A Fuzzy Logic Based Soft Computing Approach in CBIR System Using Incremental Filtering Feature Selection to Identify Patterns

1 V. Yadaiah and 2 Dr. R. Vivekanandam, Rajaram Jatothu

1,2 Research Scholar, Department of Computer Science and Engineering, Sri Satya Sai University of Technology and Medical Sciences, Sehore, M.P., India.
2 Professor, Director in Muthayammal Eng. College, Rasipuram, Namakkal, T.N., India

Abstract

Content based Image Retrieval (CBIR) may be a set of techniques for retrieving semantically-relevant pictures from an image database supported automatically-derived image options. Generally, in CBIR systems, the visual features are described at low-level. They are simply rigid mathematical measures that cannot influence the inherent subjectivity and fogginess of individual’s understandings and perceptions. As a result, there is a niche between low-level features and high-level semantics. We have a tendency to are witnessing the era of massive information computing where computing the resources is turning into the most bottleneck to handle those massive datasets. With in the case of high dimensional data where every view of information is of high spatiality, feature selection is important for additional rising the clustering and classification results.

In this paper, we have a tendency to propose a new feature selection method is Incremental Filtering Feature Selection (IFFS) algorithm that employs the Fuzzy Rough Set for choosing best subset of features and for effective grouping of huge volumes of data, respectively. We introduce a new system of visual features extraction and matching by using Fuzzy Logic (FL). FL is a powerful tool that deals with reasoning algorithms used to emulate human thinking and decision making in machines. An in depth experimental comparison of the proposed method and other methods are done. The performance of the proposed model yields promising results on the feature selection, and retrieval accuracy in the field of Content based Image Retrieval.

Keywords: Content Based Image Retrieval, Fuzzy Logic, Fuzzy Color, and Incremental Filtering Feature Selection.

INTRODUCTION

Very massive collections of images are growing quickly because of arrival of cheaper storage devices and also the internet. Finding an image from a huge set of images is very challenging task. One solution to this problem is to label images manually. But it is too expensive, time consuming and not feasible for several applications. Moreover, the labeling process depends on the semantic accuracy in describing the image. Therefore, many content based image retrieval systems are developed to extract low levels features for describing the image content [1].

A typical content-based retrieval system (as in Fig.1) is split into 2 stages: off-line feature extraction and on-line image retrieval [2]. In off-line stage, the system mechanically extracts visual attributes of every image in the database based on its pixel values and stores them in a different database inside the system, known as a feature database. In on-line stage, the user will submit a query example to the retrieval system. The system represents this example with a feature vector. The distances (i.e., similarities) between the feature vectors of the query example and those of the media in the feature database are then computed and ranked. The system ranks the search results and then returns the results which are most similar to the query examples.

Image data is vague in nature and in content-based retrieval this property creates some issues like [3]:

1. Descriptions of image contents typically involve inexact and subjective concepts.
2. Typically imprecision and vagueness exist in descriptions of the images and in some of the visual features.
3. User’s needs to image retrieval could also be naturally fuzzy.

Fuzzy Logic (FL) is used in CBIR system because it is the character of of image data, and also the nature of human perception and thinking process. So, it can minimize semantic gap between high level semantic and low level image features. Also, it is robust to the noise and intensity modification in the images. Finally, the users are interested in results according to similarity (closeness) instead of equality (exactness).

In [4], a color histogram representation, called Fuzzy Color Histogram (FCH), is presented by considering the color similarity of each pixel’s color associated to all the histogram bins through fuzzy-set membership function. An approach for computing the membership values based on fuzzy-means algorithm is developed. The proposed FCH is further exploited in the application of image indexing and retrieval. Konstantinidis et al [5] proposed a fuzzy linking system for color histogram creation in L*a*b* color space. It contains 10 bins, and 27 rules used to derive the final histogram. Kucuktunc et al [6] proposed a fuzzy linking system for color histogram creation in L*a*b* color space. Their system contains 15 bins, and 27 rules used to derive the final histogram.
Feature selection for large-scale data sets has been conceived as a significant dimensional reduction technique in machine learning. It aims to improve the accuracy and performance of classifiers by removing redundant features and selecting informative features from the data. In the process of feature selection, feature evaluation criteria are used to evaluate the quality of the candidate subsets. For a feature subset, different evaluation criteria may give different results as there are five kinds of evaluation criteria such as; distance measures, information measures, dependency measures, consistency measures, and classification error rate measures. The first four evaluation criteria are used to evaluate feature subsets according to inherent characteristics of the data. The last one relies on a classification algorithm to evaluate and select useful features and is usually used to improve the classification performance, but it is time-consuming. Hence, there is a need for an efficient algorithm for selecting informative features over large scale datasets, by decomposing large datasets and fusing it.

Clustering is one of the primary tasks used in pattern recognition and data mining communities to search large databases for various applications, and so, clustering algorithms that scale well to big data are important and useful. Cluster analysis is a method of clustering data sets with the most similarity in the same cluster and the greatest dissimilarity between different clusters. Clustering algorithms can be used to uncover unknown relations existing in a set of unlabeled data. In general, clustering methods can be divided into two categories; probability model-based approaches, and non-parametric approaches. Probability model-based approaches assume that the data set follows a mixture of probability distributions so that the expectation-maximization (EM) algorithm can be used as an estimation method for clustering [7].

In non-parametric approaches, a clustering method may be based on an objective function of similarity or dissimilarity measures. As a result, partitional clustering is generally used. The most frequently used partitional methods are k-means [8] and fuzzy c-means (FCM) [9]. Most of the current clustering algorithms depend on iterative procedures to find local or global optimal solutions in high dimensional datasets. Many experiments with different algorithms have to be performed to find these solutions and to study the influence of different dataset features. Hence, clustering algorithms have a high intrinsic time complexity.

LITERATURE REVIEW

Many works have been done in this direction by various researchers in the past. Among them, Jensen et al[10] investigated that attribute reduction is a fuzzy rough set theory based feature selection by presenting a dependency function based reduct and designed a heuristic algorithm to search for one of the deducts. However, it has been proven to be not convergent on many real datasets and the algorithm was restructured with efficient termination criteria to achieve the convergence on all the datasets by Bhatt et al [11]. Uniform representations of approximation spaces and their information measures were formed [12], introducing probability into fuzzy approximation space, a theory about fuzzy probabilistic approximation spaces.

In [13] defined a conditional entropy based on fuzzy rough sets to characterize the dependency function-based reduct and then
used the entropy to develop a feature selection algorithm. Aboul Ella Hassanien [14], introduced a hybridization scheme that combines the advantages of fuzzy sets and rough sets in conjunction with statistical feature extraction technique to classify the breast cancer images. Chen et al [15] introduced Gaussian kernel into fuzzy-rough sets for computing fuzzy similarity relation and developed a novel method of attribute reduction with kernel tricks. Cornelis [16], presented a generalization of the classical rough set framework for data-based attribute selection and reduction using fuzzy tolerance relations and introduced the concept of fuzzy decision reducts, dependent on an increasing attribute subset measure.

Zhao [17], invented a special case of fuzzy-rough sets (FRS) named fuzzy variable precision rough sets (FVPRSs) by combining FRS and VPRS. They employed the discernibility matrix approach to investigate the structure of attribute reductions in FVPRS and developed an algorithm to find all reductions and obtained one near-optimal attribute reduction. In [18] proposed forward approximation accelerator for combining sample reduction and dimensionality reduction and the strategy enhanced the heuristic fuzzy-rough feature selection algorithms dealing with larger data sets. Zeng et al [19], presented fuzzy rough set approaches for incremental feature selection on Hybrid Information System (HIS) to preserve information in dynamic and hybrid environment and proposed a novel hamming distance that can deal with different types of data and applied into Gaussian kernel with FRS with updating features when a new feature is added or an old one is deleted.

Clustering has been successfully applied to the analysis of datasets from several fields such as; image processing, pattern recognition, analysis of microarray data in bioinformatics, credit card behavior modeling, etc, in order to provide valuable knowledge within these fields. Kultur et al [20]. One of the most widely used fuzzy clustering methods is the Fuzzy C-Means (FCM) algorithm [21]. Some parallelization efforts have been done in the literature for FCM algorithm to deal with large datasets. In [22], redefined a clustering technique called Fuzzy Minimal to enhance the classification of large datasets. They revealed that there is a linear speed-up of Parallel Fuzzy Minimal (PFM) when compared to the sequential counterpart version, keeping very good classification quality. Havens [23] extended FCM clustering to very large data. They compared methods that are based on sampling followed by non-iterative extension and incremental techniques that make one sequential pass through subsets of the data and kernelized versions of FCM that provide approximations based on sampling, including three proposed algorithms. Also, they presented a set of recommendations for the use of different very large FCM clustering schemes.

Kwok [24], proposed an algorithm named Parallel Fuzzy C-Means (PFCM), that is meant to run on parallel computers of the Single Program Multiple Data (SPMD) model type with the Message Passing Interface (MPI) and enforced PFCM to cluster huge data set and evaluated in terms of parallelization capability and quantifiability. In [25], proposed a PFCM algorithm for image segmentation and evaluated against sequential algorithm by splitting the computations among the processors and minimizing the necessity for accessing external storage, and enhanced the performance and potency of image segmentation task. In [26] proposed an efficient method to cluster data points of all the images at a time. The gray level histogram is used in the FCM algorithm to minimize the time for segmentation and the space required. A parallel approach is then applied to further reduce the computation time.

The remaining of the paper is organized as follows: Section 3 describes the proposed methods with the suitable algorithm. Experimental results are presented in Section 4. The paper is concluded with a mention on the future scope of this work in section 5.

THE PROPOSED CBIR SYSTEM

Firstly, we should to select the suitable color space for CBIR system. Most color spaces (e.g., RGB, CMY (K), and HSI family) are device-dependent and not perceptually uniform, but L*a*b* color space stays at the safe side away from these two issues. So, among all color spaces, the L*a*b* color space was selected because it is device-independent and perceptually uniform color area that approximates the manner that humans understand color[1]. Secondly, we build a fuzzy inference system (FIS), then we create the required algorithms for relevant image retrieval.

A. Fuzzy Inference System for Color Feature Extraction

In L*a*b* color space, L* stands for luminance, a* represents relative greenness-redness and b* represents relative blueness-yellowness. Building a fuzzy inference system (FIS) for extracting FCH, as shown in Fig. 7, is achieved by the following steps:

1) Fuzzification: When separating the three triplets of L*a*b* color space, as shown in Fig. 2, each one is fuzzified as an input variable to FIS.

L* component does not contribute in providing any distinctive color but for shades of colors: white, black, and grey. Thus, the L* component receives a lower weight with respect to the other two components of the triplet. For this reason, we further divided L* component into only three triangular-shaped fuzzy sets: Black, Gray and White, as shown in Fig. 3.
In order for CBIR to work effectively, a* and b* are subdivided into five triangular-shaped fuzzy sets, as depicted in Fig. 4 and Fig. 5. For a*, we have: Green, Greenish, Middle, Reddish, and Red. For b*, we have: Blue, Bluish, Middle, Yellowish, and Yellow. The reason for which the middle fuzzy set exists both in a* and b*, is that in order to represent black, grey and white as seen in L*, then a* and b* must be very close to the middle of their regions; this is a well-known fact about the L*a*b* space.

2) Defuzzification: The output variable of our FIS, which represents 2-D Fuzzy Colored Image (FCI), is split into 15 equally spiltted trapezoidal-shaped fuzzy sets, as shown in Fig. 6. So, the output variable consists of only 15 bins roughly representing the subsequent colors: Black, Gray, Red, Red-Orange, Orange, Yellow-Orange, Yellow, Yellow-Green, Green, Blue-Green, Blue, Blue-Violet, Violet, Red-Violet, and White.

3) Knowledge base (fuzzy IF-THEN rules): It is used for mapping from a pixel of three fuzzy inputs (L*, a*, and b*) to a pixel of only one fuzzy output. Our proposed FIS has 75 rules established through empirical conclusion. Some of these rules are listed below (as in Fig. 6 and Fig. 7).

- IF (L is Black) and (a is Middle) and (b is Middle) THEN (FCH is Black).
- IF (L is Gray) and (a is Red) and (b is Middle) THEN (FCH is Red).
- IF (L is White) and (a is Green) and (b is Blue) THEN (FCH is Cyan).
This representation takes into account the uncertainty presents in the extraction process of features and consequently, increases the precision rate in the image retrieval process.

B. Off-line Feature Extraction Algorithm

From given an image, FCH can be extracted using the algorithm illustrated in Fig. 8. In this algorithm, the query image is read, and then resized to 50 × 50 pixels (aspect ratio saved). After that, it is converted from the default RGB color space to a color space appropriate for CBIR system (L*a*b* color space). Then, it is normalized and entered to the previously built fuzzy inference system (FIS) for extracting the 2-D fuzzy colored image (FCI). The FCH is calculated from this 2-D fuzzy colored image by subdividing it into 15 bins.

C. On-line Fuzzy Features Matching Algorithm

After obtaining the fuzzy color histogram as a visual feature of the query image using our FIS, described previously, we need to compare it with the FCHs of all images in the image database to specify the degree of similarity, and then retrieve the most relevant (similar) images to the user. To achieve this goal, there are many fuzzy similarity measures. The similarity measures used in our proposed system is called Min-max ratio. According to this measure, the similarity (S(A,B)) between two fuzzy sets is given by Eq. 1.

\[
S(P,Q) = \frac{\sum_{i=1}^{N} \min(\mu_P(i), \mu_Q(i))}{\sum_{i=1}^{N} \max(\mu_P(i), \mu_Q(i))}
\]

Where are the membership values of the \(i\)th bin of histograms HA and HB, respectively. For an identical pair of fuzzy sets, the memberships are equal and the similarity value will be equal to 1.

D. Feature Selection

In this section, we discuss the proposed feature selection algorithm called incremental filtering feature selection (IF2S) algorithm based on fuzzy rough set for effective classification. This algorithm consists of three phases. In phase I, we have used the fuzzy rough set theory for selecting the suitable subsets. In phase II, the most relevant features are selected based on mutual information. Finally, the selected features are confirmed according to the feature ranking process. The following section provides the preliminary concept of fuzzy relations and fuzzy rough sets as shown in Fig. 8, Fig. 9 and Fig. 10.

I. Fuzzy Rough Set based Subset Selection

This section deliberates about the basic concept and definition of fuzzy relations and fuzzy rough sets. The proposed feature selection algorithm uses the fuzzy rough set for effective subset selection.

1. Fuzzy Relations

The fuzzy relations demonstrate the relationship between two sets or elements of the given values. Let is a non-empty universe and the fuzzy power set is represent as \((\times)\). In which, \(\times\) is a power set of the given relation and is indicates the fuzzy relation on \(\times\) if \(\in (\times)\), where \((, )\), measures the strength of the relationship between \(\in\) and \(\in\).
FUZZY ROUGH SETS

According to [27], a set of lower and upper approximation operators of a fuzzy set $X$, which is based on fuzzy relation $FR$ is defined, for each $x \in U$ as in Eq. 2 and Eq. 3.

$$FRX(x) = \inf_{y \in U} \max\{1 - FR(x,y), X(y)\}$$  \hspace{1cm} (2)

$$FRX(x) = \sup_{y \in U} \{ FR(x,y) , X(y) \}$$  \hspace{1cm} (3)

To measure the degree of certainly belonging to and the degree of $x$ possibly belonging to, respectively on which the fuzzy rough set of $X$ is defined by $FRX$, $FRX^c$. The earlier works on fuzzy rough sets mainly focused on constructing the approximations of fuzzy sets along the line of $FR$ and $FR^c$. In this paper, we have used the existing fuzzy rough set based feature reduction [28] technique for effective subset selection in the first phase of the proposed feature selection. The proposed feature selection algorithm uses the feature dependency score of the features in a subset during a particular time interval.

Algorithm: Incremental Filtering Feature Selection (IFFS)

Algorithm Input

**Input:** Data

**Output:** Best Feature Subset, Optimal Feature Set, Selected Features
Step 1: Read the input data
Step 2: Initialize δ = 0.45, Best Feature Subset (BFS) = {}, OFS ← ∅
Step 3: Apply fuzzy rough set for feature subset selection

**Phase I: Fuzzy Rough Set based Subset Selection**
Step 3.1: For each feature from F
Step 3.2: Find the fuzzy relationship (FRX) for every two features using equation 1 and 2.
Step 3.3: If FRX(Xi,Yi) > Threshold then add these features Xi and Yi into BFS
Step 3.4: Repeat step 3 until δ = 0.95
Step 3.5: If BFS (fi) > Threshold then add the feature fi OFS.
Step 3.6: Repeat step 5 until BFS reaches empty.
Step 4: Calculate the Joint Mutual Information Value for feature selection

**Phase II: Joint Mutual Information based feature selection**
Step 4.1: For each feature of OFS
Step 4.2: Calculate the Joint Mutual Information value using the equation 10 and 11.
Step 4.3: If JMI (fi, fj) > Threshold then add these two features into the feature set (FS).
Step 4.4: Repeat the steps 4.2 and 4.3 until OFS is empty.
Step 5: Call the Dynamic Ranking function for ranking the features

**Phase III: Dynamic Feature Ranking**
Step 5.1: For each feature of FS
Step 5.2: Calculate the Feature Dependency Score (FDS) using the equation 12 for a specified time period t1 and t2.
Step 5.3: Sort the features based on FDS in descending order.
Step 5.4: For each feature of OFS
Step 5.5: If FDS value of (OFS(Si, Sj)) > Threshold then add the features Si into SF.
Step 5.6: Repeat the steps 5.4 & 5.5 until the set OFS is empty.
Step 6: Display the selected features.

**II. Mutual Information based Feature Selection**
Optimal feature selection from the subset in Phase II of the proposed feature selection algorithm is based on Mutual Information [29]. In the proposed work, the necessary features are selected from the subsets, which are selected by the fuzzy rough set based subset selection. Finally, the features are placed into two subsets such as Si and Sj based on the dependency between the different features present in the subset. The principles of information theory are discussed focusing on entropy and mutual information and explain the reasons for employing them in feature selection. The entropy of random variable measures the uncertainty of features and an average amount of information [30]. The entropy of a discrete random variable \( X = (x_1, x_2, ..., x_N) \) is denoted by \( E(X) \), Where \( x_i \) refers to the possible values that \( X \) can take \( E(X) \) Eq. 4.

\[
E(X) = \sum_{i=1}^{N} p(x_i) \log p(x_i) \tag{4}
\]

Where \( p(x_i) \) is the probability mass function, The value of \( p(x_i) \), when \( X \) is discrete is Eq. 5

\[
p(x_i) = \frac{\text{No. of instants with value } x_i}{\text{Total no. of instants (N)}} \tag{5}
\]

**III. Dynamic Feature Ranking**
In the proposed work, we have ranked the features based on the time, the mutual information value of each feature and the joint mutual information value of a pair of features. Features’ uncertainties are tackled using the joint mutual information values between the features. The mean value of is calculated for each subset of the dataset such as Si, Sj, SK and Sl. And the mutual information values are considered in the range between 0 and 1. Also, the dependency of the features in a subset is represented based on time. The information gain values between any two features with more values and normalizes its values to the range \([0, 1]\) with value 1, which indicates that knowledge of complete prediction and the value 0 indicates that X and Y are independent. Moreover, it considers a pair of features symmetrically. Entropy-based measures require nominal features; also it is possible to apply for measuring the correlations between continuous features as well when the values are discretized properly. Therefore, it is necessary to use in this work for better ranking based on their relations. At this time, correlation based feature selection [29] is utilized, which uses the best-first strategy search method for calculating the merit of feature subset. However, there is a necessity to fix the stopping criteria, due to this strictly needed constrain correlation between features, which is calculated based on Symmetrical Uncertainty according to Chen et al [29]. Feature Dependency Score (FDS) is calculated during the time interval \(< t_1, t_2 >\) as follows in Eq. 6.

\[
\text{FDS} = \frac{2.0 \times X}{\left[ \frac{M(Si,<t_1,t_2>) + M(Sj,<t_1,t_2>) - M(Si,<t_1,t_2>,Sj,<t_1,t_2>) - M(Si,<t_1,t_2>,Sj,<t_1,t_2>)}{M(Si,<t_1,t_2>) + M(Sj,<t_1,t_2>)} \right]} \tag{6}
\]

**RESULTS AND DISCUSSION**
The proposed approach has been evaluated by experiments on WANG Database.
A. Evaluation Metrics

This section describes in detail about the evaluation metrics, which are used in this work for measuring the performance of the proposed system. Classification accuracy is one of the most popular metrics in the classifier evaluation. It is the proportion of the number of true positives and true negatives obtained by the classification algorithm in the total number of instances, as given by Eq. 7.

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \tag{7}
\]

Where TN, TP, FP and FN represent the number of true negatives, true positives, false positives and false negatives, respectively.

Also, Clustering accuracy has been used as cluster validation metric to judge the quality of the cluster formation algorithm. Clustering Accuracy is defined as follows

\[
\text{Clustering Accuracy} = \frac{\text{No. of Correct Count}}{(\text{Total No. of instance} - \text{Sample})} * 100 \tag{8}
\]

B. Experimental Results

We conducted several experiments to demonstrate the effectiveness of the proposed algorithms. This Section analyzes the Incremental Filtering Feature Selection (IFFS) Algorithm. We focus on the computational features of the Temporal Interval based Fuzzy Minimal Clustering algorithm and how it can be designed for handling large datasets. To guarantee the correctness of our algorithms, a quality comparison between the results obtained by other existing algorithms is also provided. The experiments are developed on a Windows-based machine with 4 GB DDR3 memory and Intel Core I5 2.6 MHz processor using Matlab 6.0 release 12.

I. Dataset Description

We selected the Wang datasets, the improvement in prediction and classification accuracy is because the proposed TIFMC has a minimal rough set which is obtained by applying the clustering algorithm. Thus, the experimental results show that the proposed system provides better classification accuracy.

The classic color histogram is computed using statistical system, where the movement from one bin to the neighboring one occurs suddenly. Therefore, any small change in illumination results in a large change in this part of color histogram, as shown in Fig. 11. Therefore, the performance of the CBIR system will decrease to the minimum. There is no existence of this problem in the FCH, because it does not care of illumination change (all intensities of colors are represented by only one bin using fuzzy membership function), as shown in Fig. 12.

Figure 11. Similarity between the bus-1 image and its 15% illuminated copy using the classic color

Figure 12. Similarity between the bus-1 image histogram and its 15% illuminated copy using the FCH.

The conventional 3-D color histogram suffers from large distances between perceptually very similar images, and also suffers from dispersion. For example, the top nine images relevant to rose-1 image have only 55% of mean, and 17% of standard deviation. So, if we assigned a threshold of 80%, which indicates to very similar images, the number of relevant images retrieved is only one!, as shown in Fig. 15. Suppose that the number of relevant images is X, then the recall measure, which is the fraction of relevant images returned by the query, will be 1/X. As a result, the performance of the conventional system is not good.

Our proposed system does not have the previously discussed problem, i.e., the distance measures are proportional to the perceptual similarity of the relevant images, and there is no dispersion exists. So, if we assigned a threshold of 80%, which indicates to very similar images, the number of relevant images retrieved is seven images, as shown in Fig. 13. Suppose that the number of relevant images is X, then the recall measure will be
7/X. Intuitively, 7/X recall measure of fuzzy system is larger than 1/X recall measure of the conventional system; therefore, we can deduce that the fuzzy system has much better performance than the conventional system.

Practically, we have proven that using multiple experiments applied on several images of WANG database, as shown in Table. 1.

<table>
<thead>
<tr>
<th>No. of Similar Images</th>
<th>Similarity Degree (%)</th>
<th>Classic Color Method</th>
<th>Fuzzy Color Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.65</td>
<td>0.91</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.59</td>
<td>0.9</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.51</td>
<td>0.89</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.48</td>
<td>0.83</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.45</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>0.82</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.43</td>
<td>0.8</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.41</td>
<td>0.79</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 13.** Graph of Similarity Degree (%) comparison between Classic and Fuzzy color methods

**CONCLUSION AND FUTURE WORK**

From large variety of experiments applied on images of WANG database, we have concluded the following results:

1. The FCH is robust to the noise and illumination changes in the images. As a result, it guarantees the retrieval of the images relevant to the query image despite the presence of the noise and the change of the illumination. Therefore, the recall measure interestingly increases.
2. Even though the FCH is a vector of only 15 elements, it has improved the CBIR system performance, because it is computed logically not statistically.
3. The FCH minimizes the size of the features database and decreases the computational cost, because it is one-dimensional descriptor with only 15 bins. In the conventional system, the color histogram is three-dimensional with more than 1000 bins. So, the size of the fuzzy color histogram is smaller than the color histogram of the conventional system at ration of (0.015).
4. The perceptually relevant images have very small distance measures between them and the query image, and they do not suffer from dispersion because features extraction depends on perception (fuzzy) not on measure.
It is easy for users of the CBIR system to understand and directly modify the FCH of the query image, then submit the modified FCH to the CBIR system again as feedback, because the human thinking process is applied in the FCH extraction. Therefore, the FCH seems very easy for users to understand and modify.

A new feature selection method, Incremental Filtering Feature Selection (IFFS) Algorithm, and a new clustering algorithm, that employs the Fuzzy Rough Set for selecting an optimal subset of features and for effective grouping of a large volume of data, respectively. An extensive experimental comparison of the proposed method and other methods are done. The performance of the proposed algorithms yields promising results on the feature selection for effective improvements in classification accuracy. Future works in this direction could be the introduction of intelligent agents and fuzzy rules for effective decision making.

REFERENCES


