Evaluation of Fuzzy Rough Set Feature Selection for Content Based Image Retrieval System with Noisy Images

Maryam Shahabi Lotfabadi
School of Engineering &
Information Technology,
Murdoch University
90 South Street, Murdoch
Western Australia 6150
M.Shahabi@Murdoch.edu.au

Mohd Fairuz Shiratuddin
School of Engineering &
Information Technology,
Murdoch University
90 South Street, Murdoch
Western Australia 6150
F.Shiratuddin@Murdoch.edu.au

Kok Wai Wong
School of Engineering &
Information Technology,
Murdoch University
90 South Street, Murdoch
Western Australia 6150
K.Wong@Murdoch.edu.au

ABSTRACT

In this paper Fuzzy Rough Set is used for feature selection in the Content Based Image Retrieval system. Noisy query images are fed to this Content Based Image Retrieval system and the results are compared with four other feature selection methods. The four other feature selection methods are Genetic Algorithm, Information Gain, OneR and Principle Component Analysis. The main objective of this paper is to evaluate the rules which are extracted from fuzzy rough set and determine whether these rules which are used for training the Support Vector Machine can deal with noisy query images as well as the original queried images. To evaluate the Fuzzy Rough set feature selection, we use 10 sematic group images from COREL database which we have purposely placed some defect by adding Gaussian, Poisson and Salt and Pepper noises of different magnitudes. As a result, the proposed method performed better in term of accuracies in most of the different types of noise when compared to the other four feature selection methods.

Keywords

Fuzzy Rough Set, Content Based Image Retrieval system, Noise.

1. INTRODUCTION

Content Based Image Retrieval (CBIR) systems have been one of the important research areas [1, 2]. Many researchers work on different parts of the CBIR systems to improve them, such as classifiers [3], relevance feedback [4], retrieval process [5], features [6] and etc. In this paper, working with noisy queried images is investigated and an efficient method for dealing with noisy query images is proposed.

Image noise is a random (not present in the object image) variation of brightness or colour information in them, and is usually an aspect of electronic noise [7, 8]. Noise represents unwanted information which can deteriorates image quality [9, 10]. It can be produced by the sensor and circuitry of a scanner or digital camera. Image noise can also originate in film grain and in the unavoidable shot noise of an ideal photon detector [11]. Image noise is an undesirable by product of image capture that adds spurious and extraneous information [7, 9].

Fuzzy Rough Set has produced good results when used as a feature selection in the COREL image database [12-15] and other databases [16], as compared to other feature selection methods. In this paper, we want to evaluate the performance of Fuzzy Rough Set feature selection for noisy images. We want to know the semantic rules which were extracted from Fuzzy Rough Set can work with noisy images, and which of these rules can recognise noisy images and allocate them to their related semantic groups.

In this paper, a COREL image dataset is used to obtain the experimental results. This image dataset does not have any noise, so we added noise to the queried image to compare the performance of the Fuzzy Rough Set feature selection with other feature selection methods in a noisy environment. The main purpose of doing this is to evaluate the effect of noise on the feature selection techniques. Three types of noise that are the Gaussian noise, Poisson noise, and salt and pepper noise are used. Using the Matlab software from Mathworks, the three noises were added to the queried images.

In this paper, one of the feature selection methods used to compare with the proposed method is Genetic Algorithm, as Genetic Algorithm is one of the soft computing methods that has demonstrated effective feature selection capability [17, 18]. In addition, Information Gain, OneR and Principle Component Analysis are well-known feature selection methods, and many researchers used these methods for their feature selection tasks. Therefore, it is essential to compare the proposed method with them.

The paper is organised as follows: in section 2, the different kinds of noise are presented. In section 3, the stages of our proposed work are shown. Section 4 and 5 present the experimental results and conclusion respectively.

2. IMAGE NOISE

For a better understanding of the difference between these three kinds of noise, a brief discussion of each type of noise is provided below. Figure 1 shows the three types of noises added to the original image (a), with (b), (c) and (d) showing the different effect each type of noise can produce.

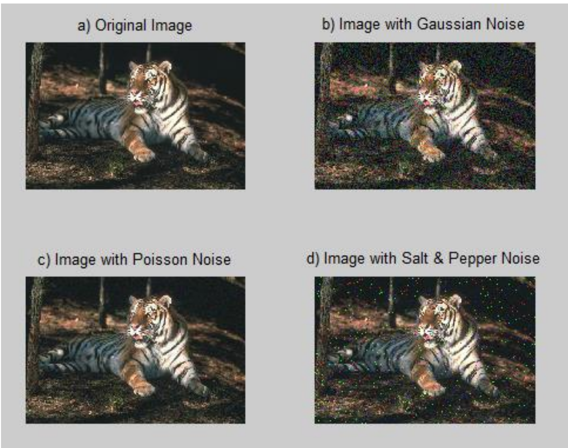

Figure 1. Original image (a) and images with Gaussian, Poisson and Salt & Pepper noises added (b, c and d)

2.1. Gaussian Noise

Gaussian noise represents statistical noise having the probability density function (PDF) equals to that of the normal distribution, which is also known as the Gaussian distribution [19]. In other words, the values that the noise can take on are Gaussian-distributed. In this paper, Gaussian white noise of mean m and variance v was added to the queried images. The Mean noise and variance are 0 and 0.01 respectively for an image with Gaussian Noise (see Figure 1 in (b)).

2.2. Poisson Noise

The Poisson noise, also known as Photon noise, is a basic form of uncertainty associated with the measurement of light, inherent to the quantized nature of light and the independence of photon detections [20]. Its expected magnitude is signal dependent and constitutes the dominant source of image noise except in low light conditions. Matlab syntax generates the Poisson noise from the data instead of adding artificial noise to the data. Figure 1 in (c) shows an image with Poisson noise.

2.3. Salt and Pepper Noise

Impulsive noise is sometimes called the Salt and Pepper noise or Spike noise [21]. An image containing Salt and Pepper noise will have dark pixels in its bright regions and bright pixels in ITS dark regions [21, 22]. Section (d) in Figure 1, represents an image with Salt and pepper noise. The noise density in this image is 0.02.

3. THE PROPOSED FRAMEWORK

In this section, the fuzzy rough set feature selection method used in this paper is briefly described. This session will also present how the proposed feature selection method fit into the stages of our previous proposed CBIR system.

3.1. Fuzzy Rough Set

For the purpose of the research, the algorithm used in [23] was selected and as shown in Figure 2.

C, the set of all conditional features;
D, the set of decision features.
(1) $R \leftarrow \{\}$; $\gamma'_{best} = 0$; $\gamma'_{prev} = 0$
(2) Do
(3) $T \leftarrow R$
(4) $\gamma'_{prev} = \gamma'_{best}$
(5) $\forall x \in (C - R)$
(6) $IF\ \gamma'_{R\cup\{x\}}(D) > \gamma'_{T}(D)$
(7) $T \leftarrow R \cup \{x\}$
(8) $\gamma'_{best} = \gamma'_{T}(D)$
(9) $R \leftarrow T$
(10) $until\ \gamma'_{best} == \gamma'_{prev}$
(11) $return\ R$

Figure 2. The Fuzzy Rough Set Feature Selection Algorithm

This algorithm employs the dependency function γ', to choose which features are added to the current reduced candidate. The dependency function is defined as follows:

$$\gamma'_P(Q) = \frac{\sum_{x\epsilon U} \mu_{POS_P(Q)}(x)}{|U|}$$

The function is determined by the fuzzy cardinality of $\mu_{POS_P(Q)}(x)$ divided by the total number of objects in the universe. The membership of an object $\in U$, belonging to the fuzzy positive region can be defined by:

$$\mu_{POS_P(Q)}(x) = sup_{X\in U/Q}\ \mu_{P\_X(x)}$$

Object x does not belong to the positive region only if the equivalence class it belongs to is not a constituent of the positive region.

Fuzzy lower and upper approximations are defined as [24]:

$$\mu_{P\_X(x)} = sup_{F\epsilon U/P}\, min(\mu_F(x), inf_{y\epsilon U}\, max\{1 - \mu_F(y), \mu_X(y)\}$$

$$\mu_{P^-X(x)} = \sup_{F\epsilon U/P} \min(\mu_F(x), \sup_{y\epsilon U} \min\{\mu_F(y), \mu_X(y)\}$$

During the implementation, not all $y\epsilon U$ need to be considered. Only those where $\mu_F(y)$ is non-zero, i.e. where object y is a fuzzy member of (fuzzy) equivalence class F. $< P\_X, P^-X >$ is called a fuzzy rough set [25].

The algorithm terminated when the addition of any remaining feature does not increase the dependency.

3.2. Stages of The Content Based Image Retrieval System

Figure 3 shows the stages of the proposed image retrieval system used in this research. In Figure 3, the system has training and testing phase. Firstly, in the training phase, the shape, colour and texture features of the image database are extracted. The important features are then selected by using a Fuzzy Rough Set method. Semantic rules are then generated with these features. After that, the SVM classifier is built using these semantic rules.

Still referring to Figure 3, in the testing phase, user feeds the noisy queried image to the system. The system extracted the noisy queried image features and gave these features to the SVM classifier which is built into the training phase. This classifier will extract the relevant images based on the noisy queried image provided.

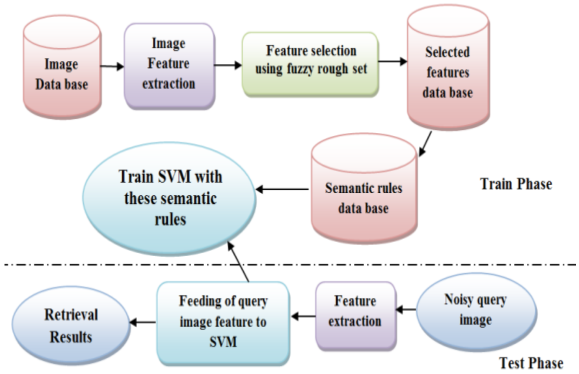

Figure 3. Stages of Proposed Image Retrieval System

4. EXPERIMENTAL RESULTS

In this section, the results that compare the four feature selection methods with the proposed retrieval system by using the generated noisy images modified based on the three types of noise. To investigate the function of the image retrieval system based on the above mentioned methods, we use the COREL image database containing one thousand images. In this database, images are classified into ten semantic groups. The groups are Africans, beach, bus, flower, mountains, elephant, horse, food, dinosaur, and building.

4.1. Precision-Recall Graph

Recall equals to the number of the related retrieval images to the number of the related images available in the image database. The precision equals to the number of the related retrieval images to all the retrieval images [1, 26]. Figure 4, 5 and 6 show the precision-recall graphs for ten semantic groups with Gaussian noise, Poisson noise, and Salt and Pepper noise respectively, that is used for measuring the efficiency of the proposed retrieval system.

From the graphs, we observe that the proposed retrieval system achieved better results than the other four systems in all three kinds of noise. The reason for this is better feature extraction algorithm has been applied in the training phase to save appropriate and eliminate useless image features (see Figure 3). With these useful features, the system can train the SVM classifier with more accuracy and semantic rules.

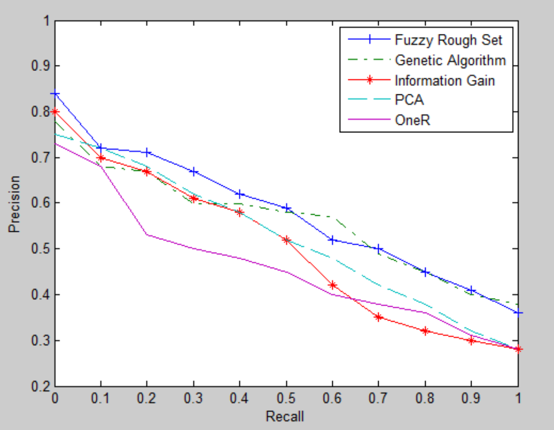

Figure 4. Precision-Recall Graph with Gaussian Noise

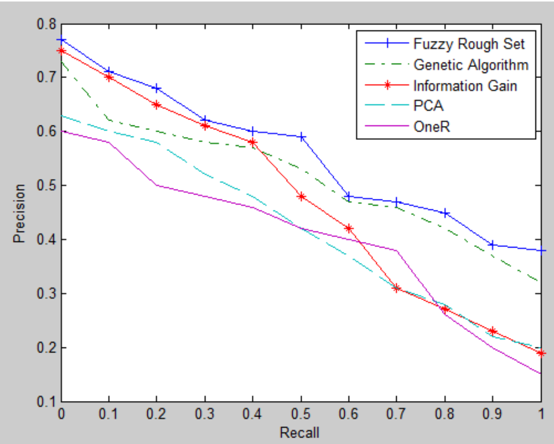

Figure 5. Precision-Recall graph with Poisson Noise

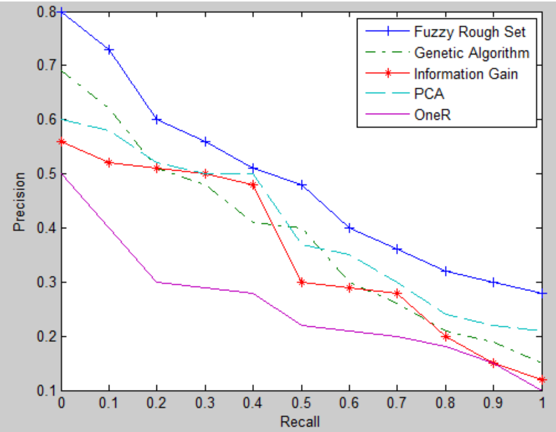

Figure 6. Precision-Recall graph with Salt and Pepper Noise

The reasons behind the superiority of our proposed feature selection method are:

1) Rough Set theory is a useful method for describing and modelling vagueness in ill-defined environments. However, Rough Set cannot work in a continuous environment such as image features, so these features should be discretised first. Discretization may influence the retrieval results. Because of that, the use of Fuzzy Rough Set has many advantages. Firstly, it can work with continuous data. Also, the use of the membership function of a Fuzzy Set has many advantages in the definition, analysis, and operation of fuzzy concepts. By combining these methods (Rough Set and Fuzzy Set), we can use the advantages of both in our image retrieval system.

2) The rules extracted from Fuzzy Rough Set feature selection are semantic rules, which used for training the Support Vector Machine classifier.

3) One of the features of support vector machine is that it can perform well with noisier data.

4.2. The Investigation of The Retrieval Accuracy

To investigate the total accuracy of the above mentioned retrieval systems, 60 noisy images are fed into the system as queried images. That means 60 query images with Gaussian noise (Mean=0 and Variance= 0.01), 60 query images with Gaussian noise (Mean=0 and Variance= 0.02) and etc. ND for Salt and Pepper noise is referring to the Noise Density. Three different noise densities are used in the Salt and Pepper noise in the experimental results. The average of the retrieval accuracy is calculated for each system with three types of noise. Table 1 shows the results. As expected, the results in most cases are better using our proposed feature selection method.

The results extracted from Table 1 are as follows:

- The image retrieval system which used Fuzzy Rough Set for feature selection in their methodology had better results compared to the other retrieval systems which used other feature selection methods in their methodology.
- Overall, most of the feature selection methods had better results with the Salt and Pepper noise.
- When the mean and the variance of the Gaussian noise were increased, the retrieval accuracy of all retrieval systems decreased because the mean and the variance highly influence the query image features. However the Fuzzy Rough Set achieved better results compare to other methods in this situation.
- The Genetic Algorithm had the worst result with Poisson noise compared to other types of noises.

4.3. The Image Comparison of the Retrieval Systems

In the last test, we show that the retrieval results for the queried Flower image (Figure 7, 8, and 9). The first, second, and up to the fifth row in Figure 10, 11 and 12 is related to Fuzzy Rough Set, Genetic Algorithm, PCA, Information Gain, and OneR respectively. Figure 7 is a queried image with Gaussian noise applied to it (mean=0 and variance 0.01), Figure 8 is a queried image with Salt and Pepper noise applied to it (noise density 0.02) and Figure 9 is a queried image with Poisson noise applied to it. Referring to Figure 10, 11 and 12, the retrieval system with the Fuzzy Rough Set method shows more related output images to the user. The first left image in Figure 10, 11 and 12 matched closely to the queried image.

Figure 7. Query Image with Gaussian Noise (M=0, V=0.01)

Figure 8. Query Image with Salt and Pepper Noise (ND=0.02)

Figure 9. Query Image with Poisson Noise

Table 1. Accuracy of Retrieval with Three Kinds of Noise

Feature Selection Methods	Gaussian Noise			Salt & Pepper Noise			Poisson Noise
	M=0,V=0.01	M=0,V=0.02	M=0.01,V=0.02	ND=0.01	ND=0.02	ND=0.03	
Fuzzy Rough	%92.6	%91.1	%88.32	%93.2	%91.56	%82.1	%92.3
Genetic Algorithm	%90.5	%91.7	%81.38	%87.4	%82.9	%79.4	%78.4
PCA	%91.2	%81.4	%70.6	%88.5	%82	%80.3	%75.3
Information Gain	%91.67	%80.7	%67.3	%93.4	%81.3	%78.2	%75.7
One R	%78.2	%70.3	%68.74	%85.1	%82	%76.3	%71.6

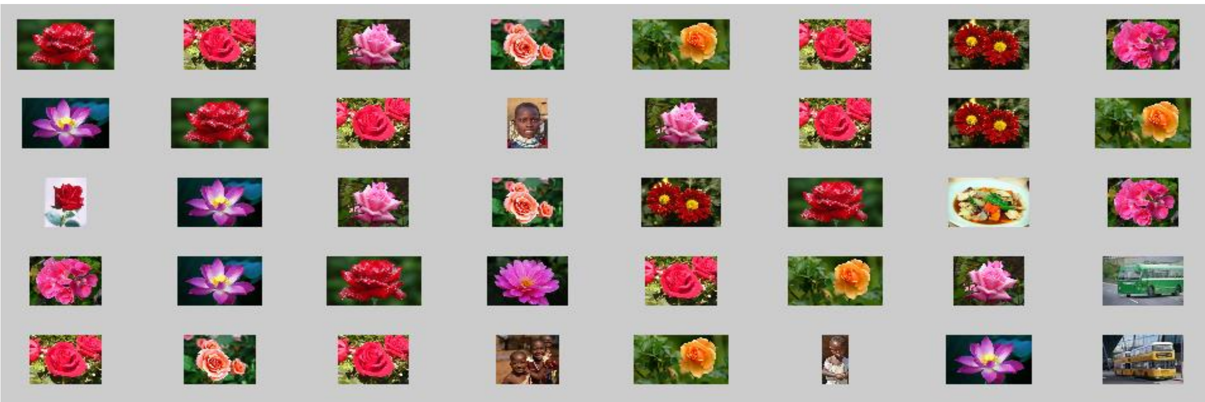

Figure 10. Retrieved Images According to: First Raw- Fuzzy Rough Set, Second Raw- Genetic Algorithm, Third Raw- PCA, Fourth Raw- Information Gain, Fifth Raw- OneR

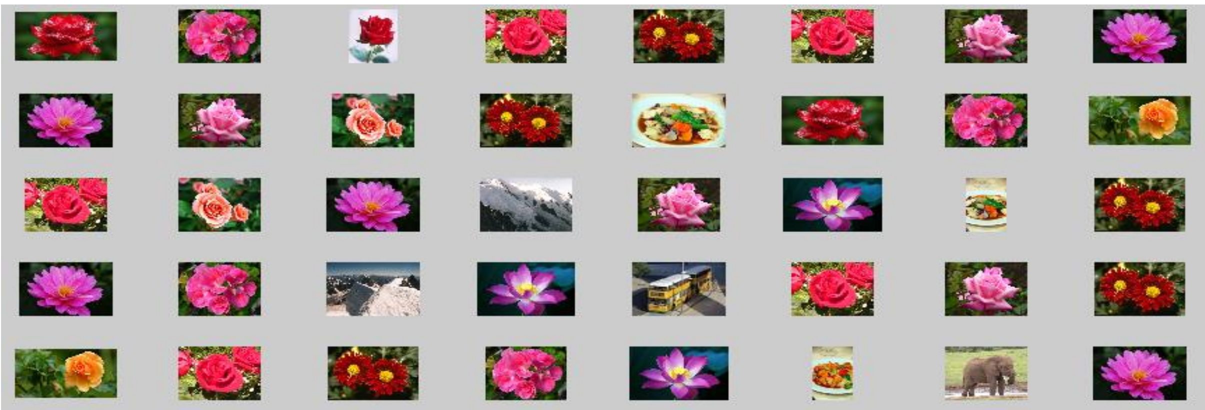

Figure 11. Retrieved Images According to: First Raw- Fuzzy Rough Set, Second Raw- Genetic Algorithm, Third Raw- PCA, Fourth Raw- Information Gain, Fifth Raw- OneR

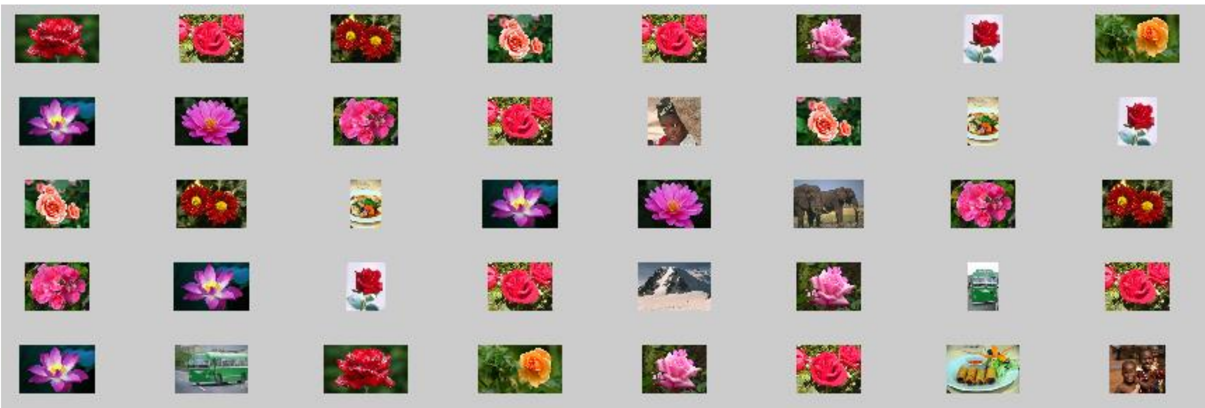

Figure 12. Retrieved Images According to: First Raw- Fuzzy Rough Set, Second Raw- Genetic Algorithm, Third Raw- PCA, Fourth Raw- Information Gain, Fifth Raw- OneR

5. CONCLUSION

In this paper, Fuzzy Rough Set feature selection is evaluated in a noisy environment. Gaussian noise, Poisson noise and Salt and Pepper noise were used to estimate the Fuzzy Rough Set feature selection accuracy in a Content Based Image Retrieval system. In the experimental results, Fuzzy Rough Set feature selection was compared with four other feature selection methods. These four feature selection methods are Genetic Algorithm, Information Gain, OneR and Principle Component Analysis. From the experimental results with a noisy queried image, it can be observed that Content Based Image Retrieval system using Fuzzy Rough Set feature selection has better retrieval accuracy and Precision Recall graph, when compared to the other four retrieval systems. The drawbacks of these four feature selection methods described in this paper are as follows: (1) In PCA, the computation of the eigenvectors might be infeasible for very high dimensional data, (2) The OneR algorithm is topologically unstable, (3) The Genetic Algorithm cannot find the best features and stuck in a local maximum hence the best features are not guaranteed. Furthermore, it increases the computational complexity and (4) Information Gain has a problem when it is applied to features that can take on a large number of distinct values. Based on these drawbacks, the four retrieval systems cannot achieve better results than our proposed feature selection method.

By utilising Fuzzy Rough Set feature selection method, the proposed system has the advantage that it deals efficiently with an image feature environment that is noisy, vague and uncertain. In addition, rules extracted from the selecting features of the Fuzzy Rough Set feature selection are semantic and can train the classifier properly. An important advantage of this work is training the SVM with semantic rules that can separate the relevant images from irrelevant ones more accurately.

6. REFERENCES

[1] T. Dharani and I. L. Aroquiaraj, "A survey on content based image retrieval," in *2013 International Conference on Pattern Recognition, Informatics and Medical Engineering (PRIME)*, 2013, pp. 485-490.

[2] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys,* vol. 40, pp. 1-60, 2008.

[3] Z. Liang, Y. Zhuang, Y. Yang, and J. Xiao, "Retrieval-based cartoon gesture recognition and applications via semi-supervised heterogeneous classifiers learning," *Pattern Recognition,* vol. 46, pp. 412-423, 2013.

[4] X.-Y. Wang, B.-B. Zhang, and H.-Y. Yang, "Active SVM-based relevance feedback using multiple classifiers ensemble and features reweighting," *Engineering Applications of Artificial Intelligence,* vol. 26, pp. 368-381, 2013.

[5] S. Mukhopadhyay, J. K. Dash, and R. Das Gupta, "Content-based texture image retrieval using fuzzy class membership," *Pattern Recognition Letters,* vol. 34, pp. 646-654, 2013.

[6] N. Cherin, F. Cordier, and M. Melkemi, "Texture Mapping of Images with Arbitrary Contours," *Journal of WSCG,* vol. 21, pp. 193-204, 2013.

[7] L. Chen, X. Huang, J. Tian, and X. Fu, "Blind noisy image quality evaluation using a deformable ant colony algorithm," *Optics & Laser Technology,* vol. 57, pp. 265-270, 2014.

[8] L. G. Jaime, E. E. Kerre, M. Nachtegael, and H. Bustince, "Consensus image method for unknown noise removal," *Knowledge-Based Systems,* 2013.

[9] P. Frosini and C. Landi, "Corrigendum to "Persistent Betti numbers for a noise tolerant shape-based approach to image retrieval"," *Pattern Recognition Letters,* vol. 34, pp. 1320-1321, 2013.

[10] P. Walek, J. Jan, P. Ourednicek, J. Skotakova, and I. Jira, "Preprocessing for Quantitative Statistical Noise Analysis of MDCT Brain Images Reconstructed Using Hybrid Iterative (iDose) Algorithm," *Journal of WSCG,* vol. 20, pp. 73-80, 2012.

[11] P. Frosini and C. Landi, "Persistent Betti numbers for a noise tolerant shape-based approach to image retrieval," *Pattern Recognition Letters,* vol. 34, pp. 863-872, 2013.

[12] M. Shahabi Lotfabadi, M. F. Shiratuddin, and K. W. Wong, "Using fuzzy rough feature selection for image retrieval system," in *2013 IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing (CIMSIVP)*, 2013, pp. 42-48.

[13] M. Shahabi Lotfabadi, M. F. Shiratuddin, and K. W. Wong, "Feature Decreasing Methods Using Fuzzy Rough Set based on Mutual Information," presented at the Proceeding of the 8th IEEE Conference on Industrial Electronics and Applications (ICIEA 2013), 2013.

[14] M. Shahabi Lotfabadi, M. F. Shiratuddin, and K. W. Wong, "Using Rough Set Theory to Improve Content Based Image Retrieval System," presented at the IEEE. Eleventh Postgraduate Electrical Engineering and Computing Symposium (PEECS2012), 2012.

[15] M. Shahabi Lotfabadi, M. F. Shiratuddin, and K. W. Wong, "Utilising Fuzzy Rough Set based on Mutual Information Decreasing Method for Feature Reduction in an Image Retrieval System," in *Proceedings of the International Joint Conferences on Computer, Information, and Systems Sciences, and Engineering (CISSE 2012)*, 2012.

[16] M. Shahabi Lotfabadi and A. M. Eftekhari Moghadam, "The Comparison of Different Feature Decreasing Methods Base on Rough Sets and Principal Component Analysis for Extraction of Valuable Features and Data Classifying Accuracy Increasing," in *2010 First International Conference on Integrated Intelligent Computing (ICIIC 2010)*, Bangalore, India 2010, pp. 108-113.

[17] C.-F. Tsai, W. Eberle, and C.-Y. Chu, "Genetic algorithms in feature and instance selection," *Knowledge-Based Systems,* vol. 39, pp. 240-247, 2013.

[18] S. Li, H. Wu, D. Wan, and J. Zhu, "An effective feature selection method for hyperspectral image classification based on genetic algorithm and support vector machine," *Knowledge-Based Systems,* vol. 24, pp. 40-48, 2011.

[19] L. Liu, L. Zeng, K. Shen, and X. Luan, "Exploiting local intensity information in Chan–Vese model for noisy image segmentation," *Signal Processing,* vol. 93, pp. 2709-2721, 2013.

[20] M. Setayesh, M. Zhang, and M. Johnston, "A novel particle swarm optimisation approach to detecting continuous, thin and

smooth edges in noisy images," *Information Sciences*, vol. 246, pp. 28-51, 2013.

[21] Y. Li, J. Sun, and H. Luo, "A neuro-fuzzy network based impulse noise filtering for gray scale images," *Neurocomputing*, vol. 127, pp. 190-199, 2014.

[22] C.-H. Son, H. Choo, and H.-M. Park, "Image-pair-based deblurring with spatially varying norms and noisy image updating," *Journal of Visual Communication and Image Representation*, vol. 24, pp. 1303-1315, 2013.

[23] R. Jensen and Q. Shen, "Fuzzy-rough sets for descriptive *dimensionality* reduction," *Fuzzy Systems, Proceedings of the 2001 IEEE International Conference* pp. 29-34, 2002.

[24] Q. Hu, L. Zhang, S. An, D. Zhang, and D. Yu, "On Robust Fuzzy Rough Set Models," *IEEE Transactions on Fuzzy Systems*, vol. 20, pp. 636-651, 2012.

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

[26] D. Cerra and M. Datcu, "A fast compression-based similarity measure with applications to content-based image retrieval," *Journal of Visual Communication and Image Representation*, vol. 23, pp. 293-302, 2012.