EFFICIENT ZONE BASED FEATURE EXTRATION ALGORITHM FOR HANDWRITTEN NUMERAL RECOGNITION OF FOUR POPULAR SOUTH INDIAN SCRIPTS

[1]S.V. Rajashekararadhya, [2]Dr P. Vanaja Ranjan

[1]Research Scholar, Department of Electrical and Electronics Engineering, CEG, Anna University

[2]Asst.Professor, Department of Electrical and Electronics Engineering, CEG, Anna University

Chennai, India

Email: [1]svr_aradhya@yahoo.co.in, [2]vanajar@annauniv.edu

ABSTRACT

Character recognition is the important area in image processing and pattern recognition fields. Handwritten character recognition has received extensive attention in academic and production fields. The recognition system can be either on-line or off-line. Off-line handwriting recognition is the subfield of optical character recognition. India is a multi-lingual and multi-script country, where eighteen official scripts are accepted and have over hundred regional languages. In this paper we propose Zone centroid and Image centroid based Distance metric feature extraction system. The character centroid is computed and the image (character/numeral) is further divided in to n equal zones. Average distance from the character centroid to the each pixel present in the zone is computed. Similarly zone centroid is computed and average distance from the zone centroid to each pixel present in the zone is computed. We repeated this procedure for all the zones/grids/boxes present in the numeral image. There could be some zones that are empty, and then the value of that particular zone image value in the feature vector is zero. Finally 2*n such features are extracted. Nearest neighbor and Feed forward back propagation neural network classifiers are used for subsequent classification and recognition purpose. We obtained 99 %, 99%, 96% and 95 % recognition rate for Kannada, Telugu, Tamil and Malayalam numerals respectively.

Keywords: *Indian Scripts, Image Processing, Neural Network, Nearest Neighbor, Distance feature*

1. INTRODUCTION

Machine simulation of human function has been a very challenging research area due to the tremendous advent of digital computers. Character recognition is the important area in image processing and pattern recognition fields. The aim of character recognition is to translate human readable characters to machine readable characters. Optical character recognition is a process of translation of human readable characters to machine readable characters in optically scanned and digitized text. Handwritten character recognition (HCR) has received extensive attention in academic and production fields. The recognition system can be either on-line or off-line. In on-line handwriting recognition words are generally written on a pressure sensitive surface (digital tablet PCs) from which real time information, such as the order of the stroke made by the writer is obtained and preserved. This is significantly different to off-line handwriting recognition where no dynamic information is available [1]. Off-line handwriting recognition is the process of finding letters and words are present in digital image of handwritten text. It is the subfield of optical character recognition(OCR). Several methods of recognition of English, Latin, Arabic, Chinese scripts are excellently reviewed in [1, 2, 3, 4].

There are five major stages in the HCR problem: Image preprocessing, segmentation, feature extraction, training and recognition and post processing. Research in HCR is popular for various practical application potential such as reading aid for the blind, bank cheques, vehicle number plate,

automatic pin code reading of postal mail to sort. There is a lot of demand on Indian scripts character recognition and a review of the OCR work done on Indian language is excellently reviewed in [5]. In [6] a survey on feature extraction methods for character recognition is reviewed. Feature extraction method includes Template matching, Deformable templates, Unitary Image transforms, Graph description, Projection Histograms, Contour profiles, Zoning, Geometric moment invariants, Zernike Moments, Spline curve approximation, Fourier descriptors , Gradient feature and Gabor feature.

India is a multi-lingual and multi-script country comprising of eighteen official languages, namely Assamese, Bangla, English, Gujarati, Hindi, Kankanai, Kannada, Kashmiri, Malayalam, Marathi, Nepali, Oriya, Punjabi, Rajasthani, Sanskrit, Tamil, Telugu and Urdu. Recognition of handwritten Indian scripts is difficult because of the presence of numerals, vowels, consonants, vowel modifiers and compound characters.

We will now briefly review the few important works done towards HCR with reference to the Indian language scripts. In [7] Grid based feature extraction method is used to recognize the handwritten Bangla numerals using multifier classifier. The accuracy of 97.23% is reported. Image is divided into n zones. For each zone four types of features are extracted. The first feature is extracted from each zone by finding total foreground pixels of the zone and then divided this number by the total number of foreground pixels falling in the entire image. The second feature is extracted by finding average orientation of the edge pixels, falling in each of the zones. Third and fourth features are extracted by finding slopes and intercept of the part of the image falling on each of the zone. This procedure is sequentially repeated for entire zone present in image. Recognition of conjunctive Bangla Characters by Artificial Neural Network is reported in [8]. Recognition of Devanagari characters using gradient features and fuzzy-neural network are reported in [9] and [10] respectively. We found curvature feature for recognizing Oriya characters in [11].

In [12] zone/grid based feature extraction for handwritten Hindi numerals is reported. For extracting the features, each character image is divided in to 24 zones. By considering the bottom left corner of the image as absolute reference, the vector distance for each pixel present in the zone is computed and normalized vector distance is then computed by dividing the sum of vector distances of all black pixels present in a zone with their total number. Similarly this procedure is repeated for all the zones to obtain 24 feature set. Character recognition for Telugu scripts using multi-resolution analysis and associative memory is reported in [14]. Fuzzy technique and neural network based off-line Tamil character recognition are found in [15] and in this work handwritten numerals are preprocessed and segmented into primitives. These primitives are measured and labeled using fuzzy logic. Strings of a character formed from these labeled primitives. To recognize the handwritten characters, conventional string matching is performed.

In [16] author is described a scheme for extracting features from the gray scale images of the handwritten characters on their state-space map with eight directional space variations. In [17] for feature computation, the bounding box of a numeral image is segmented into blocks and the directional features are computed in each of the blocks. These blocks are then down sampled by a Gaussian filter and the features obtained from the down sampled blocks are fed to a modified quadratic classifier for recognition. Off-line HCR for numeral recognition is also found in [18-22].

Recognition of isolated handwritten Kannada numerals based on image fusion method is found in [18]. The numeral is normalized to fit into a size of 32x32 pixels. The binary numeral image is divided into 64 zones of equal size, each zone being of size 4x4 pixels. Then the total foreground pixels of the image are computed. Sequentially total foreground pixels of each zone are computed. If zone total pixel density is greater than 5% of total foreground pixels, 1 is stored for that particular zone, other wise 0 is stored.

In [19] Zone and Distance metric based feature extraction is used. The character centroid is computed and the image is further divided in to n equal zones. Average distance from the character centroid to the each pixel present in the zone is computed. This procedure is repeated for all the zones present in the numeral image. Finally n such features are extracted for classification and recognition.

In [20] zone and vertical projection distance metric is used. The preprocessed numeral image (50x50) is fed to the feature extraction module. The image is divided into 25 zones (each zone size is 10x10). Average pixel distance of the each column present in the zone is computed vertically. Hence 10 distance features are obtained for zone. This

procedure is repeated sequentially for all the zones. Finally 250 features are extracted for recognition. In [21] zone centroid is computed and the image is further divided in to n equal zones. Average distance from the zone centroid to the each pixel present in the zone is computed. This procedure is repeated for all the zones present in the numeral image. Finally n such features are extracted for classification and recognition.

From the above literature survey it is clear that not much work done on south Indian scripts. This motivated us to work towards south Indian scripts. Selection of feature extraction method is also a most important factor for achieving efficient character recognition. In this paper we propose a simple and efficient zone based feature extraction algorithm. Nearest neighbor and feed forward back propagation neural network classifiers are used for recognition and classification of numeral image. We have tested our method for Kannada, Telugu, Tamil and Malayalam numerals.

The rest of the paper is organized into six sections. In section 2 we will briefly explain about the overview of Kannada, Telugu, Tamil and Malayalam scripts. In section 3 we will briefly explain about the data collection and preprocessing. In section 4 we will discuss about the proposed feature extraction method. Section 5 describes the numeral classification and recognition using nearest neighbor classifier and feed forward back propagation neural network. Section 6 describes the experimental results and comparative study and finally, conclusion is given in section 7.

2. OVERVIEW OF KANNADA, TELUGU, TAMIL AND MALAYALAM SCRIPTS

In this section, we will explain the properties of four popular South Indian scripts. Most of the Indian scripts are originated from Brahmi script through various transformations. Writing style of Indian scripts considered in this paper is from left to right, and the concept of upper/lower case is not applicable to these scripts.

Kannada is one of the major Dravidian languages of Southern India and one of the earliest languages evidenced epigraphically in India and spoken by about 50 million people in the Indian state of Karnataka, Tamil Nadu, Andra Pradesh and Maharasthra. The script has 49 characters in its alphasyllabary and is phonetic. The characters are classified into three categories: swaras(vowels), vyanjans(consonants) and yogavaahas (part vowel, part consonants). The script also includes 10 different Kannada numerals of the decimal number system.

Telugu is the Dravidian language and it is the third most popular script in India. It is the official language of the southern Indian state, Andhra Pradesh and also spoken by neighboring states. Telugu is also spoken in Bahrain, Fiji, Malaysia, Mauritius, Singapore and the UAE. The Telugu script is closely related to the Kannada script. Telugu is a syllabic language. Similar to most languages of India, each symbol in Telugu script represents a complete syllable. Officially, there are 10 numerals, 18 vowels, 36 consonants, and three dual symbols.

Tamil is a Dravidian language and is one of the oldest languages in the world. It is the official language of the Indian state of Tamil Nadu, and also has official status in Sri Lanka, Malaysia and Singapore. Tamil script has 10 numerals, 12 vowels, 18 consonants and five grantha letters. The script however is syllabic and not alphabetic. The complete script therefore consists of 31 letters in their independent form, and an additional 216 combining letters representing every possible combination of a vowel and a consonant.

Malayalam is a Dravidian language and it is the eighth most popular scripts in India and spoken by about 30 million people in the Indian state of Kerala. Both the language and the writing system are closely related to Tamil. However Malayalam has its own script. The script has 16 vowels, 37 consonants and 10 numerals.

The challenging part of Indian handwritten character recognition is the distinction between the similar shaped components. A very small variation between two characters or numerals leads to recognition complexity and degree of recognition accuracy. The style of writing characters is highly different and they come in various sizes and shapes. Same numeral may take different shapes and conversely two or more different numerals of a script may take similar shape.

3. DATA COLLECTION AND PREPROCESSING

Data collection for the experiment has been done from the different individuals. Currently we are developing datasets for Kannada, Telugu, Tamil and Malayalam numerals.

We have collected 2000 Kannada numeral samples from 200 different writers. Writers were provided with the plain A4 sheet and each writer has asked to write Kannada numerals from 0 to 9 for one time. The database is totally unconstrained and has been created for validating the recognition system. Similarly we have collected 2000 Telugu numeral samples from 200 different writers, 1500 Tamil numeral samples from 150 writers and 1500 Malayalam numeral samples from 150 different writers. The collected documents are scanned using HP-scan jet 5400c at 300dpi which is usually a low noise and good quality image. The digitized images are stored as binary images in BMP format. A sample of Kannada, Telugu, Tamil and Malayalam handwritten numerals from the data set are shown from figure 1 to figure 4 respectively.

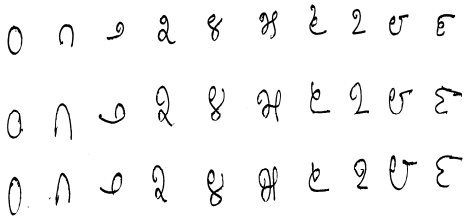

Figure 1. Sample handwritten Kannada numerals 0 to 9

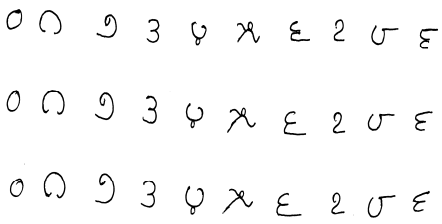

Figure 2. Sample handwritten Telugu numerals 0 to 9

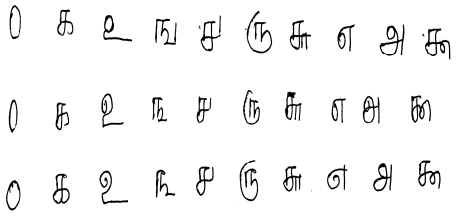

Figure 3 Sample handwritten Tamil numerals 0 to 9

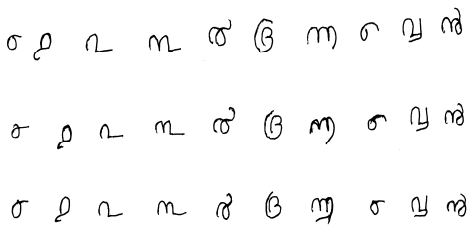

Figure 4 Sample handwritten Malayalam numerals 0 to 9

Preprocessing includes the steps that are necessary to bring the input data into an acceptable form for feature extraction. The raw data, depending on the data acquisition type, is subjected to a number of preliminary processing stages. Preprocessing stage involves noise reduction, slant correction, size normalization and thinning. Among these size normalization and thinning are very important. Normalization is required as the size of the numeral varies from person to person and even with the same person from time to time. The input numeral image is normalized to size 50x50 after finding the bounding box of each handwritten numeral image.

Thinning provides a tremendous reduction in data size, thinning extracts the shape information of the characters. It can be considered as conversion of off-line handwriting to almost on-line data. Thinning is the process of reducing thickness of each line of pattern to just a single pixel. In this research work, we have used morphology based thinning algorithm for better symbol representation. The detail information about the thinning algorithm is available in [23].

Thus the reduced pattern is known as the skeleton and is close to the medial axes, which preserves the topology of the image. figure 5 shows the steps involved in our method as par preprocessing is considered.

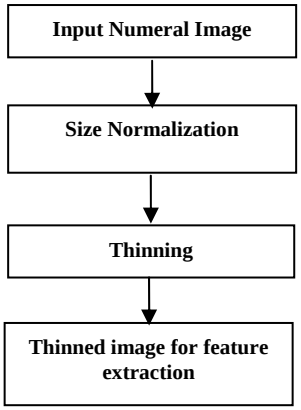

Figure 5. Preprocessing of the input numeral image

4. PROPOSED FEATURE EXTRACTION METHOD

For extracting the feature, the zone based hybrid approach is proposed. The most important aspect of handwriting recognition scheme is the selection of good feature set, which is reasonably invariant with respect to shape variations caused by various writing styles. The major advantage of this approach stems from its robustness to small variation, ease of implementation and provides good recognition rate. Zone based feature extraction method provides good result even when certain preprocessing steps like filtering, smoothing and slant removing are not considered. In this section, we explain the concept of feature extraction method used for extracting features for efficient classification and recognition. The following paragraph explains in detail about the feature extraction methodology.

The character centroid is computed and the image (character/numeral) is further divided in to fifty equal parts as shown in figure. 6. Average distance from the character centroid to the each pixel present in the zone is to be computed. Similarly zone centroid is computed and average distance from the zone centroid to each pixel present in the zone is to be computed. We repeated this procedure for all the zones/grids/boxes present in the numeral image. There could be some zones that are empty, and then the value of that particular zone image value in the feature vector is zero. Finally 100 such features are used for feature extraction.

For classification and recognition nearest neighbor classifier and feed forward back propagation neural network classifiers are used. Algorithm 1 provides Image centroid and Zone (ICZ) based distance metric feature extraction system [19]. Algorithm 2 provides Zone Centroid and Zone (ZCZ) based Distance metric feature extraction system [21]. Algorithm 3 provides the proposed hybrid feature extraction system. (ICZ + ZCZ). The following are the algorithms to show the working procedure of our feature extraction methods.

Algorithm 1: Image Centroid and Zone (ICZ) based Distance metric feature extraction system.

Input: Preprocessed numeral image

Output: Features for Classification and Recognition

Method Begins

Step 1: Compute the input image centroid

Step 2: Divide the input image in to **n** equal zones.

Step 3: Compute the distance between the image centroid to each pixel present in the zone.

Step 4: Repeat step 3 for the entire pixel present in the zone.

Step 5: Compute average distance between these points.

Step 6: Repeat this procedure sequentially for the entire zone.

Step 7: Finally, **n** such features will be obtained for classification and recognition.

Method Ends

Algorithm 2: Zone Centroid and Zone (ZCZ) based Distance metric feature extraction system.

Input: Preprocessed numeral image

Output: Features for Classification and Recognition

Method Begins

Step 1: Divide the input image in to **n** equal zones.

Step 2: Compute the zone centroid

Step 3: Compute the distance between the zone centroid to each pixel present in the zone.

Step 4: Repeat step 3 for the entire pixel present in the zone.

Step 5: Compute average distance between these points.

Step 6: Repeat this procedure sequentially for the entire zone.

Step 7: Finally, **n** such features will be obtained for classification and recognition.

Method Ends

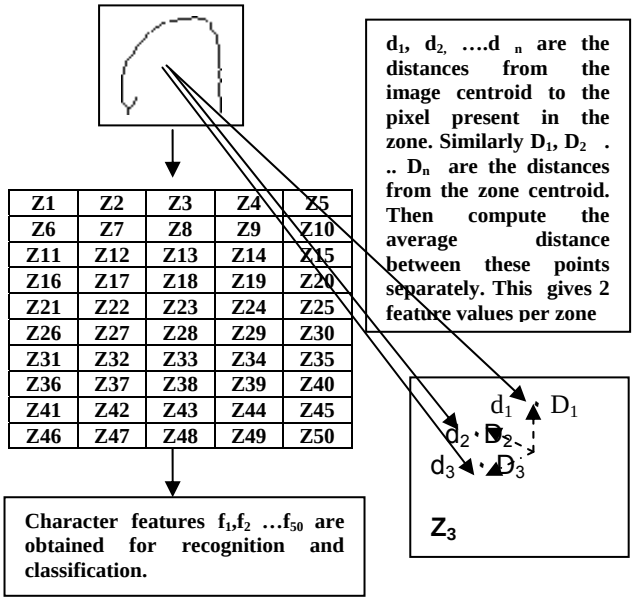

Figure 6. Procedure for extracting features from the numeral image

Proposed Algorithm 3: ICZ + ZCZ based Distance metric feature extraction system.

Input: Preprocessed numeral image

Output: Features for Classification and Recognition

Method Begins

Step 1: Compute the input image centroid
Step 2: Divide the input image into **n** equal zones.
Step 3: Compute the distance between the image centroid to each pixel present in the zone.
Step 4: Repeat step 3 for the entire pixel present in the zone.
Step 5: Compute average distance between these points.
Step 6: Compute the zone centroid
Step 7: Compute the distance between the zone centroid to each pixel present in the zone.
Step 8: Repeat step 7 for the entire pixel present in the zone
Step 9: Compute average distance between these points.
Step 10: Repeat the steps 3-9 sequentially for the entire zone.
Step 11: Finally, **2*n** such features will be obtained for classification and recognition.

Method Ends

5. CLASSIFICATION AND RECOGNITION

For large-scale pattern matching, a long-employed approach is the nearest- neighbor classifier. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the actual classification phase, the same features as before are computed for the test samples. Distances from the new vector to all stored vectors are computed. Then Classification and recognition is achieved on the basis of similarity measurement.

An Artificial Neural Network (ANN) is a computational model widely used in pattern recognition. It has been used extensively both for the recognition of non-Indian as well as Indian digits. Recognition of handwritten numeral is a very complex problem. Feed forward back propagation network is used for subsequent recognition and classification of numeral image.

5.1. Network structure and number of hidden layer nodes

The recognition performance of Back Propagation network will highly depend on the structure of the network and training algorithm. Feed forward back propagation neural network has been selected to train the network. The number of nodes in input, hidden and output layers will determine the network structure. All the neurons of one layer are fully interconnected with all neurons of its just preceding and just succeeding layers (if any). The network consists of 50 nodes in the input layer (corresponding to one feature in each of the 50 zones), 80 neurons in the hidden layer. The output layer has 10 neurons corresponding to 10 numerals. Therefore only the number of hidden nodes needs to be determined.

The number of hidden nodes will heavily influence the network performance. Insufficient hidden nodes will cause under fitting where the network cannot recognition the numeral because there are not sufficient adjustable parameters to model the input-output relationship. Excessive

hidden nodes will cause over fitting where the network fails to generalize. There is no theoretical development based on which, the optimal number of neurons in the hidden layer can be determined.

There are several rules of thumb for deciding the number of neurons in the hidden layer.

- The number of hidden neuron should be less than twice the input layer size.
- The number of hidden neuron should be in the range between the size of the input layer and the size of the output layer.
- Finding the minimum number of epochs taken to recognize a character and recognition efficiency of training as well as testing samples.
- The number of hidden neurons should be 2/3 of the input layer size, plus the size of the output layer.

We followed these rules to arrive at 80 neurons for the hidden layer.

5.2. Transfer function and performance function

Since our desired outputs must be ranged between 0 to 1, so we have selected log sigmoid as the transfer function for both hidden and output layer. We have used 'Mean Squared Error' (MSE) as performance parameter function. MSE is the average squared error between the network outputs and the target outputs. During training, the weights of the network are iteratively adjusted to minimize the function. We adopt 'Gradient descent back propagation' as a learning algorithm. The algorithm updates weights according to gradient descent momentum and adaptive learning rate. The values we have used to set training parameters are learning rate to 0.1, momentum factor to 0.8 and performance goal to 0.01.

6. EXPERIMENTAL RESULTS AND COMPARATIVE STUDY

6.1. Experimental results

In order to evaluate the performance of the proposed method, we consider handwritten Kannada, Telugu, Tamil and Malayalam Numerals. We have collected handwriting of 200 individual writers and total of 2000 samples are considered for Kannada numerals. Here 1000 samples are used for training purpose and remaining 1000 samples are used for testing. Similarly we have collected handwriting of 200 individual writers and total of 2000 samples are considered for Telugu numerals. Here 1000 Telugu numeral samples are considered for training and 1000 Telugu numeral samples are considered for testing purpose. Similarly we have considered 1500 samples for each Tamil and Telugu numeral. For recognition and classification purpose nearest neighbor classifier and feed forward back propagation network classifiers are used. The output layer has 10 neurons corresponding to 10 numerals and neurons values are numeric.

Table 1, Table 2 shows the results for Kannada numeral for different testing samples using nearest neighbor and neural network classifier respectively. Table 3 and Table 4 gives results for Telugu numeral. Table 5 and Table 6 gives results for Tamil numeral. Table 7 and Table 8 gives results for Malayalam Numerals.

Table 1 Kannada Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Nearest neighbor classifier					
Kannada Numerals	Testing samples				
	200	**400**	**600**	**800**	**1000**
0	100	100	100	100	100
1	100	100	100	100	100
2	100	100	100	98.75	99
3	95	97.5	95	93.75	91
4	100	97.5	98.33	96.25	96
5	100	97.5	91.67	91.25	90
6	95	87.5	88.33	90	90
7	100	97.5	98.33	93.75	91
8	100	100	100	97.50	97
9	100	100	98.33	98.75	93
Average Recognition	**99**	**97.75**	**97**	**96**	**94.70**

Table 2 Kannada Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Neural network					
Kannada Numerals	Testing samples				
	200	**400**	**600**	**800**	**1000**
0	100	100	100	100	100
1	100	100	96.67	97.50	96
2	100	100	100	100	99
3	100	100	98.33	92.50	90
4	100	97.5	98.33	97.50	97
5	95	95	95	91.25	92

6	100	95	90	90.00	88
7	95	90	91.67	93.75	91
8	100	100	98.33	97.50	96
9	95	97.5	96.67	96.25	97
Average Recognition	**98.5**	**97.5**	**96.50**	**95.63**	**94.60**

Table 3 Telugu Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Nearest neighbor classifier					
Telugu Numerals	Testing samples				
	200	**400**	**600**	**800**	**1000**
0	100	100	100	100	100
1	100	100	100	98.75	96
2	100	100	98.33	95	96
3	95	97.5	96.67	96.25	95
4	100	100	100.00	97.50	98
5	95	97.5	95.00	95.00	93
6	100	100	100	100	96
7	100	95	95.00	95	91
8	100	100	100	100	100
9	100	95	96.67	96.25	97
Average Recognition	**99**	**98.50**	**98.17**	**97.38**	**96.20**

Table 4 Telugu Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Neural network					
Telugu Numerals	Testing samples				
	200	**400**	**600**	**800**	**1000**
0	100	100	98.33	96.25	96
1	100	100	100	100	99
2	100	95	93.33	92.50	94
3	95	97.5	96.67	97.50	97
4	100	100	100.00	97.50	98
5	95	95	91.67	93.75	91
6	100	97.5	96.67	97.50	98
7	85	87.5	91.67	87.50	82
8	100	97.5	93.33	92.50	94
9	100	97.5	98.33	98.75	98
Average Recognition	**97.50**	**96.75**	**96**	**95.38**	**94.70**

Table 5 Tamil Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Nearest neighbor classifier					
Tamil Numerals	Testing samples				
	100	**200**	**300**	**400**	**500**
0	100	100	96.67	97.50	98
1	90	90	90	90	92
2	100	95	86.67	90	90
3	90	90	90.00	90	88
4	100	95	93.33	92.50	94
5	90	90	86.67	82.50	86
6	70	70	76.67	72.50	62
7	100	100	100	100	100
8	100	100	96.67	92.50	94
9	90	85	90	90	84
Average Recognition	**93**	**91.5**	**90.67**	**89.75**	**88.8**

Table 6 Tamil Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Neural network					
Tamil Numerals	Testing samples				
	100	**200**	**300**	**400**	**500**
0	100	100	96.67	97.50	98
1	70	70	73.33	80	76
2	100	100	86.67	90	90
3	100	95	96.67	97.50	96
4	100	95	96.67	95	96
5	90	95	96.67	90	92
6	100	95	90	82.50	78
7	100	95	96.67	97.50	98
8	100	100	100	97.50	98
9	100	100	96.67	92.50	88
Average Recognition	**96**	**94.5**	**93**	**92**	**91**

Table 7 Malayalam Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Nearest neighbor classifier					
Malayalam Numerals	Testing samples				
	100	**200**	**300**	**400**	**500**
0	100	100	96.67	97.50	96
1	100	100	100	100	100
2	100	100	96.67	97.50	98
3	100	95	93.33	90.00	84
4	100	80	76.67	77.50	78
5	100	95	96.67	95.00	94
6	70	85	86.67	87.50	88
7	80	90	93.33	92.50	94
8	100	100	100	100	98
9	100	100	100	95	94
Average Recognition	**95**	**94.5**	**94**	**93.25**	**92.4**

Table 8 Malayalam Handwritten Numeral Recognition Results for Different Testing Samples

(ICZ + ZCZ) Neural network					
Malayalam Numerals	Testing samples				
	100	**200**	**300**	**400**	**500**
0	100	95	93.33	95	88
1	100	100	100	97.50	96
2	90	95	96.67	95	96
3	100	95	93.33	92.50	94

4	70	80	73.33	72.50	76
5	100	100	96.67	97.50	98
6	100	100	100	100	98
7	90	80	83.33	80	84
8	100	95	96.67	97.50	98
9	90	95	90	87.50	82
Average Recognition	**94**	**93.5**	**92.33**	**91.5**	**91**

6.2. COMPARATIVE STUDY

We compared our results with some of the recently published work on off-line handwritten numerals of Indian scripts. In [12] for Hindi numeral, zone based feature extraction algorithm is used and obtained 95% recognition accuracy. We have implemented the same feature extraction algorithm and we tested for 2000 Kannada numeral samples. We obtained recognition of 97.5%. We obtained 99% recognition accuracy for the proposed algorithm. Table 9 provides the comparative results of our previous work and the present work of this paper for a common data set. Similarly table 10, table 11 and table 12 provides comparative results for Telugu, Tamil and Malayalam numerals respectively.

Table 9. Comparative Results for Kannada Numerals

Kannada numerals data set size = 2000		
Feature extraction method	Classifier	Recognition rate (%)
Zoning [12]	Nearest neighbor	97.5
Zoning [12]	Neural network	96
Zoning [ICZ] [19] This paper	Nearest neighbor	97.5
Zoning [ICZ] [22]	Neural network	98
Zoning[21] [ZCZ]	Nearest neighbor	95.5
Zoning [ZCZ] This paper	Neural network	96
Proposed method **ICZ + ZCZ**	Nearest neighbor	**99**
Proposed method **ICZ + ZCZ**	Neural network	**98.5**

Table 10. Comparative Results for Telugu Numerals

Telugu numerals data set size = 2000		
Feature extraction method	Classifier	Recognition rate (%)
Zoning [ICZ] [19] This paper	Nearest neighbor	97.5
Zoning [ICZ] This paper	Neural network	96.5
Zoning [ZCZ] This paper	Nearest neighbor	96
Zoning [ZCZ] This paper	Neural network	94.5
Proposed method **ICZ + ZCZ**	Neural network	**99**
Proposed method **ICZ + ZCZ**	Nearest neighbor	**97.5**

Table 11. Comparative Results for Tamil Numerals

Tamil numerals data set size = 1500		
Feature extraction method	Classifier	Recognition rate (%)
Zoning [ICZ] This paper	Nearest neighbor	88
Zoning [ICZ] This paper	Neural network	89
Zoning [ZCZ] This paper	Nearest neighbor	83
Zoning [21] [ZCZ]	Neural network	95
Proposed method **ICZ + ZCZ**	Nearest neighbor	**93**
Proposed method **ICZ + ZCZ**	Neural network	**96**

Table 12. Comparative Results for Malayalam Numerals

Malayalam numerals data set size =1500		
Feature extraction method	Classifier	Recognition rate (%)
Zoning [ICZ] This paper	Nearest neighbor	94
Zoning [ICZ] This paper	Neural network	92
Zoning [ZCZ] This paper	Nearest neighbor	90
Zoning [ZCZ] This paper	Neural network	92
Proposed method **ICZ + ZCZ**	Nearest neighbor	**95**
Proposed method **ICZ + ZCZ**	Neural network	**94**

7. CONCLUSION

In this paper we have presented a hybrid type Zone based feature extraction algorithm for the recognition of four popular Indian numeral scripts. Nearest neighbor and Feed forward back-propagation neural network classifiers are used for classification and recognition. We have obtained 99% recognition rate for Kannada and Telugu numerals. For Tamil and Malayalam numerals we have obtained 96% and 95% respectively. Using zone based feature extraction, we have achieved good results even when certain preprocessing steps like filtering, smoothing and slant removing are not considered.

Our future work aims to improve classifier to achieve still better recognition rate and also to develop new zone based feature extractions algorithms, which provides efficient results. Also we plan to extend our work to other Indian numeral scripts. Also effective implementation of multiple classifier system is one of our future research directions.

REFERENCES

[1] R. Plamondon and S. N. Srihari, "On-line and off- line handwritten character recognition: A comprehensive survey", *IEEE. Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 1, pp. 63-84, 2000.

[2] Nafiz. Arica and 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, 2001.

[3] Liana M. Lorigo and Venu Govindaraju, "Offline Arabic handwriting recognition: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 5, pp. 712-724, 2006.

[4] G. Nagy, "Chinese character recognition, A twenty five years retrospective", *Proceedings of ICPR*, pp. 109-114, 1988.

[5] U. Pal, B. B. Chaudhuri, "Indian Script Character recognition: A survey", *Pattern Recognition*, vol. 37, pp. 1887-1899, 2004.

[6] Anil.K.Jain and Torfinn Taxt, "Feature extraction methods for character recognition-A Survey", *Pattern Recognition*, vol. 29, no. 4, pp. 641-662., 1996,

[7] A. Majumdar and B.B. Chaudhuri, "Printed and handwritten Bangla numeral recognition using multiple classifier outputs", *Proceedings of the first IEEE ICSIP06*, Vol. 1, pp. 190-195, 2006.

[8] Abdur Rahim, Shuvabranta Saha, Mahfuzur and Abdus sattar, "Recognition of conjunctive Bangla characters by artificial neural network", *International Conference on Information and Communication Technology*, pp. 96-99, 2007.

[9] U. Pal, N.Sharma, T. Wakabayashi and F. Kimura, "Off line handwritten character recognition of Devanagiri Scripts'', *Ninth International Conference on Document Analysis as Recognition (ICDAR 2007)*, pp.496-500, 2007.

[10] P.M. Patil, T.R. Sontakke, "Rotation scale and translation invariant handwritten Devanagiri numeral character recognition using fuzzy neural network", Elsevier, Pattern Recognition, vol. 40, pp. 2110-2117. 2007.

[11] U. Pal, T. Wakabayashi and F. Kimura, " A system for off-line Oriya handwritten character recognition using curvature feature ", *10th International Conference on Information Technology*, IEEE, pp.227-229, 2007.

[12] M. Hanmandlu, J. Grover, V. K. Madasu and S. Vasikarla, "Input fuzzy for the recognition of handwritten Hindi numeral:", *International Conference on Informational Technology*, vol. 2, pp. 208-213., 2007.

[14] Arun. K. Pujari, C. Dhanunjaya, Naidu, M. Sreenivasa, Rao and B.C. Jinaga, " An Intelligent character recognizer for Telugu scripts using multiresolution analysis and associative memory", Elsevier, *Image vision Computing*, pp.1221-1227, 2004.

[15] R.M.Suresh, S.Arumugam, "Fuzzy technique based recognition of handwritten characters", *Elsevier, Image Vision Computing*, Vol. 25, pp. 230-239, 2007

[16] Lajish. V. L, " Handwritten character using gray- scale based state-space parameters and class modular neural network ", *IEEE International Conference on signal processing, Comunication and Networking*, pp. 374-379, 2008.

[17] U. Pal, T. Wakabayashi and F. Kimura, "Handwritten numeral recognition of six popular scripts", *Ninth International conference on Document Analysis and Recognition ICDAR 07*, Vol.2, pp.749-753, 2007.

[18] G.G. Rajaput and Mallikarjun Hangarge, "Recognition of isolated handwritten Kannada numerals based on image fusion method: ", *PReMI07*, LNCS.4815, pp.153-160, 2007.

[19]S.V. Rajashekararadhya and P. Vanaja Ranjan, "Isolated handwritten Kannada digit recognition: A novel approach", *Proceedings of the International Conference on Cognition and Recognition"*, pp.134-140, 2008.

[20] S.V. Rajashekararadhya, P. Vanaja Ranjan and V.N. Manjunath Aradhya, "Isolated handwritten Kannada and Tamil numeral recognition: A novel approach", *First International Conference on Emerging Trends in Engineering and Technology ICETET 08*, pp.1192-1195, 2008.

[21] S.V. Rajashekararadhya, and P. Vanaja Ranjan ,"Handwritten numeral recognition of three popular South Indian scripts: A novel approach:", *Proceedings of the second International Conference on information processing ICIP*, pp.162-167, 2008.

[22] S.V. Rajashekararadhya, and P. Vanaja Ranjan ,"Neural network based handwritten numeral recognition of Kannada and Telugu scripts", TENCON 2008 Hyderabad – in press.

[23] Rafael C. Gonzalez, Richard E. woods and Steven L. Eddins, *Digital Image Processing using MATLAB*, Pearson Education, Dorling Kindersley, South Asia, 2004.