Review of Automatic Handwritten Kannada Character Recognition Technique Using Neural Network

[1]Mukesh Kumar,[2] Dr.Jeeetendra Sheethlani
[1]Department of Computer Science
SSSUTMS, Sehore

Abstract—Data processing and management is common now a days. In this paper, automatic processing of forms written in Kannada language is considered. A suitable pre-processing technique is presented for extracting handwritten characters. Principal Component Analysis (PCA) and Histogram of oriented Gradients (HoG) are used for feature extraction. These features are fed to multilayer feed forward back propagation neural network for classification. Only 57 characters are used for recognition. Performances of two features are compared for different number of classes. HoG is found to have better recognition accuracy than PCA as number of classes increased. This is implemented in Visual Studio 2010 using Open CV library.

IndexTerms —Back Propagation Neural Network, Form Processing, Histogram of Gradients, Kannada Script, Principal Component Analysis.

I. INTRODUCTION

The automation of handwritten form processing is attracting intensive research interest due to its wide application and reduction of manual work. In Indian context, many organizations will collect the data on paper based forms. Automatic processing of these forms is a process of capturing the information stored in the forms and converting it into electronic (machine readable) format. Handwritten character recognition immensely to the advancement of automation process. This is classified into offline and online recognition. For Automatic Form Processing (AFP) offline recognition method is used since it involves automatic conversion of text in an image into letter codes which are usable within computer and text processing applications [1]. AFP includes complete scan of a form using scanner. The scanned image then undergoes various pre-processing operations, character segmentation and recognition of handwritten characters. India is a multi-lingual and multi-script country comprising of eighteen official languages, Kannada is one among them. Several works has been done for the recognition of handwritten Kannada characters. The major pre processing steps of AFP includes edge detection, morphological operations to make it suitable for segmentation. Segmentation separates the image text documents into lines, words and the characters. Thungamani M and RamakanthKumar P [2] discuss two segmentation techniques such as classical approach and holistic approach. In classical approach, the input image is segmented into sub images. In holistic approach, the characters are recognized without dissection. Mamatha H.R and Srikanatamurthy K [3], proposed a segmentation scheme using projection profiles. Morphological operations are used to remove the noise. After this text lines are extracted using horizontal projection profile, words and characters are extracted using vertical projection profiles. India is a multi-lingual and multi-script country comprising of eighte official languages, Kannada is one among them. Several works has been done for the recognition of handwritten Kannada characters. The major pre processing steps of AFP includes edge detection, morphological operations to make it suitable for segmentation. Segmentation separates the image text documents into lines, words and the characters. Thungamani M and RamakanthKumar P [2] discuss two segmentation techniques such as classical approach and holistic approach. In classical approach, the input image is segmented into sub images. In holistic approach, the characters are recognized without dissection. Mamatha H.R and Srikanatamurthy K [3], proposed a segmentation scheme using projection profiles. Morphological operations are used to remove the noise. After this text lines are extracted using horizontal projection profile, words and characters.

Kannada script has large number of character set. This may reduce the recognition accuracy and increase the computational cost. To avoid this problem an algorithm has been proposed to reduce symbol set [4], where the vowel modifiers (kagunitha) and consonant modifiers (vattakshara) which are not connected to base characters are considered as separate classes. Devanagari script has similar characteristics as Kannada script like vowel modifiers, consonant conjuncts etc,. The recognition of this script consists of three phases: segmentation, decomposition, i.e,. decomposing a composite character into base part and modifier parts; and recognition [5]. Only small subsets of compound characters (upper and lower signs) are considered for the recognition. Many Arabic letters also share common primary shapes, which differs only in the number of dots and the dots or above or below the primary shape. A survey on offline recognition of Arabic handwriting recognition is presented in [6]. Different segmentation, feature extraction and recognition engines used for OCR are also discussed. An overview of character recognition methodologies with respect to the offline character recognition systems such as pre-processing, segmentation, representation, recognition and post-processing methods are presented in [7]. Shape based features such as Fourier descriptors

and chain codes are used for the recognition of handwritten Kannada characters (vowels and numerals) are discussed in [8]. Support Vector Machine (SVM) is used for recognition purpose and an accuracy of 95% is obtained. A brief survey on offline recognition of Devanagari script is presented in [9]. Performance of different feature extraction techniques using different classifiers is tabulated. Gradient and PCA based features with PCA, SVM and Neural Network classifiers are found to have better recognition accuracy. Development of two databases for two popular Indian scripts Devanagari and Bangla for numeral recognition is presented in [10]. This uses a multistage cascade recognition scheme using wavelet-based multi-resolution representations and multi layer perceptron (MLP) classifiers to achieve higher recogniti accuracy. This is then used to the recognition of mixed numerals for three Indian scripts such as Devanagari, Bangla and English. Unconstrained handwritten recognition of Kannada characters using very large dataset of 200 samples using ridgelet transform is discussed in [11]. To reduce the dimension of feature vector PCA is used. It is found that ridgelet features offered promising result than PCA. A zone based method for the recognition of handwritten Kannada vowels and consonants is presented in [12]. Character image is divided into 64 non-overlapped zones and from each zone crack codes are computed. SVM is used as classifier and an accuracy of 87.24% is achieved. Literature records few papers on Kannada character recognition. Choice of methods for feature extraction is important for achieving efficient character recognition for large classes. In this paper, Kannada handwriting recognition for automatic form processing is considered. PCA and HoG are used for feature extraction. Performances of features are compared for 57 classes.

II. KANNADA SCRIPT

Kannada alphabets were developed from th descendents of Brahmi script such as Kadamba and Chalukya scripts. Handwritten character recognition of Kannada characters is a very challenging task because of its large dataset, shape similarity among characters an non-uniqueness in the representation of diacritics. Kannada script has 49 primary characters: 15 vowels and 34 consonants as shown in Fig. 1(a) and Fig. 1(b). Each of the vowels, modify primary consonants to form a compound character or kagunita as shown in Fig. 1(c). Additionally a consonants emphasis glyph called consonant conjuncts shown in Fig. 1(d) [vattakshara], exists for each of the 34 consonant. Thus, total number of possible combinations is as shown in TABLE

C D E F G H IÄ J K L M N O CA CB

(a) Vowels

PÀ R UÀ WÀ Y

ZÀ bÀ d gÀhÄ k

l oÀ qÀ qsÀ t

vÀ xÀ zÀ zsÀ £À

¥À ¥sÀ § ¨sÀ ªÀÄ

AiÀÄ gÀ ® ªÀ ±À µA ¸À ºÀ ¼À

(b) Consonants

PÀ PÁ Q QÃ PÀÄ PÀÆ PÀÈ PÉ PÉÃ PÉÊ PÉÆ PÉÆÃ

PË PÀA PÀB

(c) Vowel modified consonants

Ì Í Î Ï Ð Ñ Ò Ó Ô Õ Ö × Ø Ù Ú Û Ü Ý Þ ß à á â ã ä å ç è é

ê ë ì í î

(d) Consonant conjuncts

Fig. 1: Basic Kannada character set

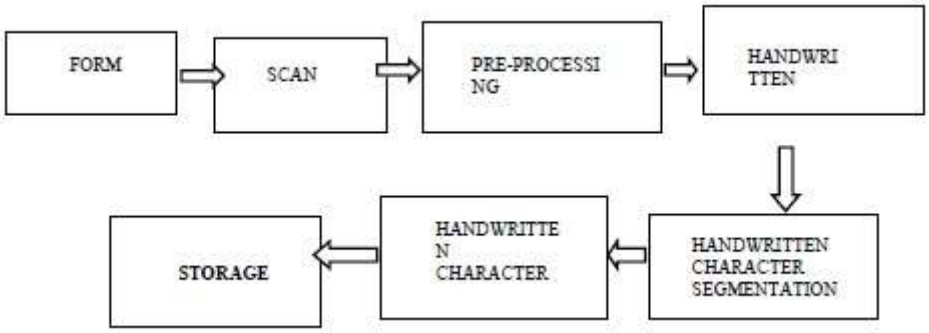
Table I

Maximum possible combinations of Kannada characters

Character Type	V (Vowel)	C (Consonant)	CV (Kagunita)	CCV (tÆÚ)	CCCV (vÀì÷å)	N	Total
Possible combinations	15	34	510	17340	589560	10	589585

III. FORM PROCESSING METHODOLOGY

Automatic Form Processing system involves Image acquisition using scanner, pre-processing of scanned form, only handwritten character extraction, handwritten character segmentation, character recognition and storage as shown in The template of the Birth certificate is created with all the required data fields. The applicant is then instructed to fill the form in Kannada language with all the base characters in the upper box and conjuncts in the lower box. The filled form is then scanned using flatbed scanning.



BLOCK DIAGRAM OF AUTOMATIC FORM PROCESS

IV. EXPERIMENTAL RESULTS

The segmented characters are used for feature extraction using PCA and HoG. Performance of these feature for different number of classes are compared using neural network with back propagation learning. 100 samples per class are used for recognition purpose. Neural network parameters are listed in TABLE 2. Recognition accuracy of PCA and HoG for different number of classes are listed in TABLE 3.

Table 2: Neural Network Training Parameters

Number of Input nodes	270
No. of Hidden layer	1
No. of Hidden nodes	125
Training epochs	30000
Goal achieved	10e-6
No. of Output nodes	57

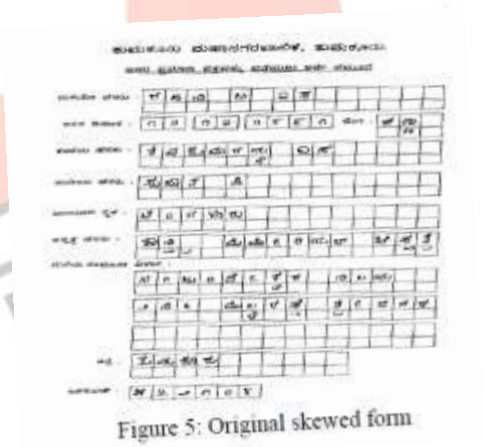


Figure 5: Original skewed form

Figure 6: Skew corrected Image

Figure 6: Skew corrected Image

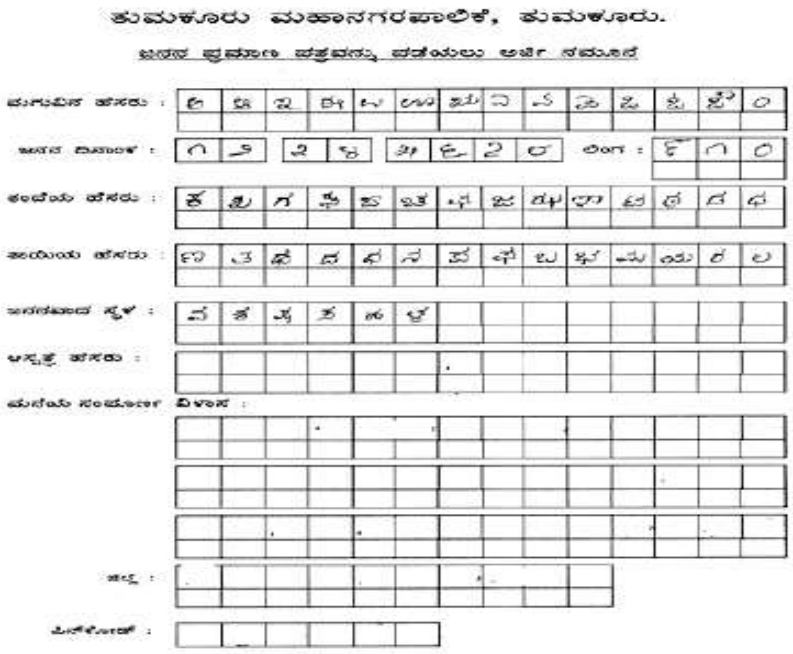


Figure 8: Dilated form

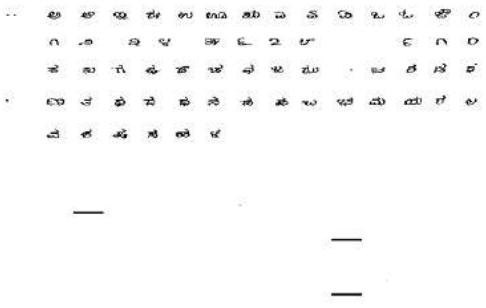


Figure 12: Hole filled Image

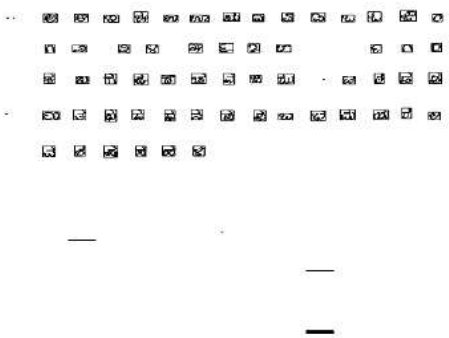


Figure 13: Segmented characters

IV. CONCLUSION

In this paper, Kannada character recognition with application to automatic form processing is presented. Only 57 characters are considered for processing. Using suitable pre-processing techniques only handwritten characters are extracted. PCA and HoG are used for feature extraction. Performance of features is compared using neural network classifier. Results of pre-processing implemented using Visual Studio using OpenCV library are presented. HoG is found to have better recognition accuracy than PCA for large number of classes.

V. REFERENCES

[1] Afef Kacem, Asma Saidani, Abdel Belaid, "A system for an automatic reading of student information sheets", International Conference on Document Analysis and Recognition, pp 1265-1269, 2011.
[2] M. Thungamani and P. Ramakanth Kumar, "A Survey Methods and Strategies in Handwritten Kannada Character Segmentation", International Journal of Science Research, Vol 1, Issue 1, June 2012.
[3] H.R Mamatha and and K. Srikantamurthy, "Morphological operations and projection profile based segmentation of handwritten Kannada document", International Journal of Applied Information Systems (IJAIS), Vol 4, No.5, October 2012.
[4] Nethravathi B, Archana C.P, Shashikiran K, A.G Ramakrishnan and VIjay Kumar, "Creation of huge annotated database for Tamil and Kannada OHR", 12th IEEE International Conference on Frontiers in Handwriting Recognition, Nov 2010, Pages 415-420.
[5] R M. K. Sinha and H. N. Mahabala, "Machine Recognition of Devanagai Script", IEEE Transactions on Systems, Man and Cybernetics, Vol. Smc 9, No 8, August 1979.
[6] Liana M. Lorigo and Venu Govindaraju, "Offline Arabic handwriting recognition: A survey", IEEE Transactions on Pattern Analysis and machine Intelligence, Vol 28, No.5, May 2006.
[7] Vengatesan K., and S. Selvarajan"Improved T-Cluster based scheme for combination gene scale expression data" International Conference on Radar, Communication and Computing (ICRCC), pp. 131-136. IEEE (2012).
[8] Kalaivanan M., and K. Vengatesan." Recommendation system based on statistical analysis of ranking from user. International Conference on Information Communication and Embedded Systems (ICICES), pp.479-484, IEEE, (2013).
[9] K. Vengatesan, S. Selvarajan: The performance Analysis of Microarray Data using Occurrence Clustering. International Journal of Mathematical Science and Engineering, Vol.3 (2) .pp 69-75 (2014).
[10] K Vengatesan, V Karuppuchamy, S Pragadeeswaran, A Selvaraj," FAST Clustering Algorithm for Maximizing the Feature Selection in High Dimensional Data", Volume – 4, Issue-2, International Journal of Mathematical Sciences and Engineering (IJMSE), December 2015