



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

Eighth Century Tamil Character Recognition

R. Prabavathi¹, R. Karthikeyan^{*2}

Assistant Professor, Dept. of CSE, Jerusalem College of Engineering, Chennai, India¹

Assistant Professor, Dept. of CSE, Bharath University, Chennai, Tamil Nadu, India²

*Corresponding Author,

ABSTRACT: Ancient Tamil character recognition is an important research and application area on pattern recognition theory, which plays an important role on realizing document. The main objective of the project is to determine characters from any given text of A-Z and Tamil consonants. A system of English and Ancient Eighth Century Tamil character recognition based on point feature extraction of the character. Eighth Century Tamil characters are recognized using Multilayer Feed Forward Neural Network has been described in our work. Pre-processing is the first step which is done using thresholding algorithm. Feature extraction forms the basic underlying part of recognizing each character during testing. Characters are classified by Back Propagation Neural Network. An Artificial Neural Network has been created and trained to diagnose a single character. A Feature Vector of each character has been used to train the Neural Network. This system will be suitable for converting handwritten documents into structural text format. The handwritten character recognition system with better-quality recognition rates will be eminently suitable for several applications including document reading and conversion of any handwritten document into structural text format.

KEYWORDS: Tamil stone inscriptions; Neural network, Median filter

I. INTRODUCTION

The ancient Tamil character recognition system includes three stages: image processing, feature extractor, and classifier. The process of ancient Tamil character recognition involves extraction of some defined characteristics called features to classify an unknown character into one of the known classes[1]. Pre-processing is primarily used to reduce variations of handwritten characters. The pre-processing stage comprises three steps: Binarization, Noise Removal. A feature extractor is essential for efficient data representation and extracting meaningful features for later processing. The extracted features are then fed to the neural network for classification. This paper demonstrates the use of neural networks for developing a system that can recognize hand-written English alphabets and ancient Tamil character. In this system, each English alphabet is ancient Tamil character represented by binary values that are used as input to a simple feature extraction system, whose output is fed to our neural network system[2]. We have proposed and developed a scheme for recognizing hand written English alphabets and ancient Tamil character. Neural network was found superior than SVM as it can provide multiple solutions. Experimental results shown that the machine has successfully recognized the alphabets with the average accuracy of 82.5%, which significant and may be acceptable in some applications.

II. PREPROCESSING

A. Binarization

Binarization is the process of converting of the pixels of an image in to binary form.Extraction of foreground from the background is called as thresholding[3]. Typically two peaks comprise the histogram gray-scale values of a document image: a high peak analogous to the white background and a smaller peak corresponding to the foreground. Fixing the threshold value is determining the one optimal value between the peaks of gray scale values. Each value of the threshold is tried and the one that maximizes the criterion is chosen from the two classes regarded as the foreground and background points.

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

B. Noise Removal

The presence of noise can cost the efficiency of the character recognition system[4]. Noise may be due the poor quality of the input image, but whatever is the cause of its presence it should be removed before further processing. We have used adaptive median filtering for the removal of thr noise from the image.

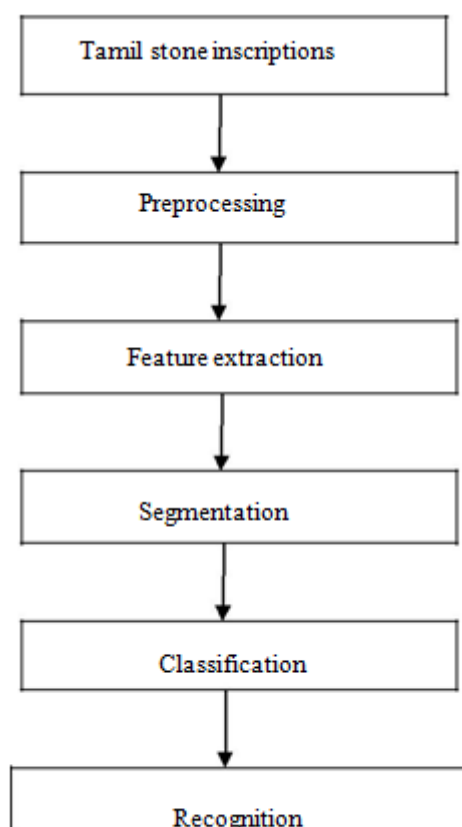


Fig 1. Block diagram

Adaptive Median filter:

Median filtering follows this basic prescription. The median filter is normally used to reduce noise in an image, somewhat like the mean filter[5]. However, it often does a better job than the mean filter of preserving useful detail in the image. This class of filter belongs to the class of edge preserving smoothing filters which are non-linear filters. This means that for two images $A(x)$ and $B(x)$:

$$\text{median}[A(x) \ B(x)] \neq \text{median}[A(x)] \ \text{Median}[B(x)]$$

These filters smooths the data while keeping the small and sharp details[6]. The median is just the middle value of all the values of the pixels in the neighborhood. Note that this is not the same as the average (or mean); instead, the median has half the Values in the neighborhood larger and half smaller. The median is a stronger "central indicator" than the average[7]. In particular, the median is hardly affected by a small number of discrepant values among the pixels in the neighborhood. Consequently, median filtering is very effective at removing various kinds of noise.



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

III. FEATURE EXTRACTION

In this stage, each pre-processed sample is subjected to feature extraction. the main features are:

- 1.color.
- 2.texture.
- 3.edges.

A. Histograms and Color Features

Initially, Calculate percentage of color present in image. An image I is a set of pixels. At each pixel Measure some m-dimensional property.

$f: R \rightarrow [0, 1]^m$

Create a finite partition of M .

B_k are subsets of M

B_k are called bins and k is the label of the bin. Convert the bins to histogram using the indicator function where x is element of the image.

B. Texture Features

Texture has no precise definition. Texture is a tactile or visual characteristic of a surface[8]. Texture primitives (or texture elements, texels) are building blocks of a texture. Texel is a small geometric pattern that is repeated frequently on some surface resulting in a texture. Segment an image into regions with the same texture, i.e. as a complement to gray level or color. Recognize or classify objects based on their texture. Find edges in an image, i.e. where the texture Changes "shape from texture" object detection, compression, synthesis.

Types of textures:

Strong Texture

Spatial interactions between primitives are somewhat regular. frequency of occurrence of primitive pairs in some spatial relationship used for description

C. Edge Information

Images are discrete functions indicating the light intensity of a scene. Edges correspond to large discontinuities in the Image. Edges are detected using the gradient, The gradient is a vector with magnitude in the u and v directions equal to the respective partial derivatives. the partial derivative of a discrete function are calculated using the Taylor's series.

IV. SEGMENTATION

Segmentation is a process of distinguishing lines, words, and even characters of a hand written or machine printed document, a crucial step as it extracts the meaningful regions for analysis[9]. There exist many sophisticated approaches for segmenting the region of interest. For handwritten document, this is quite difficult.

A. Word and Character Segmentation.

The process of word segmentation succeeds the line separation task. Most of the word segmentation issues usually concentrate on discerning the gaps between the characters to distinguish the words from one another other. This process of discriminating words emerged from the notion that the spaces between words are usually larger than the spaces between the characters. There are not many approaches to word segmentation issues dealt in the literature. In

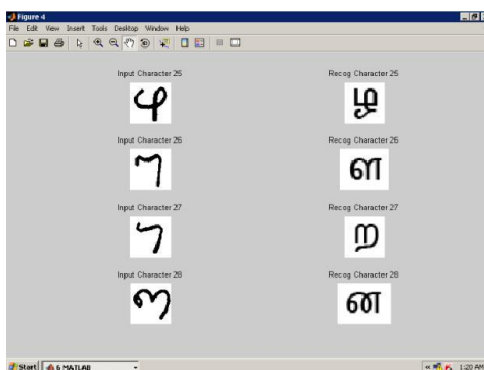
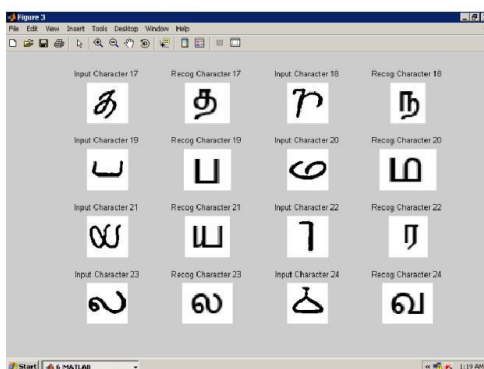
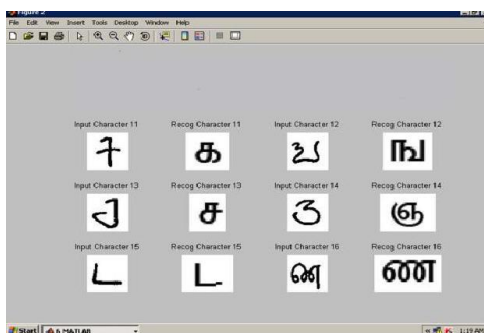
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

spite of all these perceived conceptions, exemptions are quiet common due to flourishes in writing styles with leading and trailing ligatures. Alternative methods not depending on the one-dimensional distance between components, incorporates cues that humans use. Meticulous examination of the variation of spacing between the adjacent characters as a function of the corresponding characters themselves helps reveal the writing style of the author, in terms of spacing. The segmentation scheme comprises the notion of expecting greater spaces between characters with leading and trailing ligatures[10]. Recognizing the words themselves in textual lines can itself help lead to isolation of words. Segmentation of words in to its constituent characters is touted by most recognition methods. Features like ligatures and concavity are used for determining the segmentation points.

The algorithm exploits the caps between character segments and heights of character segments too.



Output 1, 2, 3 shows the validation of Tamil stone inscriptions through Neural network classification. The proposed algorithm helps in Tamil character recognition which gives improved accuracy of recognition. Neural network compute the error derivative of the weights (EW)[11]. In other words, it must calculate how the error changes as each weight is



International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

increased or decreased slightly. The back propagation algorithm is the most widely used method for determining the EW[12]. The back-propagation algorithm is easiest to understand if all the units in the network are linear.

The algorithm computes each EW by first computing the EA, the rate at which the error changes as the activity level of a unit is changed. For output units, the EA is simply the difference between the actual and the desired output[13]. To compute the EA for a hidden unit in the layer just before the output layer, we first identify all the weights between that hidden unit and the output units to which it is connected. We then multiply those weights by the EAs of those output units and add the products. This sum equals the EA for the chosen hidden unit[14]. After calculating all the EAs in the hidden layer just before the output layer, we can compute in like fashion the EAs for other layers, moving from layer to layer in a direction opposite to the way activities propagate through the network. This is what gives back propagation its name. Once the EA has been computed for a unit, it is straight forward to compute the EW for each incoming connection of the unit[15]. The EW is the product of the EA and the activity through the incoming connection.

VI. CONCLUSION

The performance of character recognition is dependent on the accuracy of stroke recognition. The results obtained for recognition of eight century characters show that reliable classification is possible using neural network. The results also indicate the scope for further improvement, especially in the case of Tamil character recognition. Future work is directed towards incorporating a database of words for spell check at word level.

REFERENCES

- [1] S.RajaKumar, Dr.V.Subbiah Bharathi."Eighth century tamil character recognition".IEEE Transcation.
- [2] Bhuvanewari B., Hari R., Vasuki R., Suguna, "Antioxidant and antihepatotoxic activities of ethanolic extract of Solanum torvum", Asian Journal of Pharmaceutical and Clinical Research, ISSN : 0974-2441, 5(S3) (2012) pp. 147-150.
- [3] S.RajaKumar, Dr.V.Subbiah Bharathi."century identification and recognition of ancient tamil character recognition"
- [4] Subha Palaneeswari M., Abraham Sam Rajan P.M., Silambanan S., Jothimalar, "Blood lead in end-stage renal disease (ESRD) patients who were on maintenance haemodialysis", Journal of Clinical and Diagnostic Research, ISSN : 0973 - 709X, 6(10) (2012) pp.1633-1635.
- [5] K.Baranikumar, Dr. J.P.Ananth Ancient tamil character recognition using Gaussian filter.
- [6] Sathyanarayana H.P., Premkumar S., Manjula W.S., "Assessment of maximum voluntary bite force in adults with normal occlusion and different types of malocclusions", Journal of Contemporary Dental Practice, ISSN : 1526-3711, 13(4) (2012) pp.534-538.
- [7] Xiaoqing Ding, Tao Wu. "Character Independent Font Recognition on a Single Chinese Character". IEEE Transactions on pattern analysis and machine intelligence, Vol.29, NO.2, February 2007.
- [8] Seethalakshmi et.al, "Optical Character Recognition for printed Tamil text using Unicode", Journal of Zhejiang University SCIENCE, Vol. 6A No. 11, 2005.
- [9] Sukumaran V.G., Bharadwaj N., "Ceramics in dental applications", Trends in Biomaterials and Artificial Organs, ISSN : 0971-1198, 20(1) (2006) pp.7-11.
- [10] R.Ramanathan, N.Valliappan, S. Pon Mathavan, M.Gayathri, R.Priya, K.P.Soman "Generalised and Channel Independent SVM based Robust Decoders for Wireless Applications" to be published in the Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing – ARTCom 2009, India, 27-28 Oct 2009
- [11] Selva Kumar S., Ram Krishna Rao M., Balasubramanian M.P., "Chemopreventive effects of Indigofera aspalathoides on 20-methylcholanthrene induced fibrosarcoma in rats", International Journal of Cancer Research, ISSN : ISSN: 1811-9727, 7(2) (2011) pp.144-151.
- [12] R.Ramanathan, Arun.S.Nair, V.Vidhyasagar, N.Sriram, K.P.Soman "A Support Vector Machines Approach for Efficient Facial Expression Recognition" to be published in the Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing – ARTCom 2009, India, 27-28 Oct 2009
- [13] Mahalakshmi K., Prabhakar J., Sukumaran V.G., "Antibacterial activity of Triphala, GTP & Curcumin on Enterococci faecalis", Biomedicine, ISSN : 0970 2067, 26(Mar-4) (2012) pp. 43-46.
- [14] R.Ramanathan, P.A.Rohini, G.Dharshana., K.P.Soman "Investigation and Development of Methods to Solve Multi-Class Classification Problems " to be published in the Proceedings of International Conference on Advances in Recent Technologies in Communication and Computing – ARTCom 2009, India, 27-28 Oct 2009.
- [15] Selva Kumar S., Ram Krishna Rao M., Deepak Kumar R., Panwar S., Prasad C.S., "Biocontrol by plant growth promoting rhizobacteria against black scurf and stem canker disease of potato caused by Rhizoctonia solani", Archives of Phytopathology and Plant Protection, ISSN : 0323-5408, 46(4) (2013) pp.487-502.
- [16] Dr. T.Nalini, Dr.K.Manivannan, Vaishnavi Moorthy, Efficient Remote Data Possession Checking In Critical Information Infrastructures Ensuring Data Storage Security In Cloud Computing, International Journal of Innovative Research in Computer and Communication Engineering, ISSN (Online): 2320 – 9801, pp 12-18, Vol. 1, Issue 1, March 2013



ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 3, Issue 2, February 2015

- [17]Dr.K.P.Kaliyamurthie, Cervical Cancer Screening and Classification Using Acoustic Shadowing, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 1659-1662, Volume 1, Issue 8, October 2013
- [18] Dr.K.P.Kaliyamurthie, Automated Information Retrieval System Using Correlation Based Multi- Document Summarization Method, International Journal of Innovative Research in Computer and Communication Engineering, ISSN(Online): 2320-9801,pp 4328-4335, Vol. 1, Issue 7, September 2014
- [19] Dr.K.P.Kaliyamurthie, D.Parameswari, Modelling Cloud Storage, International Journal of Innovative Research in Computer and Communication Engineering, ISSN: 2249-2651,pp 1-5, Volume1 Issue3 Number2–Dec2011
- [20]Dr.K.P.Kaliyamurthie,D.Parameswari, Implementation of Customized Greencall Algorithm for Energy-Efficient Of Wireless LANs, International Journal of Innovative Research in Computer and Communication Engineering, ISSN: 2249-2615,pp 15-21, Volume1 Issue 1 Number2-Aug 2011