Neighborhood granules and rough rule-base in tracking

Debarati Bhunia Chakraborty & Sankar K. Pal

Natural Computing
An International Journal

ISSN 1567-7818

Nat Comput
DOI 10.1007/s11047-015-9493-6

Nat Comput
DOI 10.1007/s11047-015-9493-6

Neighborhood granules and rough rule-base in tracking

Debarati Bhunia Chakraborty[1] · Sankar K. Pal[1]

Abstract This paper deals with several new methodologies and concepts in the area of rough set theoretic granular computing which are then applied in video tracking. A new concept of neighborhood granule formation over images is introduced here. These granules are of arbitrary shapes and sizes unlike other existing granulation techniques and hence more natural. The concept of rough-rule base is used for video tracking to deal with the uncertainties and incompleteness as well as to gain in computation time. A new neighborhood granular rough rule base is formulated which proves to be effective in reducing the indiscernibility of the rule-base. This new rule-base provides more accurate results in the task of tracking. Two indices to evaluate the performance of tracking are defined. These indices do not need ground truth information or any estimation technique like the other existing ones. All these features are demonstrated with suitable experimental results.

Keywords Neighborhood rough sets · Granular computing · Rough rule-base · Video tracking

1 Introduction

Granulation is a basic step of human cognition system. It is a natural process of interpretation in human mind. This phenomenon was first introduced to machine learning by Zadeh (1997). The concept of information granules is used effectively in several areas of machine learning. Different ways of forming granules are introduced, e.g., crisp granules, fuzzy granules, rough-fuzzy granules, neighborhood granules. The crisp granules makes the computation much faster. But, in real life the information system is not always crisply separable rather overlapping in nature. Hence, fuzzy and rough-fuzzy granules appear to be more effective to deal with. In case of fuzzy granules the selection of the membership functions is crucial. The concept of neighborhood granules has been introduced (Hu et al. 2008) later on. These granules deal with both of the numerical and categorical features by considering the neighborhoods within a range to be in the same granule. Here, overlapping is an inherent property of this granulation and no degree of overlapping needs to be defined.

The predefined parameters in all the existing granulation approaches are those related to the sizes or shapes of the granules based on which the granules are formed. But, natural granulation is arbitrary and it does not have any fixed shape or size. In this article an attempt of forming such granules in images is made. The similarities in both color and spatial feature spaces are considered during this formulation.

Rough sets (Pawlak 1992) provide an effective framework of granular computing. This theory deals with uncertainties or incompleteness of knowledge arising from the limited discernibility of objects in the domain of discourse. Its key concepts are those of 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 (Swirniaski 2001; Komorouski et al. 1999), image processing (Mushrif and Ray 2008; Pal et al. 2005; Sen and Pal 2009; Meher and Pal 2011), data mining and knowledge discovery (Komorouski et al. 1999; Pedrycz and Song

✉ Debarati Bhunia Chakraborty
debarati.earth@gmail.com

Sankar K. Pal
sankar@isical.ac.in

1 Center for Soft Computing Research, Indian Statistical Institute, Kolkata 700108, India

Published online: 30 May 2015

2014). Rough rule-base has been effective in feature reduction and computation gain. Rough set has been used recently to deal with certain problems of image processing. Here uncertainties arising from grayness and spatial ambiguities are handled using the concept of set approximation (Pal et al. 2005; Sen and Pal 2009). However its merit in video processing has not yet been explored.

Video processing is an important part of computer vision. Tracking of moving object(s) from video sequences is one of the basic steps of video processing. Object tracking is required in several fields of computer vision, e.g: surveillance, gesture recognition. This problem has been studied over decades and there exist several literature (Yilmaz et al. 2006; Maggio and Cavallaro 2010). Some of the approaches are partially supervised (Comaniciu et al. 2003; Wang et al. 2012), that is, initial object/ background model is known and some of them are unsupervised (Heikkila and Pietikainen 2006; Pal and Chakraborty 2013). One may note that in video tracking the complete information is not available. This makes the prediction difficult as there exist several uncertainties, e.g, changes in shapes/ sizes of moving object(s), changes in motion of the object(s), and changes in nos. of object(s). The rough rule base can be effectively used to deal with these incompleteness and uncertainties and have gain in computation time.

According to the existing literatures the rough rule-base so far is used over data points only. In the present article, the concept of granular rough rule-base is introduced where both the processing and decision making are done in granular level. As the granulation and rough rule base make a process faster individually, it is expected that the performance will get even better if these two are applied together. The proposed concept of neighborhood granulation is used here in the granular level.

The features usually used for video tracking are edge, texture, shape, color (e.g., RGB, YCrCb) etc. Information accumulated from multiple cameras are also used for better performance, but it is costlier. Kinect sensor has recently become a popular RGB and depth (D) sensor for video processing (Janoch et al. 2011; Lai et al. 2011; Sinthanayothin et al. 2012) as it provides cheap but quality datasets. It is obvious that, the depth of moving object(s) in a sequence always has lower values than that of its background. So, addition of this feature will give more accurate result in tracking with less complexity, as this feature is obtained from the sensor itself. The experiments are conducted over videos obtained by kinect sensor with different types of hand movements.

Two indices for performance evaluation of tracking are defined here. One of them is by incorporating the merits of D feature and the other one is by using the merits of the new neighborhood granulation. The main advantages of these indices are these don't require any ground truth or estimation like the other existing indices (Yilmaz et al. 2006; Maggio and Cavallaro 2010).

The novelty of the article mainly lies in: (i) defining a way of forming neighborhood granules in images to make it closer to natural partitioning, (ii) introducing the concept of rough rule base in tracking, (iii) formation of neighborhood granular rough rule-base to reduce indiscernibility, and (iv) defining two new indices to evaluate the performance of tracking which work without any ground truth estimation.

The article is organized as follows. In Sect. 2 brief introductions to rough set, rough rule-base and neighborhood rough set are given along with their relevance in video tracking. In Sect. 3 the new concept of neighborhood granulation is introduced. In Sect. 4 the proposed method of tracking is discussed. The concept of neighborhood granular rough rule-base is defined in this process. In Sect. 5 two new indices to evaluate tracking are defined. In Sect. 6 all the concepts and their effectiveness are experimentally demonstrated along with suitable comparisons.

2 Neighborhood rough set

Here we describe the basic concept of rough sets and neighborhood rough sets in brief.

2.1 Concept of 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.

All the attributes in $\mathcal{A}$ may not be necessary for classification. Certain attributes in it may be redundant and can be eliminated without losing the essential discriminatory information. The procedure of eliminating those redundant equivalence relation is discussed below.

2.1.1 Reduction of knowledge

Let $\mathbf{R}$ be a family of equivalent relations and let $R \in \mathbf{R}$. R will be said *discernable* in $\mathbf{R}$ if $IND(\mathbf{R}) = IND(\mathbf{R} - \{R\})$, otherwise R is *indiscernible* in $\mathbf{R}$. The family of $\mathbf{R}$ is *independent* if each $R \in \mathbf{R}$ is indiscernible in $\mathbf{R}$, otherwise it is dependent. Suppose, $\mathbf{P} \subseteq \mathbf{R}$, $\mathbf{P}$ is independent and $IND(\mathbf{P}) = IND(\mathbf{R})$; then, $\mathbf{P}$ is called a *reduct* of $\mathbf{R}$. $\mathbf{R}$ may have more than one reduct. The set of all indispensable relations in $\mathbf{R}$ is called the *core* of $\mathbf{R}$. It can be shown that $CORE(\mathbf{R}) = \bigcap RED(\mathbf{R})$. This is how the knowledge can be reducted from a knowledge base.

2.2 Concept of neighborhood rough set

In case of real life data sets there are nominal as well as numerical attributes. In numerical feature space, the concept of neighborhood plays an important role (Hu et al. 2008; Du et al. 2011). Overlapping is also a usual property in real life numerical feature space. The concept of neighborhood rough sets (NRS) comes from these properties of data sets in heterogeneous feature space. This concept was introduced by Hu et al. (2008). The neighborhood of a point $x_i \in U$ is denoted as $\aleph(x_i)$ and is expressed as:

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

In Equation 1, $\triangle$ is a distance function and δ is a threshold dependent on the dataset which may be the same for all points or may vary. So, the granules, defined here around points x_i with its neighborhood $\aleph(x_i)$, are overlapping. Therefore, unlike PaRS a data point may belong to more than one granule in NRS. In other words, one granule can't solely represent only its constituting points.

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

$$\begin{aligned} \underline{R_1}X &= \{x \in U : \aleph(x) \subseteq X\} \\ \overline{R_1}X &= \{x \in U : \aleph(x) \cap X \neq \emptyset\} \\ \underline{R_2}X &= \{\aleph(x) \in U : \aleph(x) \subseteq X\} \\ \overline{R_2}X &= \{\aleph(x) \in U : \aleph(x) \cap X \neq \emptyset\} \end{aligned} \tag{2}$$

From Eq. 2 it can be said that 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, and the set becomes a classical rough set (PaRS). That is, NRS is a generalization of PaRS.

2.3 Relevance of NRS in video

Neighborhood rough set is useful due to its property of incorporating the neighbors' information in heterogeneous feature space. Here, there is no need to define the size of granules. The granules will be automatically formed according to the feature threshold value (δ) and the distance function [$\triangle$ in Eq. (1)]. Since the granules are overlapping, they are used for modeling the overlapping characteristics of data. The basic difference with fuzzy set theoretic modeling is that no membership function is needed, neither any fuzzification/ defuzzification required.

In case of video sequences the moving objects and background are not always crisply separable in all the frames. There is no well defined criterion in a certain feature space to make the object-background separated. Collating the information from several feature space should be more effective for this task. Hence, performing computation in heterogenous feature space seems to be more appropriate. Moreover. there is no need to define a particular threshold to separate object and background. Only the nearness needs to be defined in forming the granules. That is, how much nearer the points have to be in the same granule. Accordingly, the size of the granules is determined incorporating the signal distribution statistics.

The common property of moving object(s) in any sequence is motion. How slow or fast the object moves, all the pixels within the moving object are supposed to have higher and similar temporal values compared to its background. These can be well reflected while forming the neighborhood granules.

Time and accuracy are two major concerns of tracking. Decision making becomes much faster if there are predefined rules. Construction of rule-base using rough set is a popular approach. It contains selection of features, reduction of features and construction of rules. Incorporation of neighborhood granules to form the rule base and make the decision is expected to make the process more faster. The accuracy will not be much hampered if the granules keep the useful information intact.

In the following sections we describe our new concepts of forming neighborhood granules, rough rule base for tracking and quantitative indices. These are followed by experimental results.

3 Neighborhood granules in images

As discussed earlier granular computing is quite effective to reduce the computational time of a system. But it may lead to loss information too. Hence, meaningful granulation is very important for granular computing. Neighborhood granulation incorporates some extra information (neighborness) during granule formation. We are looking for a meaningful granulation in images. Consideration of granules of equal size or predefined shape may not be much effective in that sense. The main concept of our approach here is that the granules in an image should be more natural or in other words like the way that of a human eye. That is, the segments that are roughly distinguished by a human eye in an image should be the granules. The concepts of spatio-color and spatio-temporal neighborhood granules are introduced to serve this purpose.

3.1 Formation of spatio-color granules

The spatial and color nearness are incorporated to form spatio-color granules in still images. A granule $\aleph(x_i)$ over a point x_i is represented formed as

$$\aleph_{sp-clr}(x_i) = \cup x_j \in U : \\ x_i \text{ and } x_j \text{ binary connected over} \\ |color(x_j) - color(x_i)| < Thr \tag{3}$$

It means the granules around x_i contains the pixels where the color difference from that of x_i is $< Thr$ and which is binary connected over this condition. Thr is the color nearness threshold. That is how much similar color will fall into a bin. This way only the neighborhood points (spatial and color based) form granules.

Here, the concept of formation of neighborhood granules is a bit different from the conventional one, that is, as in Eq. (1). Unlike the previous approaches of forming neighborhood granules neither the shapes and sizes are predefined here, nor the granules are formed over every point. The granules here are of arbitrary shapes and sized according to the similarities and x_i does not belong to all of the data points. Here, x_i is such a point which is not already contained by any other granules. This way, the unnecessary overlapping granules and complexity can be avoided. Here the overlapping only occurs when the points in the granules have some common neighborhood properties. That is, the overlapping granules are formed depending on the nature of the dataset but not always. This way, the granulation decreases the complexity without being deviated from its objective. Its effectiveness in images is shown below.

It is seen from Fig. 2 that, the granules are not of fixed shapes or size rather they represent meaningful segments in images. The main advantages of this granulation over the conventional multilevel segmentation or clustering are that,

- no predefined no. of classes is needed
- no image threshold needs to be defined or found out

3.2 Formation of spatio-temporal granules

Temporal information plays the most important role in video processing. Frame difference extraction, computation of median/ mean over frames are a few of the most popular approaches of temporal information extraction. Here another is introduced where instead of giving equal importance to all the current and its previous frames as is done in other approaches, the highest importance is given to the current frame. All the change information from current to its all previous frames are computed and kept as a third dimension of the image. If the current frame (f_t) is of size $M \times N$ and its previous P-frames $(f_{t-p} : p = 1\ldots P)$ are considered then, the changed information matrix $(Temp\_Val)$ in w.r.t. f_t as shown in Eq. (4) is of size $M \times N \times P$. Then compute median over $Temp\_Val$ matrix (see Eq. (4)).

$$Temp\_Val_p = |f_t - f_{t-p}| \forall p \in P \tag{4a}$$

$$Temp\_Val\_Med = Median(Temp\_Val_1, .., Temp\_Val_P) \tag{4b}$$

The spatio-temporal granules are formed over the values of $Temp\_Val\_Med$ in Eq. (4) according to Eq. (3).

The spatio-color and spatio-temporal granules contain the useful information without keeping all the pixel level information. Dealing with these granules seem to be much more convenient in video processing. These granules are used for decision making in the proposed approach discussed in the next section.

4 Tracking using neighborhood granules and rule-base

A brief overview of the proposed algorithm for tracking given here in Fig. 1. It is seen from Fig. 1 that the current and its previous P no.s of frames are given as an input here. It is a partial supervised approach, in the sense that initial labeled dataset are given from which the rule-base is generated. The spatio-color and spatio-temporal granules are the inputs given to the rule-base (see blocks 1 and 2 in Fig. 1). The ways of forming granules are shown in Sect. 3. The foreground segmentation (blocks 3 in Fig. 1) is performed depending on the rule-base and it is tracked. The parameters of the rule base gets updated depending on the tracking result.

The details of the generation of rough rule-bases and method for tracking are described in this Section.

4.1 Rough rule-base for tracking: criterion

The proposed method of tracking is based on generating a rule base depending on some observations and making

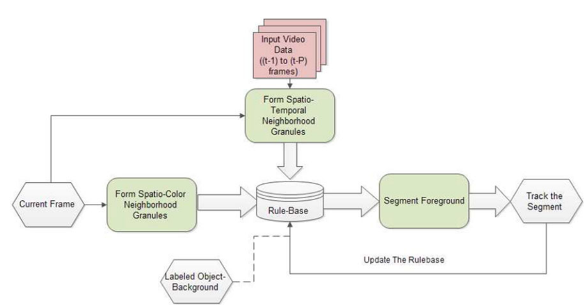

Fig. 1 Block diagram of the proposed tracking method

decisions using it. The conditional attributes of the decision making system are spatial, temporal and color features.

The decision making with the help of a rule base is a faster approach than other computational ones. The speed of the system is a very important phenomenon in video processing, but the uncertainty should also be handled simultaneously. The rough rule-base approach is chosen here to deal with both of the factors. In the following we describe two methods of rule generation, one is in pixel level and the other in granular level.

This is partially supervised approach, where, the initial object/ background should be labeled. The rule-base is formed with the help of this information and observation of *P* nos. of initial frames including the labeled one.

4.2 Rule-base in pixel level

To generate rule base in pixel level, the initial object(s)/ background in the first frame need to be labeled. Once the rule base is formed unknown object(s) in the sequences can be tracked. Let the rule base be generated from the first *P* frames. The observed conditional features are:

- Temporal features: Frame difference in RGB-D- feature space or $(Temp\_Val_1)$ form Eq. 4, denoted as T_{RGB} and T_D respectively.
- Color features: The RGB-D values present in the object model, denoted as RGB_V, D_V.
- Spatial features: The probable location of the object region in the current frame, (generally taken as 1.5 times of the difference region between successive frames), denoted as Sp_L.

The model for each conditional attribute is generated by observing initial *P* frames. The decision attributes are object *O* and background *B*.

The decision table and rule generation is shown in Table 1. The first two conditional attributes have the cardinality of 2 (0-1) which signify change (1) or not change (0), and the last three conditional attributes are also of cardinality two: within the model (W) and outside (Ou). The observation for different cases are shown here.

Twelve different cases are shown in Table 1. The cases are as follows:

1. An ideal background pixels.
2. A background pixel with same features that of an object.
3. An object pixel which moves slower than estimated.
4. A similar colored background pixel that was inside the object in the previous frame.
5. An object pixel, with changing depth (movement towards or away from camera) along with its movement.
6. A moving object pixel moves within the object area That is, object to object pixel.
7. A background pixel within region of interest with RGB value change, that is an object to background pixel.
8. An object pixel which starts to move from the current frame onwards.
9. A noisy background pixel.
10. An object pixel moving in similar RGB-D background.
11. An object pixel which moves faster than estimated.
12. An ideal object pixel or background to object pixel.

Table 1 Rule generation for object background separation

U	T_{RGB}	T_D	RGB_V	D_V	Sp_L	*Decision*
1	0	0	*Ou*	*Ou*	*Ou*	*B*
2	0	0	*W*	*W*	*Ou*	*B*
3	0	0	*W*	*W*	*Ou*	*O*
4	1	0	*W*	*W*	*W*	*B*
5	1	1	*Ou*	*Ou*	*W*	*O*
6	1	1	*W*	*Ou*	*W*	*O*
7	1	0	*Ou*	*W*	*W*	*B*
8	1	1	*Ou*	*Ou*	*Ou*	*O*
9	1	1	*Ou*	*Ou*	*Ou*	*B*
10	0	1	*W*	*Ou*	*W*	*O*
11	1	1	*W*	*W*	*Ou*	*O*
12	1	1	*W*	*W*	*W*	*O*

It can be noticed from Table 1 that, *Rules* 2, 3 and 8, 9 are inconsistent. That means the decision taken according to the rules may not be true. Remaining seven rules are true. So, the dependency between the condition and decision attributes is: 8/12.

To eliminate the superfluous attributes the core of the conditional attributes is computed. It is seen that, elimination of T_{RGB} does not make the rule base indiscernible or reduce the dependency. Hence, T_{RGB} is not a core. Whereas rules 4 and 12 become indiscernible by the elimination of T_D. Individual elimination of RGB_V or D_V does not affect the dependency, but together it affects. Hence, these two attributes can be in reduct. Sp_L is also a core. Based on the concept of core the rules that lead to maximum correct decisions are defined as follows:

1. If $T_D = 1$ and $RGB_V = W$ or $D_V = W$ and $Sp_L = W$ then decision = *O*.
2. If $T_D = 0$ and $RGB_V = Ou$ or $D_V = Ou$ and $Sp_L = Ou$ then decision = *B*.

This is a faster process of tracking. This method is well suited for ideal object-background separation. But, when there is some other challenges or the cases like 2, 3, 8, 9 occurs, this method fails. In search of having more accurate process or to increase the dependency between the condition and decision attributes the concept of rule-base with neighborhood granules is introduced with keeping all of the input information same.

4.3 Rule base with neighborhood granules

The objective of this section is to decrease the no. of misclassified pixels and increase the dependency between condition and decision attributes keeping the input information same. A way of reducing the indiscernibility in the rule base is by increasing the cardinality of the attributes. This can be done if overlapping phenomenon is taken into account.

As discussed earlier proper granulation makes a process faster and accurate. The studies in Sect. 3 shows how effective the spatio-color neighborhood granulation can be to reduce the cardinality of feature space without losing its generality. To utilize the merit of this granulation (in which overlapping is an inherent property) all of the inputs, conditional attributes, decision attributes are defined in neighborhood granular level.

The spatio-color granules in the current frame is given as the input in the rule base. That is, the conditions are checked and decisions are taken over these granules. The set of conditional attributes are the reducts obtained from Sect. 4.2, i.e., $\{T_D, RGB_V, D_V, Sp_L\}$. The conditional attributes in granular level are defined as: (i) temporal granules formed over the *Temp_Val_Med* in D feature space according to Eq. (1) denoted as $T_{\aleph D}$, (ii) the RGB-D granules, i. e., the color granules present in object/background model, denoted as $RGB_\aleph$ and $D_\aleph$, (iii) estimated spatial location SpL which is computed over the spatio-temporal granules (discussed in Sect. 3.2). Generally selected as 1.5 times over the moving granules.

As the conditional attributes form neighborhood granules here, the overlapping phenomenon plays a very useful role. That is, none of the conditional granules are defined crisply rather every individual granule includes its overlapping granules too. An input granule may not always certainly belong or does not belong to a certain conditional granule. But, the input granules or their overlapping granules but can have non-zero intersection with the conditional granules or their overlapping ones. Then these cases are considered as the partial belongingness (PB). This leads to automatic incorporation of the small changes in object-background which is quite obvious in videos.

This way, the cardinality of each conditional attribute becomes three: the input granule (i) belongs (Be), (ii)does not belong (NB), (iii)partially belongs (PB) to the conditional granules. Therefore, as the cardinality of the conditional attributes get increased, as a result of which the inconsistency in the rule base may get reduced. The rule base is as follows considering the twelve cases of 4.2.

It can be seen from Table 2 that, rules 2, 3 are no more inconsistent whereas rules 8,9 are still inconsistent. The physical reflection of the rule base is that, all the cases except noise can now be classified. The dependency between condition and decision attributes is: 10/12.

It can be said that the incorporation of neighborhood granules into the rule-base makes it more robust. It is expected to have more accurate result.

5 Indices to evaluate tracking

There exists several indices to evaluate the performance of tracking, but, most of them are either based on ground truth or the trajectory. The cases under consideration have an extra feature- depth. The depth values using kinect sensor are acquired in such a way that, the objects closer to the sensor have less values, and the values increase along with the distance from the sensor. This property is used to evaluate the performance of tracking. Two indices are defined here using the merits of D-values in tracking. The inputs given for the evaluation are the D-featured video sequence and the moving object segmented sequence.

Two indices, so defined below, are namely E_I and Gr_I. E_I explores the merits of edge neighborhood in D feature space, while Gr_I incorporates the merits of neighborhood granulation in RGB-D feature space.

5.1 E_I index based on edge-neighborhood

The edge of the moving object segment is considered here. As discussed earlier, it is obvious that the moving object is always in front of the corresponding background, hence it has lower depth values. It is obvious that, if a neighborhood

Table 2 Rule generation for object background separation using neighborhood granules

U	$T_{\aleph D}$	$RGB_\aleph$	$D_\aleph$	Sp_L	*Decision*
$\aleph_1$	*NB*	*NB*	*NB*	*NB*	*B*
$\aleph_2$	*NB*	*Be*	*Be*	*NB*	*B*
$\aleph_3$	*PB*	*Be*	*Be*	*NB*	*O*
$\aleph_4$	*PB*	*Be*	*Be*	*PB*	*B*
$\aleph_5$	*Be*	*Be*	*PB*	*PB*	*O*
$\aleph_6$	*PB*	*Be*	*Be*	*Be*	*O*
$\aleph_7$	*NB*	*NB*	*Be*	*Be*	*B*
$\aleph_8$	*Be*	*NB*	*NB*	*NB*	*B*
$\aleph_9$	*Be*	*NB*	*NB*	*NB*	*O*
$\aleph_{10}$	*PB*	*PB*	*Be*	*Be*	*O*
$\aleph_{11}$	*Be*	*Be*	*Be*	*NB*	*O*
$\aleph_{12}$	*Be*	*Be*	*Be*	*Be*	*O*

of a edge pixel has lower or same D-value then, the pixel will belong to the object otherwise background.

Let, E_{ep} be an edge pixel and let, $[x_i], i \in w \times w$ be the set of the neighborhood pixels within $w \times w$ window around E_{ep}. If the segmentation is correct then, the D-values of the foreground pixels (FS) within the window will be same or lower than that of E_{ep}, otherwise higher. The false positive (FP) and false negative (FN) values for edge pixel E_{ep} are accordingly computed as:

$$FP_{ep} = \sum_{i \in w \times w} logical(ep_i > E_{ep}, \quad if\ ep_i \in FS) \tag{5a}$$

$$FN_{ep} = \sum_{i \in w \times w} logical(ep_i < E_{ep}, \quad if\ ep_i \notin FS) \tag{5b}$$

If there are EP no. of edge pixels in a frame, then E_I index is defined incorporating the values obtained form Eq. (5) as:

$$E_I = \frac{\sum_{ep \in EP} \frac{FN_{ep}+FP_{ep}}{w \times w}}{EP} \tag{6}$$

Note: It can be noted that, the indices can reflect over segmentations and under segmentations. In case of over segmentation, the edge pixels in the over segmented regions will always have the same or marginally lower values with respect to its surroundings and the FP will be higher for the pixels, which will increase the E_I too. Similarly in case of under segmentation the FN values will be higher as well as the E_I index. Hence, higher values of this indices reflect less accuracy.

5.2 Gr_I index based on inter–intra granular statistics

The neighborhood granules are formed incorporating spatial and temporal neighborhood information. As the moving object always have lower D-values than its background, the difference values should always be negative in case of moving object granules. However, it may not occur in practical scenario. The area containing a moving object in a frame may not totally be replaced by the background in the next frame. Hence, previous P frames have been considered to incorporate the better effectiveness of D-feature space in temporal domain.

The internal compactness of the 3-D neighborhood granules in D-feature space reflects their effectiveness in the respective sets. The deviation or range of values among the granules in a set shows the accuracy of the set formation. These phenomena are considered while defining the measure.

The internal deviation within a granule can't be more than Thr as all of them are neighborhood granules. But each set contains several granules. Ideally, the parameter values of different granules in a set should not deviate much from each other. But, they differ when the scenario changes. The objective of this measure is to detect such cases. The derivation of Gr_I is as follows:

Let, the sets of internal granular deviations of object(s) ($ObInter$) and background ($BckInter$) be defined respectively as:

$$ObInter = \{vo_i : vo_i = mean(\aleph_{O_i})\} \tag{7a}$$

$$BckInter = \{vb_i : vb_i = mean(\aleph_{B_i})\}. \tag{7b}$$

In Eq. (7), $\aleph_{O_i}$ and $\aleph_{B_i}$ are the i^{th} object and background granules. The sets $ObInter$ and $BckInter$ contains the inter-granular mean values (vo_i and vb_i) for all the granules in the corresponding object and background sets. The intra granular means (mno, mnb) and maximum deviations (mxo, mxb), i.e., the means and deviations among the granules in respective sets are computed for each set as:

$$mno = mean(ObInter) \tag{8a}$$

$$mxo = maxdev(ObInter) \tag{8b}$$

$$mnb = mean(BckInter) \tag{8c}$$

$$mxb = maxdev(BckInter) \tag{8d}$$

In Eq. (8), the value(s) of mno are supposed to be higher (object granules have higher D-difference values) in object set(s). The maximum deviations in any of the sets should be lower if the sets are correctly formed. Hence, mxo and mxb are expected to have lower value. The values of mnb should be low (ideally 0) as those are of background granules. The effective deviations with respect to mean for each of the sets are computed as:

$$mso = \frac{mno}{mxo} \tag{9a}$$

$$msb = \frac{mxb}{mnb} \tag{9b}$$

In Eq. (9), mso is supposed to be higher and msb lower. Lower values of mso signifies there is background part incorporated in the object set i.e., over tracking happens. Whereas higher msb reflects there are object granules in background i.e., under tracking. The Gr_I index is defined incorporating these characteristics as:

$$Gr_I = \frac{msb}{mso} \tag{10}$$

It can be seen from Eq. (10), higher values of Gr_I signifies either higher msb or lower mso or both, all of which reflect mis-tracking. This means that the higher the value of Gr_I the less accurate the tracking is.

6 Results and discussions

In the present section the effectiveness of the concepts defined before are demonstrated experimentally. The experiment is characterized by establishing the effectiveness of:

- proposed neighborhood granulation over images
- rough rule base for video tracking, both in granular and pixel levels
- visual and quantitative performance of the proposed method w.r.t other popular methods
- new neighborhood granular rough rule base in tracking over crisp granular and pixel level rule bases
- the proposed indices for evaluating the quality of tracking.

6.1 Effectiveness of the proposed neighborhood granulation

The effectiveness of the proposed spatio-color granulation is demonstrated here. For doing so, several color and gray-level still images were given as the input for granulation. The results over two of such image, namely Lenna and Peppers are shown in Table 3 and Fig. 2 both in RGB and gray level feature spaces for different values of threshold *Thr* (in Eq. (3)). The corresponding *Beta* (Pal et al. 2000) and *DB* (Davies and Bouldin 1979) indices are also given. Both the visual and quantitative results show that the neighborhood granules lead to meaningful segments with different shapes and sizes. As expected the lower value of *Thr* granulates the images more close to those in pixel level ($Thr = 10$ Fig. 2d), and the nos. of segments or levels get reduced with increase in *Thr*. However, the accuracy get affected, that is, *Beta*-value decreases and *DB*-value increases with increase in *Thr* (granule size). In other words, dealing with less nos. of segments make a process faster at the cost of the accuracy. Here, the *Thr* value of 30 gives a compromise between the accuracy and nos. of segments. The rest of the experiments are conducted with $Thr = 30$.

A few experimental results are shown to establish how effective the proposed granulation (NGr) is to form meaningful segments compared to other segmentation methods. Visual (Fig. 3) and quantitative (Table 4) comparisons of this granulation to two popular multilevel segmentation and clustering methods, namely C-means (Kanungo et al. 2002) and fuzzy C-means (FCM) (Chuanga et al. 2006) are shown here.

From Table 4 it is seen that the NGr gives higher *Beta* values and lower *DB* values which reflect its more accuracy in segmentation. These quantitative results are also validated visually as in Fig. 3.

Table 3 NGr accuracy with variation in *Thr* in *Color* and *Gray* leveled images

Thr	*No. of Levels*	*Beta*	*DB*
Lena Gray			
50	7	33.32	17.58
30	16	15.88	43.35
10	67	108.74	1.37
Lena Color			
50	37	6.6	36.6
30	169	22.5	20.2
10	3662	110	2
Peppers Gray			
50	3	10.07	90.02
30	9	7.29	15.5
10	67	95.27	1.8
Peppers Color			
50	48	7.78	28.39
30	231	20.74	10.5
10	4911	132.26	1.63

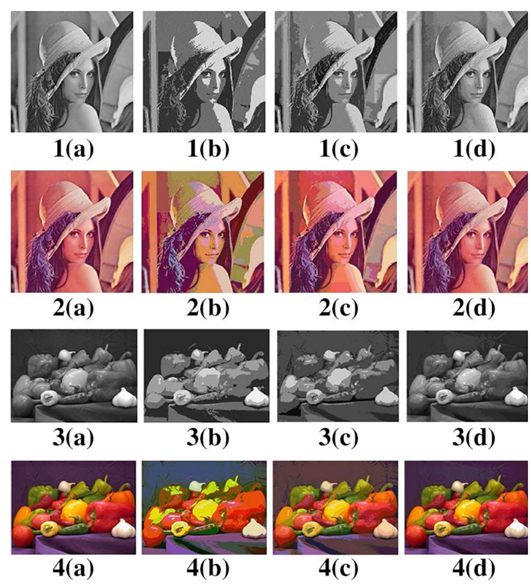

Fig. 2 Granulation over 1,2: *Lena* and 3,4: *Peppers* images **a** Original, granulation with *Thr*: **b**50, **c** 30, and **d** 10

Here, NGr is seen to provide natural granulation with less number of segments.

Videos are colles of frames or still images, hence the aforesaid concept is applied over the frames of videos. The results of tracking are shown in the next section.

6.2 Results of tracking

The method of tracking is implemented with several RGB-D video sequences obtained by kinect sensor. The datasets are obtained from ChaLearn (2011). All of the frames of the sequences are of size 240×320 pixels. There are several sequences with different types of movement of human hands which are found to be properly tracked. The tracking results over three such types of sequences are shown here, as example, where the value of P (The nos. of previous frames) is chosen as 6.

In Fig. 4 there are three types of hand movements. In the first case (sequence M_1) single hand/ left hand is moving with changing shapes, sizes and directions. In the second case (sequence M_7) both of the hands are moving, gets overlapped with each other, gets overlapped with similar colored still object (head). In the third case (sequence M_2) both of the hands are moving in front of similar colored background with changing shapes and sizes. All the scenarios are successfully tracked by the proposed method NGrRB.

6.3 Comparisons between neighborhood granular level and pixel level methods

The time and accuracy comparisons between neighborhood granular and pixel level approaches of the proposed method i.e., NGrRB and PRB are shown in Table 5. The accuracy is measured based on the distance between the centroids (CD) of the ground truth and the obtained foreground segment of the respective frames. The less centroid distance reflects more accuracy. The time given here is the average CPU time in second needed to process a frame.

It is seen from Table 5 that, the pixel level method is a bit faster but is less accurate. These results lead to the theoretical conclusion that the PRB provides less indiscernibility than that of NGrRB, thereby producing inferior performance.

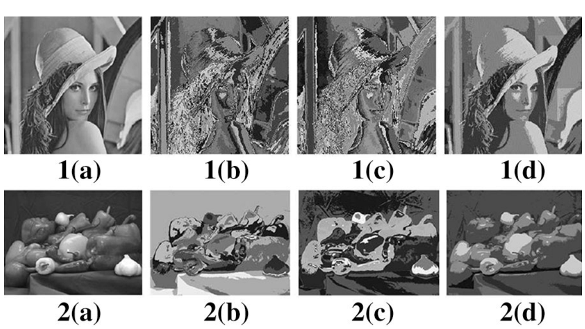

Fig. 3 Visual Comparison Over **1**: *Lena* and **2**: *Peppers* Images **a**: Original and Segmentation Using **b** $K-means$, **c** FCM, **d** NGr With $Thr = 30$

Table 4 *DB*-index and *Beta*-index of $C-Means$, FCM and NGr for *'Lena' Image* and *'Peppers' Image* sequences

Method	*No. of Levels*	*Beta*	*DB*
Lena			
$C-Means$	16	15.88	43.35
FCM	16	14.17	30.26
NGr	16	31.6	17.43
Peppers			
$C-Means$	9	6.5	17.3
FCM	9	7.29	15.5
NGr	9	15.32	7.33

Fig. 4 The results of tracking **a** frame no.s 14, 22, 33, 43 from sequence M_1 **b** 7, 14, 24, 30 from sequence M_7, **c** frame no.s 12, 16, 41, 43 from M_2 sequence

It is shown in Table 1 that rules 2 and 3 are indiscernible, i.e., the pixels that are moving slower than expected can't be correctly classified by PRB, but this can be overcome with the help of NGrRB (see Table 2). One of such visual example of frame no. 12 from M_2 sequence is shown in Fig. 5. Some parts of the hands are moving slower in this scenario. Here, both the hands can't be detected entirely as the foreground segment by PRB (see Fig. 5a) i.e., misclassification occurs, whereas the segmentation is accurately done with NGrRB (see Fig. 5b), as expected.

In a part of the experiment, we have demonstrated the advancement of taking neighborhood granules over crisp granules. The algorithm is executed with crisp granules of window sizes 5×5 and 7×7. The time and accuracy performance for the aforesaid three sequences are listed in Table 6. As expected, the computational time reduces in case of crisp granular rule base (CGRB) but leads to loss of information and hence poorer results. Moreover, CGRB is dealing with the similar rule base than that of PRB, hence has the same indiscernibility. These two factors together make the CGRB based system even less accurate than by PRB. Further as investigation says larger the size of crisp granules faster is the process and less is the accuracy not

Table 5 Time and accuracy comparisons between NGrRB and PRB methods

Sequence	Method	Avg. CPU Time/ Frames	*CD*
*M*_1	*NGrRB*	0.237	10.03
*M*_1	*PRB*	0.215	14.59
*M*_7	*NGrRB*	0.308	15.52
*M*_7	*PRB*	0.276	16.96
*M*_2	*NGrRB*	0.251	6.54
*M*_2	*PRB*	0.21	7.062

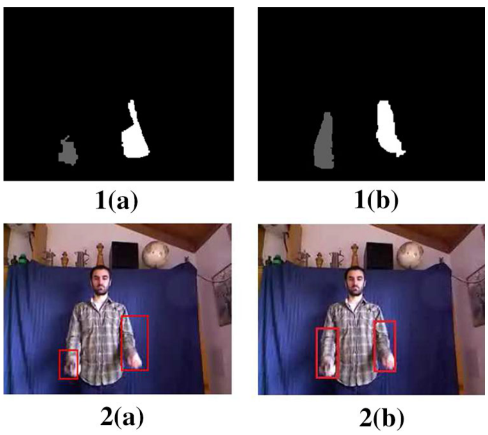

Fig. 5 Visual Comparison between **a** PRB and **b** NGrRB: **1** Foreground Segments, **2** Tracking Results

Table 6 Time and accuracy comparisons between NGrRB and CGRB

Sequence	Method	Avg. CPU Time/ Frames	*CD*
*M*_1	*NGrRB*	0.237	10.03
*M*_1	$CGRG_{5\times5}$	0.255	15.61
*M*_1	$CGRG_{7\times7}$	0.203	17.98
*M*_7	*NGrRB*	0.308	15.52
*M*_7	$CGRG_{5\times5}$	0.245	17.15
*M*_7	$CGRG_{7\times7}$	0.198	17.98
*M*_2	*NGrRB*	0.251	6.54
*M*_2	$CGRG_{5\times5}$	0.189	7.75
*M*_2	$CGRG_{7\times7}$	0.123	8.92

only compared to NGrRB, but also to PRB. Whereas NGrRB keeps a balance between time and accuracy.

6.4 Comparisons with other methods

Comparative study of the proposed NGrRB method is done with two popular partially supervised methods, namely, kernel based mean-shift tracking (MS) (Comaniciu et al. 2003) , and PLS-tracking (Wang et al. 2012).

Quantitative comparisons are shown in Table 7 with the same metrics as in Table 5 for the aforesaid three

Table 7 Time and accuracy comparisons between NGrRB, PLS and MS methods

Sequence	Metric	NGrRB	PLS	MS
*M*_1	Average centroid distance	10.03	11.42	15.06
*M*_1	CPU time/frame (s)	0.237	0.252	0.310
*M*_2	Average centroid distance	6.54	7.21	8.86
*M*_2	CPU time/frame (s)	0.251	0.323	0.330
*M*_7	Average centroid distance	15.52	16.32	17.22
*M*_7	CPU time/frame (s)	0.308	0.341	0.395

sequences. It can be noticed from the table that, the NGrRB method consumes less time compared to the other two methods. The performance of PLS is nearly equal to that of NGrRB though it needs slightly more time, whereas MS performs the worst. Visual results support well the performance as inferred by the centroid distance.

6.5 Validation of proposed measures

The effectiveness of the proposed Gr_I [see Eq. (10)] and E_I [see Eq. (6)] indices is shown here along with examples. Their values obtained for tracking of *M*_1 sequence with the three methods (MS, PLS and NGrRB) are shown graphically in Fig. 7. The comparative visual results for frame nos. 14, 22, 33, 43 are there in Fig. 6 as example. In case of fame no. 14, MS gives the worst result and overt-tracking takes place (Fig. 61a), whereas in case of PLS and NGrRB the performances are satisfactory. These are well reflected by Gr_I index (Fig. 6b). Over-tracking by MS includes the extra object which doesn't belong to the object spatio-temporal granule, and hence increases the inter-intra granular deviations in the object set resulting in higher Gr_I index. But, in case of E_I index (Fig. 6a), the extra inclusion (part of the head) by MS is a static part of the object. Hence it also has lower depth values as compared to its background, and by the fact the index E_I can't

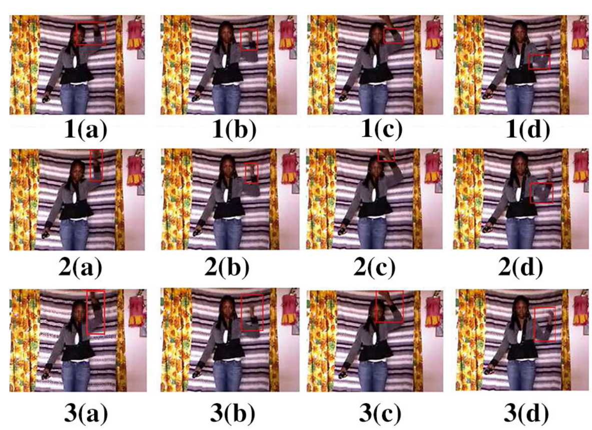

Fig. 6 Tracking Results for Frame nos. **a** 14, **b** 22, **c** 33, **d** 43 of *M*_1 Sequence, **1** MS-method, **2** PLS method and **3** NGrRB method

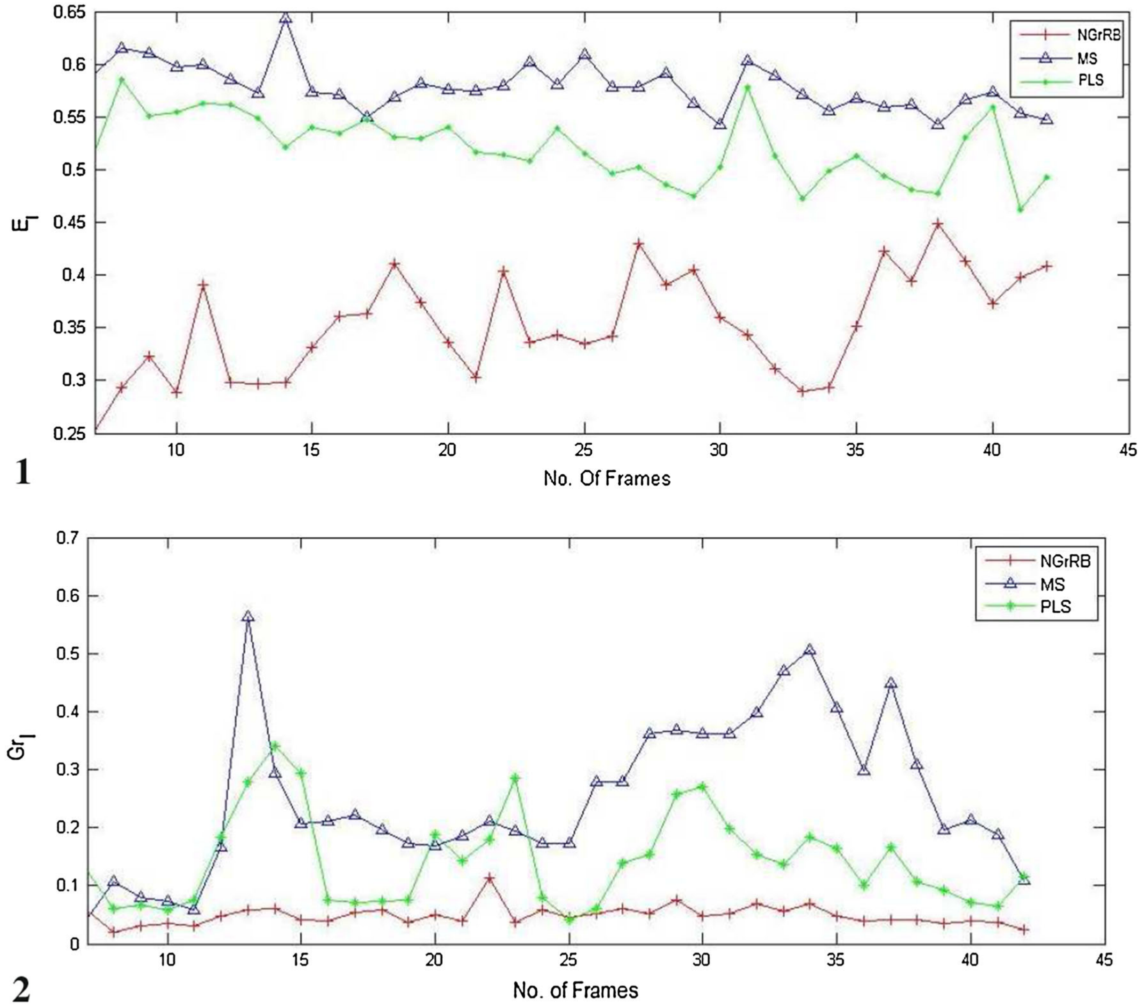

Fig. 7 Quantitative comparisons among **1** E_I Indices, **2** Gr_I indices

sharply reflect the mis-tracking. In the same way, in case of frame no. 22 only a small part of the moving left hand is being tracked by MS and PLS whereas the total hand is being tracked by NGrRB. The Gr_I and E_I are the lowest for it. Similar conclusions hold good for the other cases (e.g., frame nos. 33 and 43) Hence, it can be said that, the proposed indices can effectively evaluate the performance of tracking.

7 Conclusions

The objective of this investigation is to introduce a new way of forming natural granules, demonstrating how the concept of rough-rule base can be used for tracking over this granulation, and defining two new indices for evaluation of the tracking.

The method of formation of neighborhood granules proves to be pretty effective for meaningful natural partitioning of images with arbitrary shapes and sizes of regions. It reduces the nos. of levels in images. These phenomenon can be applied to other problems of image processing, e.g., segmentation, compression.

The use of rough rule base in video tracking significantly reduces the time requirement. This technique also provides good results in relatively simpler cases; thereby making it suitable for tracking over less complex scenarios with effective time improvement.

The concept of neighborhood granular rough rule base proves to be effective in reducing the indiscernibility compared to pixel leveled one. This way it provides more accurate decision-making. It may have wide application in other areas for decision science where indiscernibility is prominent.

The proposed indices for performance evaluation of tracking are able to detect less accurate or mis-tracked results without ground truth information.

Among the three granulation methods, NGrRB (neighbourhood granule based)is superior to PRB (pixel level) and CGRB (crisp granule based) in terms tracking performance as it models the overlapping characteristics of regions. From computation point of view, CGRB takes least time.

Acknowledgments S. K. Pal acknowledges the J. C. Bose National Fellowship, Government of India and Indian National Academy of Engineering (INAE) Chair Professorship.

References

ChaLearn (2011) ChaLearn Gesture Dataset (CGD 2011). California

Chuanga KS, Tzenga HL, Chena S, Wua J, Chenc TJ (2006) Fuzzy c-means clustering with spatial information for image segmentation. Comput Med Imaging Graph 30(1)

Comaniciu D, Ramesh V, Meer P (2003) Kernel-based object tracking. IEEE Trans Pattern Anal Mach Intell 25:564–577

Davies DL, Bouldin DW (1979) A cluster separation measure. IEEE Trans Pattern Anal Mach Intell 1:224–227

Du Y, Hu Q, Zhu P, Ma P (2011) Rule learning for classification based on neighborhood covering reduction. Inf Sci 181(24):5457–5467

Heikkila M, Pietikainen M (2006) A texture-based method for modeling the background and detecting moving objects. IEEE Trans Pattern Anal Mach Intell 28:657–662

Hu Q, Yu D, Liu J, Wu C (2008) Neighborhood rough set based heterogeneous feature subset selection. Inf Sci 178(18):3577–3594

Janoch A, Karayev S, Jia Y, Barron JT, Fritz M, Saenko K, Darrell T (2011) A category-level 3-d object dataset: putting the kinect to work. In: ICCV Workshop on Consumer Depth Cameras for Computer Vision

Kanungo T, Mount DM, Netanyahu NS, Piatko CD, Silverman R, Wu A (2002) An efficient k-means clustering algorithm: analysis and implementation. IEEE Trans Pattern Anal Mach Intell 24(7): 881–892

Komorouski J, Pawlak Z, Polkowski L, Skowron A (1999) Rough sets: a tutorial. In: Pal SK, Skowron A (eds) Rough fuzzy hybridization: a new trend in decision-making. Springer, Singapore, pp 3–98

Lai K, Bo L, Ren X, Fox D (2011) A large-scale hierarchical multi-view rgb-d object dataset. In: ICRA. (pp. 1817–1824)

Maggio E, Cavallaro A (2010) Video tracking: theory and practice. Wiley, New York

Meher SK, Pal SK (2011) Rough-wavelet granular space and classification of multispectral remote sensing image. Appl Soft Comput 11(8):5662–5673

Mushrif MM, Ray AK (2008) Color image segmentation: rough-set theoretic approach. Pattern Recogn Lett 29(4):483–493

Pal SK, Ghosh A, Shankar BU (2000) Segmentation of remotely sensed images with fuzzy thresholding, and quantitative evaluation. Int J Remote Sens 21:2269–2300

Pal SK, Shankar BU, Mitra P (2005) Granular computing, rough entropy and object extraction. Pattern Recogn Lett 26:2509–2517

Pal SK, Chakraborty D (2013) Unsupervised tracking, roughness and quantitative indices. Fundam Inform (IOS Press) 124(1–2):63–90

Pawlak Z (1992) Rough sets: theoretical aspects of reasoning about data. Kluwer Academic Publishers, Norwell, MA

Pedrycz W, Song M (2014) A granulation of linguistic information in ahp decision-making problems. Inf Fusion 17:93–101

Sen D, Pal SK (2009) Generalized rough sets, entropy, and image ambiguity measures. IEEE Trans Syst Man Cyberns Part B 39: 117–128

Sinthanayothin C, Wongwaen N, Bholsithi W (2012) Skeleton tracking using kinect sensor and displaying in 3d virtual scene. Int J Adv Comput Technol (IJACT) 4(11):213–223

Swirniaski RW (2001) Rough sets methods in feature reduction and classification. Int J Appl Math Comput Sci 11:565–582

Wang Q, Chen F, Xu W, Yang MH (2012) Object tracking via partial least squares analysis. IEEE Trans Image Proc 21(10):4454–4465

Yilmaz A, Javed O, Shah M (2006) Object tracking: a survey. ACM Comput Surv 38(4):1264–1291

Zadeh LA (1997) Toward a theory of fuzzy information granulation and its centrality in human reasoning and fuzzy logic. Fuzzy Sets Syst 90(2):111–127