

Image Segmentation for Different Color Spaces using Dynamic Histogram based Rough-Fuzzy Clustering Algorithm

E.Venkateswara Reddy
Research Scholar
Acharya Nagarjuna University
Guntur

E.S.Reddy, Ph.D
Principal
University College of Engg
ANU, Guntur

ABSTRACT

This paper describes a comparative study of color image segmentation for various color spaces such as RGB, YUV, XYZ, Lab, HSV, YCC and CMYK using Dynamic Histogram based Rough Fuzzy C Means (DHRFCM). The proposed algorithm DHRFCM is based on modified Rough Fuzzy C Means (RFCM), which is further divided into three stages. In the pre-processing stage, convert RGB into required color space and then select the initial seed points by constructing histogram. In the next phase, use the rough sets to reduce the seed point selection and then apply Fuzzy C-Means algorithm to segment the given color image. The proposed algorithm DHRFCM produces an efficient segmentation for color images when compared with RFCM and also the unsupervised DHRFCM algorithm is compared with different clustering validity indices such as Davies-Bouldin (DB index), Rand index, silhouette index and Jaccard index and their computational time for various color spaces. Experimental results shows that the proposed method perform well and improve the segmentation results in the vague areas of the image.

General Terms

Image Processing, Color Image Segmentation, Validity indices

Keywords

Rough Sets, Dynamic Histogram Fuzzy C-Means algorithm, Rough Fuzzy C-Means algorithm, histogram, RGB, YUV, HSV, XYZ, LAB, CMYK conversions

1. INTRODUCTION

Image segmentation is one of the most challenging tasks in image analysis and a one of the very important pre-processing steps in low level computer vision as well as high level computer vision. It is also useful in the field of pattern recognition. In recent years, representing uncertainty in spatial data has become more and more of a concerning. Spatial data is often associated with uncertainty because of measurement inaccuracy, sampling discrepancy, outdated data sources, or other errors [1][2]. Uncertainty can be caused by our limited perception or understanding of reality (e.g., limitations of the observation equipment; limited resources to collect, store, transform, analyze, or understand data). It can also be inherent in nature (e.g., due to prejudice). Moreover, sensors (e.g., acoustic, chemical, electromagnetic, mechanical, optical radiation and thermal sensors) are often used to collect data in applications such as environment surveillance, security, and measurement inaccuracies, sampling frequency of the sensors, deviation caused by a rapid change of the measured property over time (e.g., drift, noise), wireless transmission errors, or network latencies. It is getting increasingly important to specify how large the uncertainty in this data is, and,

consequently, how large the uncertainty in the information obtained from this data is. Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images [5][6][7]. Image Segmentation is becoming more important for medical diagnosis process. Currently, development an efficient computer aided diagnosis system that assist the radiologist has thus become very interest, the aim being not to replace the radiologist but to over a second opinion [3, 4]. Consequently, the need of efficient research on features extracted and their role to the classification results makes researchers to select features randomly as input to their systems. In image segmentation an image is divided into different regions with similar features. There are many different types of approaches of image segmentation. Edge-based method, region-based techniques and threshold-based techniques and so on. Images are partitioned according to their global feature distribution by clustering based image segmentation methods. In this paper, a image segmentation method based on K-means using rough set theory is proposed, in which pixels are clustered according to the intensity and spatial features and then clusters are combined to get the results of final segmentation. The paper is organized as follows. In section 2 color models and conversions are described. In section 3 rough set based fuzzy c means algorithm is presented. In section 4 the proposed method of Histogram based fuzzy C-Means clustering algorithm is presented. In section 5 experimental results and in section 6 some conclusions have been made

2. COLOR MODELS

The purpose of a color model is to facilitate the specification of colors in some standard generally accepted way. In essence, a color model is a specification of a 3-D coordinate system and a subspace within that system where each color is represented by a single point. Each industry that uses color employs the most suitable color model. For example, the RGB color model is used in computer graphics, YUV or YCbCr are used in video systems, PhotoYCC is used in Photo CD production and so on. Transferring color information from one industry to another requires transformation from one set of values to another. Intel IPP provides a wide number of functions to convert different color spaces to RGB and vice versa. In the RGB model, each color appears as a combination of red, green, and blue. This model is called additive, and the colors are called primary colors. The primary colors can be added to produce the secondary colors of light (see Figure "Primary and Secondary Colors for RGB and CMYK Models") - magenta (red plus blue), cyan (green plus blue), and yellow (red plus green). The combination of red, green, and blue at full intensities makes white.

3. ROUGH SET CONCEPTS

Rough Set Theory was firstly introduced by Pawlak in 1982 [2][3], and is a valuable mathematical tool for dealing with vagueness and uncertainty [4]. Similar or indiscernibility relation is the mathematical basis of the Rough Set theory. The key concept of rough set theory is the approximate equality of sets in a given approximation space [2][3]. An approximation space A is an ordered pair (U, R) , where U is a certain set called universe, and that equivalence relation $R \subset U \times U$ is a binary relation called indiscernibility relation; if $x, y \in U$ any $(x, y) \in R$, this means that x and y are indistinguishable in A ; equivalence classes of the relation R are called elementary sets (atoms) in A (an empty set is also elementary), and the set of all atoms in A is denoted by U/R . In the Rough Set approach, any vague concept is characterized by a pair of precise concepts, that is the lower and upper approximation of the vague concept (Pawlak, 1997). Let $X \subseteq U$ be a subset of U , then the lower and upper approximation of X in A are respectively denoted as:

$$\underline{A}(X) = \{x \in U : [x]_R \subset X\},$$

$$\bar{A}(X) = \{x \in U : [x]_R \cap X \neq \emptyset\},$$

Where $[x]_R$ denotes the equivalence class of the relation R containing element x . In addition, the set $BN_A(X) = \bar{A}(X) - \underline{A}(X)$ is called a boundary of X in A . If set X is roughly definable in A it means that we can describe the set X with some "approximation" by defining its lower and upper approximation in A . The upper approximation $\bar{A}(X)$ means the least definable set in A containing the objects that possibly belong to the concept, whereas the lower approximation $\underline{A}(X)$.

3.1 Rough Set Based Fuzzy C-Means

Algorithm

Specific steps of the RFCM clustering algorithm are given as follows:

Step1: Randomly initialize the number of clusters to c , where $2 \leq c \leq \sqrt{n}$ and n is number of image points

Step2: Randomly chooses c pixels from the image data set to be cluster centers

Step3: Optimize the initial cluster centers set by Rough Sets

Step4: Set step variable $t=0$, and a small positive number ϵ .

Step5: Calculate (at $t=0$) or update (at $t>0$) the membership

matrix $U = \{u_{k,x}\}$ using $C_i = \sum_{j=1}^n U_{ij}^m X_j / \sum_{j=1}^n U_{ij}^m$

Step6: Update the cluster centers

$$U_{ij} = \frac{1}{\sum_{l=1}^k x_j \in R_{wi} \left(\frac{d_{ij}^2}{d_{il}^2} \right) \frac{1}{m-1}}$$

Step7: Compute the corresponding Xie-Beni index

Step8: Repeat step 5-8 until $\|XB^{t+1} - XB^t\| < \epsilon$

Step9: Return the best XB and best center positions

4. RELATED WORK

In the existed rough set based fuzzy c means algorithm, initial seed points will consider randomly. This will take time consuming and will not get more accuracy. Instead of selecting seed points randomly, we select initial seed points by constructing the histogram for three components and then split each component as 'C' groups (required no of clusters C). Consider peak values of every group as initial seed point. The proposed Method, Dynamic Histogram based RFCM consists of the following steps

Algorithm DHRFCM

Step1: Read an RGB image as input

Step2: Convert RGB image into Required Color Space (HSV, YCC, CMYK, Lab, etc.).

Step3: Split Image into n partitions

Step4: Construct histogram for each partition, and then split it into 'C' blocks (C is no of clusters required).

Step5: Consider peak values of every block as initial seed point of that particular cluster.

Step6: Apply Rough sets concepts to reduce the seed points

Step7: Use Fuzzy C-Means (FCM) algorithm to get the segmented image

Step8: Find the cluster validity indices such as DB index, Rand index, XB index, etc, to identify the best segmentation results

5. EXPERIMENTAL RESULTS

In this section, experimental results on real images are described in detail. In these experiments, the number of different types of object elements in each image from manual analysis was considered as the number of clusters to be referenced. The Xie-Beni index value has been utilized throughout to evaluate the quality of the classification for all algorithms. All experiments were implemented on PC with 1.8GHz Pentium IV processor using MATLAB (version9.0).

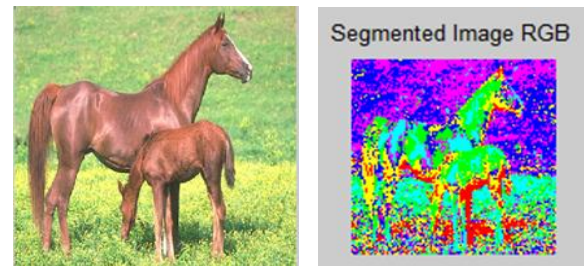


Fig 1. (a) Original Image Corel - 744 (b) Segmented in RGB

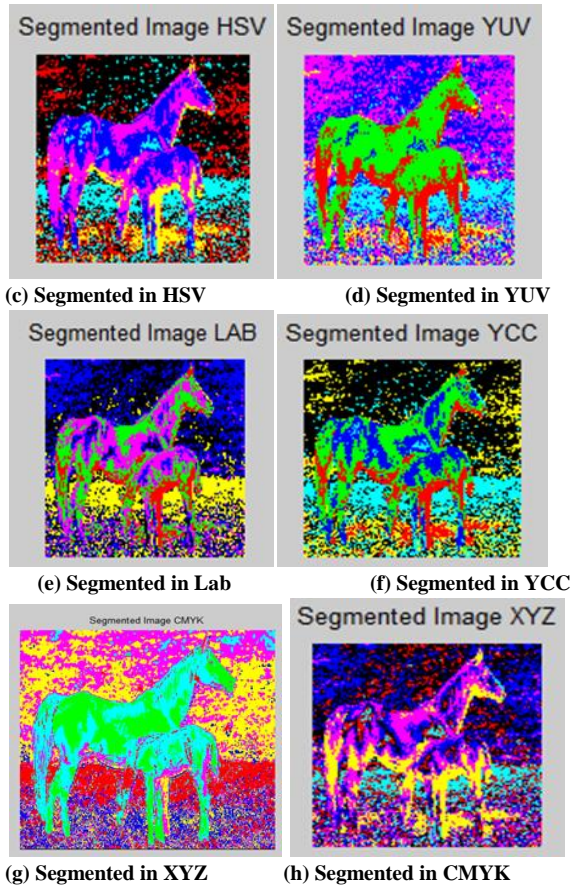


Table 1. Segmentation Time for different Color Models of the Image: Corel - 744

Color Model	Time in Sec
RGB	47.849976
HSV	30.783493
YUV	35.007284
LAB	49.635199
XYZ	65.500038
YCC	64.521090
CMYK	31.921300

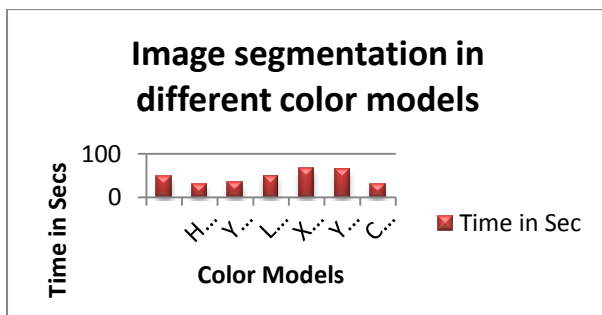


Fig 2. Segmentation Time for different Color Models

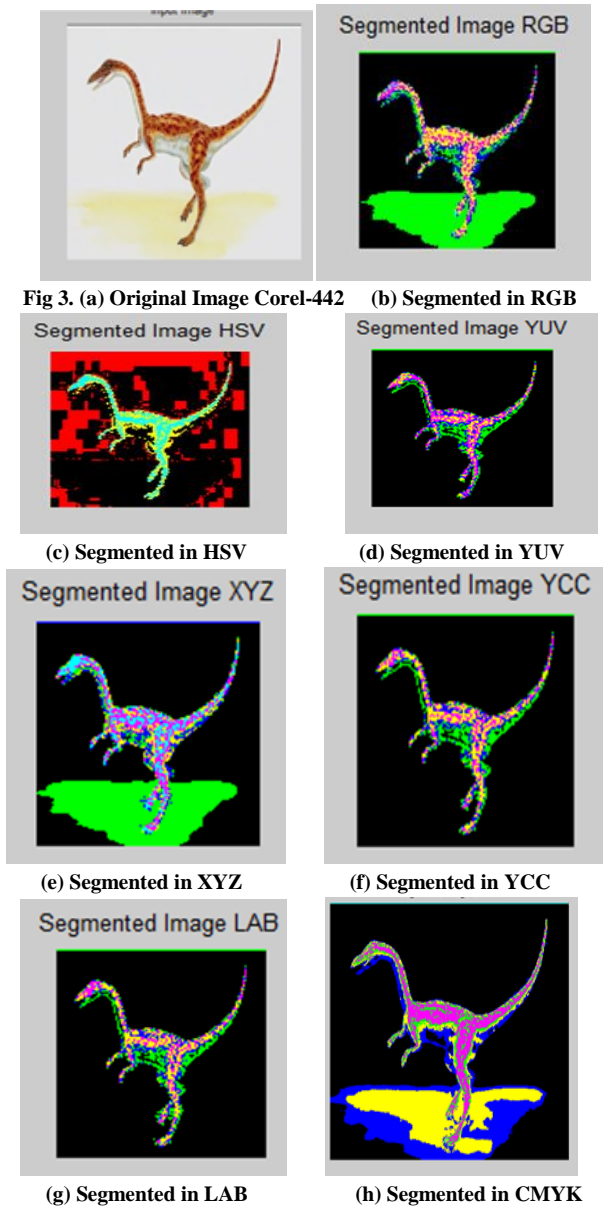


Table 2. Segmentation Time for different Color Models of the Image: Corel - 442

Color Model	Time in Sec
RGB	44.641166
HSV	33.496686
YUV	51.033043
LAB	41.160160
XYZ	53.374113
YCC	42.118179
CMYK	39.482357

Image Segmentation in different color Models

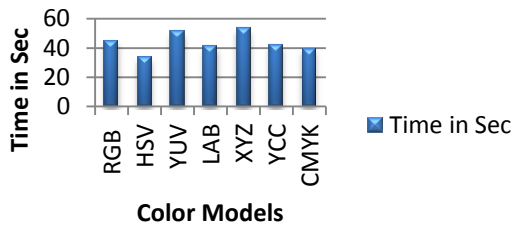


Fig 4. Segmentation Time for different Color Models

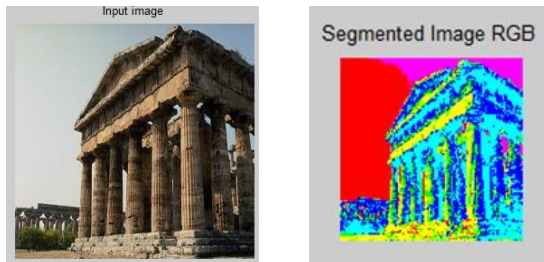
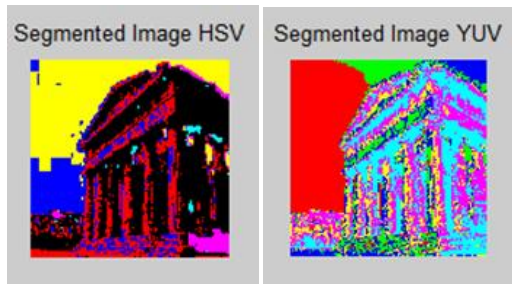
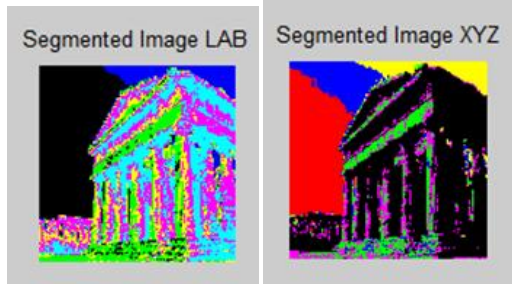


Fig 5. (a) Original Image Corel-228 (b) Segmented in RGB



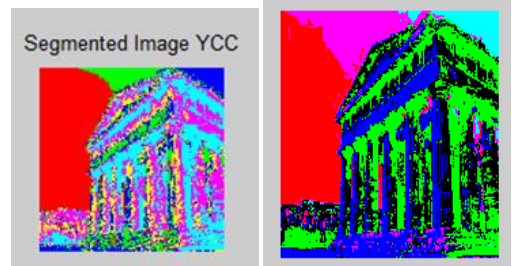
(c) Segmented in HSV

(d) Segmented in YUV



(e) Segmented in LAB

(f) Segmented in XYZ



(g) Segmented in YCC

(h) Segmented in CMYK

Table 3. Segmentation Time for different Color Models of the Image: Corel - 228

Color Model	Time in Sec
RGB	52.725633
HSV	30.187466
YUV	45.177639
LAB	44.520543
XYZ	38.011824
YCC	54.768642
CMYK	37.989592

Image Segmentation in different color Models

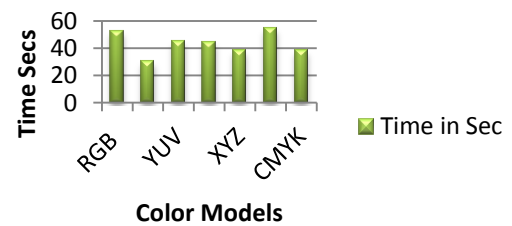


Fig 6. Segmentation Time for different Color Models

Table 4. Various Validity Indices for Iris Data set using Rough Fuzzy C Means Clustering with Euclidean distance

No of Clusters	DB index	Jaccard Index	Silhouette index	Rand index
2	1.1827	0.72886	0.61289	0.92886
3	1.1335	0.69586	0.55282	0.87973
4	0.85632	0.62817	0.42217	0.85423
5	0.76193	0.55673	0.4899	0.83884
6	0.75073	0.62719	0.3895	0.85351
7	0.52537	0.55544	0.46448	0.8511
8	0.40347	0.46883	0.33198	0.80555
9	0.35298	0.35674	0.34265	0.78523
10	0.27834	0.32514	0.20386	0.76653

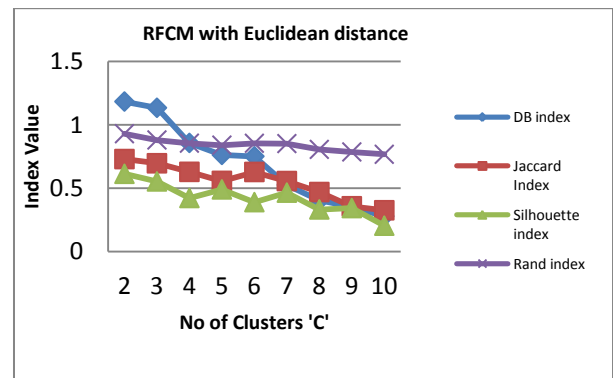


Fig 7. Validity indices value based on the Clusters

Table 5. Various Validity Indices for Iris Data set using Rough Fuzzy C Means Clustering with Correlation distance

No of clusters	DB index	Jaccard Index	Silhouette index	Rand index
2	0.63023	0.81348	0.91289	0.98238
3	0.57214	0.79971	0.80063	0.92671
4	0.55692	0.69478	0.67956	0.89021
5	0.46453	0.61424	0.62613	0.86813
6	0.44844	0.64443	0.60588	0.87311
7	0.41334	0.52032	0.60468	0.83839
8	0.37384	0.50665	0.51232	0.83062
9	0.30453	0.43922	0.43228	0.81011
10	0.30188	0.33923	0.34472	0.77933

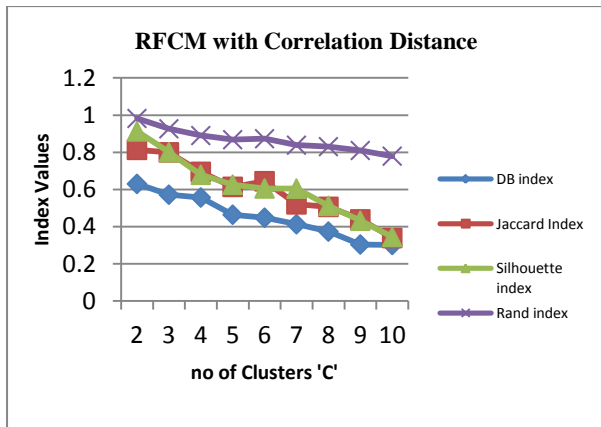


Fig 8: Validity indices value based on the Clusters

Table 6. Various Validity Indices for Iris Data set using Dynamic Histogram based RFCM Clustering with Euclidean distance

No of Clusters	DB index	Jaccard Index	Silhouette index	Rand index
2	0.51842	0.54886	0.62283	0.57123
3	0.40561	0.52586	0.56283	0.54674
4	0.45632	0.49817	0.42217	0.49145
5	0.36193	0.45674	0.48553	0.43861
6	0.25073	0.35714	0.34952	0.39382
7	0.22536	0.32544	0.46234	0.32167
8	0.10356	0.23863	0.32784	0.23162
9	0.09218	0.15684	0.23265	0.19234
10	0.08834	0.12614	0.10386	0.11346

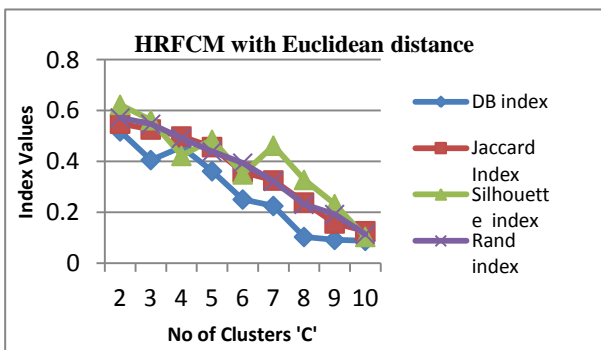


Fig 9: Validity indices value based on the Clusters

Table 7. Various Validity Indices for Iris Data set using Dynamic Histogram based RFCM Clustering with Correlation distance

No of Clusters	DB index	Jaccard Index	Silhouette index	Rand index
2	0.41845	0.54182	0.72478	0.61231
3	0.38745	0.42154	0.52314	0.56178
4	0.35123	0.49812	0.50012	0.49123
5	0.29654	0.38513	0.42378	0.41789
6	0.19643	0.24178	0.32167	0.38156
7	0.11262	0.19436	0.21378	0.31445
8	0.08934	0.21468	0.19236	0.29134
9	0.04623	0.09256	0.25618	0.19231
10	0.02674	0.10245	0.11673	0.09267

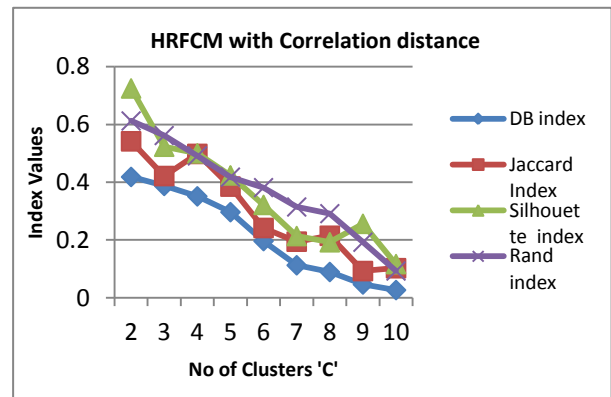


Fig 10: Validity indices value based on the Clusters

6. CONCLUSION

The developed approach contributes a hybrid methodology, which integrates judiciously rough sets and fuzzy c-means algorithm. This formulation is geared towards maximizing the utility of both rough sets and fuzzy sets with respect to knowledge discovery tasks. The proposed method is applied on several real data sets from UCI Machine Learning Repository and Images from Corel, BSDS data set. Some of the cluster validity indices such as DB index, Rand index, silhouette index, and Jaccard index for evaluating the quality of the proposed algorithm are also considered. The effectiveness of the proposed algorithm is demonstrated, along with a comparison with other related algorithm like RFCM and proved better performer by considering various indices given. Experimental results shows that the proposed method perform well and improve the segmentation results in the vague areas of the image. The proposed algorithm proves that HSV and CMYK color images will give better segmentation results when compared with other color models.

7. REFERENCES

- [1] Skowron A., Stepaniuk, J. : Information granules in distributed environment. In: Zhong, N., Skowron, A., Ohsuga, S. . (eds.) RSFDGr C 1999. LNCS (LNAI) Vol. 1711, pp. 357–365. Springer, Heidelberg (1999).
- [2] Wei Zhang, Yu-zhu Zhang, Cheng Li, "A new hybrid algorithm for image segmentation based on rough sets and enhanced fuzzy c-means clustering", IEEE, International Conference on Automation and Logistics, August 2009.

- [3] Milinda M. Mushrif, Ajoy K. Ray, “Color image segmentation: Rough-set theoretic approach”, Pattern Recognition Letters, Elsevier, 2008
- [4] S Mohapatra, D Patra, “Unsupervised Leukocyte Image Segmentation Using Rough Fuzzy Clustering”, ISRN Artificial Intelligence, 2012
- [5] Hassanien, A. E., A. Abraham, J. F. Peters, G. Schaefer, and C. Henry. 2009. Rough sets and near sets in medical imaging: A review. *IEEE Transactions on Information Technology in Biomedicine* 13(6):955–968. Digital object identifier: 10.1109/TITB.2009.2017017.
- [6] Beaubouef, Theresa, Frederick E Petry, and Gurdial Arora. 1998. Information-theoretic measures of uncertainty for rough sets and rough relational databases. *Information Sciences* 109(1-4):185–195.
- [7] Martin, D.,C. Fowlkes, D. Tal, and J. Malik. 2001. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In proceedings of the 8th international conference on computer vision, vol. 2, 416–423.
- [8] Bhatt, R B, and M Gopal. 2004. Frid: fuzzy-rough interactive dichotomizers. In Proceedings of the IEEE international conference on fuzzy systems, 1337–1342.
- [9] N. Mac Parthalain, R. Jensen, and Q. Shen. Comparing Fuzzy -Rough and Fuzzy Entropy-assisted Fuzzy-Rough Feature Selection. Proceedings of the 6th Annual Workshop on Computational Intelligence (UKCI’06), 2006.
- [10] Z.Bao, B.Han, and S. Wu. A novel clustering algorithm based on variable precision rough-fuzzy sets, Proceedings of the International conference on intelligent computing, ICIC 2006, Kunming, China, August 16-19, pp. 284-289, 2006.
- [11] S. Asharaf and M.N. Murty. An adaptive rough fuzzy single pass algorithm for clustering large data sets, *Pattern Recognition*, vol. 36, no. 12, pp. 3015–3018, 2004.
- [12] W.H. Au and K.C.C. Chan. An Effective Algorithm for Discovering Fuzzy Rules in Relational Databases. Proceedings of the 7th IEEE International Conference on Fuzzy Systems, pp. 1314–1319, 1998.
- [13] Banerjee, M., Mitra, S., Pal, S. K.: Rough-Fuzzy MLP: Knowledge Encoding and Classification.
- [14] S Mohapatra, D Patra, “Unsupervised Leukocyte Image Segmentation Using Rough Fuzzy Clustering”, ISRN Artificial Intelligence, 2012
- [15] Jensen, R., Shen, Q.: Fuzzy-Rough Attribute Reduction with Application to Web Categorization, *Fuzzy Sets and Systems*, 141, 2004, 469–485.
- [16] Krishnapuram, R., Keller, J. M.: A Possibilistic Approach to Clustering, *IEEE Transactions on Fuzzy Systems*, 1(2), 1993, 98–110.
- [17] E.V.Reddy, Dr. E.S.Reddy “Image Segmentation using Rough set based Fuzzy K Means clustering algorithm”, Global Journals Inc, GJCST Volume 13 Issue-6 Version 1.0 ISSN : Online: 0975-4172 Print: 0975-4350
- [18] E.V.Reddy, G.V.Suresh, Dr. E.S.Reddy “rough set analysis for uncertain data classification”, *IEEE Explore*, 2010.
- [19] E.V.Reddy, G.V.Suresh “uncertain data classification using rough set theory” advances in intelligent and soft computing series: 4240 of Springer.