Research Journal of Applied Sciences, Engineering and Technology 12(1): 27-36, 2016
DOI:10.19026/rjaset.12.2300
ISSN: 2040-7459; e-ISSN: 2040-7467

Submitted: July 2, 2015 Accepted: August 15, 2015 Published: January 05, 2016

Research Article

Recognition of Multi-lingual Handwritten Numerals Using Partial Derivatives

[1]K.N. Saravanan and [2]R. Anitha,
[1]Christ University, Bangalore, Karnataka,
[2]Muthayammal Engineering College, Rasipuram, Tamilnadu, India

Abstract: The multi-font and multi-lingual handwritten numerals recognition has been a demanding requirement in this decade. This research work proposes multi-lingual handwritten numerals recognition using partial derivatives for classifying handwritten numerals of five major Indian languages. The objective of the proposed work aims at designing and developing a recognition algorithm for multilingual handwritten numerals. This objective is achieved through data collection and preprocessing which involves creation of handwritten numeral databases, data collection, round off mean aspect ratio value based representation and identification of features using partial derivatives. The features derived from partial derivatives are stored in a five dimensional column vector which yielded a recognition rate of 94.80, 95.89, 96.44, 95.81 and 92.03%, respectively for Kannada, Gurumukhi, Sindhi, Malayalam and Tamil Handwritten Numerals respectively.

Keywords: Binary numeral image, feature identification, multi-lingual handwritten numeral recognition, representation, single linkage clustering and distance metric

INTRODUCTION

Extraction of numerals from student answer scripts, identification of distances from traffic information board, recognition of numerals information from tabular forms are some of the applications of a numeral recognition system. Recent computer system and communication technologies such as software packages like word processors with multiple fonts and multi size fonts, sending and receiving electronic mail and sending messages through fax machine also have the impact of increasing the number of readers, literacy and the way of writing by the human beings. According to Plamondon and Srihari (2000) even with the evolution of technologies, the process of handwriting recognition is still challenging. Students' uses pen and paper to write the language content, equations and graphical drawings, but they are not using a note-book computer system. These systems usually have the keyboard and mouse as the interface for human-machine interaction. The limitation of the input devices like keyboard is that they have limited size of space to accommodate the symbols of a language.

For languages with small character set namely English which has 26 upper case alphabets and 26 lower case alphabets and 10 numerals, the usage of keyboard seems to be easy, whereas for languages with larger character set namely, Chinese (50000 characters), Japanese (35000 characters) and Tamil (246 characters), the usage of keyboard seems to be difficult. The Japanese language adopted from Chinese character set, uses Kanji alpha numeric which consists of 6349 characters (Tappert *et al*., 1990). These requirements lead to the generation of on-line recognition (electronic tablets) and off-line recognition system (optical recognition systems). The ultimate aim of these systems is to process handwritten data with arbitrary user written alphabets or symbols or any graphical marks'. For the purpose of recognizing the numeral in on-line, the handwritten data has to be converted into digital data by using a special pen for writing on an electronic surface.

In this system, the machine recognizes the handwritten data either while the user writes or at a later time. Some on-line recognizers are capable of even learning user writing samples of the writer, to adapt for subsequent recognition. In off-line recognition system, the machine recognizes the handwritten character after the writing has been completed. The recognition process can be performed days, months or even a year later. This system also facilitates interaction with the user without a keyboard and it would be of great help to the physically (visually) challenged people when interfaced with a text to voice synthesizer. The aim of the research work is to design and develop a recognition algorithm for multilingual handwritten numerals.

The major works reported in the literature about their used features for recognition, classification

Corresponding Author: K.N. Saravanan, Christ University, Bangalore, Karnataka, India

algorithms and recognition rate are presented in the following paragraphs.

Sung-Bae (1996) has normalized the size of the input image by 16-by-16 (which has later been compressed to 4-by-4 feature vectors and is used as global feature). This has resulted in 96.05% recognition rate. Sabaei and Faez (1997) have selected features such as Pseudo, Zernike and Legendre moments for recognition. For each Farsi numeral, 200 samples in the form of handwritten numerals had been collected from different persons. As the numerals 0, 4 and 6 have different shapes, the total number of digits for Farsi numerals is 13 and the best recognition rate achieved is about 95% when the moments of orders seems to be higher than five.

Elnagar *et al*. (1997) have proposed a method for recognizing handwritten numerals in Hindi based on structural descriptors. This process involves scanning the handwritten numeral and normalizing it to 30-by-30 pixels and then thinning it. In the second step, features like strokes and cavity have been extracted and these features have been represented syntactically. Finally, the syntactic representation of the feature is then matched against a stored set of syntactic representation prototypes and the recognition result of a maximum of 94% has been reported.

Sanossian (1998) has used three primitive features in a segment, namely, boundary distance, pixel density and line distance from centroid, to extract features in Hindi numerals and obtained an average recognition rate of 95.8%. Chen and Ng (1999) have proposed a crossing feature coding method to extract the features and a recognition accuracy of 91.3% for handwritten numerals and 99% for printed numerals has been obtained. Zhang *et al*. (2000) have extracted global features and fine segment features of handwritten numerals in eight directions and have obtained 98.5% recognition rate for handwritten numerals, using two hidden layer feed forward neural network as classifier in the recognition system. Al-Omari (2001) used the object's Centre of Gravity (COG) and angle of orientation as key features of the shape. The testing set involved only 20 numerals for each class and the rate of recognition was found to be 87.22%.

Haar Wavelets and discrete wavelet transform have been used by Mowlaei *et al*. (2002) for feature extraction and classification, which resulted in 91.81% recognition rate. Zhang *et al*. (2004) proposed a feature extraction method that is hybrid in nature for recognition of numerals. It is comprised of geometrical features, namely, end points, loops, joints, local segment features, middle line and convexity and coefficients of complex wavelet transformation that is two dimensional in nature and has resulted in 99.1% recognition rate.

Mozaffari *et al*. (2005) used the standard feature points for decomposing the numeral into its primitives. Principal Component Analysis has been used to obtain same sized global codes. Using Nearest Neighbour Classifier (NNC), 94.44% recognition rate has been achieved. Impedovo *et al*. (2006) proposed a new technique for zoning description in which best classification results have been achieved by partitioning the pattern image into M = 9 zones, if handwritten numeral digits were considered on the training sets. It has also been mentioned that the optimal zoning still outperforms the traditional zoning method on the testing sets as well.

Purkait and Chanda (2010) have proposed morphological opening and closing operation on pre-processed image and obtained 500 features. Structuring element (line) along the major, minor, vertical and horizontal directions have been used to get four different images. This method had yielded in 97.75% recognition accuracy.

Hossain *et al*. (2011) proposed rapid feature extraction method that computes the projection of each section which is formed by partitioning the image. This method resulted in 94.12% recognition accuracy using with probabilistic neural network, 94.10% accuracy using k-nearest neighbour classifier and 92.03% recognition accuracy using Feed forward back propagation neural network.

Majhi *et al*. (2011) have developed a classifier for Odiya handwritten numerals. Curvature and Image gradient features have been used to extract the features and the image has been normalized (64*64 pixels of height and width). Using these primitive features, the number of features obtained is 2,592 which has been reduced to 64, using Principal Component Analysis (PCA) technique. This feature extraction method yielded an accuracy of 98 and 94% for gradient features and curvature feature respectively.

Kartar *et al*. (2011) have used three different feature sets, namely, projection histogram, distance profile and Background Directional Distribution (BDD) resulting in the recognition rate of 99.2, 98 and 99.13%, respectively has been reported respectively, using SVM with Radial Basis Function (RBF) kernel classifier for Gurumukhi handwritten numerals, using only 150 samples. Mamatha *et al*. (2011) have used k-means clustering algorithm for classification and directional chain code method to extract the features from the resized image of size 30-by-30 pixels and obtained 96% recognition rate for Kannada numerals.

Roy *et al*. (2012) have proposed quad tree based feature set using SVM classifier and the recognition rate of 93.38% has been reported for Bangla numerals. Baheti and Kale (2013) have used affine invariant moments feature extraction approach for Gujarati numerals and the recognition rate of 94% has been reported using Support Vector Machine (SVM) as a classifier. Baheti and Kale (2013) have used zone based pixel density values as the features for Gujarati

numerals and the recognition rate of 94% has been reported using neural network as a classifier. Medhi and Kalita (2014) has proposed the features based on blobs or stems in the shape of the Assamese numerals. Decision tree has been used as a classification algorithm and has obtained 80% recognition rate.

Pirlo and Impedovo (2012) proposed an algorithm based on Voronoi diagrams and has achieved 94% accuracy for handwritten Latin numerals.

Reddy *et al*. (2012) have applied projection profiles as the primary features and normalization of individual numerals has been done using 64-by-64 pixels as image size. A total of 832 features have been used and an accuracy of 99.3% has been obtained for Off-line Assamese Language.

Bhattacharya and Chaudhuri (2009) have created two Indian scripts databases, namely, Bangla and Devanagari. After applying multilayer perceptron classifiers, the recognition accuracy of 98.2 and 99.04% has obtained for Bangla and Devanagari numerals respectively.

The objective of the proposed research work aims at, Development of handwritten numerals databases of five different Indian languages, namely, Kannada, Gurumukhi, Sindhi, Tamil and Malayalam, Finding round off mean aspect ratio value based representation scheme, Identification and extraction of features using partial derivatives and Recognition of multi lingual numerals using a distance metric.

HANDWRITTEN NUMERAL DATABASE COLLECTION FOR VARIOUS INDIAN LANGUAGES

In order to solve the problem of recognizing a pattern in which a feature vector and a classifier would be identified to automate the process of handwritten numeral recognition, it is essential to have the data that has been collected to represent the expected conditions in which they operate.

One of the challenges faced while doing recognition of handwritten numeral of various Indian languages is the non existence of standard databases. However, standard databases such as MNIST, Centre for Excellence for Document Analysis and Recognition (CEDAR) and Centre for Pattern Recognition and Machine Intelligence (CENPARMI) are available for Latin numerals. According to the Eighth Schedule to Indian Constitution, there are 22 official languages in India namely, Bengali, Assamese, Dogri, Bodo, Hindi, Kannada, Gujarati, Konkani, Kashmiri, Malayalam, Maithili, Nepali, Oriya, Santali, Marathi, Gurumukhi, Sanskrit, Tamil, Santhali, Urdu and Telugu and Urdu. Out of 22 official languages, the selected scripts for this research work are Kannada, Gurumukhi, Malayalam, Santali and Tamil.

Data collection: In the proposed work, samples handwritten were collected using a specialized tabular form. Each numeral of a particular language was written in a frame box of size of 1.5 cm×1.5 cm, on an A4 size white sheet with light gray background lines provided as separators. Each A4 size white sheet can accommodate (a maximum of 400 handwritten numerals) 25 square boxes in each column and 16 square boxes in each row and hence six white sheets have been used to collect the individual numerals. Based on the willingness of the candidate, the candidates were asked to write different numerals one or more times. The objective of the data collection had not been disclosed to the candidates who had been asked to write the numerals and was restricted to write only one numeral per square box.

Data sources: The numerals of various languages selected for these research works have been written by different categories of people namely, University and college students and research scholars. In a real life application, the type of pens used may be ink pen, gel pen or ball point pen and the colour of the ink may also vary. In the proposed research work, only blue and black colours of the ink were allowed to fill the tabular forms. The resolution used to scan the filled sheets is 300 dots per inch. Hewlett Packard scanner has been

Fig. 1: Specially designed tabular form for data collection (Kannada numeral 5)

Table 1: Representation of round off mean aspect ratio values in terms of zones for Kannada, Gurumukhi, Sindhi, Tamil and Malayalam

Languages	Number of training data for all the ten numerals	Mean aspect ratio value	Round off mean aspect ratio value	Representation of round off mean aspect ratio values in terms of zones	
				Y axis	X axis
Kannada	16200	0.93	0.9	9	10
Gurumukhi	16200	1.44	1.4	7	5
Sindhi	16200	1.38	1.4	7	5
Tamil	10800	0.94	0.9	9	10
Malayalam	14400	0.96	1.0	5	5

used for scanning and the scanned image has been stored as a binary image (Fig. 1).

Roundoff mean aspect ratio value based representation for binary numeral images: After performing image preprocessing and segmentation, image representation and feature selection of numerals play a vital role in a recognition system. Let L be a language and the numeral pattern classes of L be denoted by P_1, $P_2 \ldots P_W$, where W represents the number of pattern classes. Each pattern class P_i, is represented by a pattern vector *x*. In order to obtain 'n' common descriptor for each pattern class of a language, the round off mean aspect ratio value based zoning representation scheme has been proposed in this research work. Using the proposed representation scheme, the number of zones has been identified along both axes and the zone size would be defined for all the numerals in a language. One of the advantages of the proposed zoning representation scheme is that it could handle various handwritten styles of different writers without affecting the shape of the numerals.

Table 1 shows the language, the number of training data has been used to obtain mean aspect ratio value, mean aspect ratio value of a language, round off mean aspect ratio value and the zones along y and x axis.

Extraction of features using partial derivatives from the zones: In this research work, five primitive features using partial derivatives of a two dimensional numeral image have been identified and extracted based on the density matrix of the numerals. The density matrix of the numerals has been defined as the number of white pixels (text pixels) in a given zone. These primitive features are, zd(zd_xaxis+xincr, zd_yaxis) – zd(zd_xaxis, zd_yaxis), zd(zd_xaxis, zd_yaxis-yincr) – zd(zd_xaxis, zd_yaxis), zd (zd_xaxis + xincr,zd_yaxis + yincr) – zd(zd_xaxis, zd_yaxis), zd (zd_xaxis + xincr, zd_yaxis-yincr) -zd (zd_xaxis, zd_yaxis) and zd (zd_xaxis + xincr, zd_yaxis) + zd(zd_xaxis, zd_yaxis-yincr) + zd (zd_xaxis + xincr, zd_yaxis + yincr) + zd (zd_xaxis + xincr, zd_yaxis-yincr) – 4zd(zd_xaxis, zd_yaxis), where 'zd' stands for zone density of a numeral image and 'zd_xaxis' stands for zones x axis and 'zd_yaxis' stands for zone y axis. The values of the variable 'xincr' and 'yincr' are 1.

All the feature values are calculated and converted into absolute values. These feature values are stored for each numeral of a pattern class and the feature values are clustered based on single linkage algorithm. The mean feature value of numeral images in each cluster has been stored in the feature vector. The number of features obtained using these primitives could be varied based on the round off mean aspect ratio value of a language. For the language Kannada and Tamil, the number of zones along y axis is 9 and along the x axis is 10, so the number of features required is 395, whereas, the number of zones along y axis and x axis are 7 and 5 respectively for the languages Gurumukhi and Sindhi and the number of features is 141. Similarly, for the language Malayalam, the number of zones along both axes is 5 and hence the number of features is 97. Figure 2 to 7 shows the extraction of a feature value from Kannada numeral five.

Fig. 2: Kannada binary image (numeral 5)

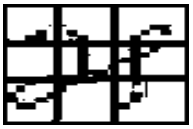

Fig. 3: Zone representation of the Kannada numeral 5 (3-by-3)

48	14	67
2	200	53
49	15	15

Fig. 4: Density matrix of the numeral in Fig. 3

-34	53	198	-143	-34	0

Fig. 5: Numeric value obtained using zd(x+1, y) – zd(x, y) feature

34	53	198	143	34	0

Fig. 6: Absolute value obtained using zd(x+1, y) – zd(x, y) feature

0.073	0.11	0.42	0.30	0.07	0

Fig. 7: Final feature value of the numeral image using zd(x+1, y) – zd(x, y) feature

Table 2: Number of clusters for Kannada numerals

S. No	Kannada language numerals	Number of training data	Number of clusters
1	೦	1620	1082
2	೧	1620	802
3	೨	1620	710
4	೩	1620	1006
5	೪	1620	1043
6	೫	1620	911
7	೬	1620	910
8	೭	1620	799
9	೮	1620	855
10	೯	1620	922

Table 3: Recognition accuracy obtained for all numerals in Kannada

Kannada numerals	೦	೧	೨	೩	೪	೫	೬	೭	೮	೯	Number of testing data	Recognition rate (%)
೦	580	0	0	0	0	0	0	0	0	0	580	100
೧	0	499	0	0	0	1	0	0	2	2	504	99.00
೨	0	0	252	2	0	0	0	0	0	0	254	99.22
೩	0	1	1	365	3	28	7	70	4	6	485	75.26
೪	0	0	0	0	624	4	1	0	0	8	637	97.96
೫	0	0	0	1	2	272	2	0	0	2	279	97.49
೬	0	0	0	1	1	0	596	5	0	32	635	93.85
೭	0	0	0	6	0	1	3	545	0	0	555	98.120
೮	0	0	0	0	0	2	0	0	430	0	432	99.547
೯	1	0	2	4	4	9	34	3	3	542	602	90.03

Training numerals algorithm:

Input : Numeral image for a language from the database

Output : Pattern vector stored in the library as a prototype

Step 1 : Enhance the binary numeral image

Step 2 : Resize the enhanced image based on the round off mean aspect ratio value of a language

Step 3 : Calculate the density matrix of the numeral based on the zones along two axes

Step 4 : Extract the features based on partial derivatives

Step 5 : Repeat the steps 1 to 4 to cluster the images using single linkage algorithm and store the class prototype in the library

End of the algorithm

Testing numerals algorithm:

Input : Single numeral image for a language from the database

Output : Classification of the numeral based on the language

Step 1 : Enhance the binary numeral image

Step 2 : Resize the enhanced image based on the round off mean aspect ratio value of a language

Step 3 : Calculate the density matrix of the numeral based on the zones along the axes

Step 4 : Extract the features based on partial derivatives.

Step 5 : Apply the distance metric between the generated features and the features stored in the library for a particular language and assign the numeral class to the nearest feature vector in the library

End of the Algorithm

EXPERIMENTAL RESULTS

The proposed feature extraction algorithm using partial derivates has been implemented using MATLAB version R2007b. Pentium Dual Core CPU, Processor of speed 2.0 GHz and 3.0 GB RAM have been used in this research work for carrying out the experiment. Testing of sample handwritten numerals has been done for the languages Kannada, Gurumukhi, Sindhi, Tamil and Malayalam and the results obtained have been discussed in the following sub sections.

Results and discussion for Kannada numerals: The samples used in this research work for Kannada language are 16,200 for training and 4,963 for testing. The number of clusters formed using the proposed training algorithm for the Kannada numerals 0 to 9 are listed in the Table 2. The confusion matrix obtained using the proposed algorithm for the Kannada numerals

0 to 9, the number of samples tested and their corresponding percentage of accuracy are listed in the Table 3.

Results and discussion for Gurumukhi numerals: The samples used in this research work for Gurumukhi language are 16,200 for training and 4,901 for testing. The number of clusters formed using the proposed training algorithm for the Gurumukhi numerals 0 to 9 are listed in the Table 4. The confusion matrix obtained using the proposed algorithm for the Gurumukhi numerals 0 to 9, the number of samples tested and their corresponding percentage of accuracy are listed in the Table 5.

Table 4: Number of clusters for Gurumukhi numerals

S. No	Gurumukhi language numerals	Number of training data	Number of clusters
1	੦	1620	1008
2	੧	1620	797
3	੨	1620	852
4	੩	1620	876
5	੪	1620	868
6	੫	1620	853
7	੬	1620	959
8	੭	1620	864
9	੮	1620	885
10	੯	1620	874

Table 5: Recognition accuracy obtained for all numerals in Gurumukhi

Gurumukhi numerals	੦	੧	੨	੩	੪	੫	੬	੭	੮	੯	Number of testing data	Recognition rate (%)
੦	580	0	0	0	0	0	0	0	0	0	580	100
੧	0	392	0	0	0	0	0	0	0	0	392	100
੨	0	2	245	5	1	0	0	0	0	0	253	96.84
੩	0	0	10	451	1	15	1	0	0	0	478	94.35
੪	0	3	0	1	613	9	0	2	1	1	630	97.30
੫	0	1	1	1	4	199	1	1	0	2	210	94.76
੬	0	0	3	0	2	2	642	0	3	23	675	95.11
੭	0	1	1	0	0	0	0	506	0	0	508	99.60
੮	0	0	4	0	1	1	2	0	496	18	522	95.01
੯	2	4	5	0	2	1	46	1	16	576	653	88.20

Table 6: Number of clusters for Sindhi numerals

S. No	Sindhi language numerals	Number of training data	Number of clusters
1	੦	1620	1008
2	੧	1620	797
3	੨	1620	836
4	੩	1620	1025
5	੪	1620	868
6	੫	1620	853
7	੬	1620	942
8	੭	1620	918
9	੮	1620	835
10	੯	1620	1022

Table 7: Recognition accuracy obtained for all numerals in Sindhi

Sindhi numerals	੦	੧	੨	੩	੪	੫	੬	੭	੮	v	Number of testing data	Recognition rate (%)
੦	580	0	0	0	0	0	0	0	0	0	580	100
੧	0	392	0	0	0	0	0	0	0	0	392	100
੨	0	7	599	1	0	4	2	0	0	1	614	97.56
੩	0	1	1	592	2	6	1	15	0	0	618	95.79
੪	0	3	1	0	612	9	0	2	1	2	630	97.14
੫	0	1	1	2	4	199	0	2	0	1	210	94.76
੬	0	1	0	2	1	3	625	0	0	4	636	98.27
੭	0	10	3	13	10	10	9	614	6	10	685	89.63
੮	0	0	3	0	0	2	0	0	502	1	508	98.81
੯	2	1	3	4	2	3	10	6	2	468	501	93.41

Results and discussion for Sindhi numerals: The samples used in this research work for Sindhi language are 16,200 for training and 5,374 for testing. The number of clusters formed using the proposed training algorithm for the Sindhi numerals 0 to 9 are listed in the Table 6. The confusion matrix obtained using the proposed algorithm for the Sindhi numerals 0 to 9, the number of samples tested and their corresponding percentage of accuracy are listed in the Table 7.

Results and discussion for Malayalam numerals: The samples used in this research work for Malayalam language are 14,400 for training and 6,474 for testing. The number of clusters formed using the proposed training algorithm for the Malayalam numerals 0 to 9 are listed in the Table 8. The confusion matrix obtained using the proposed algorithm for the Malayalam numerals 0 to 9, the number of samples tested and their corresponding percentage of accuracy are listed in the Table 9.

Results and discussion for Tamil numerals: The samples used in this research work for Tamil language are 10800 for training and 9,106 for testing. The number of clusters formed using the proposed training algorithm for the Tamil numerals 0 to 9 are listed in the

Table 8: Number of clusters for Malayalam numerals

S. No	Malayalam language numerals	Number of training data	Number of clusters
1	൦	1440	813
2	൧	1440	752
3	൨	1440	721
4	൩	1440	791
5	൪	1440	821
6	൫	1440	804
7	൬	1440	838
8	൭	1440	793
9	൮	1440	875
10	൯	1440	863

Table 9: Recognition accuracy obtained for all numerals in Malayalam

Malayalam numerals	൦	൧	൨	൩	൪	൫	൬	൭	൮	൯	Number of testing data	Recognition rate (%)
൦	756	0	0	0	0	3	0	1	0	0	760	99.47
൧	0	686	0	0	0	0	0	0	0	0	686	100
൨	0	0	694	0	0	0	0	0	0	1	695	99.85
൩	0	1	5	703	1	4	1	1	2	3	721	97.50
൪	0	6	9	1	616	1	4	2	14	0	653	94.33
൫	2	1	0	2	1	168	2	11	1	1	189	88.88
൬	0	0	0	3	1	7	746	11	2	4	774	96.38
൭	1	14	4	13	4	12	38	767	4	8	865	88.67
൮	0	0	1	3	6	1	5	5	402	0	423	95.03
൯	1	1	1	13	2	3	13	4	5	665	708	93.92

Table 10: Number of clusters for Tamil numerals

S. No	Tamil language numerals	Number of training data	Number of clusters
1	൦	1080	672
2	க	1080	594
3	உ	1080	534
4	ஈ	1080	436
5	சு	1080	700
6	ரு	1080	709
7	கூ	1080	606
8	எ	1080	569
9	அ	1080	603
10	சை	1080	719

Table 11: Recognition accuracy obtained for all numerals in Tamil

Tamil numerals	௦	க	உ	ங	ச	ஞ	ண	எ	அ	ன	Total testing data	Recognition rate (%)
௦	1119	0	0	0	0	0	0	0	0	1	1120	99.91
க	0	1192	0	1	5	5	5	0	5	34	1247	95.58
உ	0	2	1075	2	0	5	4	0	6	1	1095	98.17
ங	1	17	4	930	12	4	43	6	3	45	1065	87.32
ச	0	62	0	17	938	3	7	4	4	61	1096	85.58
ஞ	0	1	0	1	0	147	0	1	1	0	151	97.35
ண	0	21	0	10	9	2	398	0	0	71	511	77.88
எ	1	2	3	3	6	4	2	836	8	4	869	96.20
அ	0	2	2	1	4	2	5	6	915	0	937	97.65
ன	1	38	0	27	25	16	73	4	0	831	1015	81.87

Table 12: Proposed Algorithm with the existing algorithms (Comparision)

Name of the Author (s)	Languages	Extracted features	Classifier	Recognition accuracy
Rajashekararadhya and Ranjan (2009)	Kannada	Average angle calculated from character and zone centroids to the pixels	SVM	97.85%
Mamatha *et al*. (2011)		Directional features (chain code)	K-means	96%
Saravanan and Anitha (Proposed algorithm)	Kannada	Partial derivatives	Minimum distance classifier	94.80%
Kartar *et al*. (2011)	Gurumukhi	Distance Profiles	SVM	98%
Saravanan and Anitha (Proposed algorithm)	Gurumukhi	Partial derivatives	Minimum distance classifier	95.89%
Saravanan and Anitha (Proposed algorithm)	Sindhi	Partial derivatives	Minimum distance classifier	96.44%
Rajashekararadhya and Ranjan (2009)	Malayalam	Average angle calculated from character and zone centroids to the pixels	SVM	95%
Saravanan and Anitha (Proposed algorithm)	Malayalam	Partial derivatives	Minimum distance classifier	95.81%
Rajashekararadhya and Ranjan (2009)	Tamil	Average angle calculated from character and zone centroids to the pixels	SVM	95.1%
Saravanan and Anitha (Proposed algorithm)	Tamil	Partial derivatives	Minimum distance classifier	92.03%

Table 10. The confusion matrix obtained using the proposed algorithm for the Tamil numerals 0 to 9, the number of samples tested and their corresponding percentage of accuracy are listed in the Table 11.

The proposed algorithm for multilingual recognition has been compared with the existing algorithms in the literature (Table 12).

When compared to the existing algorithms, the proposed algorithm is found to give better recognition rate for Malayalam. But this is not the case for other languages as their recognition rates are found to be less when the proposed algorithm is used. The reason is that most of the proposed algorithm has been designed to recognize multilingual numerals whereas the existing algorithms have been designed to recognize numerals of a specific language only. Hence, as a first step, the objective to recognize multilingual handwritten numerals has been successfully achieved and in future, enhancement need to be done so as to improve the recognition accuracy of proposed algorithm against the existing algorithms.

CONCLUSION

The objective of the proposed work is to recognize multilingual handwritten numerals. The languages used for experimentation are Kannada, Gurumukhi, Sindhi, Malayalam and Tamil and the recognition accuracy of 94.80, 95.89, 96.44, 95.81 and 92.03%, respectively has been obtained using the proposed algorithm.

LIMITATIONS

The handwritten numerals must be single connected which means that they should not be broken or fragmented. In such a case, the algorithm treats the broken numerals as different number of numerals based upon the number of fragments. This is one of the most important limitations.

SCOPE FOR FUTURE ENHANCEMENT

The zone size applied in the proposed representation scheme is 16-by-16. The zone size could

be applied with various values and an optimum value of zone size could be identified. Single Linkage clustering algorithm and Euclidean distance metric has been applied for clustering data, for which an average of 42% of the sample data are required as prototypes. This percentage can be reduced using other clustering and distance metrics.

ACKNOWLEDGMENT

Our sincere thanks to Dr. Nangegowda and Prof. Joy Paulose of Christ University, Bangalore, India for their encouragement and constant support throughout this research work.

REFERENCES

Al-Omari, F., 2001. Handwritten Indian numeral recognition system using template matching approaches. Proceeding of the ACS/IEEE International Conference on Computer Systems and Applications. Beirut, pp: 83-88.

Baheti, M.J. and K.V. Kale, 2013. Recognition of Gujarati numerals using hybrid approach and neural networks. Proceeding of International Conference on Recent Trends in Engineering and Technology (ICRTET, 2013), 5: 12-17.

Bhattacharya, U. and B.B. Chaudhuri, 2009. Handwritten numeral databases of Indian scripts and multistage recognition of mixed numerals. IEEE T. Pattern Anal., 31(3): 444-455.

Chen, M.W. and M.H. Ng, 1999. Recognition of Unconstrained Handwritten Numerals Using Crossing Features. Proceeding of the 15th International Symposium on Signal Processing and Its Applications (ISSPA '99), 1: 283-288.

Elnagar, A., F. Al-Kharousi and S. Harous, 1997. Recognition of handwritten Hindi numerals using structural descriptors. Proceeding of the IEEE International Conference on Systems, Man and Cybernetics, Computational Cybernetics and Simulation. Orlando, FL, 2: 983-988.

Hossain, M.Z., M.A. Amin and Y. Hong, 2011. Rapid feature extraction for Bangla handwritten digit recognition. Proceeding of the International Conference on Machine Learning and Cybernatics (ICMLC, 2011). Guilin, 4: 1832-1837.

Impedovo, S., M.G. Lucchese and G. Pirlo, 2006. Optimal zoning design by genetic algorithms. IEEE T. Syst. Man Cyb., 36(5): 833-846.

Kartar, S.S., D. Renu and R. Rajneesh, 2011. Handwritten Gurmukhi numeral recognition using different feature sets. Int. J. Comput. Appl., 29(2): 20-24.

Majhi, B., J. Satpathy and M. Rout, 2011. Efficient Recognition of Odiya Numerals using Low complexity neural classifier. Proceeding of the International Conference on Energy, Automation and Signal (ICEAS, 2011), pp: 140-143.

Mamatha, H.R., K.S. Murthy, A.V. Veeksha, P.S. Vokuda and M. Lakshmi, 2011. Recognition of handwritten kannada numerals using directional features and K-means. Proceeding of the International Conference on Computational Intelligence and Communication Network (CICN, 2011). Gwalior, pp: 644-647.

Medhi, K. and S.K. Kalita, 2014. Recognition of assamese handwritten numerals using mathematical morphology. Proceeding of the IEEE International Advance Computing Conference (IACC, 2014). Gurgaon, pp: 1076-1080.

Mowlaei, A., K. Faez and A.T. Haghighat, 2002. Feature extraction with wavelet transform for recognition of isolated handwritten Farsi/Arabic characters and numerals. Proceeding of the 14th International Conference on Digital Signal Processing (DSP, 2002), 2: 923-926.

Mozaffari, S., K. Faez and M. Ziaratban, 2005. Structural decomposition and statistical description of Farsi/Arabic handwritten numeric characters. Proceeding of the 8th International Conference on Document Analysis and Recognition, 1: 237-241.

Pirlo, G. and D. Impedovo, 2012. Voronoi-based zoning design by multi-objective genetic optimization. Proceeding of the 10th IAPR International Workshop on Document Analysis Systems (DAS). Gold Cost, QLD, pp: 220-224.

Plamondon, R. and S.N. Srihari, 2000. Online and off-line handwriting recognition: A comprehensive survey. IEEE T. Pattern Anal., 22(1): 63-84.

Purkait, P. and B. Chanda, 2010. Off-line recognition of hand-written Bengali numerals using morphological features. Proceeding of the 12th International Conference on Frontiers in Handwriting Recognition. Kolkata, India, pp: 363-368.

Rajashekararadhya, S.V. and P.V. Ranjan, 2009. Support vector machine based handwritten numeral recognition of Kannada script. Proceeding of the IEEE International Advance Computing Conference, pp: 381-386.

Reddy, G.S., P. Sharma, S.R.M. Prasanna, C. Mahanta and L.N. Sharma, 2012. Combined online and offline assamese handwritten numeral recognizer. Proceeding of the National Conference on Communications (NCC, 2012). Kharagpur, pp: 1-5.

Roy, A., N. Mazumder, N. Das, R. Sarkar, S. Basu and M. Nasipuri, 2012. A new quad tree based feature set for recognition of handwritten Bangla numerals. Proceeding of the IEEE International Conference on Engineering Education: Innovative Practices Future Trends (AICERA, 2012). Kottayam, pp: 1-6.

Sabaei, M. and K. Faez, 1997. Unsupervised classification of handwritten Farsi numerals using evolution strategies. Proceeding of IEEE Region 10

Annual Conference. Speech and Image Technologies for Computing and Telecommunications (TENCON '97). Brisbane, Qld., Australia, 1: 403-406.

Sanossian, H., 1998. Feature extraction technique for Hindi numerals. Proceedings of the IEEE Signal Processing Society Workshop Neural Networks for Signal Processing VIII. Cambridge, pp: 524-530.

Sung-Bae, C., 1996. Recognition of unconstrained handwritten numerals by doubly self-organizing neural network. Proceeding of the 13th International Conference on Pattern Recognition, 4: 426-430.

Tappert, C.C., C.Y. Suen and T. Wakahara, 1990. The state of the art in online handwriting recognition. IEEE T. Pattern Anal., 12(8): 787-808.

Zhang, P., L. Chen and A.C. Kot, 2000. A floating feature detector for handwritten numeral recognition. Proceeding of the 15th International Conference on Pattern Recognition. Barcelona, 2: 553-556.

Zhang, P., T.D. Bui and C.Y. Suen, 2004. Extraction of hybrid complex wavelet features for the verification of handwritten numerals. Proceeding of 9th International Workshop on Frontiers in Handwritten Recognition, pp: 347-352.