

Kannada Handwritten Numerals Recognition and Translation Using Template Matching

Shivanand Killedar
Lecturer, Dept of EEE AITM Belagavi
shivanand.killedar@gmail.com

Satish Deshapande
Assistant Professor, G.I.T Belagavi
satishdeshpande1968@gmail.com

Abstract:- Template matching technique equates portion of input character images versus trained character images, Where Convolution is one of the method in template matching. This convolution technique can be used to recognize images, scanned documents and even handwritten scripts. In this work convolution is used to recognize the hand written numerals of Kannada by using pen, pencil, sketch and machine printed digits. The convolution method is used to differentiate 2000 numerals based on style, stokes, texture, thinness, thickness, darkness and even fonts size. The steps involved in convolution for template matching are Scanning of hand written document to get preprocessed binarized noise free images .in the next step comparison will takes place between input and template images which is totally depend upon correlation between the two. If the correlation is high matching is successful between handwritten written numeral images and trained images. The average set gives out 86.28% matching efficiency. By comparing all the results which is obtained from the work reveals that sketch handwritten is the suitable than pencil and pen in template matching.

Keywords:binarization,convolution,differentiate,templatematching, preprocessing.

1. Introduction

Kannada is the official language of the south Indian state of Karnataka, is spoken by about 60 million people. And it is originated southern Brahma lipi during Asoka period. The Kannada rudiments were originated from the Kadamba and Chalaukya scripts which are used in fifth and seventh centuries A.D. Kannada language is different from other Indian languages there is no line which connect all characters. In Kannada all the characters are isolated from each other. Kannada script is more complex due bearing of compound characters. Due to which segmentation is tough. And here no possibility of upper or lower case as in English. There are total 51 main characters out of which 16 are vowels and 35 are consonants. And even it has numerals form 0 to 9. And all character is similar and changes in its strokes [1]

The main intention is to produce a system that recognizes the character by using “quantification of the character into mathematical vector entity using the geometrical properties of the character image”.

These recognized patterns used in airport, banking sector, post offices and recognition is also done using template matching method which is based on image processing. In which each numeral is matched with trained or template images depending on which numerals may be recognized.

Only few works related to Kannada numeral recognition is done based on “Selection of a feature extraction method”. Different characteristics extraction methods have been developed for numerals belong to different Indian scripts. “Three different types of characteristics, namely, density features, moment features and descriptive component features are used”.

2. Block Diagram of Numeral Recognition Steps

Kannada hand written numeral recognition involves image input, where the input image is converted into digital image. And it is given as a input to MATLAB which results in the Preprocessing step, which involves procedures for smoothing, enhancing, and filtering, and it is given to the proposed method. In this proposed method using template matching which recognizes the character and gives output. The detailed steps involved in numeral recognition is shown in figure.2



Figure.2 Block Diagram of Numeral Recognition Steps

3. Methodology

In this methodology, the following cases were carried out using pen, pencil, sketch and printed images. Firstly the Kannada numerals from 0-9 are written on a paper using pencil by 200 samples (peoples). And these papers are scanned using scanner and converted into digital image such as jpeg. Now these images are imported in MATLAB and there is a program written in MATLAB which preprocesses the image. After preprocessing, line and character segmentation is done and each character is matched with trained character. Unknown input characters are converted into machine readable form i.e. Kannada numerals converted into English numerals.

Now similarly on the same basis the Kannada numeral for 200 people for Pencil, Sketch pen as well machine printed is carried out in MATLAB.

4. Result And Discussions

In this work experiments were carried out with database containing 2000 handwritten Kannada digits. There were 200 people of different age groups and from various fields such as business, education, industry, service, also from the students of English & Kannada medium. Each of them has written digits from 0 to 9. As discussed in this project, because of writing style and size of the digits will vary from person to person, hence it is challenging to get 100% accuracy.

From the handwritten Kannada numeral using pencil obtained result is 82.36%, for pen it is 88.01%, for sketch 93.71% accuracies, for grid (sketch) it is 86.44% and for the printed digits nearly 100% of accuracy.

It is observed from all the results that of hand written using sketch is more efficient when more compared with the pen and pencil. Which is shown in table 4.1 and graphical representation is shown in plot.4.2

4.1 Using Pencil:

For hand written Kannada numeral using pencil it is up to 82.36%.it is because of thinness and lightness of numerals by using pencil .from this technique it is not possible to achieve 100% matching accuracy. The input pencil image and its output is shown in figure.4.1 and 4.2

Input Image:



Fig.4.1 Numerals Written Using Pencil

Result:

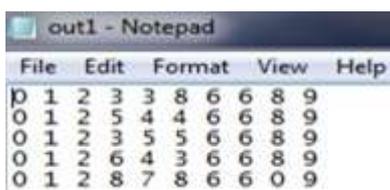


Fig.4.2 Result for numerals written using pencil

4.2 Using Pen:

For hand written Kannada numeral using pen it is up to 88.01%. This is due to increase in thickness and increase in darkness of numerals, as compared to pencil pen hand written letters are more efficient. But still it is not possible to achieve 100% matching accuracy. Because of numerals written using pen are neither thick nor thin. The Input image written using pen and its output shown in figure.4.3 and 4.4

Input Image:



Fig.4.3 Numerals Written Using Pen

Result:

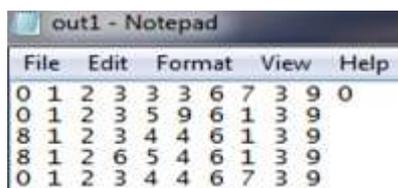


Fig.4.4 Result for Numerals Written Using Pen

4.3 Using Sketch:

For hand written Kannada numeral using sketch it is 93.71%. So as compared to pencil and pen sketch hand written is more efficient. But still 100% matching accuracy is not possible, due to uncommon shapes of hand written numerals. The Input sketch image and its output shown in figure.4.5 and 4.6

Input Image:



Fig.4.5 Numerals Written Using Sketch

Result:

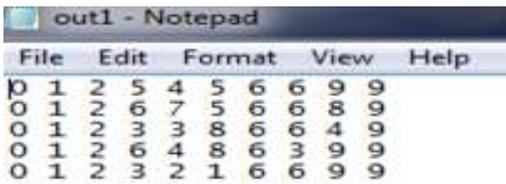


Fig.4.6 Result for Numerals Written Using Sketch

Input Image:



Fig.4.9 Numerals Written Using Grid (sketch)

4.4 Using Printed Digits:

For machine printed Kannada numeral it is nearly 100% accuracy. And because of increase in thickness and darkness of numerals using printed digits as compared with pencil, pen, and sketch matching accuracy can be increased. The Input of machine printed image and its output shown in figure.4.7 and 4.8

Input Image:



Fig.4.7 Numerals Written Using Printed Digits

Result:

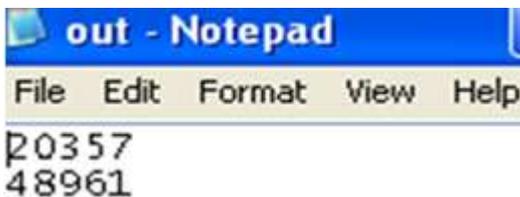


Fig.4.8 Result for Numerals Written Using printed digits

4.5 Using grid (sketch):

To retain part of the character, such as horizontal segmentation or vertical segmentation of character size 8X10 grids on the paper is considered. By considering this the writer wrote the numerals using sketch in specified grids, like one digit per grid. Here, 86.44% is achieved, because of some pixels in the line present between characters, hence while matching input image with template image less accuracy is obtained even though numerals written using sketch. The Input grid (sketch) image and its output shown in figure.4.9 and 4.10

Result:

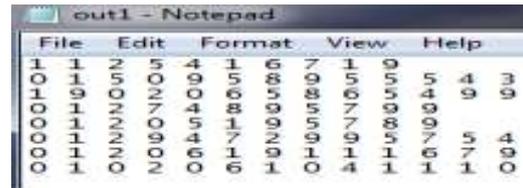
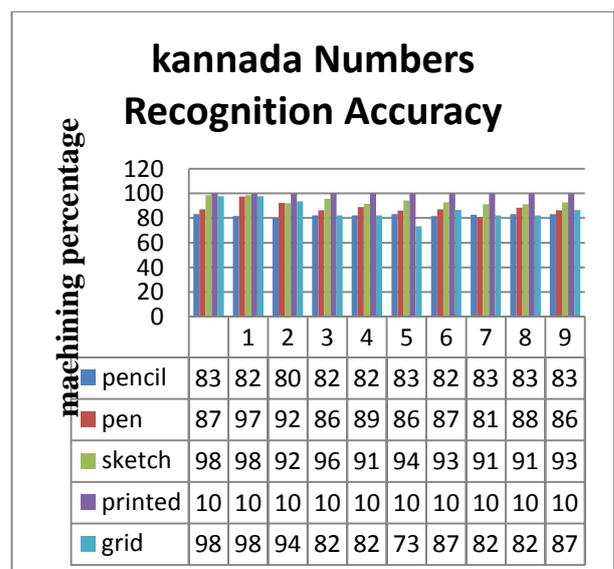


Fig.4.10 Result for Numerals Written Using Grid (Sketch)

Table 4.1: Kannada Numbers Recognition Accuracy

Kannada Numbers	No Of Samples Tested	Percentage of recognition (pencil)	Percentage of recognition (pen)	Percentage of recognition (sketch)	Percentage of recognition (printed)	Percentage of recognition (grid)
೦	200	83.3	87.1	98.4	100	97.7
೧	200	81.6	97.4	98.4	100	97.7
೨	200	80.0	92.3	92.0	100	93.7
೩	200	82.3	86.4	95.7	100	82.2
೪	200	82.3	88.4	91.4	100	82.2
೫	200	83.3	86.0	94.4	100	73.3
೬	200	81.6	87.1	92.7	100	86.6
೭	200	82.6	80.7	91.2	100	82.2
೮	200	83.3	88.4	90.6	100	82.2
೯	200	83.3	86.3	92.8	100	86.6
overall	2000	82.36	88.01	93.76	100	86.44



Plot 4.2 Kannada Numerals Recognition Accuracy

5. Summary of Result

1. Average of percentage matching for Standard Type-WRITTEN Fonts: $\approx 100\%$
2. Average of percentage matching for hand WRITTEN pencil : 82.36%
3. Average of percentage matching for hand WRITTEN pen : 88.01%
4. Average of percentage matching for hand WRITTEN sketch : 93.76%
5. Average of percentage matching for hand WRITTEN sketch (grid) : 86.44%

6. Conclusion

This work presented an approach for hand written Kannada numeral recognition under the method of template and convolution based on the line segmentation and horizontal projection with respect to computation time and threshold. it recognition the hand written Kannada numbers whenever there is multiple contents in the numeral pattern and independent of size, slant, orientation, numeral gap and translation.

References

- [1] Ashwin T.V and P.S Sastry, "A font and size independent OCR system for printed Kannada using SVM", Sadhana, vol. 27, Part 1, February 2002, pp. 35–58.
- [2] Postl. W, "Detection of linear oblique structures and skew scan in digitized Documents", ICPR, IEEE CS, 1986.
- [3] A font and size-independent OCR system for printed Kannada documents using support vector machines T. V. Ashwin, P. S. Sastry.
- [4] P. Berkes, "Handwritten Digit Recognition with Nonlinear Fisher Discriminate Analysis", Artificial Neural Networks: Formal Models and Their Applications - ICANN 2005, 2005, pp. 285–287.
- [5] Ashwin TV, Sastry P S 2002, A font and size independent OCR system for printed Kannada documents using support vector machines. Sadhana 27: 35–58
- [6] Chong Chee-Way, Raveendran P, Mukundan R 2003 A comparative analysis of algorithms for fast computation of Zernike moments. Pattern Rec. 36: 731–742
- [7] Recognition of handwritten Bengali characters: a novel multistage approach " by A.F.R. Rahman, R. Rahman, M.C. Fairhurst International Journal of Computer Applications (0975 – 8887) Volume 26– No.9, July 2011
- [8] Large Vocabulary Off-Line Handwriting Recognition: A Survey (2003) by Alessandro L. Koerich, Robert Sabourin, Ching Y. Suen Pattern Anal Applic (2003) 6: 97–121 DOI 10.1007/s10044-002-0169-3
- [9] A font and size-independent OCR system for printed Kannada documents using support vector machines T V ASHWIN and P S

SASTRY Sadhana Vol. 27, Part 1, February 2002, pp. 35–58. © Printed in India

[10] Spatial Features for Multi-Font/Multi-Size Kannada Numerals and Vowels Recognition B.V. Dhandra, 2 Mallikarjun Hangarge, 1 Gururaj Mukarambi

[11] Offline Kannada Handwritten Numeral Recognition: Holistic Approach by Vishweshwarayya C. Hallur¹ and R. S. Hegadi² Proceeding of the second international conference "emerging research in computing, information, communication and application" ERCICA 2014 ISBN:9789351072607