Handwritten Kannada Characters Recognition using Curvelet Transform

Shashikala Parameshwarppa
Department of Computer Science and Engineering,
Govt. Engg. College, Raichur
Karnataka, India.

B.V. Dhandra
Department of P.G. Studies and Research in Computer Science
Gulbarga University Gulbarga,
Karnataka, India.

ABSTRACT
The Selection of a feature extraction method for recognition of an object/character is probably the single most factors in achieving high recognition accuracy. Therefore, in this paper an effort is made to identify the Second Generation Discrete Curvelet Transform (DCTG2) as the potential features for recognition of handwritten Kannada character system .Images are made noise free by median filter and images are normalized into 64x64 pixels. Curvelet transform with different scales are applied to the input images to generate the curvelet coefficients .Then the standard deviation are computed for the curvelet coefficients to form feature vector of size 20.The total of 2800 Kannada vowels and 6800 handwritten Kannada consonants of sample images are used for classification based on the KNN classifier. To test the performance of the proposed algorithm two fold cross validation is used. The average recognition accuracy of 90.57% is obtained for handwritten basic Kannada characters respectively. The proposed algorithm is independent of thinning and skew of the character images.

Keywords
Kannada character Recognition, Curvelets, Standard deviation, KNN classifier.

1. INTRODUCTION
The optical Character Recognition plays an important role in case of document image analysis. Still the document image analysis is one of the active fields in research for Computer Scientists worldwide due to its real life applications such as mail processing, automatic data entry, bank check reading, reading of the customer filled forms and many more. 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. Many researchers have developed the character recognition systems by using template matching, spatial features, Fourier and shape descriptors, Normalized chain code, Invariant moments, central moments, Zernike moments , modified invariant moments, structural , statistical, Topological,Gabor,Zoning features combinations of these feature etc. A survey on different feature extraction methods for character recognition is reported in [1]. Different pattern classifiers like neural networks, Hidden Markov models, and Fuzzy and SVM classifiers are used. Details of these methods can be found in [2]. Any recognition system of hand printed character must address the problem of unconstrained shapes, variation in writing style, intensity strokes, scale and orientations. To simplify the task, often unconstrained character recognition systems are designed [3]. Extensive work has been carried out for recognition of characters in foreign languages like English, Chinese, Arabic. Lot of contributions can be found for printed characters compared to handwritten characters. A brief review of work done related to recognition of India script is presented below.

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%.For Kannada and English character recognition, 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. Dhandra et al. [7] have used Discrete Curvelet Transforms as feature vector for bilingual and Trilingual script (Kannada,English and telgu) Identification and reported 94.19%, 95.24% recognition accuracy using Nearest Neighbor classifier. The features used in this algorithm are derived from the Discrete Curvelet Transform (DCVT), introduced by Candes and Donoho in [8]. Here the curvelet transform function is applied on the given image and the coefficients are obtained. The obtained coefficients are used as the feature vector for recognition of Kannada vowels and Kannada consonants. Dhandra et al. [9], have proposed modified invariant moments for isolated printed multi font/size Kannada vowels and numerals recognition and reported 97.8% accuracy. Rajput et al. [10] have proposed crack codes and density of the object pixels, features of size 64 for handwritten Kannada vowels and consonants recognition and achieved the recognition accuracy of 91.02% using multi level SVM classifier with five-fold cross validation. Their accuracy is reasonably high but at the same time complexity of the algorithm is also high due to large size of the feature set. Hence, from the above it is clear that the algorithms designed for Kannada characters recognition suffers either from the recognition accuracy or from the time and space complexity. 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 structure of Kannada characters are circles, holes and curvature in nature .This observation made us to use the second generation discrete curvelet transform features to recognize the hand printed isolated Kannada characters .

Section 2 of this paper contains the data collection and pre-processing methods. Section 3 is devoted for feature extraction method and designing of the proposed algorithm for hand printed Kannada Vowels and Kannada consonant character recognition system. The experimental results

obtained are presented in Section 4. Comparative analysis is given in Section 5 and Conclusion is presented in Section 6.

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 character are not available. Hence, the data collection is made and created the own database. Totally 2800 Kannada vowels and 6800 Kannada consonant images are collected from the Varies group of people belonging to Primary Schools, High Schools and Colleges. These are scanned through a flat bed HP scanner at 300 dpi which usually yields a low noise and good quality document image. The images were 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 raw input of the digitizer typically contains noise due to erratic hand movements and inaccuracies in digitization of the actual input. The noise present in the image is removed by applying median filter. A minimum bounding box is then fitted to the isolated character. To bring uniformity among the cropped character 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.

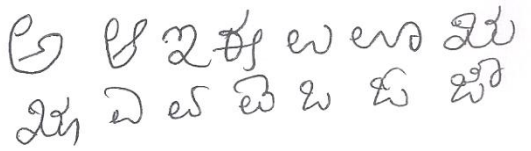


Fig. 1: Handwritten Kannada Vowels

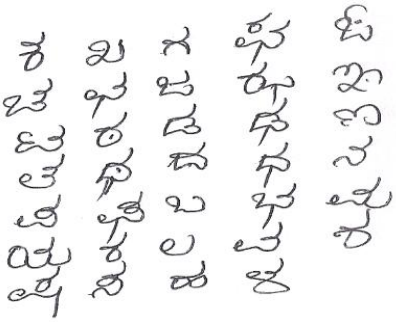


Fig. 2: Handwritten Kannada Consonants

3. FEATURE EXTRACTION

The Kannada handwritten characters have curves and straight lines, so curvelet transform is designed to extract the features, since it allows edges and other singularities along the lines in a more efficient way than other transforms. Hence, in this paper, focus is made on the Discrete Curvelet Transform with the Wrapping Technique. For extracting the features, a wrapping based discrete curvelet transform is used and it can be found in by Candes and Donoho [13]. Curvelet coefficients have different scales and angles. Two parameters are involved in the implementation of curvelet transform: number of scales and number of angles at the coarsest level. Energy of these coefficients is different for different coefficients based on angles and scales. In the proposed method 64X64 image blocks is decomposed into four scales using real-value curvelets. After the application of curvelet transform on the input image, one subband at the coarsest and one subband at the finest level of curvelet decompositions are obtained. Different subbands are obtained at each level for the other levels of curvelet decomposition. The number of wedges (subband) is $N_j = 4.2^{\lceil j/2 \rceil}$ at the scale 2^{-j} .When a scale is 1, 2, 3 and 4 then the number of wedges are 4, 8, 16 and 16 respectively. All the coefficients obtained cannot be used in the feature vector as it will increase the size of the feature vector drastically and also the time taken for feature vector formation. Hence, for extracting the potential features and also reducing the size of the feature vector for each sample, the standard deviation is obtained for the first half of the total sub bands at each of the remaining scales except scale 1. Only the first half of the total sub bands are considered, since curvelet angle at θ produces the same coefficients at the angle (θ+π) in the frequency domain. Hence, considering half of the total number of sub bands at each scale reduces the total computation time for the feature vector formation without loss of the information contained in an image. For the finest and the coarsest sub bands the standard deviation calculated is 20 used directly in the feature vector. Rectangular frequency tilling of an image with 5 levels curvelets is shown in Fig.3.The feature extraction and recognition process is given in Algorithm-3.1 & 3.2.

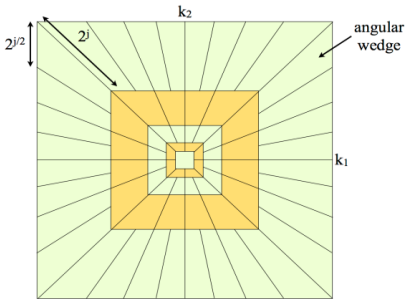


Fig.3: Rectangular frequency tilling of an image with 5 levels Curvelets

3.1 Training Phase

Algorithm-3.1: Feature Extraction Method

Input : Pre processed isolated Handwritten Kannada character image.

Output : Feature library.

Start :

1. Preprocessed image 64X64 pixels
2. Apply Wrapping based discrete Curvelet Transform on the preprocessed image.
3. Different numbers of sub bands are obtained at each level for the other levels of the curvelet decomposition.
4. The scale of 4 and angular orientations 4 are used for 'wedges'. Obtain the curvelets coefficient for each wedge.
5. Compute standard deviations of the curvelet coefficients of the first half of the total subbands (except for scale=1), obtained in step 4 to get feature set of size 20.
6. Repeat the Steps 1 to 5 for all the training images.
7. Computed standard deviations of curvelet coefficients of feature vector size 20, as

the features stored in train library in the database

End.

3.2 Testing Phase

Algorithm-3.2: Recognition of Handwritten Kannada character

Input : Isolated test character images.

Output : Recognition of the input Kannada character

Start :

1. Extract the features as obtained in Algorithm-3.1.
2. Store these feature vectors in test library database.
3. Compute the distance between the feature vectors of the test image stored in the test library and with the feature vector of the trained image stored in the train library.
4. Obtain the minimum distance computed in the step 3. Recognize the character as the label of the train image corresponding to the minimum distances.

End

4. EXPERIMENTAL RESULTS

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. For measuring the performance of an algorithm all preprocessed images are normalized to size 64x64 and experiment is carried out using wrapping based discrete curvelet transform on the preprocessed images. A total of 9600 character images of Kannada characters are classified using KNN classifier. The performance of an algorithm is tested using 2-fold cross validation. The average recognition rate for basic Kannada character is 90.57% from the experiment. The classification results of Kannada Vowels and consonants are presented in Table 1 shows the overall recognition accuracy for basic hand printed Kannada characters.

Table-1: Percentage of Recognition Accuracy for Handwritten Kannada Vowels and Consonants with KNN Classifier.

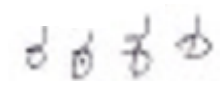
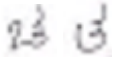
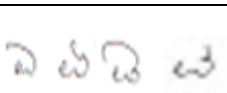
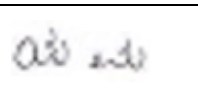
Hand written Kannada vowels and Consonants	No. of Sample Trained	No. of Sample Tested	Percentage of Recognition with KNN classifier
ಅ	100	100	89.0
ಆ	100	100	90.0
ಇ	100	100	93.7
ಈ	100	100	94.6
ಉ	100	100	90.2
ಊ	100	100	89.2
ಋ	100	100	90.0
ೠ	100	100	93.4
ಎ	100	100	92.1
ಏ	100	100	86.7
ಐ	100	100	92.0
ಒ	100	100	92.5
ಓ	100	100	90.2
ಔ	100	100	86.1
ಕ	100	100	88.1
ಖ	100	100	88.6
ಗ	100	100	91.0
ಘ	100	100	88.0
ಙ	100	100	94.1
ಚ	100	100	92.0
ಛ	100	100	90.0
ಜ	100	100	92.7
ಝ	100	100	95.2
ಞ	100	100	93.8
ಟ	100	100	89.0
ಠ	100	100	84.5
ಡ	100	100	94.0
ಢ	100	100	92.0
ಣ	100	100	92.3
ತ	100	100	90.0
ಥ	100	100	86.0
ದ	100	100	90.1

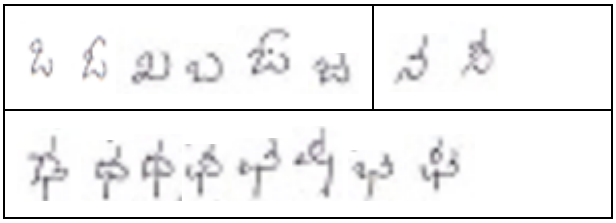
ಧ	100	100	90.0
ನ	100	100	90.1
ಪ	100	100	96.0
ಫ	100	100	94.0
ಬ	100	100	90.4
ಭ	100	100	86.0
ಮ	100	100	94.8
ಯ	100	100	92.8
ರ	100	100	88.5
ಲ	100	100	98.0
ವ	100	100	90.2
ಶ	100	100	90.2
ಷ	100	100	90.0
ಸ	100	100	86.4
ಹ	100	100	89.0
ಳ	100	100	90.0
Average Percentage Recognition Accuracy			90.57

From the Table 1, it is observed that the classification accuracy for similar shaped vowels and consonants are little less.

Similar shaped characters are grouped and shown in Table 2.

Table-2: Similar shaped characters

Group of similar shaped characters	
	
	
	



5. COMPARITIVE ANALYSIS

The Table-3 shows comparative analysis of proposed method with other methods. From the comparative study it is seen that proposed method presents the better recognition accuracy with less number of feature set size as compare to existing other methods. The Fig-4 shows character recognition rate.

Table-3: Comparative Analysis of Handwritten Kannada vowels and consonants with other existing methods.

Authors	Characters Considered	Features Computed & Dimensions	Classifier	Character Recognition Rate
Aradhya et al [4]	Handwritten vowels and Consonants	Fourier transform and PCA	PNN	68.89%
B.V.Dhandra et al [5]	Handwritten Consonants [28 classes]	Zone based Pixel density [64]	SVM	73.33%
Proposed system	Handwritten vowels and Consonants [48 classes]	Curvelet Coefficients [20]	KNN	90.57%

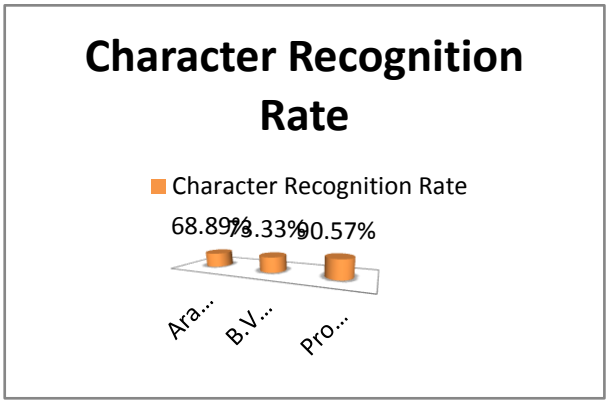


Fig 4. Graphical representation of character recognition proposed method with other existing methods.

6. CONCLUSION

An algorithm proposed here for recognition of handwritten Kannada vowels and Kannada consonants using Curvelet transform has exhibited the average percentage of recognition accuracy as 90.57% with KNN classifier with 2-fold cross validation. The proposed method has shown the encouraging results for recognition of Kannada vowels and consonants. The novelty of the proposed method is free from thinning and skew images of characters .The aim of the proposed system is to remove the confusions among similar characters and thereby increase the recognition rate . The future plan of this work is to build single OCR system for handwritten and printed

Kannada characters and proposed method is to be extended for characters written in other scripts also.

7. REFFERNCES

[1] Oivind Trier, Anil Jain, Torfinn Taxt (1996) Pattern Recognition, Vol 29, No 4, pp 641-662.

[2] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Classification, 2nd ed., Wiley-New York.

[3] R.M.K. Sinha, Scott D. Conell and Anil K. Jain, "Recognition of Unconstrained Online Devanagari Characters", Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Sept. 2000. pp. 368-371.

[4] Aradhya M., Niranjana S.K., Hemantha kumar G., (2010) "Probabilistic Neural Network based Approach for Handwritten Character Recognition" Special Issue of IJCCT, Vol. 1 Issue 2,3,4 pp. 9-13

[5] B.V. Dhandra, Mallikarjun Hangarge and Gururaj Mukarambi (2012). "A Zone Based Character Recognition Engine for Kannada and English Scripts". Elsevier Science Direct, pp. 3292-3299.

[6] Kunte Sanjeev R., Sudhaker Samuel (2006). "A simple and efficient optical character recognition system for basic symbols in printed Kannada text". *Sadhana*, Vol. 32, Part 5, pp. 521-533.

[7] B.V. Dhandra, Mallikarjun Hangarge, Vijayalaxmi M.B. and Gururaj Mukarambi (2014). "Script Identification Using Discrete Curvelet Transfirms". *IJCA*, Recent Advances in Information Technology. pp.16-20

[8] E. Cand`es, L. Demanet, D. Donoho and L. Ying, "Fast Discrete Curvelet Transforms," *Technical Report*, July 2005, pp. 761-799

[9] B.V. Dhandra, Mallikarjun Hangarge and Shashikala Parameshwarappa (2010). "Multi-Font Kannada Vowels and Numerals Recognition Based on Modified Invariant Moments", *IJCA*, Special Issue on RTIPPR (3): pp 146–151.

[10] G.G. Rajput, Rajeswari Horakeri (2013). "Unconstrained Kannada Handwritten Character Recognition Using Multi-level SVM Classifier", P. Maji et al.(Eds): PREMI 2013, LNCS 8251, pp. 204-212

[11] Dinesh Acharya U., N.V. Subba Reddy and Krishnamoorthi (2008). 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.

[12] Nagbhushan P., Pai Radhika M. (1999). Modified region decomposition method and optimal depth decomposition tree in the recognition of non –uniform sized characters – An experimentation with Kannada characters. Pattern Recognition. Letter. 20: pp. 1467-1475

[13] E.J. Cand`es and D.L. Donoho, "Curvelets–A Surprisingly Effective Non adaptive Representation for Objects with Edges," in *Curves and Surfaces*, C. Rabut, A. Cohen, and L.L. Schumaker, Ed., Vanderbilt University Press, Nashville, TN, 2000, pp. 105–120.

[14] Ashwin T.V., Sastry P.S. (2002). A Fonts and Size-Independent OCR System for Printed Kannada Documents Using Support Vector Machines. *Sadhana*, 27: pp. 35-58.

[15] J.L. Starck, E.J. Cand`es, and D.L. Donoho (2002). "The Curvelet Transform for Image Denoising", *IEEE Trans. Im. Proc.*, Vol. 11, No. 6, pp. 670-684

[16] Gonzales R.C. and Woods, R.E. (2002). Digital Image Processing 2nd Ed. Upper Saddle River, N.J.: Prentice-Hall, Inc. pp. 261-269

[17] E.J. Candès, L. Demanet, D.L. Donoho, L. Ying. (2003) "Fast Discrete Curvelet Transforms". *Multiscale Model. Simul.*, pp. 861-899.

[18] Nagabhushan P., Angadi S.A., Anami B.S. (2003). A Fuzzy Statistical Approach to Kannada Vowel Recognition based on Invariant Moments, Proceedings of NCDAR -2003, PESCE, Mandya, pp. 275-285.

[19] U. Pal, B.B. Chaudhuri (2004). Indian Script Character Recognition: A Survey. Pattern Recognition, 37 (2004), pp. 1887–1899.

[20] R. Sanjeev Kunte, R.D. Sudhaker Samuel (2007). 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.

[21] M.J. Fadili and J.L. Starck (2007). Curvelets and Ridgelets, "Encyclopedia of Complexity and System Science", pp. 1-29

[22] 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.

[23] Srikanta Murthy, Mamata H.R., Sucharita S., "Multi – font and Multi-size Kannada Character Recognition based on the Curvelet and Standard Deviation", *IJCA*, Vol 35, December 2011. pp.101-104

[24] Rampalli R., Ramakrishan, Anagari G. (2011). Fusion of Complementary Online and Offline Strategies for Recognition of Handwritten Kannada Characters, *Journal of Universal Computer Science*, 17(1): pp. 81-93.