# Using Fuzzy Rough Feature Selection for Image Retrieval System

Maryam Shahabi Lotfabadi, Mohd Fairuz Shiratuddin & Kok Wai Wong
School of Information Technology, Murdoch University
90 South Street, Murdoch Western Australia 6150
m.shahabi@murdoch.edu.au, f.shiratuddin@murdoch.edu.au & k.wong@murdoch.edu.au

**Abstract:**
Feature selection is an important step in processing the images especially for applications such as content based image retrieval. In large multimedia databases, it may not be practical to search through the entire database in order to retrieve similar images from a query. Good data structures for similarity search and indexing are needed, and the existing data structures do not scale well for the high dimensional multimedia descriptors. Thus feature selection is an important step. Fuzzy rough feature selection method has many advantages in determining the relevant features. In this paper, five feature selection methods are compared with the fuzzy rough method. These five feature selection methods are Relief-F, Information Gain, Gain Ratio, OneR and the statistical measure $\chi^2$. The main purpose of the comparison is to rank the image features and see which method provides better results. An image retrieval dataset (COREL dataset) was used in the comparison. In order to evaluate the performance of the six methods, ranking of the important features is defined. This is then used to compare with the automated ranking produced by the aforesaid feature selection methods. Results show that the retrieval system using fuzzy rough feature selection has better retrieval accuracy and provide good Precision Recall performance. The advantages of the use of fuzzy rough feature selection will also be discussed in the paper.

**Key words:** fuzzy rough set, feature selection, image retrieval system.

## 1- Introduction

The advancement in computing has produced innumerable digital images and photographs. This exponential growth has created a high demand for efficient tools for image searching, browsing and retrieval for use in various domains such as architecture, crime prevention, fashion, medicine, remote sensing, publishing, etc. This issue of large database has been addressed by an integrated framework called Content Based Image Retrieval (CBIR). Content-based image retrieval is one of the important topics in machine vision and research started as early as the 90s [1].

In general, the dimensionality of image feature vectors used in image retrieval applications is quite high. Typical feature vector dimensions can range from few tens to several hundreds. For example, a colour histogram may contain 256 bins. This high dimensionality of the feature vectors creates problems in constructing efficient data structures for search and retrieval [2]. It is well known that most of the indexing structures do not scale well when the dimensionality of the feature vector exceeds 20. For this reason, there is considerable interest in selecting important features [3].

Feature selection refers to the problem of selecting input features that are more predictive of a given outcome. Feature selection is used in many areas such as image processing, machine learning, pattern recognition and signal processing [4]. Unlike other dimensionality reduction methods, feature selectors try to preserve the original meaning of the features after reduction. Beside applying to large datasets, feature selection methods have also been applied to small and medium-sized datasets in order to locate more informative features for later use [5].

Some researchers did different feature selection methods, which include SVM, based schemes, such as [6] and [7]. In [6], Image features can be extracted using a difference of Gaussian filter followed by the Radon transform. The relevance and importance of these features are determined in a scaling support vector machine classifier, where zero weights are assigned to irrelevant variables. In [7], diffusion distance is computed over a pair of human face images, the shape descriptions of these images are built using Gabor filters that consist of a number of scales and levels.

The use of user-supplied information is essential to many existing methods for feature selection, and this approach has a significant drawback [8]. Some feature selectors require noise levels to be specified by the user beforehand, and some simply rank features leaving the user to choose their own subset. There are those that require the user to state how many features are to be chosen, or they must supply a threshold that determines when the algorithm should terminate. All of these require the user to make a decision based on their own judgement [8]. In addition, some features selection methods can only operate effectively with datasets containing discrete values and have difficulty handling noisy data. As most datasets contain real-valued features, it is necessary to perform a discretization step beforehand. In fuzzy rough feature selection method, this is typically implemented by standard fuzzification methods, enabling linguistic labels to be associated with the attributes values [9]. It also aids uncertainty modelling by allowing the possibility of the membership of a value to be assigned to more than one fuzzy label. However, membership degrees of feature values in the fuzzy sets are not exploited in the process of dimensionality reduction. By using fuzzy rough sets, it is possible to use the membership information to better guide feature selection.

The use of fuzzy rough set theory in feature selection is one approach that has been explored in the last decades [9]. Fuzzy rough feature selection can provide promising results mainly due to the following: (1) only the facts hidden in data are analysed, (2) no additional information about the data is required such as thresholds or expert knowledge on a particular domain, (3) it finds a minimal knowledge representation, (4) fuzzy rough feature selection can deal with continuous data set, (5) it has good results in noisy datasets, and (6) rules extract from fuzzy rough set are semantic.

In order to provide an insight to the use of fuzzy rough feature selection in image retrieval, this paper provides a comparison study with other feature selection methods. The feature selections methods selected in this study are from *entropy*, *statistical*, *decision tree* and *nearest neighbourhood* based feature selection methods. Information gain and gain ratio are entropy based feature selection methods [10, 11]. $\chi^2$ and OneR are based on statistical [12] and decision tree [11] respectively. In addition, Relief-F is based on nearest neighbour [13]. These five features selection methods have been used as a major part of the proposed feature selection techniques in image retrieval systems [3, 14]. The main purpose of the comparison is to rank the image features and see which method provides better results.

This paper is organised as follows: six different feature selection methods are describe in section 2. Important image features and experiment results are shown in section 3 and 4 respectively. Finally, the conclusion of this study is presented.

## 2- Feature selection methods

In this section, the six feature selection methods are briefly described. They are fuzzy rough, Relief-F, Information Gain, Gain Ratio, OneR and the statistical measure $\chi^2$.

### 2-1- Fuzzy Rough

Fuzzy rough set has been used in image retrieval system [15]. In this paper, the focus is on using fuzzy rough feature selection. For the purpose of the study, the algorithm used in [16] was selected and as shown in Figure 1.

$C$, 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 1- The fuzzy rough feature selection algorithm.

This algorithm employs the dependency function $\gamma'$, to choose which features are added to the current reduced candidate. 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 [17]:

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

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

During 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 [18].

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

### 2-2- Information Gain

The Information Gain (IG) is the expected reduction in entropy resulting from partitioning the dataset objects according to a particular feature [10]. The entropy of a labelled collection S of c objects is defined as:

$$Entropy\ (S) = \sum_{i=1}^{c} -p_i\, log_2 p_i$$

Where $p_i$ is the proportion of $S$ belonging to class $i$. Based on this, the Information Gain metric is:

$$IG(S,A) = Entropy\ (S) - \sum_{v \epsilon values(A)} \frac{|S_v|}{|S|} Entropy(S_v)$$

where $values(A)$ is the set of values for feature $A$, $S$ the set of training examples, $S_v$ the set of training objects where $A$ has the value $v$. This metric is the one used in ID3 (decision tree) for selecting the best feature to partition the data.

### 2-3- Gain Ratio

One limitation of the IG measure is that it favours features with many values. The Gain Ratio (GR) seeks to avoid this bias by incorporating another term, split information, that is sensitive to how broadly and uniformly the attribute splits the considered data [11]:

$$Split(S,A) = -\sum_{i=1}^{c} \frac{|S_i|}{|S|}\, log_2 \frac{|S_i|}{|S|}$$

Where each $S_i$ is a subset of objects generated by partitioning $S$ with the $c$-valued attribute $A$. The Gain Ratio is then defined as follows:

$$GR(S,A) = \frac{IG(S,A)}{Split(S,A)}$$

### 2-4- $\chi^2$ Measure

In this method, features are individually evaluated according to their $\chi^2$ statistic with respect to the classes [12]. For a numeric attribute, the method first requires its range to be discretised into several intervals. The $\chi^2$ value of an attribute is defined as:

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{k} \frac{(A_{ij} - E_{ij})^2}{E_{ij}}$$

where $m$ is the number of intervals, $k$ the number of classes, $A_{ij}$ the number of samples in the $i$th interval, $j$th class, and $E_{ij}$ the expected frequency of $A_ij$ $(E_{ij} = R_i * C_j/N)$ ; where $R_i$ is the number of objects in the $i$th interval, $C_j$ the number of objects in the $j$th class, $N$ the total number of objects, The larger the $\chi^2$, the more important the feature.

### 2-5- Relief-F

This is the Relief-F measure based on the original Relief measure [13]. Relief evaluates the worth of an attribute by repeatedly sampling an instance and considering the value of the given attribute for the nearest instance of the same and different class. Relief-F extends this idea to dealing with multi-class problems as well as handling noisy and incomplete data.

### 2-6- OneR

The OneR classifier learns a one-level decision tree i.e. it generates a set of rules that test one particular attribute [11]. One branch is assigned for every value of a feature and each branch is assigned to the most frequent class. The error rate is then defined as the proportion of instances that do not belong to the majority class of their corresponding branch. Features with the higher classification rates are considered to be more significant than those resulting in lower values.

## 3- Important features in image retrieval

In this section, the image features that are important in image retrieval application are defined and ranked. These features are collected and reviewed from the literature [19, 20]. However, in order to be consistent with the experiment used in this paper, only literature using Corel dataset with 1000 images in 10 semantic groups will be examined The features used in the analysis with their corresponding ordering of importance are shown in Table 1. From the literature, the most influential features are Mean Hue, Coarseness, Wavelet Moment and Directionality. Contrast and Mean intensity are the next most influential features. Although an ordering is given to Edge and Roughness, it is difficult to differentiate between them. The features identified in Table 1 are only valid for the Corel dataset and may not apply to all image datasets.

| Feature number | Feature name | Defined ordering |
|---|---|---|
| 1 | Contrast | 2 |
| 2 | Mean intensity | 2 |
| 3 | Roughness | 5 |
| 4 | Deviation intensity | 6 |
| 5 | Roundness | 6 |
| 6 | Convexity | 7 |
| 7 | Bulkiness | 6 |
| 8 | Mean hue | 1 |
| 9 | Coarseness | 1 |

| | | |
|---|---|---|
| 10 | Wavelet moments | 1 |
| 11 | Directionality | 1 |
| 12 | Standard deviation | 1 |
| 13 | Entropy | 3 |
| 14 | Euler number | 3 |
| 15 | Edge | 4 |
| 16 | Structure factor | 6 |
| 17 | Rectangularity | 7 |
| 18 | Sigma | 7 |

Table1- Important features of image

## 4- Experiment results

The results of the comparison study using feature selection methods discussed in section 2 can be seen in Table 2. All methods rate features 8 (Mean hue) and 9 (Coarseness) highly. This is in agreement with the defined ranking as shown in Table 1. Only the fuzzy rough (FR) method correctly rates features 10 (Wavelet moments), 11 (Directionality) and 12 (Standard deviation) highly. After these features, FR ranks Contrast and Mean intensity next. In fact, FR is the only method to detect the importance of these two features. The results show that the FR method is useful in producing results in line with the defined ranking. The reason is fuzzy rough feature selection use dependency function to select the important features. This function uses positive region that can deal with vague area and recognise more important features.

| **Feature** | **Defined** | **FR** | **Re** | **IG** | **GR** | **1R** | $\chi^2$ |
|---|---|---|---|---|---|---|---|
| 8 | 1 | 0.214 | 0.142 | 0.147 | 0.163 | 83.4 | 14.3 |
| 9 | 1 | 0.185 | 0.153 | 0.204 | 0.183 | 82.7 | 11.2 |
| 10 | 1 | 0.109 | 0.074 | 0.421 | 0.401 | 68.2 | 0.0 |
| 11 | 1 | 0.143 | 0.084 | 0.0 | 0.0 | 78.3 | 0.0 |
| 12 | 1 | 0.102 | 0.061 | 0.0 | 0.0 | 70.1 | 0.0 |
| 1 | 2 | 0.096 | 0.023 | 0.0 | 0.0 | 74.5 | 0.0 |
| 2 | 2 | 0.062 | 0.013 | 0.0 | 0.0 | 70.3 | 0.0 |
| 13 | 3 | 0.0 | 0.061 | 0.0 | 0.0 | 78.3 | 0.0 |
| 14 | 3 | 0.0 | 0.043 | 0.0 | 0.0 | 71.3 | 0.0 |
| 15 | 4 | 0.043 | 0.020 | 0.0 | 0.0 | 78.3 | 0.0 |
| 3 | 5 | 0.0 | 0.004 | 0.0 | 0.0 | 78.3 | 0.0 |
| 4 | 6 | 0.025 | 0.086 | 0.0 | 0.0 | 74.5 | 0.0 |
| 5 | 6 | 0.0 | 0.008 | 0.0 | 0.0 | 78.3 | 0.0 |
| 7 | 6 | 0.023 | 0.009 | 0.0 | 0.0 | 72.6 | 0.0 |
| 16 | 6 | 0.0 | 0.007 | 0.0 | 0.0 | 78.3 | 0.0 |
| 6 | 7 | 0.0 | 0.003 | 0.0 | 0.0 | 72.6 | 0.0 |
| 17 | 7 | 0.0 | 0.090 | 0.005 | 0.0 | 71.1 | 0.0 |
| 18 | 7 | 0.007 | 0.005 | 0.0 | 0.001 | 78.3 | 0.20 |

Table2- Feature ranker results for the Corel dataset.

After the feature selection methods have determined the order of the features and ranked them, it is important to verify them in an image retrieval system. The image retrieval system used in this experiment is used to deal with 10 semantic groups such as Africans, beach, bus, flower, mountains, elephant, horse, food, dinosaur, and building. 100 images from each semantic group are selected. Using precision-recall [21] and retrieval accuracy graphs [3], the retrieval accuracies of the six feature selection methods are compared. The precision-recall graph and retrieval accuracy are shown in Figure 2 and 3 respectively.

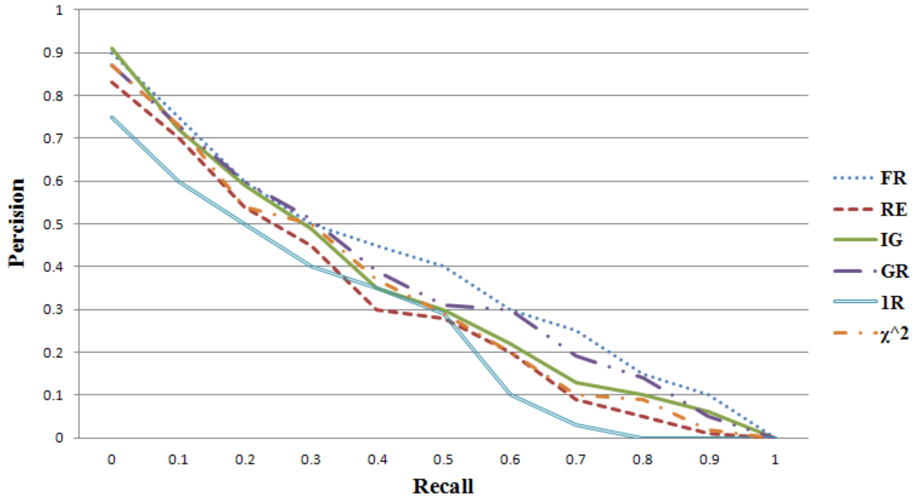


Figure 2- Precision recall graph.

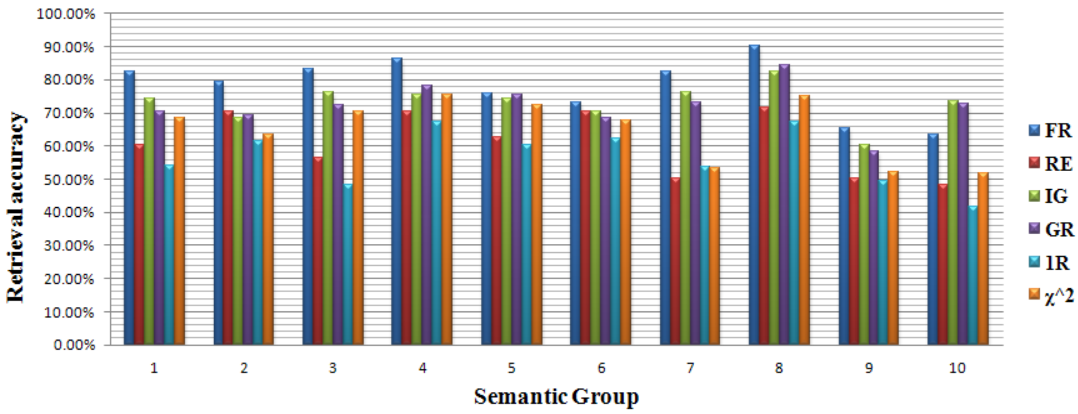


Figure 3- Retrieval accuracy graph.

From the results shown in Figures 2 and 3, fuzzy rough feature selection method has better results than the other feature selection methods. Information gain and gain ratio are in second and third position. This is interesting as both of these methods are entropy based. A feature subset with a maximum (crisp) rough set dependency has corresponding entropy of 0. Unlike these methods, the fuzzy rough method may also be applied to datasets containing real valued (instead of discrete) decision features.

Mean retrieval accuracy is 78.17% for fuzzy rough, 73.15% for Information Gain and 72.27% for Gain Ratio. The accuracies are then decreased to 65.02%, 61% and 56.57% for $\chi^2$ , Relief- F and OneR respectively.

Figure 4 show the queried bus image. The first, second, and up to the sixth row in Figure 5 is related to fuzzy rough, Information Gain, Gain Ratio, the statistical measure $\chi^2$, Relief-F and OneR respectively. Referring to Figure 5, the retrieval system with the fuzzy rough feature selection has more related output images to the user. The first left image in Figure 5 matched closely to the queried image.

Figure 4- Queried image

The reasons why fuzzy rough has better results are as follows: (1) it does not require preliminary or additional parameters to describe data; (2) works with missing values, switches among different reducts, and uses little time to generate rules; (3) can handle large amounts of quantitative and qualitative data; (4) yields easily understood decision rules supported by a set of real examples; (5) models highly nonlinear or discontinuous functional relationships and is a powerful method for characterizing complex and multidimensional patterns; and (6) discovers important facts hidden in data and expresses them in the natural language of decision rules.

## 5- Conclusion

In this paper, five different feature selection methods, Relief-F, Information Gain, Gain Ratio, OneR and the statistical measure $\chi^2$ are compared with the fuzzy rough set method. The aim of this paper is to show which feature selection method can select important features accurately and thus provides better retrieval accuracy. From the experiment results, fuzzy rough feature selection method can rank most of the image features according to the defined ranking. After that, these six methods are used as feature selection method in an image retrieval system, which is used to test the retrieval accuracy. From the results, it again demonstrates that fuzzy rough feature selection method has better retrieval accuracy when compare to the other feature selection methods.

## References


1. Cerra, D. and M. Datcu, *A fast compression-based similarity measure with applications to content-based image retrieval.* Journal of Visual Communication and Image Representation, 2012. 23(2): p. 293-302.
2. Iqbal, K., M.O. Odetayo, and A. James, *Content-based image retrieval approach for biometric security using colour, texture and shape features controlled by fuzzy heuristics.* Journal of Computer and System Sciences, 2012. 78(4): p. 1258-1277.
3. Penatti, O.v.A.B., E. Valle, and R.d.S. Torres, *Comparative study of global color and texture descriptors for web image retrieval.* Journal of Visual Communication and Image Representation, 2012. 23(2): p. 359-380.
4. Chen, Y.-S., *Classifying credit ratings for Asian banks using integrating feature selection and the CPDA-based rough sets approach.* Knowledge-Based Systems, 2012. 26(0): p. 259-270.
5. Wang, J., et al., *Rough set and scatter search metaheuristic based feature selection for credit scoring.* Expert Systems with Applications, 2012. 39(6): p. 6123-6128.
6. Zhou, H., P. Miller, and J. Zhang, *Age classification using Radon transform and entropy based scaling SVM* British Machine Vision Conference, Dundee, UK, 2011.
7. Zhou, H. and A.H. Sadka, *Combining Perceptual Features With Diffusion Distance for Face Recognition* IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews 2011. 41(5): p. 577-588.
8. Foithong, S., O. Pinngern, and B. Attachoo, *Feature subset selection wrapper based on mutual information and rough sets.* Expert Systems with Applications, 2012. 39(1): p. 574-584.
9. Derrac, J.n., et al., *Enhancing evolutionary instance selection algorithms by means of fuzzy rough set based feature selection.* Information Sciences, 2012. 186(1): p. 73-92.
10. Baranidharan, T. and D.K. Ghosh, *Medical Image Classification Using Genetic Optimized Elman Network* American Journal of Applied Sciences 2012. 9(1): p. 123-126
11. Chinpanthana, N., *Integrating Qualitative Features Selection for Semantic Image Classification with Support Vector Machine.* 3rd International Conference on Information and Financial Engineering 2011: p. 117-122.
12. Yossi Rubner, et al., *Empirical Evaluation of Dissimilarity Measures for Color and Texture.* Computer Vision and Image Understanding 2001. 84: p. 25-43.
13. Abdolhossein Sarrafzadeh, et al., *ReliefF based feature selection In Content-based image retrieval.* international of multi conference of Engineers and computer Scientists., 2012. 1.
14. Gavves, E., C.G.M. Snoek, and A.W.M. Smeulders, *Visual synonyms for landmark image retrieval.* Computer Vision and Image Understanding, 2012. 116(2): p. 238-249.
15. Maryam Shahabi Lotfabadi and A.M.E. Moghadam., *A new method for semantic image retrieval based on fuzzy rough set according to mutual information theory.* 15th national CSI conference (CSICC 2010). Tehran.Iran., 2010.
16. Shen, J.R.a.Q., *Fuzzy-rough sets for descriptive dimensionality reduction. Fuzzy Systems,* Proceedings of the 2002 IEEE International Conference, 2002: p. 29- 34.
17. Qinghua Hu, L.Z., , Shuang An, and David Zhang, and Daren Yu, *On Robust Fuzzy Rough Set Models.* IEEE Transactions On Fuzzy Systems, No. 4, 2012. 20: p. 636_651.
18. Yang, H.-L., et al., *Bipolar fuzzy rough set model on two different universes and its application.* Knowledge-Based Systems.
19. Acharya, S. and M.R.V. Devi, *Image retrieval based on visual attention model.* Procedia Engineering, 2012. 30(0): p. 542-545.
20. Bird, C.L., P.J. Elliott, and E. Griffiths, *User interfaces for content-based image retrieval.* Intelligent Image Databases, IEE Colloquium on, 1996: p. 8/1 - 8/4

21. Fanjie, M., G. Baolong, and W. Xianxiang, *Localized Image Retrieval Based on Interest Points.* Procedia Engineering, 2012. 29(0): p. 3371-3375.

Figure5- Retrieved images according to: first row- fuzzy rough, second row- Information Gain, third row- Gain Ratio, fourth row-$\chi^2$, fifth row- Relief-F, sixth row- OneR.