



RECOGNITION OF DISTORTED CHARACTER USING EDGE DETECTION ALGORITHM

S.K.Thilagavathy¹ and Dr.R.Indra Gandhi²

Research Scholar, Department of Computer Applications, GKM College of Engineering and Technology, Chennai, ¹

Professor, Department of Computer Applications, GKM College of Engineering and Technology, Chennai, Tamil India ²

ABSTRACT: Character recognition is an ever ending research application in the real world. Each character recognition should be accurate. So that it leads to understand the exact meaning and concept. Analyzing the distorted character is quite complicated work. In some unique languages like Tamil, Telugu and Malayalam the distorted character may resemble like some other character. In this paper analysis of each character is done even though it is inconsistent in shape and irrespectively distorted.

KEYWORDS: Character recognition, OCR, Edge Detection Algorithm

1. INTRODUCTION

Optical character recognition has become one of the most successful applications of technology in the field of pattern recognition and artificial intelligence. Both hand written and printed characters may be recognized, but the performance is directly dependent upon the quality of the input documents. Recognized characters which are inconsistency in shape and Irrespective of distortions are reproducing the actual characters from distorted documents based on algorithms and methods. And the steps involved in character recognition comprise pre-processing, segmentation feature extraction and classification.

2. REVIEW OF LITERATURE

According to the Line Eikvil Optical Character Recognition deals with the problem of recognizing optically processed characters. Optical recognition is performed off-line after the writing or printing has been completed, as opposed to on-line recognition where the computer recognizes the characters as they are drawn. Both hand printed and printed characters may be identified, but the performance is pursuance dependent upon the quality of the input documents. The Character recognition is classified into On-line and Off-line. The off-line is further as Single characters and Handwritten script. The Single character is divided into Printed and Handwritten. Like wise handwritten script is divided into Recognition and Verification[1]. The idea behind an OCR is to identify and analyses of a document image by dividing the page into line elements, further sub-dividing into words, and then into characters. These characters are compared with image patterns to state the probable characters. And particularly in Tamil handwritten OCR is more complicated than other related works. This is because Tamil letters have more angles and modifiers[2]. Recognition system works well for simple language like English. It has only 26 character sets. And for standard text there are 52 numbers of characters including capital and small letters. But a complex but organized language like Telugu, OCR system is still in introductory level [3]. But Dyashankar Singh, Sajay Kr. Singh and Dr.MitreyeeDutta speaks about that Character recognition process is dependent upon number of factors like various font sizes, noise, broken lines or characters etc. and these factors influence the results of recognition system[4].Based on the zone-wise character are also classified and identified. [5].Many diverse algorithms /schemes for handwritten character recognition[6,7] exist and each of these has its own merits and demerits. Some of them used Back Propagation Algorithm[8,9,10], Template matching algorithm [11] and structural analysis [11,12]etc. Here I had used the Edge Detection Algorithm to reproduce the accurate character from the distorted character.

3. MAJOR STAGES OF OCR

- 1) Pre-processing,
- 2) Segmentation
- 3) Feature Extraction,
- 4) Recognition.

In Devanagari Character Recognition, the character recognition is one of the important tasks in pattern recognition. Character recognition process depends upon number of factors like various font sizes, noise, broken lines or characters etc and these factors influence the results of recognition system. There are four different phases in optical character recognition system, namely: preprocessing stage, segmentation, feature extraction and character recognition.

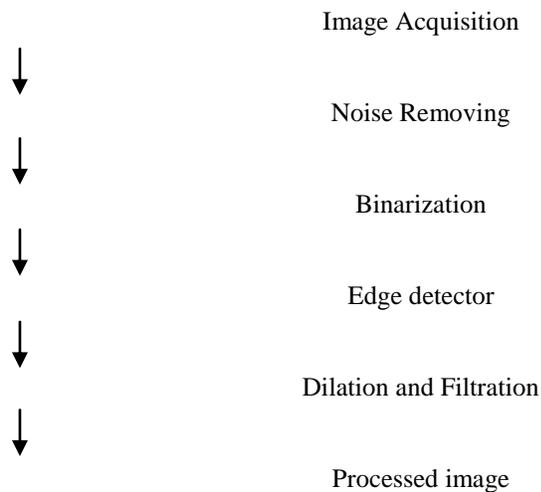


Figure1: Various Stages of Process

3.1 Preprocessing Stage

Preprocessing is an important step of applying a number of procedures for smoothing, enhancing, filtering etc, for making a image usable by subsequent algorithm in order to improve their readability for optical character recognition software. The pre-processing is a series of operations performed on scanned input image. It essentially enhances the image rendering it suitable for segmentation. The role of pre-processing is to segment the interesting pattern from the background. Generally, noise filtering, smoothing and normalization should be done in this step. The pre-processing also defines a compact representation of the pattern. Binarization process converts a gray scale image into a binary image.

3.1.1 Image Acquisition

In Image acquisition, figure 2 shows the recognition system acquires a scanned image as an input image. The image should have a specific format such as JPEG, BMT, *etc.* This image is acquired through a scanner, digital camera or any other suitable digital input device.

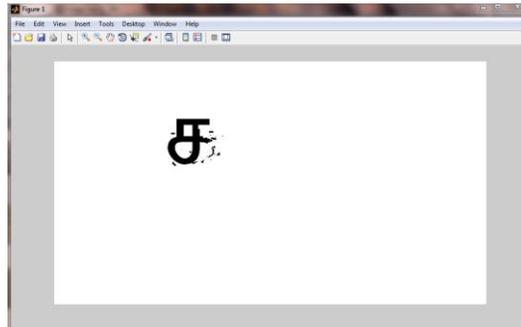


Figure 2 Image Acquisition

3.1.2 Noise Elimination

Noise in image is a major obstruction in pattern recognition errands. Noise degrades the image quality. Noise can occur at different stages like image capturing, transmission and compression. Different filters and morphological operations are available for removing image noise. Figure 3 shows the Noise elimination here. The Noise elimination is also called as smoothing. It can be used to reduce fine textured noise and to improve the quality of the Image. The techniques like morphological operations are used to connect unconnected pixels, to remove isolated pixels and also in smoothing pixels boundary.

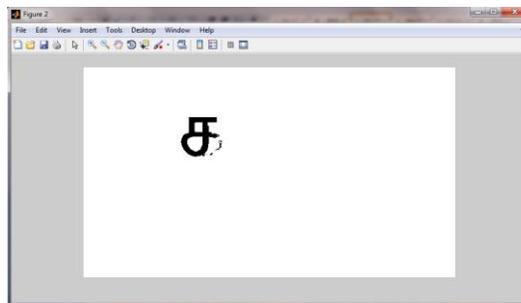


Figure 3 Smoothing

3.1.3 Binarization

Renovation of a gray-scale image into a binary image is called as binarization or thresholding. There are two approaches for conversion of gray level image to binary form; i.e. global threshold and local or adaptive threshold. Global threshold selects single threshold value based on estimation of the background level from the intensity histogram of the image. Local or adaptive threshold uses different values for each pixel according to the local area information. The purpose of binarization is to identify the extent of objects and also to concentrate on the shape analysis that is shown in figure 4.

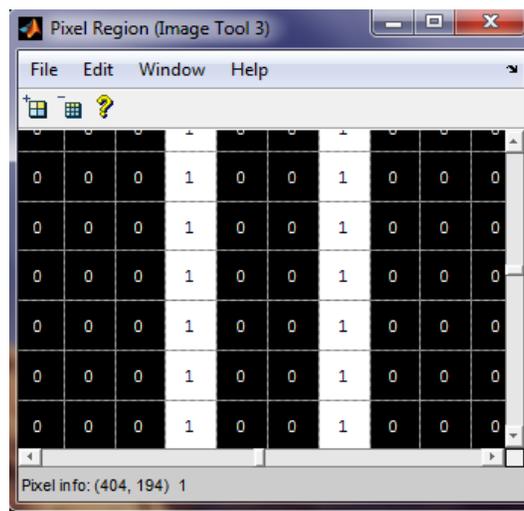


Figure 4 Binarization

3.2 SEGMENTATION

Segmentation is one of the most important and essential process that decides the success rate of character recognition system. Segmentation is the process of partitioning an image / document into disjoint and homogeneous regions. This task is attained by finding the boundaries. There are several approaches for finding the character bounds. In this stage, an image of sequence of characters is decomposed into sub-images of individual character

3.3 FEATURE EXTRACTION

In feature extraction stage each character is represented as a feature vector, which becomes its identity. The major goal of feature extraction is to extract a set of features, which maximizes the recognition rate with the least amount of elements.

In this stage, the features of the characters that are crucial for classifying them at recognition stage are extracted. This is an important stage as its effective functioning improves the recognition rate and reduces the misclassification. Diagonal feature extraction scheme for recognizing off-line handwritten characters is proposed in this work. The algorithm's are also used to obtain the extraction.

3.3.1 Character Extraction Algorithm

1. Create a Traverse List: - List of pixels which have been already traversed. This list is initially empty.
2. Scan row Pixel-by-Pixel.
3. Whenever we get a black pixel check whether the pixel is already in the traverse list, if it is simply ignore and move on else apply Edge detection Algorithm.
4. Add the List of Pixels returned by Edge detection Algorithm to Traverse List.
5. Continue the steps 2 - 5 for all rows

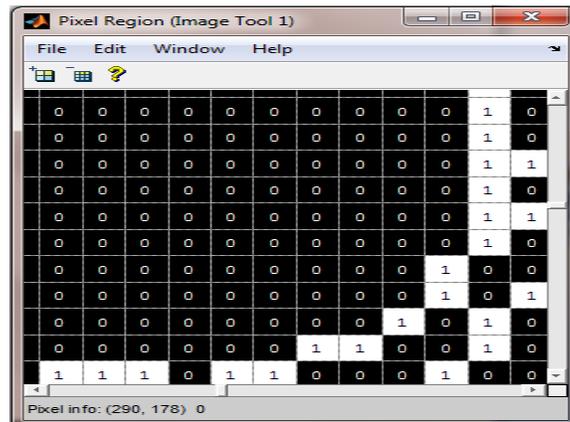


Figure 5: Character Extraction Algorithm

3.3.2 Edge Detection Algorithm

The Edge Detection Algorithm is shown in figure 6. The Edge detection algorithm has a list called traverse list. It is the list of pixel already traversed by the algorithm. The Edge Detection algorithm terminates when it has covered all the pixels of the character as every pixel's position would be in Traverse List so any further call to Edge Detection is prevented. The Edge Detection algorithm terminates when it has covered all the pixels of the Character as every pixel's position would be in Traverse List so any further call to Edge Detection is prevented.

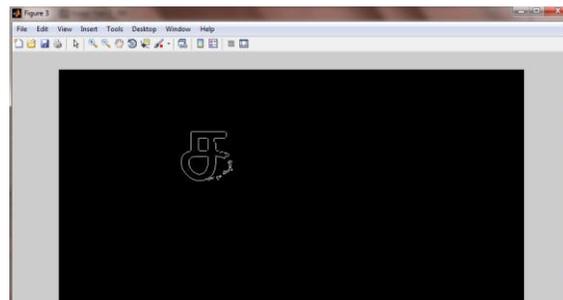


Figure 6: Edge Detection Algorithm

3.4 RECOGNITION

After processing the algorithm, recognition of the character is easily obtained. Figure 7 will display accurate result after comparing with database.

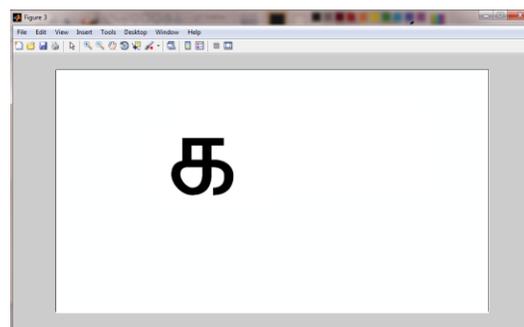


Figure 7: Recognized character from database



4. CONCLUSION

The character recognition is performed in accurately even though it is inconsistent in shape and irrespectively distorted. The character is recognized effective and reliable manner using Algorithms. The essential process like dilation and filtration are also used to find the character in effectively and excellent result was obtained.

REFERENCES

- [1] In OCR Optical Character Recognition by Line Eikvil at December 1993
- [2] A Survey on Tamil Handwritten Character Recognition using OCR Techniques by M. Antony Robert Raj, Dr.S.Abirami in David C. Wyld, et al. (Eds): CCSEA, SEA, CLOUD, DKMP, CS & IT 05, pp. 115–127, 2012.
- [3] Dyashankar Singh, Sajay Kr. Singh, Dr. (Mrs) Mitreyee Dutta, Hand Written Character Recognition Using Twelve Directional Feature Input and Neural Network – 2010 International Journal of Computer Applications (0975 – 8887) vol. 1 – No. 3.
- [4] A Survey on Handwritten and Printed Kannada Numeral Recognition Technique by Vishweshwarayya C Hallur , Avinash A Malawade and Seema December G Itagi - Vol. 3 No. 4 (December 2012) © International Journal of Advancements in Technology.
- [5] Handwritten Text Recognition System for Automatic Reading of Historical Arabic Manuscripts by M. S. Farag - International Journal of Computer Applications (0975 – 8887) Volume 60– No.13, 2012
- [6] A Complete Bangla Ocr System for Printed Characters by Md. Mahbub Alam And Dr. M. Abul Kashem Copyright © 2010 JCIT, ISSN 2078-5828 (Print), ISSN 2218-5224 (Online), Volume 01, Issue 01, Manuscript Code: 100707
- [7] A Review on the Various Techniques used for Optical Character Recognition by Pranob K Charles, V. Harish, M. Swathi, CH. Deepthi - International Journal of Engineering Research and Applications (IJERA) ISSN: 2248-9622 Vol. 2, Issue 1, Jan-Feb 2012, pp.659-662
- [8] Plamondon, R., Srihari, S. N.: On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. IEEE Trans. Patt. Anal. and Mach. Intell., Vol. 22 (2000) 63-84
- [9] Arica, N., Yarman-Vural, F.: An Overview of Character Recognition Focused on Off-line Handwriting. IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, Vol. 31 (2001) 216 – 233
- [10] On-Line Tamil Hand Written Character Recognition Using Kohonen Neural Network by S. Peyarajan and R. Indra Gandhi - Vol 02, Issue 02, July, 2011 Research Journal of Computer Systems Engineering- An International Journal
- [11] Multiple Algorithms for Handwritten Character Recognition by Jonathan J. HULL, Alan COMMIKE and Tin-Kam HO - Int. Workshop on Frontiers in Handwriting Recognition, Montreal, Canada, April 2-3, 1990.
- [12] An Approach for Structural Feature Extraction for Distorted Tamil Character Recognition by Nirase Fathima Abubacker and Indra Gandhi Raman - International Journal of Computer Applications (0975 – 8887) Volume 22– No.4, May 2011