International Journal of Combined Research & Development (IJCRD)
eISSN:2321-225X;pISSN:2321-2241 Volume: 5; Issue: 4; April -2016

Kannada Handwritten Character Segmentation using Curved Line and Recognition using Support Vector Machine

Mr. Chethan Kumar G.S
Student, Dept. of CSE
BNMIT, Bengaluru, India

Dr. Y.C Kiran
Asst. Prof. Dept of CSE
BNMIT, Bengaluru,India

Abstract - This work concentrates on developing a novel approach to segment and recognize handwritten Kannada text when a character and background area of similar color (Not Exactly Same) and if a word contains both characters as well as other symbols (special symbols, numbers etc.) A new method of generating curved segmentation paths of handwritten Kannada text line is proposed. For overlapped characters, instead of background thinning, analysis of connected components in the overlap region of two neighboring characters is carried out, and a curved segmentation path is generated by connecting a series of points that separate the connected components belonging to different characters. A multi-class Support Vector Machine (SVM) is used for the recognition purpose. SVM finds the optimal hyper-plane which maximizes the distance, the margin, between the nearest examples of both classes named Support Vectors (SVs).Experiments are performed on handwritten Kannada character images and obtained an empowering result.

Keywords – Handwritten Kannada Text , Segmentation, Recognition, SV, Feature Extraction.

INTRODUCTION

Character segmentation is one of the most important decision processes for optical character recognition (OCR). Segregating individual alphabetic characters in the script picture is regularly sufficiently critical to make a definitive commitment towards the achievement rate of the general framework. Kannada Handwritten character acknowledgment has gotten more prominent consideration in scholarly and profitable fields. The acknowledgment framework can be either online or disconnected from the net. There is huge interest for Kannada manually written character acknowledgment framework. Escalated research has been done on this and countless have been distributed on this subject amid the most recent couple of decades. The reason for this anticipate is to take transcribed Kannada words as data, process the words and fragment the words into characters utilizing bended line calculation, and to perceive the example and change the character to an embellished variant of the info utilizing the Support Vector Machine classifier. This anticipate is gone for creating programming which will be useful in perceiving characters of Kannada dialect. This anticipates is limited to Kannada characters as it were. It can be further created to perceive the characters of various dialects. It immerses the idea of neural system. Design acknowledgment is maybe the most widely recognized utilization of neural systems. The neural system is given an objective vector furthermore a vector which contains the example data, this could be a picture and manually written information. The neural system then endeavors to figure out whether the info information coordinates an example that the neural system has remembered.

In this paper, another technique for creating bended division ways is proposed for unconstrained manually written Kannada content line acknowledgment. Rather than utilizing foundation diminishing, examination of associated parts in the cover area of two neighboring characters is completed, and a bended division way is produced by basically interfacing a progression of focuses that different the associated segments fitting in with various characters. The proposed technique indicates great execution on unconstrained manually written disconnected from the net Kannada content lines

I. RELATED WORK

C. Naveena, V.N. Manjunath et al [1] proposes a new character segmentation algorithm for unconstrained handwritten Kannada scripts is presented. The proposed method is based on thinning, branch point and mixture models. The expectation-maximization (EM) algorithm is used to learn the mixture of Gaussians. We have used a cluster mean points to estimate the direction and branch point as reference points for segmenting characters.

Mamatha H R, Srikantamurthy K et al[2] proposes The authors presented a segmentation scheme for segmenting handwritten Kannada script into lines, words and typescript using morphological operations and projection profiles are proposed. The technique was tried on absolutely unconstrained manually written Kannada scripts.. Widening is a primitive morphological operation that develops or thickens objects in a paired picture and morphological operations like organizing and disintegration is done in preprocessing stages.

At first, all the associated parts in a record picture are recognized and expelled from the parallel picture utilizing associated segment examination calculation. After the consummation of first stage, the following stage is to concentrate singular content lines present in the record. Keeping in mind the end goal to concentrate singular content line, a procedure in view of projection is utilized. A projection profile is a histogram giving the quantity of ON pixels collected along parallel lines. This technique comprises of two stages. In the principal stage, scientific morphology method is utilized for expelling separated segments and developing extension between the segments. In the following stage the projection profile strategy is utilized for division of content into lines, words and characters. The strategy was tried on absolutely unconstrained written by hand Kannada scripts, which pays more test and trouble because of the intricacy included in the script.

Ganpat Singh G Rajput et al [3] the normalized character image is divided into 64 zones each of size 8x8 pixels. For each zone, from left to right and from top to bottom, the crack code, representing the line between the object pixel and the background (the crack), is generated by traversing it in anticlockwise direction. A feature vector of size 512 is obtained for each character. A multi-class SVM is used for the classification purpose. The main advantage of using SVM is sliding window concept introduced here, which gives better result for Printed Text and disadvantage of SVM classifier is it consumes more time, Since feature is extracted and comparison is done using SVM after each window movement. Manasij Venkatesh, Vikas Majjagi[4] authors (image(HMMs) to take advantage of the agglutinative nature of Kannada script, which allows them to perform implicit segmentation of characters along with recognition.

Nanxi Li, Xue Gao[5] authors presented a new method for generating curved segmentation paths is proposed for unconstrained handwritten Chinese text line recognition. Instead of using background thinning, analysis of connected components in the overlap region of two neighboring characters is carried out; a curved segmentation path is generated by simply connecting a series of points that separate the connected components belonging to different characters. This method shows good performance on unconstrained handwritten offline Chinese text lines. In handwritten Chinese text line recognition system, the segmentation path generation is an essential phase to produce a series of radicals or single characters for sequential processing. Then the generated segmentation paths are usually verified by structural or statistical rules, or even recognition information, to get final segmentation and recognition results. Thus, the segmentation path generation is very important for the performance of handwritten Kannada character recognition system.

II. PROPOSED SYSTEM

The proposed method comprises of 4 phases:
1. Pre-processing.
2. Segmentation.
3. Feature Extraction
4. Classification and Recognition

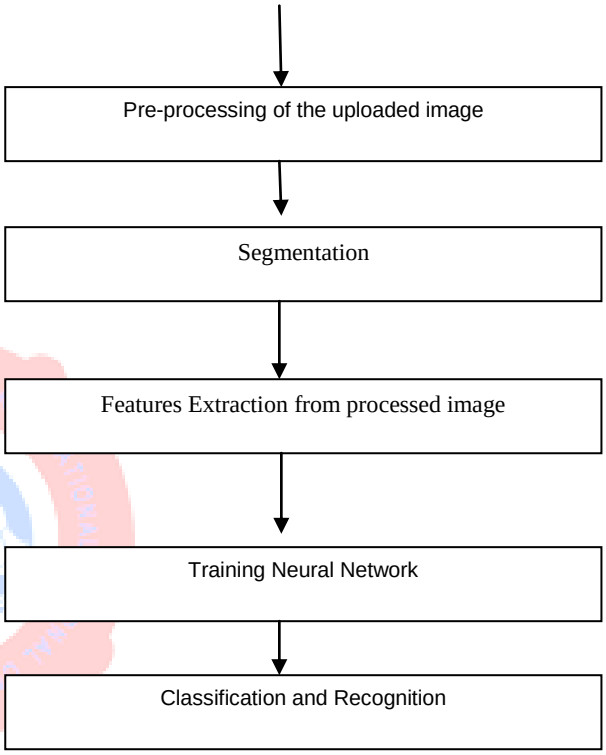


Fig 1: The block schematic diagram of the proposed model

A) Pre-Processing
This module is the initial stage for work to progress with. Here the image is selected in two ways they are:

Algorithm for Preprocessing

1) Read Input Image using open standard dialog box for retrieving files using 'imread'.

2) Convert image to gray scale image using 'imageen= graytobin(imageen)'

3) Global thresholding method is used to separate when the background and foreground character are of similar colour(Not exactly same).

B) Segmentation .
The block diagram of the proposed method is shown in Fig. 2. The method mainly includes three steps: segmentation of naturally separated characters, segmentation of overlapped

characters, and segmentation of touched characters.

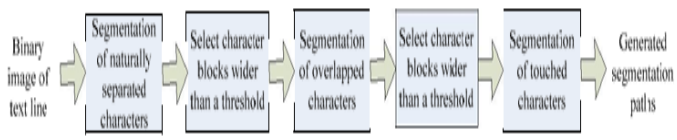


Fig.2 The block diagram of proposed method

Segmentation of naturally separated characters Histogram projection is used to segment naturally separated characters, i.e. characters that neither overlap in horizontal direction nor touch with each other. A straight line passing vertically through the mid of the zero region in histogram is a segmentation path.

Possibility 1: Components do not overlap column-wise

In this case, the following condition must be satisfied:

CimaxColumn < CjminColumn

Where CimaxColumn is the maximum column number of component i and CjminColumn is the minimum column number of the subsequent component j.

Segmentation of overlapped characters is carried out on the selected character blocks using fast connected component analysis algorithm.

Possibility 2: Components overlap partially column-wise.

For two components overlap partially, these conditions must be satisfied:
CjminColumn < CimaxColumn

CjmaxColumn > CiminColumn

CjmaxColumn > CimaxColumn

Decision: based on our proposed algorithm, partially overlapping components are considered as different objects.

Segmentation of overlapped characters

Possibility 3: Components overlap completely column-wise.

Two components are completely overlapping, if the following conditions are satisfied:

CjminColumn < CimaxColumn

CjmaxColumn > CiminColumn

CjmaxColumn < CimaxColumn

Decision: completely overlapping components are considered as different objects.

In the segmentation stage, an image of sequence of characters is decomposed into sub-images of individual character. The pre-processed input image is segmented into isolated characters by assigning a number to each character using a labelling process. This labelling provides information about number of characters in the image. Each individual character is uniformly resized into pixels

C) Feature Extraction

Every character has some elements, which assume an imperative part in example acknowledgment. Kannada characters have numerous specific components. Highlight extraction depicts the pertinent shape data contained in an example so that the errand of ordering the example is made simple by a formal strategy. Highlight extraction stage in HGCR framework examinations these Kannada character section and chooses an arrangement of elements that can be utilized to exceptionally distinguish that character fragment. Principally, this stage is heart of HGCR framework since yield relies on upon these elements. Highlight extraction is the name given to a group of methods for measuring the applicable shape data contained in an example so that the undertaking of ordering the example is made simple by a formal system. Among the diverse outline issues included in building a perceiving framework, maybe the huge one in the choice of set of elements.

Highlight extraction for exploratory information projection empowers high-dimensional information perception for better information structure understanding and for bunch investigation. In highlight extraction for order, it is attractive to concentrate high discriminative lessened dimensional highlights, which diminish the grouping computational prerequisites. Notwithstanding, include extraction criteria for exploratory information projection routinely plan to minimize a mistake capacity, for example, the mean square blunder or the entomb design separation distinction while highlight extraction criteria for order expect to expand class separability as could reasonably be expected. Henceforth, the ideal extricated highlights (with respect to a particular rule) ascertained for exploratory information projections are not as a matter of course the ideal components in regards to class

separability and the other way around. Specifically, two or more classes may have key elements that are comparable. In addition, highlight extraction for exploratory information projection is utilized for a few dimensional information perception, while grouping as a rule needs more than a few elements. Subsequently, highlight extraction ideal models for exploratory information projection are not by and large utilized for grouping and the other way around.

In the wake of extricating the components, the information ought to be spoken to in one of two routes, either as a limit or as a complete area. At the point when the attention is on outside shape qualities, for example, corners and articulations then limit representation is fitting. While provincial representation is suitable when the emphasis is on interior properties, for example, compositions or skeletal shape. In a few applications like character acknowledgment these representations coincide, which regularly require calculation taking into account limit shape and in addition skeletons and other inner properties. As far as character acknowledgment descriptors, for example, openings and coves are effective components that separate one a player in the character from another. This portrayal additionally called highlight determination, manages removing highlights which results in some quantitative data of hobby or elements that are essential for separating one class of articles from another.

Water Reservoir Method

This sort of highlight extraction use water repository rule. In the event that the water is administered from any side top, base, right, left, then the hole districts of the segment where the water will be put away are considered as supply [6]. The supply region is acquired when digit is not associated. For the most part four repositories are shaped:

• Top reservoir: Reservoir acquired just when water is poured from top of the digit.

• Bottom reservoir: Reservoir acquired just when water is poured from base of the digit.

• Left reservoir: Reservoir got just when water is poured from left of the digit.

• Right reservoir: Reservoir got just when water is poured from right of the digit.

In the wake of getting the areas, the proportion of the repository of the pixels where water is put away is considered as aggregate region and is processed in each of the four headings and henceforth put away as a component vector.

Fill Hole Density [7] A hole is an arrangement of pixel touches the outskirt such that it isolates the foundation into two districts. To keep away from this, it required to fill particular region to availability of foundation pixels. A 4connected neighborhood availability guideline [8] is utilized for network of the foundation pixel. It says that the pixel that touches the inside pixel and associated on a level plane or vertically to the middle pixel are said to be neighbor. Here 3x3 window is considered with set of 0's and 1's.The circling range of the digit is loaded with ON pixels [8]. The proportion of the fill opening thickness with aggregate zone appraisal is taken as highlight.

D) Classification and Recognition using SVM

A Support Vector Machine (SVM) is a classifier formally defined by a separating hyper-plane. SVM is used for finding an optimal hyper-plane which is then used for categorizes new examples for the given labeled training data. Only need to take into account in which sense the hyper-plane obtained optimal. For a linearly separable set of 2D-points which belong to one of two classes, we need to find a straight line that separating them.

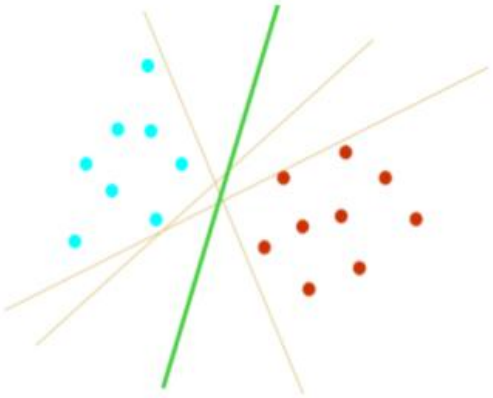


Fig 3 Optimal separating hyperplane

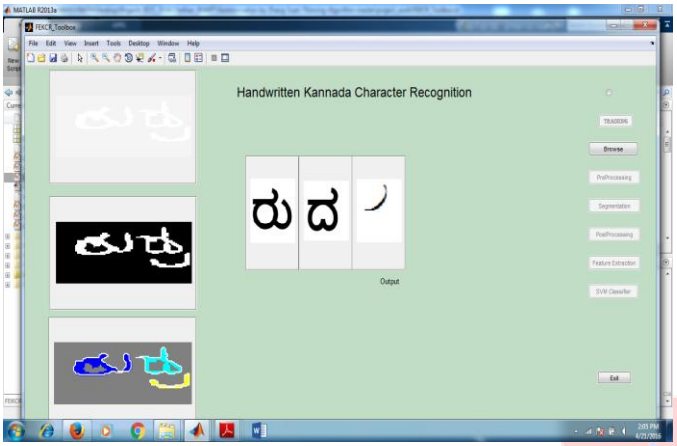
Consider the example in Fig 3 Here, there are numerous conceivable straight classifiers that can isolate the information, yet there is stand out that augments the edge (amplifies the separation in the middle of it and the closest information purpose of every class). This direct classifier is termed the ideal isolating hyper-plane. Instinctively, we would anticipate that this limit will sum up well instead of the other conceivable limits

Experiment Results

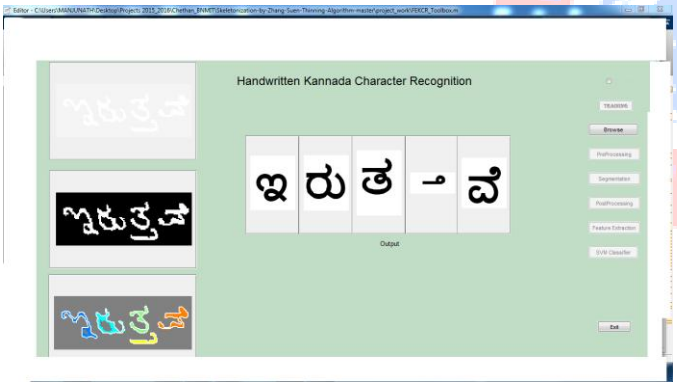
In order to assess the results, experiment is conducted on scanned handwritten Kannada text words if the scanned image is of binary image preprocessing steps like global thresholding is used to separate the character background and foreground and curved segmentation algorithm is applied on the images and feature is extracted by using the different algorithms like water reservoir method, projection based feature extraction and fill hole density method is used. This

data is used for classification and recognition using multi class SVM. We tested for different unconstrained handwritten Kannada images and obtained empowering result. Experiment results are shown in the below snapshots 1 and 2.

Snapshot 1:



SnapShot 2:



IV CONCLUSION

In this paper we presented a new method for generating curved segmentation paths for unconstrained handwritten Kannada characters. A connected component analysis algorithm is used for segmenting handwritten characters and feature extraction of segmented characters is carried out using feature extraction algorithms. Classification and recognition of characters is done using SVM classifier. Experiments show that significant improvements are achieved in both correctness and validity for generating segmentation paths and recognition of characters. Our future work aims to develop segmentation method for connected characters and improve SVM classifier to achieve still better recognition rate which provides efficient results. Also we plan to extend our work to other Indian numeral scripts

REFERENCES

[1] C. Naveena , V.N. Manjunath Aradhya,"Handwritten Character Segmentation for Kannada Scripts" IEEE World Congress on Information and Communication Technologies 2012.

[2] Mamatha H R, Srikantamurthy K, "Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document" International Journal of Applied Information Systems Volume 4– No.5,October 2012.

[3] Ganpat Singh G Rajput, Rajeshwari Horakeri, "Zone based Handwritten Kannada Character Recognition Using Crack code and SVM" International Conference on Advances in Computing, Communications and Informatics (ICACCl) 2013.

[4] Manasij Venkatesh, Vikas Majjagi, and Deepu Vijayasenan, "Implicit segmentation of Kannada characters in offline handwriting recognition using hidden Markov models" 16 Oct 2014

[5] Nanxi Li, Xue Gao, Lianwen Jin , " Curved segmentation path generation for unconstrained handwritten Chinese text lines" pp.501-505,IEEE 2008

[6] Water Reservoir Based Approach For Touching Numeral Segmentation U. Pal, A. Belaïd,And C. Choisy , Springer 2003.

[7] U Ravi Babu, Dr. Y Venkateswarlu, Aneel Kumar Chintha,"Handwritten Digit Recognition Using KNearest Neighbor Classifier" 2014 World Congress On Computing And Communication Technologies,Vol 4,Pages.60-65.

[8] R.C Gonzal, R.E.Woods, "Digital Image Processing", Pearson education, 2002