ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

Basic Kannada Handwritten Character Recognition System using Shape Based and Transform Domain Features

Shashikala Parameshwarappa[1], B.V.Dhandra[2]

Assistant Professor, Department of Computer Science and Engineering College, Govt.Engg.College, Raichur, Karnataka, India[1]

Professor& Chairman, Department of P.G.Studies and Research in Computer Science, Gulbarga University Kalaburagi, Karnataka, India[2]

Abstract: This paper presents the recognition of basic Kannada handwritten Characters based on shape features such as normalized chain code and transform domain based features as wavelet filters. In Kannada character set some characters are similar in shape; Images are made noise free by median filter and images are normalized into 64x64 dimensions. A normalized chain code and two-dimensional discrete wavelet transforms are proposed to extract as a feature vector of size 22 from the normalized binary images of size 64x64. Experiments are performed on handwritten Kannada characters consisting of 9600 images with 200 samples for each character per class. The total number of classes is 48. The SVM classifier with 2 fold cross validation is applied for classification of handwritten basic Kannada characters .These features are fed to the SVM classifier for classification of character images. To test the performance of the proposed algorithm two fold cross validation is used. The average recognition accuracy of 90.09% is obtained for Kannada vowels and consonants respectively. The experiment has given encouraging results with less number of feature set. The proposed algorithm is independent of thinning.

Keywords: Basic Kannada character Recognition; Normalized chain code, wavelet transform features; SVM classifier.

1. INTRODUCTION

Handwritten character Recognition is one of the important areas in pattern recognition field because it provides solution for document classifications, mail processing, automatic data entry, bank check reading, reading of the customer filled forms and manymore. Advancement of e-technology has made the revolution on all fields in general and document automation in particular. This revolution made to develop an OCR system for every languages and scripts for printed and hand printed documents to process automatically. Most of the works related to handwritten character recognition are done in English, Chinese, Japanese and Arabic. The task is more complicated for Indian languages due to complexity involved in the shape and similarities of the characters. The brief summary of the literature is presented below. Several feature extraction techniques are found in the literature for Kannada character recognition. The feature extraction techniques are spatial features, Fourier and shape descriptors, Normalized chain code, Invariant moments, central moments, Zernike moments, modified invariant moments, structural, Statistical, Topological, Template Matching, Gabor and Zoning features and their combinations and so on.

Rakesh Rampalli et al. [1] have proposed fusion of complementary online and offline strategies for recognition of Kannada characters and reported the recognition accuracy of 89.7% with 295 classes.

Niranjan et al. [2] have proposed Fisher Linear Discriminate Analysis for unconstrained handwritten Kannada character recognition and reported the recognition accuracy of 57% using angle distance measures. Ragha et al. [3] have used moment based features for recognition of Kagunita (the Kannada compound characters resulting from the consonant and vowel combination).These features are extracted using Gabor wavelets from the dynamically pre-processed original image. Multi-Layer Perception with Back propagation Neural Networks are employed for character classification. Average recognition rate of 86% is reported for vowels and 65% for consonants. Aradhya et al. [4] have proposed Fourier transform and principal component analysis technique for handwritten vowels and consonants of Kannada character recognition and achieved the recognition accuracy of 68.89%. Dhandra et al. [5] have used zone based pixel density feature set of size 64 and achieved the 73.33% recognition accuracy for Kannada consonants using SVM classifier. Sanjeev Kunte et al. [6] have proposed an OCR system for the recognition of basic characters of printed Kannada text, which works for different font size and font style. Each image is characterized by using Hu's invariant and Zernike moments. They have achieved the recognition accuracy of 96.8% with Neural Network classifier. G.G.Rajput et al. [7] have proposed Zone based method for recognition of

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

handwritten Kannada characters. A feature vector of size 512 is obtained for each character. A multi-class SVM is used for the classification purpose and yielded 87.24%recognition accuracy. S.A.Angadi et al. [8] have used structural features for recognition of handwritten is used Kannada vowels and consonants. They have achieved the recognition accuracy of 89.84 % and 85.14% for handwritten Kannada vowels and Consonants obtained using SVM classifier. N.Sharma et al. [9] have considered the directional chain code information of the contour points of the characters as features set for recognition of the handwritten Devanagari numerals and characters by using quadratic classifier. Their recognition accuracy is 98.86% and 80.36% respectively. G.Raju et al [14] have proposed wavelet packet as feature set for recognition of handwritten Malayalam (one of the south Indian languages) characters. Feed forward neural network architecture for classification and obtained the recognition accuracy of 90%.

Hence, from the above it is clear that the algorithms designed for Kannada characters recognitions suffers either from the recognition accuracy or from time and space complexity due to similar shape of the Kannada characters. From the literature survey, it is evident that still handwritten character/numeral recognition is a fascinating area of research to design a single optical character recognition (OCR) system

Hence, there is a need to develop an efficient algorithm to recognize the Kannada characters effectively with minimum number of features. The observation on characters reveals that the structures of Kannada characters are circles, holes and curvature in nature. This observation made us to think in the proposed study an attempt is made to use the Normalized chain code and wavelet based features for the recognition of hand written basic Kannada characters.

The data collection and pre-processing methods are presented in Section 2. Section 3 is devoted for feature extraction method and designing of the proposed algorithm for hand written Kannada Vowels and Kannada consonant character recognition system. The experimental results, discussion and comparative analysis are presented in Section 4. Conclusion is presented in Section 5.

2. DATA COLLECTION AND PREPROCESSING

It is observed that, to validate and verify the results of the proposed algorithm the standard databases for handwritten Kannada characters are not available in the literature. Hence, the data collection is made and created our own database. Totally 2800 Kannada vowels and 6800 Kannada consonant images are collected from the various groups of people belonging to Primary Schools, High Schools and Colleges.

This collected data set is digitized by scanner on a flat bed HP scanner at 300 dpi which usually yields a low noise and good quality document image. The images are cropped up manually and stored as gray scale images. Binarization of image is performed using Otsu's global thresholding method and is stored in bmp file format. The digitized images typically contain noise due to erratic hand movements and inaccuracies in digitization of the actual input. The noises present in the images are removed by applying median filter. Then a minimum bounding box is fitted to the isolated character. To bring uniformity among the character images the cropped image is normalized to 64x64 pixels. A Sample image of the handwritten Kannada Vowels and Kannada consonant is shown in Fig. 1 and Fig. 2.

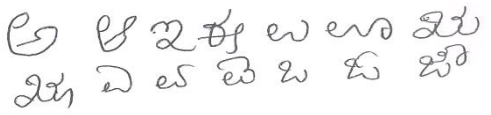

Fig. 1: Handwritten Kannada Vowels

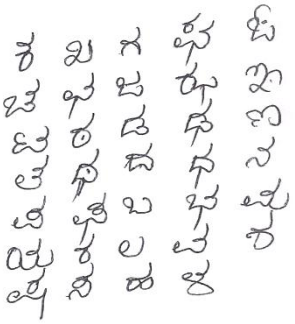

Fig. 2: Handwritten Kannada Consonants

3. FEATURE EXTRACTION

The feature extraction is described about the characteristics of an image. It is one of the most important components for any recognition system, since the classification/recognition accuracy is depending on the features. Two well known feature extraction methods 1&2. Those are Normalized chain code and wavelet decomposition for handwritten basic Kannada characters recognition system is proposed. A brief description about Normalized chain code and wavelet decomposition is given below.

3.1. Chain Code.

Chain codes are the standard input format for numerous shape analyses and are used to represent the boundary of an object by a connected sequence of straight line segments in specified direction. Chain code representation provides compact form of object representation (dimensionality reduction) and is translation invariant. The object is traversed clockwise or anticlockwise manner. Chain codes are one of the shape representations which are used to represent a boundary is based on 4-connectivity or 8-connectivity of the segments [10]. The direction of each segment is coded by using a numbering scheme as shown in Fig 3. According to [11], chain code is generated by a boundary of an object in a anticlockwise direction and assigning a direction to the segments connecting every pair of pixels.

DOI 10.17148/IJARCCE.2015.47113

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

Chain Code method's work as follows:

Start

1. Choose a starting pixel anywhere from the object boundary. There must be an adjoining boundary pixel at one of the eight locations surrounding the current boundary pixel as shown in Figure 3.
2. If the pixel found is located at the right of the current location or pixel, a code "0" is assigned.
3. If the pixel found is directly to the upper at the right, a code "1" is assigned. Similarly code "2" & code "3" is assigned as shown in Fig.3a.
4. The process of locating the next boundary pixel and assigning a code is repeated until we came back to our first location.

End.

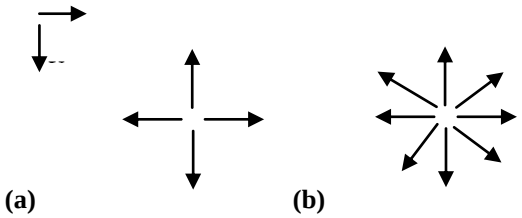

Fig. 3 Direction numbers for (a) 4-directional chain codes, (b) 8-directional chain code

The process of finding the boundary pixel and assigning a code is shown in Fig 4.

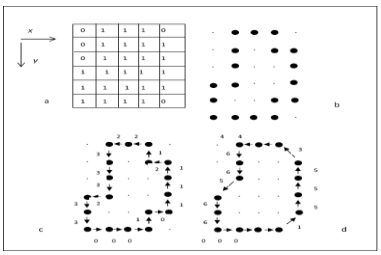

Fig. 4 (a)&(b) A 4-connected object and its boundary; c & d) Obtaining the chain code from the object in (a & b) with (c) for 4- connected and (d) for 8-connected

It is observed that the chain code for different characters has different length code and length of each chain code depends on the size of the character image. More ever length of chain code is very high in case of certain character image. We have solved this problem by normalizing the chain code values as explained below. The following chain code is generated for Kannada Vowel by traversing it in anticlockwise direction

V1= [1 2 2 0 0 0 0 0 0 0 6 7 6 6 7 6 6 6 6 6 7 6 6 6 6 5 6 6 6 5 5 6 5 4 4 5 4 5 4 5 4 4 4 5 4 4 4 4 3 4 4 4 3 4 4 2 3 3 4 3 4 4 3 2 3 3 2 3 1 1 1 1 4 4 4 6 6 6 6 2 2 2 2 4 4 2 3 2 3 2 2 2 2 2 2 2 2 2 1 1 1 4 7 7 7 3 3 2 2 1 1 5 5 5]

The frequency of the codes 0, 1, 2…..7. for vector V1 is given below in V2.

V2= [7 10 23 13 25 10 21 6].

The normalized frequency, represented by vector V3, is computed by using the formula

$$V3 = V2 / \left| V1 \right| \text{ where } \left| V1 \right| = \sum V2$$

The resultant normalized frequency of the chain code for V3 is

V3= [0.06087 0.08696 0.20000 0.11304 0.21739 0.08696 0.18261 0.05217]

Hence, the desired feature vector of size 08 is V3. The algorithm for computing normalized chain codes is given section 3.3

3.2. Wavelet Decomposition

Wavelets transform results in sub images corresponding to smooth component in the three directions horizontal ,vertical and diagonal .The information content of smooth component and high pass filtered components in the three directions should unique feature of an image. This feature can be characterized with number of zero crossing of wavelet coefficient in each sub image. The count of zero-crossing in all sub images together shall be used as a character feature. The algorithm for computing Feature extraction is given below.

Training Phase:

3.3 Algorithm: Feature Extraction

Method:

Input: pre-processed Isolated Basic Kannada Character image.

Output: Feature library.

Step 1.

Start:

1.1 Trace the boundary in counterclockwise Direction and generate 8 dimensional chain codes 0 to 7 (Fig.4).

1.2 Compute the frequency of the codes 0 to 7.

1.3 Divide frequency of each code by sum of the frequencies. To obtain feature vector of length 08.

1.4 Repeat the step 1.1 to step 1.3 for all sample images representing Basic Kannada Character considered for training.

Step 2.

2.1 Apply two level forward wavelet packet transform using db4 filter.

2.2 For each sub band count the number of zero crossing row wise, column wise, daignal wise.

2.3 Obtain the feature vector of length 14.

2.4 Repeat the step 2.1 to step 2.3 for all sample images representing Kannada character considered for training.

Step 3.

Store the computed normalized chain code from step 1 and wavelet decomposition coefficient from step 2 as the features stored in train library in the database.

End.

The eight features of chain code and fourteen features of wavelet decomposition coefficients are given to the SVM classifier for classification.

Testing Phase:

3.4 Algorithm: Recognition of basic Kannada Characters

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

Input: Isolated test Kannada image.
Output: Recognition of the Basic Kannada Character.
Start:
Step l**:** Extract the features by using Algorithm 3.3.
Step 2**:** Compute the distance between the feature vectors of the test image and with the feature vector of the trained image stored in knowledge base.
Step 3: Minimum distance computed in the above step 2 is the recognized Basic Kannada character.
End.

4. EXPERIMENTAL RESULTS

The proposed method is implemented using Intel Core 2 Quad processor @ 2.66 GHz machine and MATLAB 2012b.The proposed algorithm is executed on a database of 2800 Kannada vowels and 6800 isolated handwritten Kannada consonants images, with 200 images representing each character per class. Considering 50% for training and 50% for testing. For measuring the performance of an algorithm all pre processed images are normalized to size 64x64 and experiment is carried out using normalized chain codes and transform domain based features as wavelet filters features on pre processed images. A total of 9600 character images of Kannada characters are classified using SVM classifier. The performance of a proposed algorithm is tested using 2-fold cross validation. The average recognition rate for basic Kannada character is 90.07%.The misclassification is mainly due to characters that are similar in shape. The results exhibited in the table 1 are encouraging for the handwritten Kannada vowels and consonants recognition.

Table-1: Percentage of Recognition Accuracy for Handwritten Kannada Vowels and Consonants with SVM Classifier with 2- fold cross Validation

Training Samples=4800, Test samples=4800 Number of Features = 22			
Handwritten Kannada Vowels and Consonants	**No. of samples Trained**	**No. of samples Test**	**Percentage of Recognition Accuracy with SVM Classifier**
	100	100	89
	100	100	90.0
	100	100	93.7
	100	100	94.6
	100	100	90.2
	100	100	89.2
	100	100	90.0
	100	100	93.4
	100	100	92.1
	100	100	86.7
	100	100	92.0
	100	100	92.5
	100	100	83.2
	100	100	86.1
	100	100	82.1
	100	100	88,6
	100	100	91.0
	100	100	88.0
	100	100	94.1
	100	100	86.0
	100	100	90.0
	100	100	92.7
	100	100	95.2
	100	100	93.8
	100	100	89.0
	100	100	84.5
	100	100	94.0
	100	100	92.0
	100	100	92.3
	100	100	90.0
	100	100	86.0
	100	100	90.1
	100	100	90.0
	100	100	90.1
	100	100	86.0
	100	100	94.0
	100	100	90.4
	100	100	86.0
	100	100	94.8
	100	100	92.8
	100	100	88.5
	100	100	98.0
	100	100	90.2
	100	100	90.2
	100	100	90.0
	100	100	85.0
	100	100	86.4
	100	100	89.0
Average Percentage of Recognition Accuracy			**90.09**

From the Table-1author can conclude that the possibility of misclassification is higher among the vowels and consonants that have the similar shape mentioned in below

Figure 4.For example the vowel is recognized as 8 times and is recognized as 6 times.

Fig 4. Some similar shaped handwritten vowels and consonants

A comparative study of the proposed method with three existing methods proposed by Aradhya M and Niranjana S.K [4],.Rajput [7]et al. and S.A.Angadi [8] et al. with respect to Character recognition rate, type of features, classifier used is presented in Table-2

DOI 10.17148/IJARCCE.2015.47113

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 4, Issue 7, July 2015

Table -2: Comparative Analysis of Handwritten Kannada Vowels and Consonants with other existing methods

Authors	**Features computed& Dimension**	**Classifier used**	**Character Recognition Rate**
Aradhya M And Niranjanas.k[4]	Fourier Transform & PCA features.	PNN classifier.	68.89%
G.G.Rajput[7]	Zone based crack code[512]	Multi class SVM classifier	87.24%
S.A.Angadi[8]	Structual features (44)	SVM classifier	82.49%
Proposed method	Normalised Chain code and Wavelet packet (22)	SVM classifier.	90.09%

5. CONCLUSION

An algorithm proposed here for recognition of handwritten Kannada vowels and Kannada consonants using Normalized chain code and wavelet transform based features. The proposed algorithm gives encouraging results to meet the objectives of handwritten Kannada OCR system compared with other existing methods. An algorithm has exhibited the average percentage of recognition accuracy as 90.09% with SVM classifier with 2-fold cross validation. The aim of the proposed system is to remove the confusions among similar shape characters. The novelty of the proposed method is free from thinning.

In future, author will extend it for south Indian character recognition system. The extension of the proposed work is to meet the objective of recognition of similar shaped Kannada characters. There is a scope for improving the methodology by using better features, which will be explored, in future works to improve the recognition rate and test with other classifiers.

REFFERNCES

1. Rampalli R., Ramkrishnan, Angarai G., "Fusion of Complementary Online and offline Strategies for recognition of Handwritten Kannada Characters "Journal of Universal Computer Science, 17(1) . pp 81-93. 2011
2. Niranjan S. K, Vijaya Kumar, Hemantha Kumar "unconstrained handwritten Kannada character recognition" International Journal of Database Theory and Application, Vol.2, No. 4, pp 290-301,2009
3. Ragha, L. R., Sasikumar, M.: "Feature Analysis for Handwritten Kannada Kagunita Recognition ". International Journal of Computer Theory and Engineering, IAC-SIT 3(1), pp. 1793-8201 ,2011
4. Aradhya M.,Niranjana S.K.,Hemantha kumar G., "Probabilistic Neural Network based Approach for Handwritten Character Recognition" Special Issue of IJCCT, Vol.1 Issue 2,3,4 pp.9-13,2010
5. B.V. Dhandra, Mallikarjun Hangarge and Gururaj Mukarambi. "A Zone Based Character Recognition Engine for Kannada and English Scripts". Elsevier Science Direct, pp. 3292-3299. 2012
6. Kunte Sanjeev R., Sudhaker Samuel. "A simple and efficient optical character recognition system for basic symbols in printed Kannada text". Sadhana, Vol. 32, Part 5, pp. 521-533. ,2006
7. G.G. Rajput,Rajeswari Horakeri , " Zone based Handwritten Character Recognition using crack code and SVM " ,International Conference on advances in Computing, Communication and Informatics(ICACC)-2013
8. S.A.Angadi and Sharanabasavaraj.H.Angadi "structural features for recognition of hand written kannada character based on svm" International Journal of ComputerScience, Engineering and Information Technology (IJCSEIT), Vol. 5,No.2, ,pp.25-32 ,April 2015
9. N. Sharma, U. Pal, F. Kimura, and S. Pal, "Recognition of Off-Line Handwritten Devanagari Characters Using Quadratic Classifier," ICVGIP 2006, LNCS 4338, pp. 805 –816, 2006.
10. Gonzales R.C and Woods, R. E. "Digital Image Processing "2nd Ed. Upper Saddle River, N. J, : Prentice- Hall, Inc. (2002)
11. H. Freeman, Computer Processing of Line Drawings, Computing Surveys, Vol. 6, 57-97
12. Dinesh Acharya U., N.V. Subba Reddy and Krishnamoorthi, " Hierarchical Recognition System for Machine Printed Kannada Characters. IJCSNS LNCS International Journal of Computer Science and Network Security Vol. 8 No. 11, pp. 44-53. 2008
13. S.Correia , J.Mcarvallo and R. Sabourin, " On the performance of wavelets for Handwritten Numerals Recognition ", In ICPR 2002, quebac-canada August 2002
14. G.Raju,K.Revathy, " wave packets in the recognition of Isolated handwritten Malalam Characters" , Proceedings of the world Congress on Engineering Vol 1, July 24, 2007
15. Ashwin T.V., Sastry P.S. "A Fonts and Size-Independent OCR System for Printed Kannada Documents Using Support Vector Machines". Sadhana, 27: pp. 35-58.,2002
16. J.L. Starck, E.J. Cand`es, and D.L. Donoho "The Curvelet Transform for Image Denoising", IEEE Trans. Im. Proc., Vol. 11, No. 6, pp. 670-684, 2002
17. Gonzales R.C. and Woods, R.E.. Digital Image Processing 2nd Ed. Upper Saddle River, N.J.: Prentice- Hall, Inc. pp. 261-269, 2002
18. E.J. Candès, L. Demanet, D.L. Donoho, L. Ying. "Fast Discrete Curvelet Transforms". Multiscale Model. Simul., pp. 861-899. 2003
19. Nagabhushan P., Angadi S.A., Anami B.S, .". A Fuzzy Statistical Approach to Kannada Vowel Recognition based on Invariant Moments," Proceedings of NCDAR -2003, PESCE, Mandya, pp. 275-285. 2003
20. U. Pal, B.B. Chaudhuri , "Indian Script Character Recognition: A Survey. Pattern Recognition, 37" (2004), pp. 1887–1899. 2004.
21. R. Sanjeev Kunte, R.D. Sudhaker Samuel ,". An OCR System for Printed Kannada Text Using Two-stage Multi-network Classification Approach Employing Wavelet Features," International Conference on Computational Intelligence and Multimedia Applications, pp. 349-355. 2007
22. M.J. Fadili and J.L. Starck ," Curvelets and Ridgelets, Encyclopedia of Complexity and System Science", pp. 1-29 ,2007
23. Karthik Sheshadri, Pavan Kumar, T. Ambekar, Deeksha Padma Prasad and Dr. Ramakanth P. Kumar (2010). An OCR System for Printed Kannada using K-Means Clustering. IEEE International Conference on Industrial Technology (ICIT), pp. 183-187.
24. Kumar,S.Kumar, A , Kalyan, S.:(2010)"Kannada Character Recognition System using Neural Network". National Journal on Internet Computing. pp. 33-35