

# A Fuzzy Reasoning Technique for Pattern Recognition

Ashoka H N

Department of Electrical Engg,  
UBDT College of Engg.  
Davnagere-577 004, INDIA

Manjaiah D H

Department of Computer  
science, Mangalore University  
Mangalore, INDIA

Rabindranath Bera

Department of Electronics and  
Communication Engg.,SMIT,  
Sikkim, INDIA

## ABSTRACT

This paper presents a fuzzy reasoning technique for pattern classification. The binary image of a pattern is partitioned into number of overlapping and non-overlapping regions for feature extraction. A set of features extracted to represent the binary images are used in the classifier design. The knowledge base (KB) is build using the statistical information involved in the interclass region features. The unknown samples features are matched with KB and assigned a membership value by the fuzzy reasoning technique for their classification. On experimentation with the handwritten numeral data, found good classification and recognition rate.

## General Terms

Pattern Recognition, Fuzzy classification, Optical Character Recognition

## Keywords

Feature extraction, Knowledge base, Fuzzy reasoning

## 1. INTRODUCTION

Pattern recognition (PR) find its applications in most of the engineering and medical field like automatic image recognition, document processing, medical diagnosis, biomedicine, remote sensing, etc. The recent increase in processing capacity of computer, made an exhaustive research in the PR field to simulate human reading and recognition. The PR system classifies assigns the input pattern to the exact class. It includes data collection, pre-processing, feature extraction and classifier subsystems. Different models of subsystems are proposed by the researchers to increase accuracy and recognition speed for PR system. The design of PR system classifiers usually based on the set of features extracted and the availability of training samples. Different feature extraction, selection, and classification approaches are listed in the review article [1].

A 40-50 years of work in the optical character recognition(OCR), human still outperform even the most powerful off-line character recognition system developed so far due extensive variations in handwriting styles like size, shape, fonts, etc[2,3]. This variation increases the complexity of recognition system and reduces the performance of it. The variation involved in the pattern is usually addressed by the feature extractor. Feature extractor generates a set of various characteristics associated with pattern regions to reduce dimensionality of pattern and for the representation.

The zoning technique of statistical feature extraction method is extensively used in the handwritten character recognition system [4]. In this technique, the image fitted in a window is divided into number of regions and the density of black runs pattern or ones in binary images are transformed as features using different techniques [5-7]. Hanmandlu and Murthy [5] extracted 24 features from the binary image of a character by

measuring the coordinate distance of ones in each zone. They devised modified membership function containing structural parameters in addition to mean and variance. These parameters are estimated by defining an objective function consisting of the entropy function constrained upon the square of the error between average membership function and the unity. On experimentation with handwritten Hindi and English numerals, they reported overall recognition rate of 95% and 98.4% respectively by the fuzzy classifier. Benne et.al [8] proposed fast size normalization and thinning free recognition system for isolated handwritten numerals. They extracted structural features for the handwritten numerals representation used Euclidian distance criterion and K-nearest neighbor classifier and reported overall accuracy of 95.40%, 90.20%, and 98.40% for Kannada, Devanagari and Telugu numerals respectively.

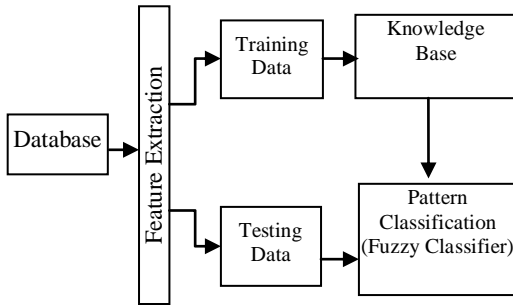
Many other researchers [9-12] proposed different classifier for pattern classification. Neural network based classifiers require more number of samples for training to build the knowledge base. They take more time to learn compared to statistical and fuzzy classifier. Fuzzy classification depends on the set of rules generated and on the training samples. Generating a set of rules for pattern recognition is difficult task. In view of this, an attempt is made to develop a novel fuzzy reasoning technique using the statistical information of training samples for pattern recognition system.

## 2. METHODOLOGY

The proposed methodology for pattern recognition system is given in Figure 1. The numeral database consists binary samples fixed in a window of particular size is partitioned into overlapping and non-overlapping regions for feature extraction. A set of features are extracted to represent the sample and it is called feature vector. Feature vectors such extracted for the database samples are divided into training and testing data. The training samples feature vectors are used to construct the KB. The class label of each training and testing sample is assigned by the fuzzy reasoning technique.

## 3. FEATURE EXTRACTION

Feature extraction is an important step in pattern recognition system to represent the input pattern. A size normalized binary images of patterns used in this experimentation are fitted in a window of size 15x15. Fig.2 shows the binary image of 10<sup>th</sup> sample of numeral class 4. In the number 104 on top of the binary image, the last digit indicates the numeral class 4 and the other two digits (10) indicate the sample number in that class. The binary image is portioned into 16 regions like horizontal, vertical, rectangular, diagonals and squares for feature extraction. Fig.3 shows the shape of partitioned region with their identification number.



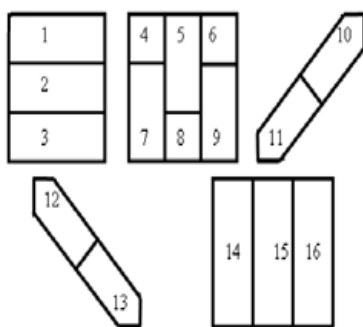
**Fig 1: Methodology adopted in the Pattern recognition system**

The shapes of these regions are so selected that most of them have portion black run (ones) of a samples. However, for some samples there could be regions without ones and such regions will have a zero feature values. The number of ones in each region is transformed into a real value as the feature value.

```

104 :
0000000000000110
0000000000010010
000000001000010
000000010000010
000000100000100
000010000000001
000010000000000
000001000000010
001000000000010
010000000000010
000011111000010
111100000111011
000000000000100
0000000000001000
0000000000001000
    
```

**Fig 2: Binary image of 10<sup>th</sup> sample of numeral class 4.**



**Fig 3: Partitioned 16 regions with their identification number.**

The feature value of a region is computed from the density of 1<sup>s</sup> in that region. The normalized feature value,  $f_r$  of  $r^{\text{th}}$  region is computed by the Eq. (1).

$$f_r = \frac{1}{q_r} \sum_{n=1}^{q_r} P_{ij} \quad (1)$$

Where,  $f_r$  = feature value of  $r^{\text{th}}$  region,  $q_r$  = total number of cells  $r^{\text{th}}$  region,  $P_{ij}$  = the value of  $i^{\text{th}}$  row  $j^{\text{th}}$  column cell of a matrix 'p'.

The feature value of a region depends on the total number of one's present in the region and not on their position. Therefore, changes in the position of one's due to variations in writing styles do not affect feature value and this increases the robustness of feature extraction method.

#### 4. CLASSIFIER DESIGN

The statistical information involved in the training set feature vectors are used to construct the KB. The KB of each region is framed by computing minimum ( $f_{r \text{ min}}^k$ ) and maximum value ( $f_{r \text{ max}}^k$ ) using the mean and standard deviation of each region. The mean feature value ( $\bar{f}_r^k$ ) and standard deviation ( $\sigma_r^k$ ) of  $k^{\text{th}}$  class  $r^{\text{th}}$  region is computed by the Eq. (2) and Eq.(3) respectively.

$$\bar{f}_r^k = \frac{1}{m^k} \sum_{i=1}^{m^k} f_{ri}^k \quad (2)$$

$$\sigma_r^k = \sqrt{\frac{1}{m^k} \sum_{i=1}^{m^k} (f_{ri}^k - \bar{f}_r^k)^2} \quad (3)$$

$r = 1, 2, \dots, n$ , and  $i = 1, 2, \dots, m^k$ ;

Where,  $n$  = total number of regions,  $m^k$  = number samples in  $k^{\text{th}}$  class,  $f_{ri}^k$  is the  $k^{\text{th}}$  class  $r^{\text{th}}$  feature value of  $i^{\text{th}}$  sample,  $\bar{f}_r^k$  = mean feature value of  $r^{\text{th}}$  region and  $\sigma_r^k$  = standard deviation of  $k^{\text{th}}$  class  $r^{\text{th}}$  region feature.

To clutch the variability involved in writing, a deviator 'α' is used. The minimum and maximum value of  $k^{\text{th}}$  class  $r^{\text{th}}$  region are computed by the Eq.(4) and Eq.(5) respectively using the mean values. Fig.4 shows the constructed fuzzy triangle for  $k^{\text{th}}$  class  $r^{\text{th}}$  region. The height of a triangle is fixed to a maximum possible value one.

$$f_{r \text{ min}}^k = \bar{f}_r^k - \alpha \sigma_r^k \quad (4)$$

$$f_{r \text{ max}}^k = \bar{f}_r^k + \alpha \sigma_r^k \quad (5)$$

#### 5. FUZZY REASONING TECHNIQUE

To classify the unknown sample, the feature extracted from a particular region  $r$ ,  $f_r$  is matched with the corresponding region fuzzy triangles of all the classes to assign the membership value by the fuzzy reasoning technique. Fuzzy reasoning is made based on feature value position in the base

of a fuzzy triangle. If the unknown sample  $r^{\text{th}}$  region feature value  $f_r$  is less than the mean value of the region, then the membership value for class  $k$  is assigned by the Eq. (6) and if the  $f_r$  value is greater than mean feature value, then the membership value is assigned by the Eq. (7). If the region feature value  $f_r$  is equal to the mean feature value of that region, then the membership value is assigned to a maximum and equal to the height of a fuzzy triangle.

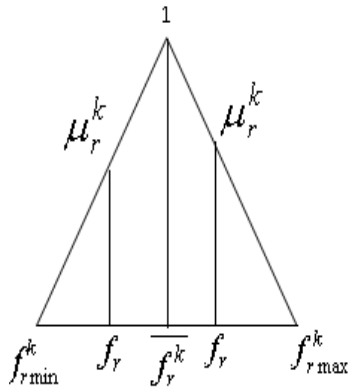


Fig 4:  $k^{\text{th}}$  class  $r^{\text{th}}$  region fuzzy KB triangle

$$\mu_r^k = \frac{(f_r^k - f_r)}{(f_r^k_{\max} - f_r^k)} \quad (6)$$

$$\mu_r^k = \frac{(f_r^k - f_r)}{(f_r^k - f_r^k_{\min})} \quad (7)$$

The membership values of each region feature for all classes are found and used in the unknown sample classification. The algebraic sum of membership values corresponds to a class  $k$  is computed by the Eq.8.

$$\mu^k = \sum_{r=1}^n \mu_r^k \quad (8)$$

The class having highest algebraic sum of the membership value ( $\mu^k$ ) is assigned to the unknown sample.

## 6. RESULTS AND DISCUSSION

The proposed algorithm is tested on handwritten numeral database. The database used in the experimentation has 440 samples were preprocessed and divided into training data and

testing data. The numeral database has 10 numeral classes and each has 22 samples in both the training and testing data. A feature vectors of samples used in the experimentation are obtained by the feature extraction method discussed in the section 2 and computed mean feature vector for each of the numeral class from their training samples. The mean feature vectors computed for the 10 numeral classes are given in Fig. 5. It is observed that inter class mean feature values of a regions are different. This difference is necessary and it describes the features distinctiveness for classification. The statistical information involved in the samples feature vectors is computed to construct the KB. The mean, minimum, and maximum feature values and the standard deviation of each region are computed for all the numeral classes. An example, the statistical values computed for the 0<sup>th</sup> numeral class is shown given Fig.6. These values are used for the construction of 0<sup>th</sup> class KB.

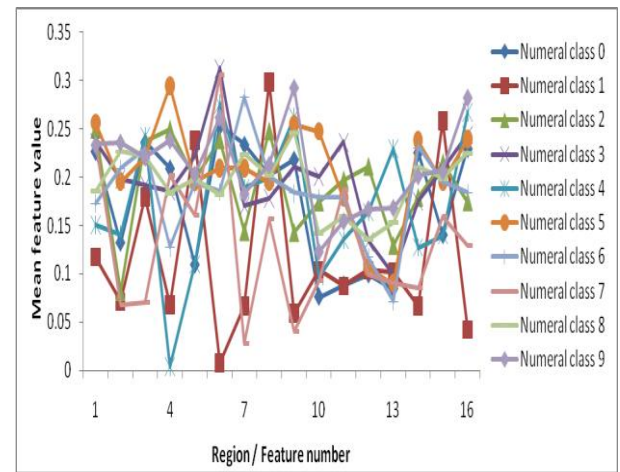


Fig 5: Mean feature vectors of 10 numeral classes

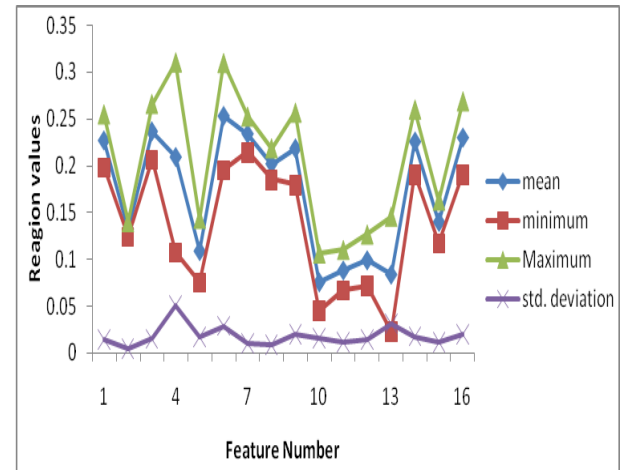
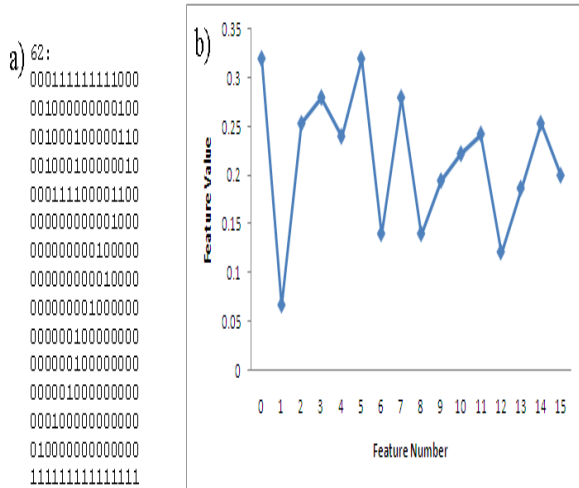


Fig 6: Statistical values of 0<sup>th</sup> numeral class

**Table.1 Membership value of features for numeral class 0-9**

Feature Number	Numeral Class									
	0	1	2	3	4	5	6	7	8	9
0	0	0	0.337953	0	0	0.05774	0	0	0	0
1	0	0.666287	0.601031	0	0	0	0	0.84552	0	0
2	0.432045	0.338755	0.623767	0	0.875996	0.197548	0.009645	0	0.110743	0
3	0.302001	0	0.719575	0.295469	0	0.87592	0.163087	0	0	0
4	0	0.939408	0.434676	0.735071	0	0	0	0	0	0
5	0	0	0.399702	0.931763	0.48783	0	0	0.785519	0	0
6	0	0	0.954347	0.632576	0	0	0	0	0	0
7	0	0.906264	0.674999	0	0	0	0.189908	0.113469	0	0
8	0	0	0.893399	0	0	0	0.31507	0	0	0
9	0	0	0.804108	0.938571	0	0.139118	0.542251	0	0.269106	0
10	0	0	0.630726	0.807719	0	0	0.637559	0.409195	0.439647	0
11	0	0	0.756723	0	0.154666	0	0	0.210791	0	0
12	0.398605	0.807025	0.886595	0.630725	0	0.334461	0.219741	0.566987	0.67082	0.421393
13	0	0	0.870506	0.866157	0	0.155321	0.330011	0	0.514045	0
14	0	0.935543	0.107827	0.063612	0	0	0	0	0	0
15	0.246438	0	0.476494	0	0	0	0.659575	0	0.329767	0
<b>Total Membership Value</b>	<b>1.37909</b>	<b>4.593282</b>	<b>10.17243</b>	<b>5.901665</b>	<b>1.518493</b>	<b>1.760108</b>	<b>3.066845</b>	<b>2.931482</b>	<b>2.334129</b>	<b>0.421393</b>



**Fig.7: Unknown numeral sample a) Binary image of numeral '2' b) Feature vector of sample numeral '2'.**

The system performance is evaluated by the classification and recognition rate. In order to recognize the unknown samples, their feature vectors are extracted by the method discussed in section 2. Fig.7 shows the unknown sample. Fig.7 (a) is the

binary image of sixth sample of numeral class two and Fig.8 (b) shows its feature vector. The features vectors of unknown samples are compared with the KB feature vectors of all the classes. Depending on the feature value, a fuzzy reasoning is made and assigned the membership value for their matching. An example, the feature vector of a sample given in Fig.7 (b) is matched with the KB. The membership values are assigned for different classes features are tabulated in the Table1. It is seen that, for most of the features corresponds to numeral class two have high membership value compare to other classes. The algebraic sum of membership values of each class is computed to recognize the unknown sample. The class 2 has highest membership value and therefore its class is assigned to unknown sample. When the unknown sample feature value does not lie within the range of reference class features region, a zero membership value is assigned to them. In Table 1, majority cells have zero values. The cells having zero values corresponds to incorrect classes and cells corresponds to the correct class have non-zero value. When the feature vectors of training samples are classified by the proposed technique, found 100% classification rate. Similarly, the feature vectors of unknown (testing) samples are matched with the KB and recognized by the fuzzy reasoning technique. It is found that all the samples were recognized correctly and found 100% recognition rate.

## 7. CONCLUSIONS

From the above, it is observed that the fuzzy reasoning technique can be designed by using the statistical information involved in the training data samples. Constructed KB uses minimum training samples and the fuzzy reasoning technique is simple and does not require the usual steps involved in the fuzzy classification technique. On experimentation with off-line handwritten numerals, the proposed fuzzy reasoning technique results good classification and recognition rate.

## 7. REFERENCES

- [1] Anil K. Jain, Robert P.W. Duin, and Jianchang Mao, "Statistical Pattern Recognition: A Review" IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 1, pp 4-37, January 2000.
- [2] Hiromichi Fujisawa, "Forty years of research in character and document recognition - an industrial perspective" Pattern Recognition, Vol.41, pp.2435 – 2446, 2008.
- [3] Nafiz Arica, Fatos T. Yarman-Vural, "An Overview of Character Recognition Focused on Off-line Handwriting", IEEE Transactions on system Man. Cybernetics-Part C: Applications and Reviews, Vol. 31 no. 2, pp.216-233, May 2001.
- [4] Umapada Pala, Partha Pratim Royb, Nilamadhaba Tripathya and Josep Lladósb, "Multi-oriented Bangla and Devnagari text recognition", Pattern Recognition, Vol. 43, No.12, 2010, pp 4124-4136.
- [5] Hanmandlu M., and Murthy O.V.R., "Fuzzy model based recognition of handwritten numerals," Pattern Recognition, Vol. 40, 2007, pp 1840 – 1854.
- [6] Rajashekaradhy S.V. and Vanaja Ranjan P., "Efficient zone based feature extraction algorithm for handwritten numeral recognition of four popular south Indian scripts", Journal of Theoretical and Applied Information Technology, 2005, pp 1171- 1181.
- [7] Xuewen Wang, Xiaoqing Ding, Changsong Liu, "Gabor filters-based feature extraction for character recognition", Pattern Recognition, vol.38, 2005,pp 369 – 379.
- [8] Benne R.G., Dhandra B.V. and Mallikarjun Hangarge, "Tri-scripts handwritten numeral recognition: a novel approach" Advances in Computational Research, ISSN:0975–3273, Vol. 1, Issue 2, 2009, pp-47-51.
- [9] Enwang Zhou, Alireza Khotanz, "Fuzzy classifier design using genetic algorithms" Pattern Recognition Vol.40, 2007, pp 3401 – 3414,
- [10] Yi-Chung Hu, "Fuzzy integral-based perceptron for two-class pattern classification problems", Information Sciences, Vol. 177, 2007, pp1673–1686.
- [11] José Vélez, Ángel Sánchez, Belén Moreno, José L. Esteban, "Fuzzy shape-memory snakes for the automatic off-line signature verification problem" , Fuzzy Sets and Systems Vol.160, 2009, pp182 – 197.
- [12] Ashoka H.N., Manjaiah D.H., Rabindranath Bera, "Feature Extraction Technique for Neural Network Based Pattern Recognition," International Journal on Computer Science and Engineering, Vol.4 No. 03, March 2012, pp 331-339.