DESIGN AND IMPLEMENTATION OF OPTICAL CHARACTER RECOGNITION USING TEMPLATE MATCHING FOR MULTI FONTS / SIZE

Nikhil Rajiv Pai[1], Vijaykumar S. Kolkure[2]

[1]M.E. (Electronics, Appeared), Department Of Electronics Engineering,Bharatratna Indira Gandhi College of Engineering, Affiliated to Solapur University, Solapur, Maharashtra, India.
[2]Assistant Professor, Department Of Electronics Engineering, Bharatratna Indira Gandhi College of Engineering, Affiliated to Solapur University, Solapur, Maharashtra, India.

Abstract

Optical character recognition (OCR) is an efficient way of converting scanned image into machine code which can further edit. There are variety of methods have been implemented in the field of character recognition. This paper proposes Optical character recognition by using Template Matching. The templates formed, having variety of fonts and size .In this proposed system, Image pre-processing, Feature extraction and classification algorithms have been implemented so as to build an excellent character recognition technique for different scripts .Result of this approach is also discussed in this paper. This system is implemented in Matlab.

Keywords- OCR, Feature Extraction, Classification

---***---

1. INTRODUCTION

Basically Optical character recognition (OCR) having two main types: 1.Off-line Character Recognition 2.On-line Character Recognition. On-line character recognition is simple and effective way for recognition, as the input to this system is data which is collected online. These systems recognize a character while the user is writing with an online writing device. Off-line character recognition simply known as OCR. Off-line character recognition is somewhat difficult task, as this system has to recognize handwritten characters. The handwritten characters having different size and fonts. It varies from person to person. Therefore it is somewhat complicated to identify a character. This system proposes an effective approach 'Template Matching'. In this variety of templates have been formed with variety of fonts and size. This method can recognize a character irrespective of its style, size and font. The input scanned image is compared with standard templates. The most closely matched template is declared as matched template. Thus character can recognized. The input scanned image undergoes the process of pre-processing including binarization, segmentation, image resizing etc.

1.1 Binarization

In this stage, the input scanned color image is converted into grayscale image. The color image is having parameters such as, R-G-B (Red-Green- Blue). Firstly the R-G-B values of input scanned image are calculated. After calculation of R-G-B values, 30% of Red, 59% of Green and 11 % of blue values are separated. The separated values are added together .The obtained addition is the expected grayscale value. Then the grayscale histogram of input scanned image F(x, y) is calculated. After binarization, the binarized image is forwarded for segmentation, so as to separate the sub parts of an input image.

1.2 Image Resizing

After the process of binarization, segmentation etc., the image should be rearranged in all directions. The image resizing is done from left, right, top, bottom side of input image. Once the resizing is done, then templates of different fonts, styles, size are formed.

1.3 Template Formation and Matching

This is the base of this approach. In this, different templates are designed in such a way that it is having variety of font size and styles. Due to this variety, characters can recognize by the system irrespective of its font size and font style.

Whatever the input characters to be recognized are compared with these standard templates. First of all, area of template is calculated from all directions. It includes calculation of distance between border and boundary of characters. This calculation of distance between border and boundary of characters is done for both templates and characters to be recognized. Whenever the character is to be recognized, a respective character is marked from all directions that means from top, bottom, right, left directions. This marking specifies the boundaries of that respective character. Then that character is compared with different templates. This comparison is done for the distance between border and boundary of characters, which has been already calculated. Suppose, the distance between border and boundary of a test character is Td and the same for reference template is Rd.

Then during comparison, every time the difference between each Rd and Td is calculated. The minimum difference between Td and Rd indicates likely matching elements. Thus the template having minimum distance is considered as best matched template. In this way a particular character gets recognized.

2. RESULTS AND DISCUSSION

Here Optical Character Recognition technique uses Template Matching approach. This system is implemented in Matlab. Here we have considered 5 different cases. In this, each character set having different font size, styles etc. The purpose of considering such different character sets is to test the accuracy of OCR.

n i k h i l k a v i t a r a j i v p a i
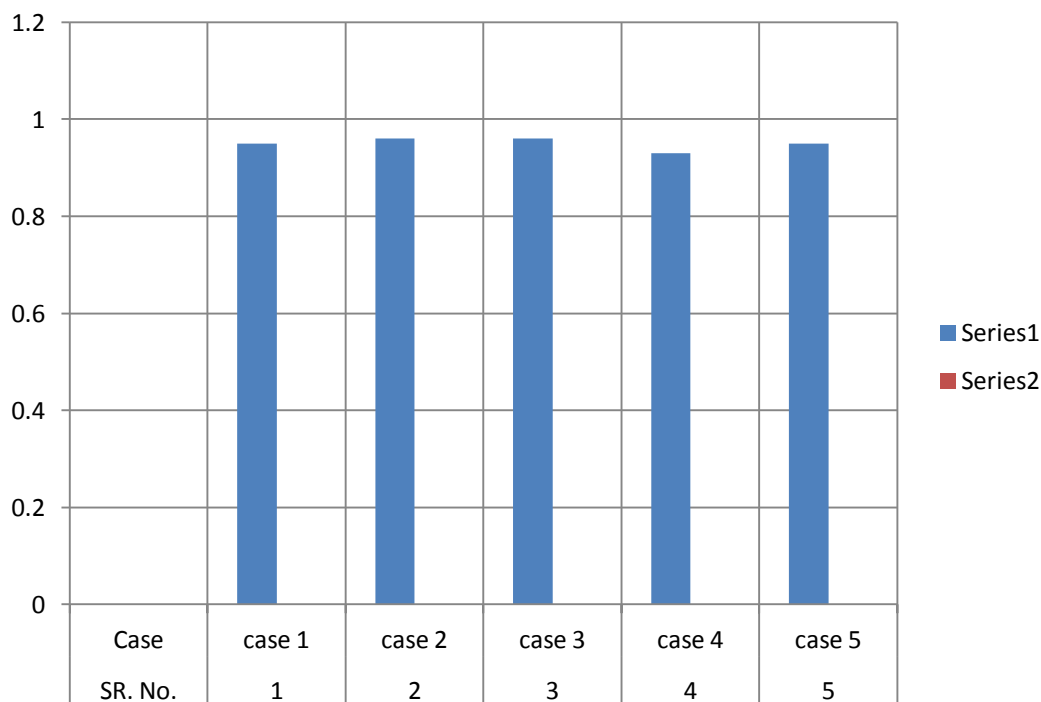
The result of above mentioned cases are shown below.

Case 1: constant font size and font style
N i k h i l K A V i t a R a J I V p a i

Case 2: variable font style and constant font size
n i k h i l k a v i t a r a j i v p a i

Case 3: variable font size and constant font style
N i k h i l *K A V* i t a **R** a J I V p a i

Case 4: Variable font style and font size

N I K H I L K A V I T A R A J I V P A I

Case 5: Italic, Bold font style and constant font size

Fig 1: Results of Experiment

Table 1: Recognition rate of different cases

Sr. No.	Case	Recognition Rate
01	Case 1	95%
02	Case 2	96%
03	Case 3	96%
04	Case 4	93%
05	Case 5	95%

Here we have considered different cases for recognition. In all cases, the characters are differs in font style and size. As multi fonts/ size templates have been designed in this approach, the above characters get compared with the standard templates. By comparing with the templates, it has been easy task to recognize a character for the system. Therefore in all cases accuracy of OCR is above 90%.

3. CONCLUSION

For different cases of characters, OCR system has been implemented. As this proposed system is designed with the standard templates, characters can recognize effectively and more efficiently by comparing them with templates. OCR using template matching can identify the characters more accurately irrespective of its font size and style. For mentioned cases OCR gets recognition rate above 90%. Thus Optical Character Recognition using Template Matching can become a smooth way for character recognition.

REFERENCES

[1]. Jagruti Chandarana, Mayank Kapadia Optical Character Recognition International Journal of Emerging Technology and Advanced Engineering Volume 4, Issue 5, May 2014

[2]. Binod Kumar Prasad, GoutamSanyal A model Approach to Off-line English Character Recognition International Journal of Scientific and Research Publications, Volume 2, Issue 6, June 2012

[3]. SandeepTiwari , Shivangi Mishra, Priyank Bhatia, Praveen Km. Yadav[May 2013] Optical Character Recognition using MATLAB International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE)Volume 2, Issue 5, May 2013

[4]. Majida Ali Abed Hamid Ali Abed Alasadi Simplifying Handwritten Characters Recognition Using a Particle Swarm Optimization Approach European Academic Research, Volume I, Issue 5/ August 2013

[5]. Mahesh Goyani, Harsh Dani, Chahna Dixit Handwritten Character Recognition – A Comprehensive Review International Journal of Research in Computer and Communication Technology, Volume 2, Issue 9, September - 2013

[6]. Sushree Sangita Patnaik and Anup Kumar Panda Particle Swarm Optimization and Bacterial Foraging OptimizationTechniques for Optimal Current Harmonic Mitigation by Employing Active Power Filter Applied Computational Intelligence and Soft Computing Volume 2012, Article ID 897127.

[7]. Pritpal Singh, Sumit Budhiraja Feature Extraction and Classification Techniques in O.C.R. Systems for Handwritten Gurumukhi Script – A Survey International Journal of Engineering Research and Applications (IJERA) Volume 1, Issue 4, pp. 1736-1739.

[8]. Lipi Shah, Ripal Patel, Shreyal Patel, Jay Maniar Skew Detection and Correction for Gujarati Printed and Handwritten Character using Linear Regression International Journal of Advanced Research in Computer Science and Software Engineering
Volume 4, Issue 1, January 2014

[9]. Ritesh Kapoor, Sonia Gupta, and C.M. Sharma, "Multi-font/size character recognition and document scanning," Int. J. of Computer Application, volume 23, no.1, pp. 21-24, 2011.

BIOGRAPHIES

Nikhil Rajiv Pai [1] currently pursuing M.E. (Electronics) From Bharatratna Indira Gandhi College of Engineering, Solapur, Maharashtra, India. His area of interest is image processing, MATlab.

Vijaykumar S. Kolkure [2] has completed M.E. (Electronics) from W.C. Sangli, Maharashtra, India. He has 10 years of teaching experience. Currently he is working as Assistant Professor at Bharatratna Indira Gandhi College of Engineering, Solapur, Maharashtra, India. His area of interest is Image Processing, Video Processing.