# Efficient Tamil Numbers Identification Using Canny Edge Detection and Neural Network Algorithm

**P. Selvakumar\*, S. Hari Ganesh\*\***

**ABSTRACT**

Character recognition is a perpetually finishing research application in this present reality. Every character recognition should be exact. With the goal that it prompts comprehend the accurate importance and idea. Dissecting the misshaped character is quite complicated work. Edge is a fundamental component of picture. The picture edges incorporate rich data that is extremely huge for acquiring the image characteristic by item recognition. Edge recognition alludes to the procedure of recognizing and finding sharp discontinuities in an image. So, edge location is a key stride in picture investigation and it is the key of comprehending numerous complex problems. In this paper examination of every character is done despite the fact that it is conflicting fit as a fiddle and irrespectively bended. Tamil characters for numerical are broke down in this work. Shrewd edge discovery parts a noteworthy piece of this character recognition and neural system is utilized to enhance order exactness.

*Keywords:* OCR, Character Recognition, Border Estimation, Detecting Edges.

## 1. INTRODUCTION

Character recognition has an extraordinary potential in information and word handling for occasion, computerized postal address and ZIP code perusing, information obtaining in bank checks, preparing of archieved institutional records, and so on. In the Recent years, optical character recognition(OCR) has picked up a force following the requirement for changing over the checked pictures into PC unmistakable organizations, for example, content records has expanded applications. OCR is a standout amongst the most entrancing and testing zones of example recognition with different useful applications. The procedure of character recognition includes extraction of characterized qualities called components to order an obscure character into one of the known classes. In this manner, OCR includes two procedures: 1.feature extraction 2.classification. The procedure of character recognition turns out to be extremely intense on account of Indian dialects like Tamil, Kannada, Arabic and Bangla [1] [3-5]. Numerous between class conditions exist in Tamil. In Tamil dialect, numerous letters resemble the other alike. So order turns into a major test [11].

With rise of the advanced substance, requirement for the improvement of an elite OCR motor has ended up key. OCR research works have been embraced by a few scientists who go for building up an elite OCR cruise.

The thought behind an OCR is to distinguish and break down a record picture by isolating the page into line components, further sub-partitioning into words, and after that into characters. These characters are contrasted with picture designs with anticipate the plausible characters. Recognition of characters should be possible either from printed reports or from transcribed records. Written by hand record recognition should be possible disconnected

\*     Research Scholar, Department of Computer Science, Research & Development Centre, Bharathiar University Coimbatore, Tamilnadu, India, *Email: selvagacalr@gmail.com*

\*\*    Assistant Professor, Department of Computer Science, HH The Rajah's College (Autonomous) Pudukkottai, Tamilnadu, India, *Email: hariganesh17@gmail.com*

from the net or on the web. Disconnected from the net character recognition is more confused than on the web. Specifically, Tamil manually written OCR is more confused than other related works. This is on account of Tamil letters have more points and modifiers.

Also, Tamil script contains huge number of character sets. An aggregate of 247 characters; comprising of 216 compound characters, 18 consonants, 12 vowels and one uncommon character. Challenges that inquire about face amid recognition procedure are because of the bends in the characters, number of strokes and gaps, sliding characters, varying composition styles so on. The strides required in character recognition involve pre-handling, division, highlight extraction and order. We propose the proficient Tamil textual style for numbers recognition utilizing Edge highlights and neural system based arrangement.

## 2.   RELATED STUDY

The Character Recognition manages the issue of perceiving optically processed characters [6]. Optical acknowledgment is performed disconnected from the net after the written work or printing has been finished, rather than on-line recognition where the PC perceives the characters as they are drawn [7]. Both hand printed and printed characters possibly recognized, however the execution is compatibility subordinate upon the nature of the info archives. The Character acknowledgment is grouped into On-line and Off-line [8]. The disconnected from the net is further as Single characters and handwritten script. The Single character is separated into Printed and Handwritten. Like savvy manually written script is isolated into Recognition and Verification. The thought behind an OCR is to distinguish and investigations of a report picture by separating the page into line components, further sub-partitioning into words, and after that into characters.

These characters are contrasted with picture designs with state the plausible characters. Furthermore, especially in Tamil manually written OCR is more confused than other related works. This is on the grounds that Tamil letters have more edges and modifiers. Acknowledgment framework functions admirably for straightforward dialect like English. It has just 26 character sets. Also, for standard content there are 52 quantities of characters including capital and little letters. Be that as it may, a complex however sorted out dialect like Telugu, OCR framework is still in early on level. In any case, Dyashankar Singh, Sajay Kr. Singh and Dr. Mitreyee Dutta talks about that Character acknowledgment procedure is endless supply of components like different text dimensions, clamor, broken lines or characters and so on and these variables impact the aftereffects of acknowledgment framework [2]. Taking into account the zone-wise character are additionally arranged and recognized. Numerous assorted calculations/plans for written by hand character acknowledgment exist and each of these has its own particular benefits and negative marks [9]. Some of them utilized Back Propagation Algorithm, Template coordinating calculation and auxiliary examination and so on. Here I had utilized the Edge Detection Algorithm to duplicate the precise character from the mutilated character [10].

## 3.   PROPOSED WORK

### 3.1. Preprocessing Stage

Preprocessing is a critical stride of applying various methodologies for smoothing, upgrading, sifting and so forth, to make a picture usable by resulting calculation with a specific end goal to enhance their coherence for optical character acknowledgment programming. The pre-handling is a progression of operations performed on checked information picture. It basically improves the picture rendering it appropriate for division.

The part of pre-preparing is to fragment the intriguing example from the foundation. By and large, clamor sifting, smoothing and standardization ought to be done in this progression. The pre-handling additionally characterizes a conservative representation of the example. Binarization process changes over a dark scale picture into a twofold picture.

### 3.2. Image Acquisition

In Image acquisition, the following figure demonstrates the acknowledgment framework gains an examined picture as an info picture. The picture ought to have a particular configuration, for example, JPEG, BMT, and so forth. This picture is obtained through a scanner, computerized camera or some other appropriate advanced information gadget.

### 3.3. Noise Elimination

Noise in picture is a noteworthy obstacle in example acknowledgment errands. Noise corrupts the picture quality. Noise can happen at various stages like picture catching, transmission and pressure. Distinctive channels and morphological operations are accessible for evacuating picture commotion. Figure 3 demonstrates the Noise end here. The Noise end is additionally called as smoothing. It can be utilized to lessen fine textured Noise and to enhance the nature of the Image. The procedures like morphological operations are utilized to associate detached pixels, to expel confined pixels furthermore in smoothening pixels limit.
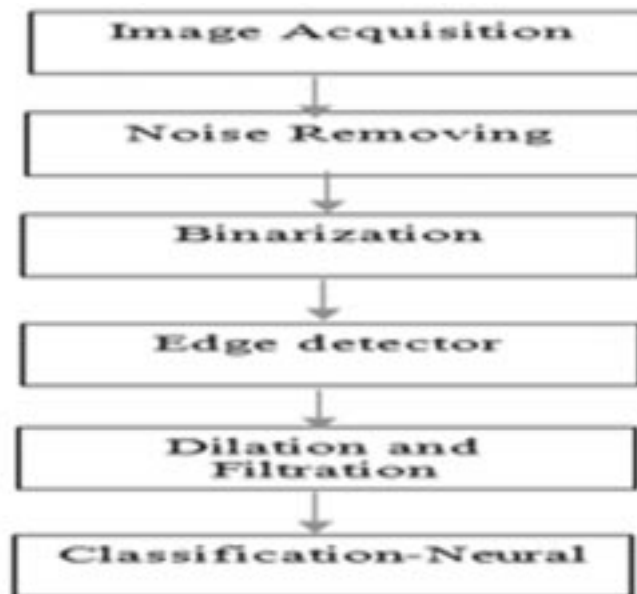


**Figure 1: Preprocessing Stages**



**Figure 2: Image Acquisition**          **Figure 3: Image Luminance**

### 3.4. Canny Edge Detection

The Canny Edge Detector is one of the most commonly used image processing tools, detecting edges in a very forceful manner. The Canny edge detector is widely considered to be the standard edge detection method in the industry. Canny saw the edge detection problem as a signal processing optimization problem, so an objective function is developed to be optimized. The solution to this problem was a rather difficult exponential function, but Canny found numerous ways to approximate and optimize the edge-searching problem. The steps in the canny edge detector are as follows:

(i)  Smoothing: Blurring of the image to remove noise.

(ii)  Finding gradients: The edges should be marked where the gradients of the image has largE magnitudes.

(iii)  Non-maximum suppression: Only local maxima should marked as edges.

(iv)  Double thresholding: Potential edges are determined by thresholding.

(v)  Edge tracking: Final edges are determined by suppressing all edges that are not connected to a very certain (strong) edge.

### 3.5. Adjusting Grey Level of a Flame Image

As the image has a wide distributed scale of the grey levels, an initial adjustment is required to redistribute using a histogram to equalize the grey levels to be spread entirely over the given image normalized in the range [0, 1]. Using the probability of the occurrence of a pixel of gray level I in the image is

$$P_x\left(i\ =\ R\ x=i\ =\frac{n_i}{n},0<i<L\right) \tag{1}$$

L – Total number of grey levels in the image.

n – Total number of pixels in the image.

$n_i$ – Total number of pixels with grey level i.

| -1 | 0 | +1 | +1 | +2 | +1 |
|----|---|----|----|----|----|
| -2 | 0 | +2 | 0  | 0  | 0  |
| -1 | 0 | +1 | -1 | -2 | -1 |
| Gx | | | Gy | | |

Figure 4. Canny Operators



Figure 5: Gray level Adjustment

The Cumulative distribution function CDF is used for redistribution and equalization given by equation (2)

$$CDF_x(i\ =\sum_{j=0}^{i} p(\ j$$

(2)

### 3.6. Smoothening the image to eliminate noise

The image considered with variation in the grayscale represented as noise is filtered using the standard convolution methods by application of a suitable mask. A Gaussian mask is selected to minimize the localization edges to preserve the edges for further detection.

### 3.7. Adjusting Thresholds for better results

The two thresholds are used to classify the pixel to belong to the edge if the magnitude of the gradient is higher than TH or if there is a path to another pixel which is above TL. All the other pixels not at the peak of the gradient are suppressed to highlight the edge.

A preliminary edge image (PEI) is formed by detecting all the pixels foreseen to belong to the edge represented with discontinuous contour. These are further eliminated by interpolation through which the gaps between the contours existing as part of the background pixels are highlighted and connected using the 4 neighborhoods.

### 3.8. Binarization and Segmentation

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
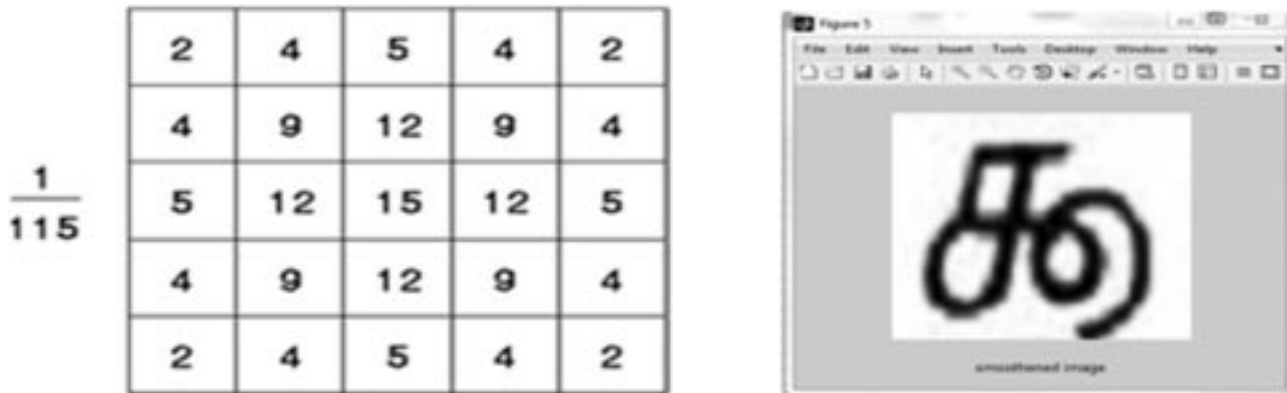


**Figure 6: Smoothened Image**
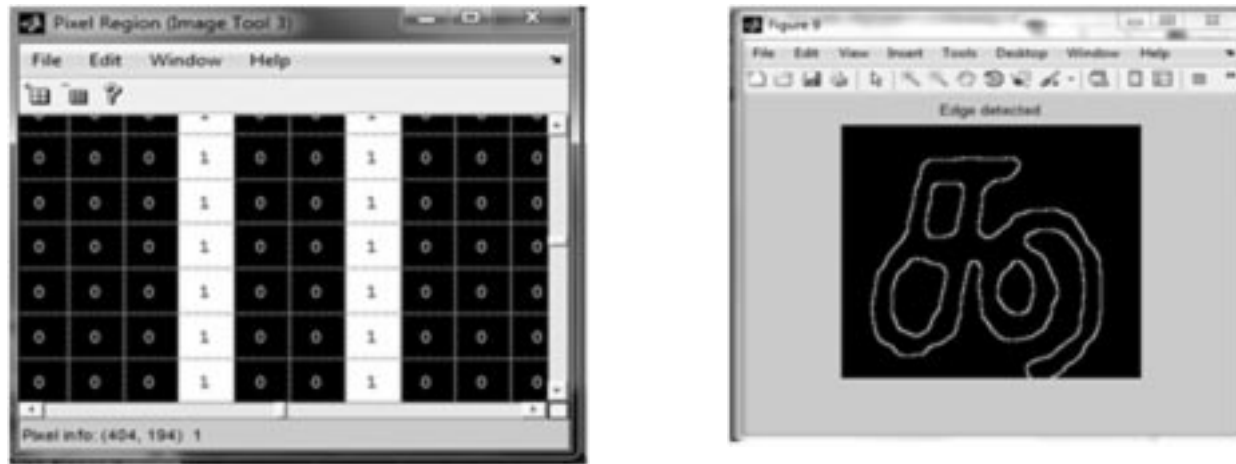


**Figure 7: Horizontal Edge Detection**

**Figure 8: Image Pixel Oriented Processing**

information. The purpose of binarization is to identify the extent of objects and also to concentrate on the shape analysis. 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.

## 6.  CONCLUSIONS

The character acknowledgment is performed in precisely despite the fact that it is conflicting fit as a fiddle and irrespectively contorted. The character is perceived compelling and dependable way utilizing Algorithms. The crucial procedure like expansion and filtration arealso used to discover the character in successfully and superb result was obtained. The proposed strategy has been connected on various obscure characters. Neural system based strategy gives the precision 85 %.

## REFERENCES

[1]    M. Antony Robert Raj, Dr.S.Abirami in David C. Wyld, et al. "A Survey on Tamil Handwritten Character Recognition using OCR Techniques" *CCSEA, SEA, CLOUD, DKMP, CS & IT*, **05**, 115–127, 2012.

[2]    Dyashankar Singh, Sajay Kr. Singh, Mitreyee Dutta., "Hand Written Character Recognition Using Twelve Directional Feature Input and Neural Network" *International Journal of Computer Applications*, **1(3)**, 2010.

[3]    Vishweshwarayya C Hallur, Avinash A Malawade and Seema December G Itagi., "A Survey on Handwritten and Printed Kannada Numeral Recognition Technique" *International Journal of Advancements in Technology*, **3(4)**, 2012.

[4]    M. S. Farag., "Handwritten Text Recognition System for Automatic Reading of Historical Arabic Manuscripts" *International Journal of Computer Applications*, **60(13)**, 2012.

[5]    Md. Mahbub Alam And Dr. M. Abul Kashem., "A Complete Bangla Ocr System for Printed Chracters" *JCIT*, **1(1)**, 2010.

[6]    Pranob K Charles,V.Harish, M.Swathi, CH. Deepthi., "A Review on the Various Techniques used for Optical Character Recognition" *International Journal of Engineering Research and Applications*, **2(1)**, 659-662, 2012.

[7]    Plamondon, R., Srihari, S. N., "On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey" *IEEE Trans. Patt. Anal. and Mach. Intell*, **22**, 63-84, 2000.

[8]    Arica, N., Yarman-Vural, F., "An Overview of Character Recognition Focused on Off-line Handwriting" *IEEE Transactions on Systems, Man, and Cybernetics*, **31**, 216 – 233, 2001.

[9]    S. Peyarajan and R. Indra Gandhi., "On-Line Tamil Hand Written Character Recognition Using Kohonen Neural Network" *Research Journal of Computer Systems Engineering- An International Journal*, **2(2)**, 2011.

[10]   Jonathan J. HULL, Alan COMMIKE and Tin-Kam HO., "Multiple Algorithms for Handwritten Character Recognition" *Int. Workshop on Frontiers in Handwriting Recognition*, 1990.

[11]   Nirase Fathima Abubacker and Indra Gandhi Raman., "An Approach for Structural Feature Extraction for Distorted Tamil Character Recognition" *International Journal of Computer Applications*, **22(4)**, 2011.