Eurasian Journal of Analytical Chemistry
ISSN: 1306-3057 OPEN ACCESS 2017 12 (4): 203-209

Tamil Character Recognition

K. Anita Davamani, S. Amudha

Received: 10 December 2016 ▪ Revised: 13 January 2017 ▪ Accepted: 12 February 2017

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: Character Recognition, Texture Features, Strong Texture, Weak Texture.

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. 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. 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.

PREPROCESSING

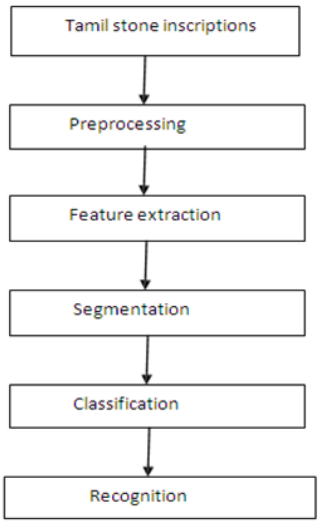
Binarization

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. 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.

K. Anita Davamani, Assistant Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai. E-mail: anitadavamani@gmail.com

S. Amudha, Assistant Professor, Department of Computer Science and Engineering, BIST, BIHER, Bharath Institute of Higher Education & Research, Selaiyur, Chennai. E-mail: amudha17s@gmail.com

Block Diagram



Noise Removal

The presence of noise can cost the efficiency of the character recognition system. 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 noise from the image.

• 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. 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):

median[A(x) B(x)]≠median[A(x)] Median[B(x)]

These filters smooths the data while keeping the small and sharp details. 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. 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.

Feature Extraction

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

- Color.
- Texture.
- Edges.

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.

$$fI : R \grave{I} IR\ 2\ \circledR M \grave{I} IR$$

Create a finite partition of M.

. .

Bk are subsets of M

Bk are called bins and k is the label of the bin. Convert the bins to histogram using the indicator function:

( ) { if ( ) bk.

X is the element of image

Bk are called bins and k is the label of the bin. Convert the bins to histogram using the indicator function.

Texture Features

Texture has no precise definition. Texture is a tactile or visual characteristic of a surface. 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.

Weak Texture

- Small spatial interactions between primitives.
- Frequencies of primitive types appearing in some neighborhood used for description.

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,

( )

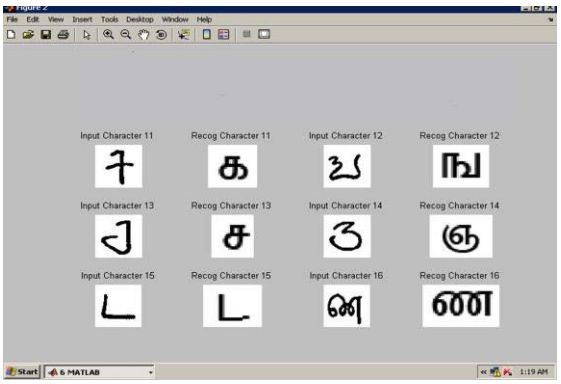
The gradient is a *vector* with magnitude in the *u* and *v* directions equal to the respective partial derivatives. the partial derivative of a discrete function are calculated using the Taylor's series.

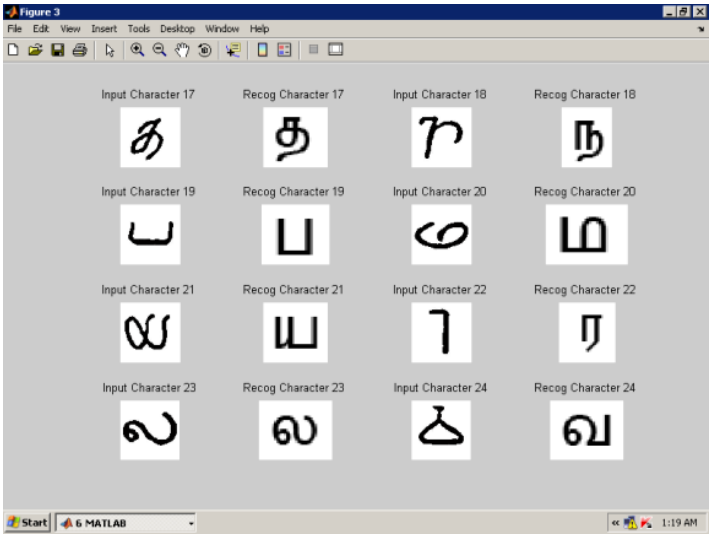
SEGMENTATION

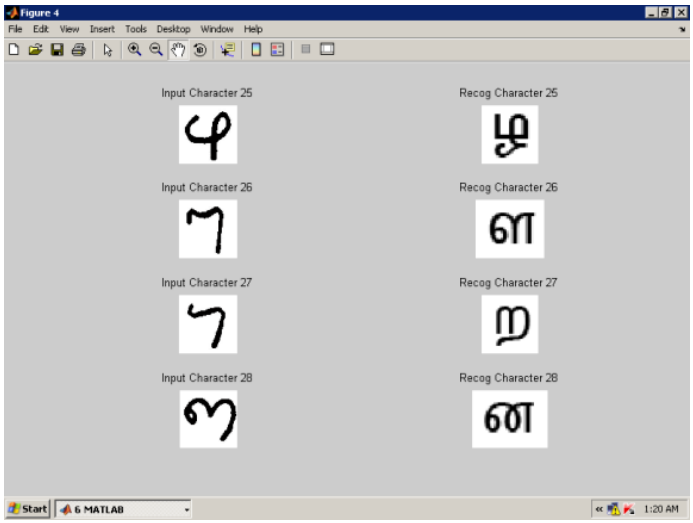
Segmentation is a process of distinguishing lines, words, and even characters of a hand written or machine printed document, a crucial step as it extracts the meaningful regions for analysis. There exist many sophisticated approaches for segmenting the region of interest. For handwritten document, this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

CLASSIFICATION AND RECOGNITION

The features are extracted and are then fed to the trained neural network for classification. The developed network is evaluated and cross validation is done to verify the results. In order to train a neural network to perform some task, we must adjust the weights of each unit in such a way that the error between the desired output and the actual output is reduced.

This process requires that the neural network compute the error derivative of the weights (EW). In other words, it must calculate how the error changes as each weight is increased or decreased slightly. The back propagation algorithm is the most widely used method for determining the EW.

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. 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. 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. The EW is the product of the EA and the activity through the incoming connection.

The steps are

1. The 21^{st} century tamil characters forms the data set. It is trained by extracting the featuers.
2. The ancient eighth century tamil character features are extracted and are stored.
3. Each input is combined with the neurons.
4. On comparison, the output is obtained.

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 classificat ion 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] Das, J., Das, M. P., & Velusamy, P. (2013). Sesbania grandiflora leaf extract mediated green synthesis of antibacterial silver nanoparticles against selected human pathogens. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, *104*, 265-270.

[2] Umanath, K.P.S.S.K., Palanikumar, K., & Selvamani, S. T. (2013). Analysis of dry sliding wear behaviour of Al6061/SiC/Al2O3 hybrid metal matrix composites. *Composites Part B: Engineering*, *53*, 159-168.

[3] Udayakumar, R., Khanaa, V., Saravanan, T., & Saritha, G. (1786). Cross layer optimization for wireless network (WIMAX). *Middle-East Journal of Scientific Research*, *16*(12), 1786-1789.

[4] Kumaravel, A., & Rangarajan, K. (2013). Algorithm for automaton specification for exploring dynamic labyrinths. *Indian Journal of Science and Technology*, *6*(5S), 4554-4559.

[5] Pieger, S., Salman, A., & Bidra, A.S. (2014). Clinical outcomes of lithium disilicate single crowns and partial fixed dental prostheses: a systematic review. *The Journal of prosthetic dentistry*, *112*(1), 22-30.

[6] Vijayaraghavan, K., Nalini, S.K., Prakash, N.U., & Madhankumar, D. (2012). One step green synthesis of silver nano/microparticles using extracts of Trachyspermum ammi and Papaver somniferum. *Colloids and Surfaces B: Biointerfaces*, *94*, 114-117.

[7] Khanaa, V., Mohanta, K., & Satheesh, B. (2013). Comparative study of uwb communications over fiber using direct and external modulations. *Indian Journal of Science and Technology*, *6*(6), 4845-4847.

[8] Khanaa, V., Thooyamani, K. P., & Udayakumar, R. (1798). Cognitive radio based network for ISM band real time embedded system. *Middle-East Journal of Scientific Research*, *16*(12), 1798-1800.

[9] Vijayaraghavan, K., Nalini, S.K., Prakash, N.U., & Madhankumar, D. (2012). Biomimetic synthesis of silver nanoparticles by aqueous extract of Syzygium aromaticum. *Materials Letters*, *75*, 33-35

[10] Caroline, M.L., Sankar, R., Indirani, R.M., & Vasudevan, S. (2009). Growth, optical, thermal and dielectric studies of an amino acid organic nonlinear optical material: l-Alanine. *Materials Chemistry and Physics*, *114*(1), 490-494.

[11] Kumaravel, A., & Pradeepa, R. (2013). Efficient molecule reduction for drug design by intelligent search methods. *International Journal of Pharma and Bio Sciences*, *4*(2), B1023-B1029.

[12] Kaviyarasu, K., Manikandan, E., Kennedy, J., Jayachandran, M., Ladchumananandasiivam, R., De Gomes, U. U., & Maaza, M. (2016). Synthesis and characterization studies of NiO nanorods for enhancing solar cell efficiency using photon upconversion materials. *Ceramics International*, *42*(7), 8385-8394.

[13] Sengottuvel, P., Satishkumar, S., & Dinakaran, D. (2013). Optimization of multiple characteristics of EDM parameters based on desirability approach and fuzzy modeling. *Procedia Engineering*, *64*, 1069-1078.

[14] Anbuselvi S., Chellaram, C., Jonesh S., Jayanthi L., & Edward J.K.P. (2009). Bioactive potential of coral associated gastropod, Trochus tentorium of Gulf of Mannar, Southeastern India. *J. Med. Sci,* 9(5), 240-244.

[15] Kaviyarasu, K., Ayeshamariam, A., Manikandan, E., Kennedy, J., Ladchumananandasivam, R., Gomes, U. U., & Maaza, M. (2016). Solution processing of CuSe quantum dots: Photocatalytic activity under RhB for UV and visible-light solar irradiation. *Materials Science and Engineering: B*, *210*, 1-9.

[16] Kumaravel, A., & Udayakumar, R. (2013). Web portal visits patterns predicted by intuitionistic fuzzy approach. *Indian Journal of Science and Technology*, *6*(5S), 4549-4553.

[17] Srinivasan, V., & Saravanan, T. (2013). Reformation and market design of power sector. *Middle-East Journal of Scientific Research*, *16*(12), 1763-1767.

[18] Kaviyarasu, K., Manikandan, E., Kennedy, J., & Maaza, M. (2015). A comparative study on the morphological features of highly ordered MgO: AgO nanocube arrays prepared via a hydrothermal method. *RSC Advances*, *5*(100), 82421-82428.

[19] Kumaravel, A., & Udhayakumarapandian, D. (2013). Consruction of meta classifiers for apple scab infections. *International Journal of Pharma and Bio Sciences*, *4*(4), B1207-B1213.

[20] Sankari, S. L., Masthan, K. M. K., Babu, N. A., Bhattacharjee, T., & Elumalai, M. (2012). Apoptosis in cancer-an update. *Asian Pacific journal of cancer prevention*, *13*(10), 4873-4878

[21] Harish, B. N., & Menezes, G. A. (2011). Antimicrobial resistance in typhoidal salmonellae. *Indian journal of medical microbiology*, *29*(3), 223-229.

[22] Manikandan, A., Manikandan, E., Meenatchi, B., Vadivel, S., Jaganathan, S. K., Ladchumananandasivam, R., & Aanand, J. S. (2017). Rare earth element (REE) lanthanum doped zinc oxide (La: ZnO) nanomaterials: synthesis structural optical and antibacterial studies. *Journal of Alloys and Compounds*, *723*, 1155-1161.

[23] Caroline, M. L., & Vasudevan, S. (2008). Growth and characterization of an organic nonlinear optical material: L-alanine alaninium nitrate. *Materials Letters*, *62*(15), 2245-2248.

[24] Saravanan T., Srinivasan V., Udayakumar R. (2013). A approach for visualization of atherosclerosis in coronary artery, Middle - East Journal of Scientific Research, 18(12), 1713-1717.

[25] Poongothai, S., Ilavarasan, R., & Karrunakaran, C.M. (2010). Simultaneous and accurate determination of vitamins B1, B6, B12 and alpha-lipoic acid in multivitamin capsule by reverse-phase high performance liquid chromatographic method. *International Journal of Pharmacy and Pharmaceutical Sciences*, *2*(4), 133-139.

[26] Udayakumar, R., Khanaa, V., & Saravanan, T. (2013). Synthesis and structural characterization of thin films of SnO 2 prepared by spray pyrolysis technique. *Indian Journal of Science and Technology*, *6*(6), 4754-4757

[27] Anbazhagan, R., Satheesh, B., & Gopalakrishnan, K. (2013). Mathematical modeling and simulation of modern cars in the role of stability analysis. *Indian Journal of Science and Technology*, *6*(5S), 4633-4641.

[28] Caroline, M.L., & Vasudevan, S. (2009). Growth and characterization of bis thiourea cadmium iodide: A semiorganic single crystal. *Materials Chemistry and Physics*, *113*(2-3), 670-674.

[29] Sharmila, S., Jeyanthi Rebecca, L., & Das, M. P. (2012). Production of Biodiesel from Chaetomorpha antennina and Gracilaria corticata. *Journal of Chemical and Pharmaceutical Research*, *4*(11), 4870-4874.

[30] Thooyamani, K.P., Khanaa, V., & Udayakumar, R. (2013). An integrated agent system for e-mail coordination using jade. *Indian Journal of Science and Technology*, *6*(6), 4758-4761.

[31] Caroline, M. L., Kandasamy, A., Mohan, R., & Vasudevan, S. (2009). Growth and characterization of dichlorobis l-proline Zn (II): A semiorganic nonlinear optical single crystal. *Journal of Crystal Growth*, *311*(4), 1161-1165.

[32] Caroline, M.L., & Vasudevan, S. (2009). Growth and characterization of L-phenylalanine nitric acid, a new organic nonlinear optical material. *Materials Letters*, *63*(1), 41-44.

[33] Kaviyarasu, K., Xolile Fuku, Genene T. Mola, E. Manikandan, J. Kennedy, and M. Maaza. Photoluminescence of well-aligned ZnO doped CeO2 nanoplatelets by a solvothermal route. *Materials Letters*, *183*(2016), 351-354.

[34] Aravind, D., & Deepa, K. (2015). Power Improvement of Photovoltaic System Using IEDM in Smart Grid Application. *Excel International Journal of Technology, Engineering and Management, 2*(2), 22-26.

[35] Naveen, N., & Rajesh Kumar, B. (2015). Automatic LPG Drum Level Pointer and SMS Booking System with Refuge. *Excel International Journal of Technology, Engineering and Management, 2*(2), 27-29.

[36] Dr.Veenadhari, S. (2016). Crop Advisor: A Software Tool for Forecasting Paddy Yield. *Bonfring International Journal of Data Mining, 6*(3), 34-38.

[37] Kondori, M.A.P., & Peashdad, M.H. (2015). Analysis of challenges and solutions in cloud computing security. *International Academic Journal of Innovative Research, 2*(8), 1-11.

[38] Kargar, M.J., & Motaghian, S. (2015). Creating Semi – Automatic Ontology on Persian Wikipedia Texts. I*nternational Academic Journal of Innovative Research, 2*(8), 38-44.

[39] Sandhiya, K., & Yamuna, L. (2014). High Performance Dual-Band Printed Doublet Design Loaded with Split Resonator Structures. *International Journal of Communication and Computer Technologies, 2*(1), 64-67.

[40] Lenin, G.J.N., Noora, J.A., Packiyalakshmi, D., Priyatharshini, S., & Thanapriya, T. (2014). Highly Directive Rectangular Patch Antenna Arrays. *International Journal of Communication and Computer Technologies, 2*(1), 68-73.

[41] Oh, T.J., & Anthony, (2017). New and Fast Emerging Advance Structure of Text Mining from Unstructured Data. *Bonfring International Journal of Industrial Engineering and Management Science, 7*(2), 13-16.

[42] Janmohammadi, P., & Babazade, M. (2015). Resource Management in the Cloud Computing Using a Method Based on Ant Colony Optimization. *International Academic Journal of Science and Engineering, 2(*6), 40-54.

[43] Farajizadeh, M., & Bakhsh, N.N. (2015). A mechanism to improve the throughput of cloud computing environments using congestion control. *International Academic Journal of Science and Engineering, 2*(7), 10-24.

[44] Saravanan, T., & Saritha, G. (2013). Buck converter with a variable number of predictive current distributing method. *Indian Journal of Science and Technology*, *6*(5S), 4583-4588.

[45] Parthasarathy, R., Ilavarasan, R., & Karrunakaran, C. M. (2009). Antidiabetic activity of Thespesia Populnea bark and leaf extract against streptozotocin induced diabetic rats. *International Journal of PharmTech Research*, *1*(4), 1069-1072.

[46] Hanirex, D. K., & Kaliyamurthie, K. P. (2013). Multi-classification approach for detecting thyroid attacks. *International Journal of Pharma and Bio Sciences,* 4(3), B1246-B1251

[47] Kandasamy, A., Mohan, R., Lydia Caroline, M., & Vasudevan, S. (2008). Nucleation kinetics, growth, solubility and dielectric studies of L-proline cadmium chloride monohydrate semi organic nonlinear optical single crystal. *Crystal Research and Technology: Journal of Experimental and Industrial Crystallography*, *43*(2), 186-192.

[48] Srinivasan, V., Saravanan, T., Udayakumar, R., & Saritha, G. (2013). Specific absorption rate in the cell phone user's head. *Middle-East Journal of Scientific Research*, *16*(12), 1748-50.

[49] Udayakumar R., Khanaa V., & Saravanan T. (2013). Chromatic dispersion compensation in optical fiber communication system and its simulation, Indian Journal of Science and Technology, 6(6), 4762-4766.

[50] Vijayaragavan, S.P., Karthik, B., Kiran, T.V.U., & Sundar Raj, M. (1990). Robotic surveillance for patient care in hospitals. *Middle-East Journal of Scientific Research*, *16*(12), 1820-1824.