

IDENTIFICATION OF TAMIL CHARACTER RECOGNITION BY USING MATLAB

¹L.Suriya Kala, ² Dr P. Thangaraj

¹Research Scholar, Mother Teresa Women's University, Kodaikanal, Tamilnadu, India

²Computer Science and Engineering, Bannari Amman Institute of Technology, Sathiyamangalam, Tamilnadu, India

ABSTRACT

The thesis describes of character recognition process of various Tamil scripts using various classifier and the work proposed noise image and segmentation process for the individual characters image of letters from each other. After the process, an important feature extraction step is used to recognize the Ancient characters accurately with hypothesis combination. Feature describes the characteristic of object uniquely for Variability method of complexity image. Noise Reduction Process with Filtering Method is one of the feature extraction, which is the process of extracting information from test data which is most relevant for classification purpose relevant the type of error with blur image of letter or characters. The technique extracts the basic components of Tamil Characters and then it can be translates into the components for additional recognition measures to the probability of blur image. Finally a recognition process system is proposed for the characters in Tamil script. The data set characters are sampled from the script automatically or tools based. The proposed Hypothesis classifier is tested on quite number samples images of letter of ancient Tamil images.

I. INTRODUCTION

During the last four decades, the field of text character recognition has been receiving significant attention for any language, from research personnel in varied disciplines such as conversion of handwritten to text image, image convert into understand format, text to speech, printed to document to an editable soft format by using any automata tool to store for long term, these sequence steps are recognition for a particular script depending upon the nature and intricacy of the character [1].

Tamil language is most popular and its techniques grammar are formulated for the recognition of a Tamil script is one of the foremost among the 16 major national languages spoken reading and writing by the South Indian people. Most of Tamil letters have circular shapes and 3x3 matrix format; in ancient days the partially due to the information that are written originally imprinted with needles on palm leaves about the novelty ideas and deliberation, a technology that favored rounded shapes of the Tamil language. The combination of alphabetical and syllabic systems is in tamil language writing. The Tamil script is used to write and speaking the Tamil Language in Tamil Nadu state of India as well as in Sri Lanka, Singapore and parts of Malaysia as we as write minority languages such as Badaga and other part of world.[2][3].

II. TECHNICAL FEATURE SOFTWARE TOOLS USED IN TAMIL LANGUAGE CONVERSION

Handwritten style is dissimilar to person to person, even though the recognition of handwritten character acquires complicated with the frequent variations involved in the shape and style of characters, diverse of writing style, overlapping the letters and the interconnection of the neighbouring characters, invisible style, incorrect formation meaning, spelling mistake etc., and it is depends upon the individual personalities. The tamil language is an unique script for understanding text of ancient to modern script as well as written script. Since, a technical development of Optical Character Recognition (OCR) is the electronic conversion of images, handwritten, printed text into machine-encoded text, scanned document, photo document and a scene-photo are superimpose of an image scheme with high quality recognition accuracy for Tamil script to minimize the difficult task and complex of reading and writing of text through this methodology. The main objective of these researches to recognize Tamil characters, this proposal methodology is done by classifying the characters into appropriate types based on features extracted from each character without any unwanted noise image, finding the hypothesis of probability of error in super image of text and risk free text from tamil.

A. Image to Text Converter

The American Standard Code for Information Interchange (ASCII) values can be recognized the Tamil characters and each of the fonts is matched with its corresponding template converted and saved as normalized text transcription languages. Another methods used by MATLAB supports standard data and image formats exchanged, including JPEG, PNG, TIFF, HDF, HDF-EOS, XLS, FITS, ASCII, BINARY files etc.,

III. LITERATURE REVIEW

A. Attempt to Recognize Handwritten Tamil Character Using Kohonen SOM:

The algorithm obtainable about the introduced for Tamil character recognition and introduce the advantage of using this Self Organizing Map (SOM) model is to capture and analysis the invariant features of the Tamil Scripts. Dissimilar to the neural network and it does not hold any hidden layer of the character. Only two layers are required for input and the output. The research is most useful for visualizing and analysis of handwritten from higher dimensional contribution of input space to lower-dimensional map space and find out the pre-classified group of final recognition [1].

B. Century Identification and Recognition of Ancient Tamil Character Recognition:

The author presents the contour let-based methods in offline text-independent ancient Tamil handwriting are identification from the stone inscription. Compared made with the 2-D wavelet and contour let-based method attains a higher accuracy, because contour let transform has capacity to capture moderately and comparability with richer directional information and significant to represent the writing style of hand writing of ancient people. From long literature there is no existed handwriting database available any data base and researcher build up a small handwriting database their conveniences and it leads to the small capacity of the experimental database setups. To make the experiment result more convincing and enlarge the database setup for future work [2].

C. Two Dimensional Principal Component Analysis for Online Tamil Character Recognition:

The researcher to attempt to apply the recently reported about the 2DPCA technique concept about the context of online character recognition of the letters and preprocessed character to form the character matrix and using the covariance matrix for constructed. The covariance matrix in 2DPCA feature extraction is compared with the conventional PCA. The experimental results indicate the recognition accuracy by using of 2DPCA with increasing of approximate 3% with PCA technique and same time being for computationally efficient. The areas of research are to develop in dimensionality reduction schemes that fall into class discriminatory characteristics of online character [3] [4].

D. Distortion Analysis of Tamil Language Characters Recognition :

Pattern recognition notion is achieved to using the normal computing techniques and neural networks for analysis of Tamil language character recognition of ancient letters. In usually, the computing conventional arithmetic algorithms are used to detect for given pattern matches. On the other hand, neural networks can tolerate noise and it will respond for the noisy patterns and it constructed with the proper architecture and trained correctly for data set and it makes as results of pattern recognition of the tamil characters letters of 'tha', 'zha' of distorted and recognized with an error [5].

E. A Survey of Methods and Strategies in Tamil Palm Script Recognition :

This investigation about an overview of the Tamil Palm manuscript Recognition systems for tamil language. The performance and analysis of various preprocessing, segmentation, extraction and classification methods has also been investigated, and it gives the picture of imaginative factor of each method of image. There is no standard solution and accuracy to identify all Tamil characters. Various methods have been introduced in each phase of the recognition process of a few character sets. Challenges still overcome in the recognition of normal, abnormal writing, slanting characters of text, joined characters, similar shaped characters, curves etc., on during recognition process[6].

F. Embedded optimal character recognition on Tamil text image using raspberry pi:

The research is to help the visually challenged people to read books and present the development of implementation for the conversion of printed Tamil text characters. The device is custody with image capture camera and captured greater than a printed page and this input is taken for camera in the hand-held device and the output is converted into speech through microphone with help of hardware interface design as well as the software implementation with hardware interfacing using Embedded Linux operating system [7].

IV. CURRENT SCENARIO TECHNOLOGY SOLUTION

There are various phases are introduced for given input image is being sophisticated to be analyzed respected algorithm and methodology. Few list of technology are used given below,

- A. **Preprocessing technique is involves the following steps:** Smoothen for given the input image, Eliminate unwanted input image, Noise Reduction of image and normalize the image for supplementary processing. Mainly the noises Reduction process used reduces with assist of two methodologies are Gaussian smoothing Method and Bilinear filtering method.
- B. **Segmentation:** Segmentation is define in term of, dividing written text or image into meaningful unit or component, such as words, sentences, paragraph and topics. The preprocessed image is divided into different segments that each segment letter converted to individual images of characters to be compared with the database for matching the signifying methodology [7][8].
- C. **Filtering:** Low filter smoothing and Advanced Median Filtering Method is dominating technique for remove the unwanted bladder image from the segmentation.

V. PROPOSED METHODOLOGY: (VARIABILITY METHOD OF COMPLEXITY IMAGE)

The proposed method recognizes the critical Tamil characters with more efficiently than the existing available method but low computational complexity is the main problem for image from the tamil letter with more accurate results. The raw unprocessed image is given as input computation in an image file (.jpg file). Based on the literal review, the most of image goes under on basic technique or refined to be any analyzed algorithm in regulate to recognize the text in given image, which are preprocessing image , segmentation and noise Reduction and Filtering and these step all follows the unwanted blur image is removed and finding the percentage through the proposal methodology of Significant level of Probability error of image and improve the image quality as output of this research, and the image is processed and given as an input to the next step. Based on the each of unique features and characters, the image are identified and classified respectively and the finally the recognized character are formed into the same alignment as the given input image to the computation tools for receiving the actual content of the image can be identified in quality and eminence image of ancient letter. The working and Flow of data in whole process is divided into eight steps and the steps are as follows Steps in fig 1:

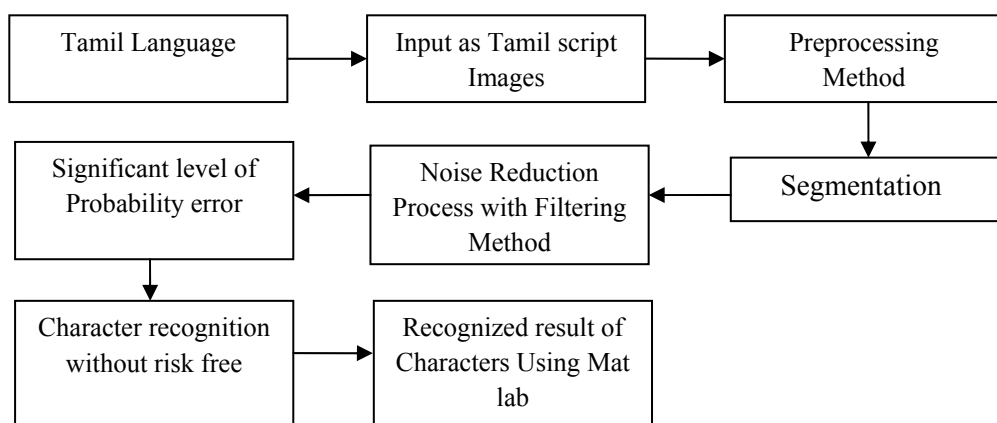


Fig 1. Over all Process of proposal methodology

As an output the recognized image letter with content identification exported to a text file and results can be used in further process of research.

VI. CONCLUSION

As this concept is a basic approach to recognize the text characters in the given image, and same can be done by different combinations to produce better results of blur image. Either the language is ancient or modern as long as, they are able to recognize the features of each letter characters it will not be a problem. In the same way we can make a diagnosis process which can open a new method to identify complex writings of text

in the objects as well as damaged documents by using of the research proposal of hypothesis and the implementation developed using of MATLAB with sample images of Tamil characters

REFERENCES

- [1] Indra Gandhi .R and Iyakutti.K , “An Attempt to Recognize Handwritten Tamil Character Using Kohonen SOM “, International Journal of Advance d Networking and Applications, Volume: 01 Issue: 03 Pages: 188-192 (2009).
- [2] Rajakumar S and Subbiah Bharathi V, “Century Identification and Recognition of Ancient Tamil Character Recognition” , International Journal of Computer Applications (0975 – 8887) Volume 26– No.4, July 2011.
- [3] Seethalakshmi r., Sreeranjani t.r., Balachandar t., “Optical Character Recognition for printed Tamil text using Unicode “,Journal of Zhejiang University SCIENCE, ISSN 1009-3095, 2005 6A(11):1297-1305.
- [4] Madanapalle Institute of Technology and Science, “Two Dimensional Principal Component Analysis for Online Tamil Character Recognition”, Conference: XI Int. Conf Frontiers in Handwriting Recognition (ICFHR 2008), At Montreal, Quebec, Canada.
- [5] Venkata Krishna Kumar s Poornima T V , “An Efficient Period Prediction System for Tamil Epigraphical Scripts Using Transductive Support Vector Machine”, International Journal of Advanced Research in Computer and Communication Engineering Vol. 3, Issue 9, September 2014.
- [6] Gowri. N and Bhasharan R, “ Distortion Analysis Of Tamil Language Characters Recognition “,IJCSI International Journal of Computer Science Issues, Vol. 8, Issue 1, January 2011 ISSN (Online): 1694-0814.
- [7] Kiruba B and Manikandan K , “A Survey of Methods and Strategies in Tamil Palm Script Recognition “,International Journal of Computer Science and Information Technologies, Vol. 7 (2) , 2016, 972-977.
- [8] Ajantha Devi V and Santhosh Baboo S , “Embedded optimal character recognition on Tamil text image using raspberry pi”, International Journal of Computer Science Trends and Technology (IJCST) – Volume 2 Issue 4, Jul-Aug 2014.
- [9] Sudha S, Jothilakshmi S , Rajasoundramani R , “Tamil Sign Language to Speech Translation “,International Journal of Science and Research (IJSR) ISSN (Online): 2319-7064, Volume 3 Issue 6, June 2014.

AUTHOR PROFILE

L. Suriya kala had completed MCA., M.Phil and working as Assistant professor in Don Bosco College, Dharmapuri, Tamil Nadu India. She is a research Scholar in the specialization of Digital Image processing at Mother Teresa Women’s University Kodaikannal, Tamil Nadu, India.

P. Thangaraj , Professor and Head ,Department of Computer Science and Engineering Bannari Amman Institute of Technology, Sathiyamangalam, TamilNadu, India.