text{SNR} = 20\log_{10} \frac{|\underline{O}_c|}{|\overline{O}_c| - |\underline{O}_c|}. \tag{10}$$

The lower value of SNR reflects the poorer performance of the filter. The performance of the filter depends on its three parameters viz., $P, \mathcal{T}_1$, and $\mathcal{T}_2$. These may be adjusted as follows.

If the value of $|\underline{O}_c|$ is much low then it can be assumed that smaller than the exact region is extracted out from δ_i and hence the value of threshold $\mathcal{T}_1$ should be lowered down.

If the noise $|\overline{O}_c| - |\underline{O}_c|$ is much higher, then it can be assumed that the extracted changed information from δ_P is more than that was required and the value of threshold $\mathcal{T}_2$ needs to be increased.

If neither the signal nor the noise gets much deviated than their expected values, but SNR is still low, then the number of previous frames P needs to be increased by two to have more information. The entire process is then repeated.

As the approximated locations and color models of the objects are roughly estimated after filtering over the P frames, these information are used for modeling the same in the next frame onward. The methods of generation of velocity profile and estimation of location of the objects as required in the process, are described in the next section.

E. Temporal Neighborhood Granules and Location Estimation

The location of the object in the next frame is estimated based on its change in location and the pattern of changes from frame to frame. This needs modeling the nature of variation in size, speed, and direction of the objects. The velocity neighborhood granules and the acceleration neighborhood granules are defined to serve this purpose. These are discussed in the following sections.

1) Velocity Neighborhood Granules ($\aleph_V$): These granules are formed to integrate the information of objects' velocity, i.e., the changes of object locations from frame to frame. It is known that velocity has two components: magnitude and direction. The signed differences between the locations (primarily the locations of the four corners of object silhouette) in consecutive frames are treated as the velocity here. Let (x_1, y_1),...,(x_P, y_P) be the P positions of a corner of an object $\overline{O}_c : c = 1, \ldots, P$). The shifts between the locations $\hat{V} \equiv (v_x, v_y)$ in consecutive frames are computed as

$$v_{x_c} = x_c - x_{c-1} \tag{11a}$$

$$v_{y_c} = y_c - y_{c-1}, \quad c = 1, \ldots, P. \tag{11b}$$

Here, the signed difference is computed in (11) to reflect the direction.

Let the set of the shift values obtained from (11) be represented as: $S_V = \hat{V}_c, \forall c = 1, \ldots, P$. There are four such sets for four corners of each object. Velocity granules ($\aleph_v$) are constituted by the shift values within the previous P number of frames, which are nearer to that of the previous frame, as follows:

$$\aleph_V = \{\hat{V}_c \in S_V : \triangle(\hat{V}_c, \hat{V}_t) < \text{Th}_V, \forall c = 1, \ldots, P\}. \tag{12}$$

There are four $\aleph_V$s for each object. The nearness threshold Th_V is chosen based on the acceleration granules, discussed in the next section.

These velocity granules ($\aleph_V$s) can easily separate the moving object(s) (moving with consistency) and noise (random movement) based on their cardinality. For example, the granules that are formed over noise should be of cardinality one, as there is no continuous shift for them.

2) Acceleration Neighborhood Granule ($\aleph_A$): The moving object(s) in a video sequence do not usually move in a uniform speed. Rather, gradual changes in shape/size, speed, direction, etc., are obvious. The acceleration granules are defined to estimate the pattern of the change in velocity. Let (v_{x_1}, v_{y_1}),...,(v_{x_P}, v_{y_P}) be the P velocities of a corner of an object. The changes in velocity of that corner, $\hat{A} \equiv (a_x, a_y)$, in consecutive frames are defined as

$$a_{x_c} = v_{x_c} - v_{x_{c-1}} \tag{13a}$$

$$a_{y_c} = v_{y_c} - v_{y_{c-1}}, \quad c = 1, \ldots, P. \tag{13b}$$

Fig. 3. Example location estimation over the output from NRS filter.

The acceleration granules are formed over the set $S_A = \{\hat{A}_c, \forall c = 1, \ldots, P\}$ as

$$\aleph_A = \{\hat{A}_c \in S_A : \triangle(\hat{A}_c, \hat{A}_t) < \text{Th}_A \quad \forall c = 1, \ldots, P\} \quad (14)$$

where Th_A is the nearness threshold, and there are four $\aleph_A$s for each object as in the case of $\aleph_V$.

The velocity and acceleration granules, thus formed for a frame, are used to estimate their probable values in the next frame. The search space in a frame is optimized with the help of these two types of granules once the object(s) are approximately known. Let the location of the objects in the tth frame be L_t. Then, the predicted location of those objects with $|\aleph_V| > P/2$ in the $(t+1)$th frame will be

$$\overline{\tilde{L_{t+1}}} = L_t \bigoplus \{\aleph_{V_r} : r \in \{\text{RU}, \text{LU}, \text{LL}, \text{LC}\}\} \bigoplus \{\aleph_{A_r} : r \in \{\text{RU}, \text{LU}, \text{LL}, \text{LC}\}\} \in \overline{O} \quad (15)$$

where RU, RL, LL, and LU represent the four corners of an object region and $\bigoplus$ stands for set summation. Note that, the region $\overline{\tilde{L}_{t+1}}$ is supposed to contain the set of spatio-color neighborhood granules in $\overline{O}$. Here $\overline{O} = \{\overline{O}_c : c \in P\}$.

An example of location estimation of the objects (obs) over the lower approximated output of Fig. 2 is shown in Fig. 3 where the estimated locations are shown in gray color. As seen in Fig. 3, there is no estimation performed for the object "ob7" as there is no continuous shift in this region. It is, therefore, detected as noise.

F. Roughness in Object Approximation

Let there be $\mathcal{M}$ number of moving objects in a sequence. Then, $\mathcal{M}$ number of estimated locations $\overline{\tilde{L}_{t+1}}$ will be there in the tth frame. The roughness of the estimated locations is measured based on how many similar color granules ($\aleph_{\text{clr}}$) are there in the approximated regions in the current frame as compared to that of the object models. Let the approximated model of the mth moving object be $\overline{\mathcal{O}}_m = \{\cup\aleph_{\text{clr}} : \aleph_{\text{clr}} \in (\overline{O_{c_m}} \cap \overline{L_{t_m}} \forall c = 1, \ldots, P)$. Then, the roughness of the object model approximation ($\tilde{R\mathcal{O}}_m$) is defined as

$$\tilde{R\mathcal{O}}_m = 1 - \frac{|\overline{\tilde{L_{t_m}}} \cap \overline{\mathcal{O}_{c_m}}|}{|\overline{\mathcal{O}_{c_m}}|}. \quad (16)$$

In (16), $|\tilde{L_{t_m}} \cap \mathcal{O}_{c_m}|$ represents the total number of similar color granules between the estimated object models and the estimated regions. If the value of $\tilde{R\mathcal{O}}_{c_m}$ is high (as determined by a predefined threshold Th_O) for the mth object, it means that there are less similar granules. In other words, it can be hypothecated that there occurs some occlusion over that object either with the background or with another object. The concept of intuitionistic entropy measure is, accordingly, introduced to deal with such issues by estimating whether a granule is intuitionally present in the object set in terms of its overlapping nature and its probability of being in the predicted location. The details of it is described in the next section.

G. Intuitionistic Entropy

This entropy is defined by incorporating the merits of both rough entropy and probabilistic entropy. The granules over which this measure is performed are those that are present in object model ($\mathcal{O}_m$) but not completely in the predicted region ($\mathcal{O}_m(\aleph_{\text{clr}}) - \tilde{L_m}(\aleph_{\text{clr}}) \cap \mathcal{O}_m(\aleph_{\text{clr}})$). It determines the presence of a spatio-color granule in the predicted object region even if it is visually absent. This is constituted by two types of uncertainty measures: NRE and NPE. The first one measures the importance of a boundary granule to the object model, whereas the second one evaluates its probability of being in the predicted region. In this way, the combination of these two intuitionally estimates the object in the next frame. The new NRE is formulated here by incorporating the characteristics of overlapping, whereas the NPE ensures the estimation.

1) Neighborhood Rough Entropy: The importance of a boundary granule $\aleph_{\text{BO}}$ to the respective object model ($\mathcal{O}_m$) is measured from this entropy. The cardinality measure of a set is redefined in this process as NRS deals with overlapping granules. This implies that same points may belong to more than one granule. Therefore, instead of counting each granule as an individual entity (such as PaRS), the part of the granule which is not already counted is considered. In this way, the repetition in counting is avoided. The overlapping cardinality (represented as $\langle . \rangle$) of a neighborhood rough set (N) is defined as

$$\langle N \rangle = \sum_{g \in G} \partial_g \quad (17)$$

where G is the total number of granules in the set N. ∂_g denotes the part of the granule not already counted. It is mathematically explained in (18) shown at the bottom of this page.

The NRE that is defined here concerns only with one (object) class, as the background class is not estimated. It is expressed as

$$\text{NRE} = -e * \text{RO} * \ln(\text{RO}) \quad (19)$$

$$\partial_g = \frac{\text{Points not in overlapping regions with already counted granules}}{\text{Total no.s of points in the granule}}. \quad (18)$$

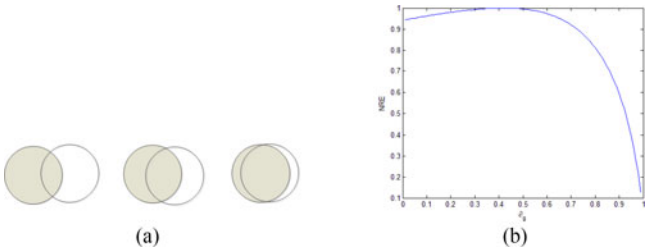

Fig. 4. Example two overlapping granules. (a) Granules with different degree of overlapping, and (b) according variation in NRE.

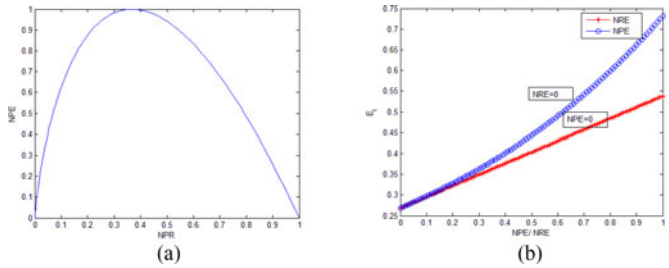

Fig. 5. Plot of (a) NPR versus NPE; (b) variation of E_I with NPE and NRE.

where the roughness RO is

$$\text{RO} = 1 - \frac{\langle \mathcal{O}_m - \aleph_{\text{BO}} \rangle}{\langle \mathcal{O}_m \rangle}. \tag{20}$$

NRE reflects the amount of difficulty in deciding whether a boundary granule $\aleph_{\text{BO}}$ belongs to the object region or not (noise). An example is shown in Fig. 4 to demonstrate the effect of ∂_g over NRE where a simple case with two granules overlapping to each other is considered. It is shown that the gray granule in Fig. 4(a) is in the lower approximation and the white one is the boundary granule. Fig. 4(a) shows that, the NRE value for a boundary granule is high if there is minimum overlapping ($0 \leq \partial_g < 0.5$). It attains the maximum value when $\partial_g = 0.5$ as the decision making over that boundary granule becomes the toughest in this scenario. The NRE decreases gradually after $\partial_g > 0.5$ as the degree of overlapping increases, and it attains its minimum value ($= 0$) once $\partial_g = 1$, that is the lower and upper approximation are the same and no more uncertainty is there. Therefore, the granules with $\partial_g < 0.5$ may be considered as noise. Note that had the study been performed according to the conventional definition of cardinality, the entropy value would have been the same for any value of ∂_g.

NRE determines the significance of $\aleph_{\text{BO}}$ to the object model. Now, its probability of being present in the estimated region $\tilde{L}_{t+1}$ is determined by NPE, as discussed in the next section.

2) Neighborhood Probabilistic Entropy: This ambiguity measure is based on the probability of occurrence of a granule, and not on the cardinalities of the approximated sets. Two types of probabilities: 1) the probability of occurrence of $\aleph_{\text{BO}}$ over the sequence and 2) its probability of being in the $\tilde{L}_{t+1}$ are considered. The computation of the second one is dependent on the first one.

Let $\aleph_{\text{BO}}$ be present $\mathcal{P}$ times in the object models over the previous P frames (i.e., $\mathcal{P} \leq P$). Then, its probability of occurrence (NPO) over the sequence will be

$$\text{NPO} = \frac{\mathcal{P}}{P}. \tag{21}$$

Let the location of $\aleph_{\text{BO}}$ in (t)th frame be l_t and its predicted location in the current frame be $\tilde{l}_{t+1} = l_t \bigoplus v_m \bigoplus a_m$ according to (15), where v_m and a_m are the mean of $\aleph_V$ and $\aleph_A$, respectively. The probability of being the granule in the predicted region (NPR) is computed based on the belongingness of $\tilde{l}_{t+1}$ in $\tilde{L}_{t+1}$ over NPO (21). Therefore, it is a conditional probability that is described as

$$\text{NPR} = \text{NPO} \times \frac{|\tilde{l}_{t+1} \bigcap \tilde{L}_{t+1}|}{|\tilde{l}_{t+1}|}. \tag{22}$$

The NPE is defined as

$$\text{NPE} = -e \times \text{NPR} \times \ln(\text{NPR}). \tag{23}$$

The characteristics of NPE versus NPR is shown in Fig. 5. It can be noticed that NPE increases initially with NPR and attains its maximum value at $1/e$, then it starts to decrease. That is, the granules with NPR $< 1/e$ may be treated as noise and not considered as probable object granules.

The intuitionistic entropy (E_I) considers the effects of both NRE and NPE. E_I is measured twice for each boundary granule $\aleph_{\text{BO}}$ (with respect to object set and that of its complement) to ensure its presence in the object set. Let NRE_1 be the value of NRE, and NPE_1 be the value of NPE after inserting $\aleph_{\text{BO}}$ in the $\underline{O}$. The object intuitionistic entropy (E_{I_O}) of the granule is defined as

$$E_{I_O} = \frac{\text{NRE}_1 + e^{\text{NPE}_1}}{e+1}. \tag{24}$$

Similarly the background intuitionistic entropy E_{I_B} is defined with the complement values of NRE_1 and NPE_1, denoted as NRE_1^c and NPE_1^c, respectively. That is

$$E_{I_B} = \frac{\text{NRE}^c + e^{\text{NPE}^c}}{e+1}. \tag{25}$$

If $E_{I_B} > E_{I_O}$, it means, the boundary granule $\aleph_{\text{BO}}$ reduces the overall ambiguity when it is in the object set. Then, $\aleph_{\text{BO}}$ is labeled as an object granule, otherwise the background (noise).

The nature of variation of E_I with NPE and NRE is shown separately in Fig. 5(b). The red "o-" line shows the variation of E_I with NPE which is exponential in nature, and the blue "*-" line shows its variation with NRE which is linear. That is, the impact of the change in NPE on E_I is higher as its probable deviation is taken into account during the formulation. This is not the case for NRE where the impact is linear.

IV. Algorithm for Tracking

The overall algorithm for tracking is described here in brief for a single moving object. This algorithm primarily consists of two parts: training and testing. These are described in Algorithms 1 and 2, respectively.

Algorithm 1: Unsupervised Object(s) Model Estimation With NRS-Filter.

INPUT: Initial P frames, Spatio-color granular threshold Th, Temporal thresholds: $\mathcal{T}_1$, $\mathcal{T}_2$
OUTPUT: Approximated object models $\underline{O}_c$ and $\overline{O}_c$: $c = 1, \ldots, P$
INITIALIZE: $\underline{O}_c = \overline{O}_c \Leftarrow \emptyset$
1: Compute temporal values δ_i and δ_p.
2: Form spatio-color neighborhood granules (see Section III-B) over *P* frames.
3: Perform rough object-background separation with NRS filter (see Section III-C) to extract out $\overline{O}_c$ and $\underline{O}_c$: $c = 1, \ldots, P-1$.
4: Measure the performance (SNR) of it as described in Section III-D
if SNR is low **then**
  while $\overline{O}_c - \underline{O}_c < \underline{O}_c$ **do**
    set $P = P + 2$; go to step 1.
  end while
end if
5: Approximate the objects' models $\overline{\mathcal{O}}$
6: Form velocity and acceleration granules ($\aleph_V$, $\aleph_A$) (see Section III-E).

Algorithm 2: Object Tracking.

INPUT: Next frame f_{t+1}, $\overline{\mathcal{O}}$, $\aleph_V$, $\aleph_A$, Th_o
OUTPUT: Object in the current frame O
INITIALIZE: $O \Leftarrow \emptyset$
1: Approximate locations $\overline{\tilde{L_{t+1}}}$ in the next frame.
2: Compute the roughness in object location approximation $\tilde{R\mathcal{O}}$.
if $\tilde{R\mathcal{O}} > \text{Th}_o$ **then**
  Compute intuitionistic entropy and estimate $\overline{\tilde{\tilde{L}}_{t+1}}$.
else
  Set $\overline{O}_{t+1} = \overline{\tilde{L}_{t+1}}$.
end if
3: Track $\overline{\tilde{L}_{t+1}}$.
4: Update $\aleph_V$ and $\aleph_A$,
5: Update $\overline{\mathcal{O}}$
6: Set $t + 1 = t$.
7: Repeat step 1 to step 4.

The same steps of these algorithms will be repeated M times for M number of moving objects in a sequence.

A. Complexity of the Algorithm

The computational complexity of this algorithm is low as the algorithm works in granular level. If there are total N number of pixels in a frame, then the complexity of granulation is $O(\frac{N}{n})$, with n degree of similarity. The computational complexity during training with P previous fames is $O(P\frac{N}{n})$ and the complexity of tracking $\mathcal{M}$ number of objects is $O(\mathcal{M}\frac{\tilde{L}}{n}\Psi)$. Ψ is the number of iterations required in the process. That is, the computational time of this algorithm is dependent on the size, number, and internal color variation in the object models. The average computational time per frame for the aforesaid sequences varies from 0.3 to 1.2 s depending on the nature of the sequences.

V. Results and Discussions

The effectiveness of the proposed algorithm is established in this section with experimental results. The key features/advantages of this algorithm lie in the following:

1) tracking multiple objects, without any prior knowledge, even with variations in shape/size or speed,
2) tracking object(s) even if it is fully/partially occluded with background or gets overlapped with other moving object(s).

The video sequences with different characteristics, e.g., indoor/outdoor surveillance [37]–[41], single/multiple moving object(s) [4], [42]–[44], body part(s) movements [45] are considered during the experiment. The algorithm is executed almost over 2500 frames in total.

The video sequences over which the results are shown here, as examples, are described below. $2b$-sequence [37] has six people moving in different directions. Sequence M_10 [45] has one lady who is moving her hands with overlapping between them. In the sequence $14\text{-}LD$ [42], one person is moving with total occlusion twice. There are five people moving in a group, get occluded to each other and background in the sequence P-15 [44]. There are three moving people who get occluded and overlapped with one another in the sequence New_Occ4 [41]. In the sequence $S2\text{-}L1$ [39], multiple people are moving in different directions, gets overlapped, and new persons appear. Sequence P-01 [38] initially contains one moving person and then one moving car appears at different velocities and with different variations in shape and size over the sequence. In A-07 sequence [40], there is initially one person moving, gets totally occluded with background and reappears, then another person appears from a different direction and moves. Two people enter one by one and get partially overlapped with change in their directions in the sequence much-132 [43]. The results that are obtained with these datasets are shown in the following sections.

A. Selection of Parameters and Initial Assumptions

The underlying assumptions during the initial labeling are: 1) no moving object should get occluded or overlapped and 2) the moving object and its respective background should be well separable in RGB feature space. Selection of parameters is a crucial issue in every tracking algorithm. In the present investigation, one of the objectives is to make the parameters adaptive, as much as possible, so that these can take the approximate values automatically depending on the nature of the movement and size of the objects in a sequence. The method of selecting the parameter P, depending on the speed of the object(s) from frame to frame for the aforesaid sequences is explained in Section V-B. The initial values of T_1 and T_2 were set as 15, and

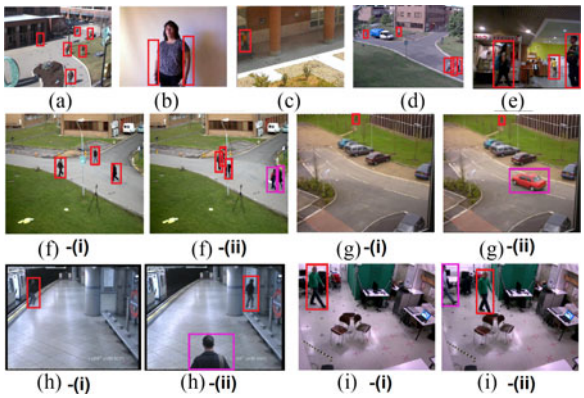

Fig. 6. Initially tagged objects of (a) 2-b sequence, (b) M_10 sequence, (c) 14-LD sequence, (d) P-15 sequence, (e) New_Occ4 sequence, (f) $S2$-$L1$ sequence, (g)P-01 sequence, (h)A-07 sequence, and (i) much-132 sequence: *(i)*initially present object(s), *(ii)* newly appeared object(s).

they get updated according to the requirement of the respective video sequence. The value of Th was set as 30, as this value is experimentally found to provide a balanced tradeoff between the accuracy and number of segments.

B. Effectiveness of Unsupervised Labeling and Estimation

The results for initial unsupervised labeling of continuous moving object(s) using NRS filter are shown in this section. The visual results in Fig. 6 show the first labeled frames with respect to the appearance of the object(s) in the respective sequences. For example, no new object(s) appear in the sequences: 2-b, M_10, 14-LD, P-15 and New_Occ4 [see Fig. 6(a)–(e)]. But new moving object(s) appear in the sequences $S2$-$L1$, P-01, A-07, much-132 [see Fig. 6(f), (g), (h), (i)]. The initially present object(s) in the respective sequences are shown in Fig. 6(f)-*(i)*, (g)-*(i)*, (h)-*(i)*, and (i)-*(i)*; and the object(s) that appeared late are shown in Fig. 6(f)-*(ii)*, (g)-*(ii)*, (h)-*(ii)*, and (i)-*(ii)*. The initially present object(s) (I) in each sequence are marked by red trackers whereas the newly appeared ones (N) are marked by violet trackers in Fig. 6.

The required number of frames (P) to label an object varies depending on the speed of the object even in the same sequence. This is shown in Table I where P-values required for labeling the initially present object(s) as well as the newly appeared object(s) for the aforesaid sequences are given. The accuracy of initial labeling is measured in terms of the root-mean-square-distance between the four corners of bounding boxes of the annotated ground truth and that of the labeled object ($\overline{O}_c$) (RMSE$_L$). The other metric that is shown in the table is accuracy in prediction in the next frame (RMSE$_P$), which is measured the same way as that of RMSE$_L$, where the distance between the predicted region in the next frame of the labeled one ($\overline{\tilde{L}_t}$ in Section III-E) and the corresponding ground truth is computed.

It can be noticed from Table I that the parameter P is selected automatically depending on the speed of the object and it varies for each object accordingly. Different types of objects in different scenarios are identified with RMSE$_L$ attaining values between 5.4 and 14.8 pixels for the aforesaid sequences. This can be interpreted as quite efficient, as evident from the results of tracking described in the next section. Similar conclusion can be drawn for the prediction task by observing the RMSE$_P$-values. These results demonstrate the effectiveness of the proposed NRS filter, and the velocity and acceleration granules for the tasks of unsupervised object labeling and location estimation, respectively.

C. Results of Tracking

The results of tracking using our proposed method is shown in this section. The effectiveness of this algorithm in dealing with the cases involving 1) variation in shape/ size (P-01, A-07), 2) appearance of new object(s) (as discussed in Section V-B), 3) nonuniform movement of object(s) ($S2$-$L1$, M_10, much-132) is shown here. The results for dealing with occlusion using the proposed intuitionistic entropy are also shown. The datasets used here characterize different types of occlusions as follows:

1) object to background (14-LD, A-07, $S2_L1$, P-15);
2) object to object ($S2$-$L1$, M_10, much-132, New_Occ4);
3) partial ($S2$-$L1$, M_10, much $-$ 132);
4) total (14-LD, A-07).

The comparative studies are carried out with seven recent partially supervised methods (where the object(s) are manually labeled in the initial frame) that are proved to be effective and robust to multiobject tracking and occlusion handling are made to demonstrate the merits of our algorithm. The parameters for these algorithms are set by following the instructions provided in the respective papers. These comparing methods alongwith their parameter values are described below.

MTJSR [16] where the templates from different image features are fused to form a robust object model and the tracking is done with particle filter. Here, the Gabor filter is set to 2 and $\pi/2$ according to the paper.

MCIO [17] where many to one (M to 1) optimization approach is performed for tracking multiple interacting objects. We set the parameters as: $M = 2$, $\sigma = 0.3$ and $g = 4$ following the paper.

OM [6] where an energy-based model of multitarget tracking is defined over all the target locations and all video frames in a given time window. Here, the values of the parameters are: $\lambda = 0.1$, $\delta = \epsilon = 0.6$, and $\beta = 0.03$ as given in the paper.

ALIEN [8] where the objects are represented by their invariant local features taken in different conditions (oversampling). Following the paper, we set the parameters as: $\lambda_T = \lambda_C = 2.5$, $l = 10$, $n_u = 8$, and $\sigma = 1$.

JORS [18] where object tracking is formulated as an optimization problem by considering both the reconstruction and classification errors in the objective function. Here, we set $N = 15$, $\beta = 0.2$, $T_1 = 0.05$, $T_2 = 0.5$, and $T_3 = 0.7$ as instructed in the paper.

DeepTrack [14] where tracking is performed using a single CNN, which is developed for learning effective feature representations of the target object in a purely online manner. Here,

TABLE I
EFFECTIVENESS OF NRS FILTER BASED UNSUPERVISED LABELING

Sequences	2-b	M_10	14-LD	P-15	New_Occ4	S2-L1-I	S2-L1-N	P-01-I	P-01-N	A-07-I	A-07-N	much-132-I	much-132-N
P	10	7	8	11	5	6	7	17	9	7	8	7	6
RMSE$_L$	11.4	9.3	5.4	12.2	7.5	12.7	15.1	6.3	12.2	5.2	14.8	11.1	9.2
RMSE$_P$	10.7	7.11	4.3	11.4	6.8	14.2	13.4	5.9	11.4	6.4	13.9	9.2	15.1

the input is locally normalized into 32×32 image patches in the first layer and 7×7 image patches in the second layer as that was there in the paper.

CNT [15] where simple two-layer convolutional networks are used for object representations and a "deep network" structure is introduced in visual tracking. Here, the size of the wrapped images and field size are set to 32×32 and 7×7, respectively, following the paper.

The visual comparative results with these seven other robust methods are shown in Fig. 7. It can be noticed from the figure that all of the algorithms perform well in multiple object tracking as well as for object-to-object overlapping/interaction handling. However, the proposed unsupervised algorithm works better when total object-to-background occlusion takes place as the newly defined intuitionistic entropy incorporates both of the overlapping and probabilistic properties of the moving object(s). It can be seen in Fig. 7(c-(iv)), (e-(iii)), (h-(ii), h-(iii)) that the unseen object(s) (totally occluded by background/other object) are almost correctly identified by the proposed algorithm whereas there are localization errors for the other algorithms. The quantitative comparative studies in terms of average RMSE (computed similarly as that of RMSE$_L$) and percentage of correctly classified frames in each sequence are shown in Tables II and III, respectively. The performances of all these eight methods throughout the aforesaid nine sequences are shown in Fig. 8. It is seen in Fig. 8 that all the eight methods perform well when there is no occlusion or overlapping, whereas the performances of the methods, namely, "MTJSR," "MCIO," "OM," "ALIEN," "JORS" deviate and attain higher values when some overlapping took place. One may note the values of RMSE for the sequences M_10, P-15, New_Occ4, S2-L1, and much-132 in Fig. 8(b)–(i) deviated for the said methods over the frames where overlapping/occlusion takes place. Values of RMSE increases for the other two methods, namely, "DeepTrack" and "CNT" when total occlusion occurs. Total object to background occlusion takes place in the sequences 14-LD and A-07 [see Fig. 8(c) and (h)]. The existing algorithms fail to estimate the totally occluded objects resulting in a higher RMSE value; however, they can again track those once they reappear (become visible). RMSE for the proposed method remains almost the same throughout the sequences (though not the best for all the cases), when no new object appears. That is, the proposed method is well capable in handling total occlusion or overlapping. These results conform to the visual results. The average RMSE (see Table II) values also reflect similar conclusions for all the sequences with multiple moving objects or with very low object-to-object interactions. This proves the effectiveness of all the existing as well as the proposed algorithms in handling such cases. The

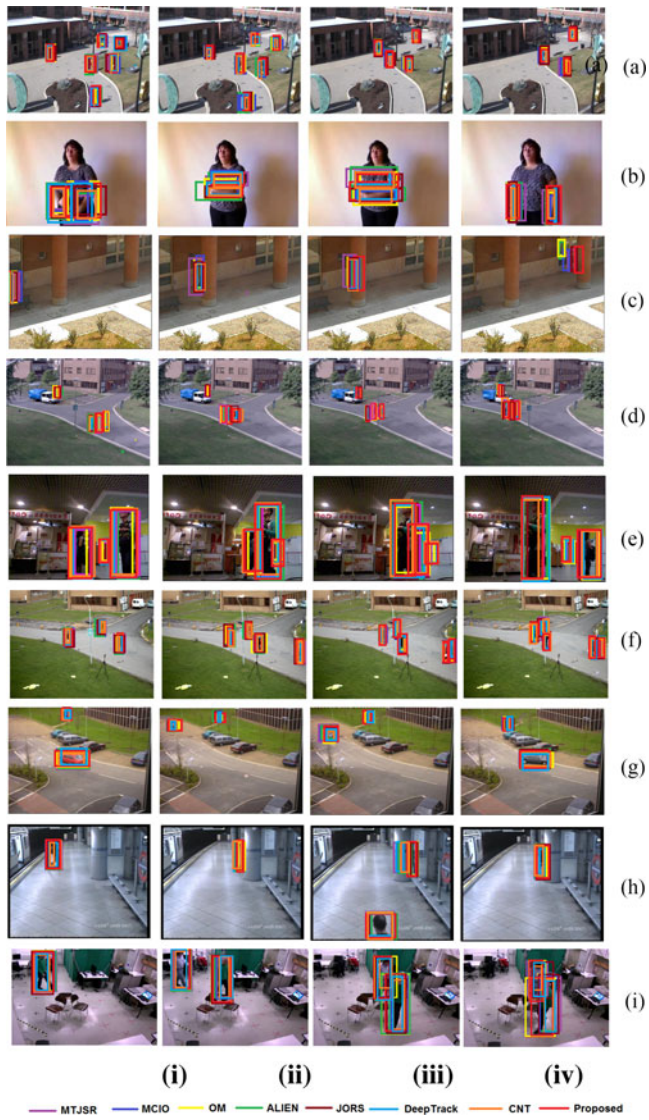

Fig. 7. Results of tracking for (a) 2-b sequence, (b) M_10 sequence, (c)14-LD sequence; (d) P-15 sequence, (e) New_Occ4 sequence, (f) S2-L1 sequence, (g) P-01 sequence, (h) A-07 sequence, and (i) much-132 sequence.

percentage of the correctly classified frames (see Table III) obtained by the algorithms for the aforesaid sequences are very high reflecting their accuracy in tracking throughout the sequences. It can also be observed that OM works the best for multitarget tracking, whereas DeepTrack for interacting objects. The proposed algorithm can estimate the locations of the totally occluded object throughout with the accuracy (RMSE) value around 3.

TABLE II
AVERAGE RMSE

Sequences	2-b	M_10	14-LD	P-15	New_Occ4	$S2$-$L1$	P-01	A-07	much-132
MTJSR	3.2	2.4	11.2	2.6	5.8	4.1	1.8	8.2	2.6
MCIO	2.3	3.7	10.7	2.3	8.2	3.9	1.2	7.8	2.4
OM	0.8	3.25	8.3	1.8	8.9	3.65	2.4	7.9	3.01
ALIEN	0.56	3.45	5.2	1.7	6.1	3.76	2.2	8.1	3.2
JORS	3.26	4.15	12.3	4.1	9.7	3.33	1.72	12.1	2.22
DeepTrack	1.23	2.41	8.3	0.8	5.1	2.88	1.47	3.9	1.68
CNT	0.97	2.58	8.1	1.6	6.21	3.08	1.13	7.8	1.91
Proposed	3.11	3.35	3.98	0.83	2.67	3.13	1.22	3.1	2.13

TABLE III
% OF CORRECTLY CLASSIFIED FRAMES

Sequences	2-b	M_10	14-LD	P-15	New_Occ4	$S2$-$L1$	P-01	A-07	much-132
MTJSR	98	97	80	82	90	93	96	86	99
MCIO	99	96	82	85	95	98	99	85	97
OM	100	98	83	84	93	97	97	87	99
ALIEN	100	97	81	84	96	98	99	90	98
JORS	94	95	79	83	92	93	97	82	98
DeepTrack	92	96	88	90	96	95	97	91	99
CNT	93	95	92	93	95	91	94	96	98
Proposed	98	97	98	96	97	98	98	98	97

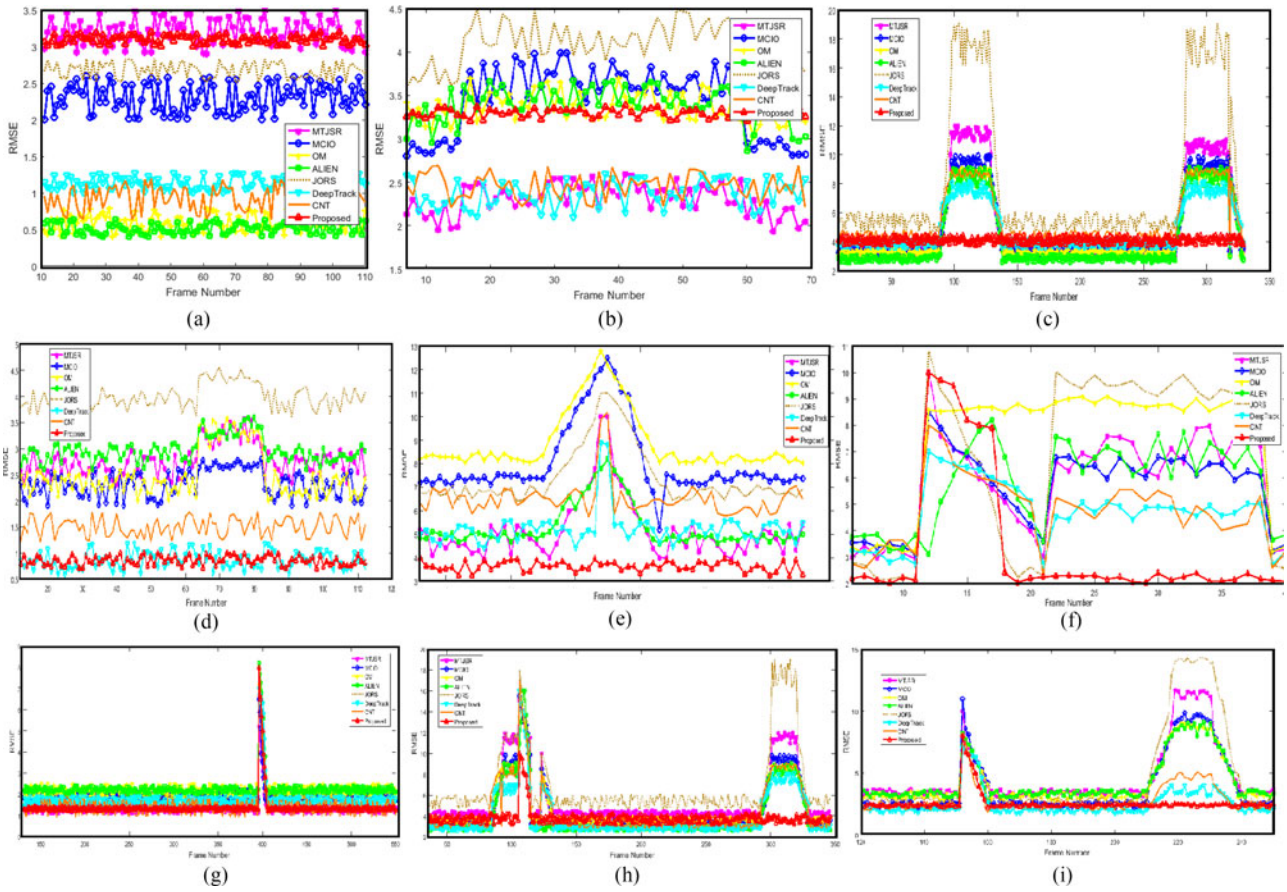

Fig. 8. RMSE value throughout the sequences: (a) 2-b sequence, (b) M_10 sequence, (c)14-LD sequence; (d) P-15 sequence, (e) New_Occ4 sequence, (f) $S2$-$L1$ sequence, (g)P-01 sequence, (h)A-07 sequence, and (i) much-132 sequence.

VI. Conclusion

This paper develops a methodology for unsupervised tracking of multiple moving object(s) considering the cases such as occlusions/ overlapping by exploring the features of NRS. The newly introduced concept of spatio-color granules over video frames is proved to be effective in reducing the complexity of the algorithm. The NRS filter designed can label the continuous moving object(s) accurately without any manual interactions. The proposed concept of intuitionistic entropy enables to handle efficiently the cases such as partial/total occlusions/overlapping. The unsupervised method, developed, provides equally good results in the task of tracking when compared with the other recent partially supervised methods, and proves to be superior in the frames where total object-to-background occlusion takes place.

The method can be further extended for real-time applications by exploring more merits of rough sets and granular computing. The proposed concepts of NRS filter and intuitionistic entropy can be applied to the other areas of signal processing where the unsupervised prediction and ambiguity handling are of concern.

Acknowledgment

The authors acknowledge Prof. D. K. Das, Jadavpur University, for his support.

References

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking : A survey," *ACM Comput. Surv.*, vol. 38, pp. 1264–1291, 2006.

[2] E. Maggio and A. Cavallaro, *Video Tracking—Theory and Practice*. Hoboken, NJ, USA: Wiley, 2010.

[3] A. Smeulders, D. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[4] Y. Wu, J. Lim, and M. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[5] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "Object tracking via partial least squares analysis," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4454–4465, Oct. 2012.

[6] A. Milan, S. Roth, and K. Schindler, "Continuous energy minimization for multitarget tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 58–72, Jan. 2014.

[7] K. Zhang, C. Nanjing, L. Zhang, and M.-H. Yang, "Fast compressive tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 10, pp. 2002–2015, Oct. 2014.

[8] F. Pernici and A. Bimbo, "Object tracking by oversampling local features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 12, pp. 2538–2551, Dec. 2014.

[9] T. Bai, Y.-F. Li, and X. Zhou, "Learning local appearances with sparse representation for robust and fast visual tracking," *IEEE Trans. Cybern.*, vol. 45, no. 4, pp. 663–675, Apr. 2015.

[10] A. Li, F. Tang, Y. Guo, and H. Tao, "Efficient discriminative nonorthogonal binary subspace with its application to visual tracking," *CoRR*, abs/1509.08383, 2015.

[11] Y. Guo, Y. Chen, F. Tang, A. Li, W. Luo, and M. Liu, "Object tracking using learned feature manifolds," *Comput. Vis. Image Understanding*, vol. 118, pp. 128–139, 2014.

[12] B. Dey and M. K. Kundu, "Robust background subtraction for network surveillance in H.264 streaming video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 10, pp. 1695–1703, Oct. 2013.

[13] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[14] H. Li, Y. Li, and F. Porikli, "Deeptrack: Learning discriminative feature representations online for robust visual tracking," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1834–1848, Apr. 2016.

[15] K. Zhang, Q. Liu, Y. Wu, and M.-H. Yang, "Robust visual tracking via convolutional networks without training," *IEEE Trans. Image Process.*, vol. 25, no. 4, pp. 1779–1792, Apr. 2016.

[16] W. Hu, W. Li, X. Zhang, and S. J. Maybank, "Single and multiple object tracking using a multi-feature joint sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 4, pp. 816–833, Apr. 2015.

[17] C. Park, T. J. Woehl, J. E. Evans, and N. D. Browning, "Minimum cost multi-way data association for optimizing multitarget tracking of interacting objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 611–624, Mar. 2015.

[18] Q. Wang, F. Chen, W. Xu, and M.-H. Yang, "Object tracking with joint optimization of representation and classification," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 25, no. 4, pp. 638–650, Apr. 2015.

[19] Z. Pawlak, *Rough Sets: Theoretical Aspects of Reasoning About Data*. Norwell, MA, USA: Kluwer, 1992.

[20] R. W. Swirniaski, "Rough sets methods in feature reduction and classification," *Int. J. Appl. Math. Comput. Sci.*, vol. 11, pp. 565–582, 2001.

[21] D. Sen and S. K. Pal, "Generalized rough sets, entropy, and image ambiguity measures," *IEEE Trans. Syst., Man, Cybern., Part B*, vol. 39, no. 1, pp. 117–128, Feb. 2009.

[22] S. K. Pal, B. U. Shankar, and P. Mitra, "Granular computing, rough entropy and object extraction," *Pattern Recognit. Lett.*, vol. 26, pp. 2509–2517, 2005.

[23] D. Sen and S. K. Pal, "Histogram thresholding using fuzzy and rough measures of association error," *IEEE Trans. Image Process.*, vol. 18, no. 4, pp. 879–888, Apr. 2009.

[24] S. K. Pal and P. Mitra, "Case generation using rough sets with fuzzy representation," *IEEE Trans. Knowl. Data Eng.*, vol. 16, no. 3, pp. 292–300, Mar. 2004.

[25] W. Pedrycz and M. Song, "A granulation of linguistic information in AHP decision-making problems," *Inf. Fusion*, vol. 17, pp. 93–101, 2014.

[26] D. Liang and D. Liu, "A novel risk decision making based on decision-theoretic rough sets under hesitant fuzzy information," *IEEE Trans. Fuzz. Syst.*, vol. 23, no. 2, pp. 237–247, Apr. 2015.

[27] H. Chen, T. Li, C. Luo, S. Horng, and G. Wang, "A decision-theoretic rough set approach for dynamic data mining," *IEEE Trans. Fuzzy Syst.*, vol. 23, no. 6, pp. 1958–1970, Dec. 2015.

[28] A. Ganivada, S. S. Ray, and S. K. Pal, "Fuzzy rough sets, and a granular neural network for unsupervised feature selection," *Neural Netw.*, vol. 48, pp. 91–108, 2013.

[29] Q. Hu, D. Yu, J. Liu, and C. Wu, "Neighborhood rough set based heterogeneous feature subset selection," *Inf. Sci.*, vol. 178, no. 18, pp. 3577–3594, 2008.

[30] Y. Du, Q. Hu, P. Zhu, and P. Ma, "Rule learning for classification based on neighborhood covering reduction," *Inf. Sci.*, vol. 181, no. 24, pp. 5457–5467, 2011.

[31] L. A. Zadeh, "Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic," *Fuzzy Sets Syst.*, vol. 90, no. 2, pp. 111–127, 1997.

[32] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Autom. Control*, vol. AC-24, no. 6, pp. 843–854, Dec. 1979.

[33] K. Okuma, A. Taleghani, O. Freitas, J. Little, and D. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *Proc. Eur. Conf. Comput. Vis.*, vol. 1, 2004, pp. 28–39.

[34] G. Shabat, Y. Shmueli, A. Bermanis, and A. Averbuch, "Accelerating particle filter using randomized multiscale and fast multipole type methods," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 7, pp. 1396–1407, Jul. 2015.

[35] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–575, May 2003.

[36] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing (3rd ed.): Principles, Algorithms, and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.

[37] E. O. W. S. B. J. Davis and V. Sharma, "Background-subtraction using contour-based fusion of thermal and visible imagery," *Comput. Vis. Image Understanding*, vol. 106, pp. 162–182, 2007.

[38] PETS-2001, *IEEE Int. Workshop Perform. Eval. Tracking Surveillance*, 2001.

[39] PETS-2009, *IEEE Int. Workshop Perform. Eval. Tracking Surveillance EC Funded CAVIAR Project*, 2009.

[40] AVSS-2007, *4th IEEE Int. Conf. Adv. Video Signal Based Surveillance*, 2007.

[41] S. Song and J. Xiao, "Tracking revisited using RGBD camera: Unified benchmark and baselines," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2013, pp. 233–240.

[42] Amsterdam Library of Ordinary Videos for evaluating visual trackers robustness, ALOV300++ Dataset, 2014.
[43] H. Possegger, S. Sternig, T. Mauthner, P. M. Roth, and H. Bischof, "Robust real-time tracking of multiple objects by volumetric mass densities," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 2395–2402.
[44] PETS-2015, *IEEE Int. Workshop Perform. Eval. Tracking Surveillance*, 2015.
[45] *ChaLearn Gesture Dataset (CGD 2011)*, ChaLearn, Berkeley, CA, USA, 2011.

Debarati Bhunia Chakraborty received the B.Tech. degree in electronics and communication engineering from Siliguri Institute of Technology, Chamta, India, in 2008, the M.Tech. degree in electronics and communication engineering from National Institute of Technology, Rourkela, India, in 2010, and the Ph.D. degree in computer science and enginnering from Jadavpur University, Kolkata, India, in 2017.

She has conducted her research work toward the Ph.D. degree at the Center for Soft Computing Research in Indian Statistical Institute during 2011–2016. She is currently working as a Research Associate at Indian Statistical Institute, Kolkata, India. Her research interests include video processing, rough sets, granular computing, soft computing, computer vision, and IoT.

Dr. Chakraborty received the prestigious Young Scientist Award by Indian Science Congress Association in 2012.

Sankar K. Pal (M'80–SM'84–F'93–LF'15) received the Ph.D. degree in radio physics and electronics from Calcutta University, Kolkata, India, in 1979 and the Ph.D. degree in electrical engineering from Imperial College of Science and Technology, London University, London, U.K., in 1982.

In 1975, he joined as a CSIR Senior Research Fellow at the Indian Statistical Institute, Kolkata, India, where he became a Full Professor in 1987, a Distinguished Scientist in 1998, and the Director in 2005. He is a Raja Ramanna Fellow of the Government of India, J. C. Bose National Fellow, and a Former Indian National Academy of Engineering Chair Professor. He founded the Machine Intelligence Unit and the Center for Soft Computing Research with the Institute in Calcutta, which are enjoying international recognition. He was with the University of California at Berkeley, Berkeley, CA, USA, and the University of Maryland at College Park, College Park, MD, USA, the NASA JSC, Houston, TX, USA, and the U.S. Naval Research Laboratory, Washington, DC, USA. He held several visiting positions in Italy, Poland, Hong Kong, and Australian Universities. He has coauthored 20 books and more than 400 research publications in the areas of pattern recognition and machine learning, image processing, data mining, web intelligence, soft computing, bioinformatics, and cognitive machines.

Prof. Pal has received several national and international awards, including the most coveted S. S. Bhatnagar Prize and the Padma Shri by the President of India, the G. D. Birla Award, the P. C. Mahalanobis Birth Centenary Gold Medal of Indian Science Congress Association by the Prime Minister of India for lifetime achievement, the NASA Tech Brief Award in USA, and the Khwarizmi International Award from the President of Iran. He has been a Distinguished Visitor of the IEEE Computer Society since 1987. He is/was on the editorial boards of 22 journals, including some IEEE transactions. He visited about 40 countries as a keynote/plenary/invited speaker. He is a Fellow of The World Academy of Sciences for the Advancement of Science in Developing Countries, International Association for Pattern Recognition, International Fuzzy Systems Association, International Rough Set Society, and all four National Academies for science/engineering in India.