International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

Recognition of Ancient Tamil Characters in Stone Inscription Using Improved Feature Extraction

M. V. Jeya Greeba[1], G.Bhuvaneswari[2]
[1]*PG Student, DMI College Of Engineering, Chennai-600123, India*
[2]*Assistant Professor, Department of CSE, DMI College Of Engineering, Chennai-600123, India.*
[1]greebamv@gmail.com, [2]bhuvankerani@gmail.com

***Abstract*— Recognition of Ancient Tamil character is one of the important task to reveal the information from the Ancient world. Feature Extraction is an important step used to recognize the Ancient characters accurately. Feature describes the characteristic of object uniquely. Feature extraction is the process of extracting information from raw data which is most relevant for classification purpose. Feature Extraction technique extracts the basic components of Tamil Characters in Stone inscription and then it translates the components for further recognition procedures. Structural Features are extracted. Presence of dots, loops and curves in Tamil character makes Feature Extraction a challenging process. The improved Feature Extraction technique will overcome these challenges and make character recognition an efficient process.**

***Keywords*—Ancient, Classification, Components, Efficient, Relevant.**

I. INTRODUCTION

This docum Tamil character recognition has always been an active field of research for computer scientists worldwide due to its useful real life applications such as automatic data entry, mail processing and form processing Character recognition is a classic problem in the field of image processing and neural networks.The script used by these inscriptions is commonly known as the Tamili script, and differs in many ways from standard AsokanBrahmi. For example, early Tamil Brahmi, unlike AsokanBrahmi, had a system to distinguish between pure consonantsand consonants with an inherent vowel.

Vatteluttu alphabet is writing system originating from the ancient Tamil people of Southern India. Developed from the Tamili (Tamil-Brahmi), Vatteluttu is one of the alphabet systems developed by Tamil people to write the Proto-Tamil language.

It is currently spoken by about 77 million people around the world with 68 million speakers residing in India mostly in the state of Tamil Nadu. It is one of the official language in India, Sri Lanka and Singapore. Tamil characters consist of small circles or loops, which are difficult to recognize.Recognizing the ancient Tamil characters that is, the Vatteluttu alphabets is a tough task for the modern generation who learn to read and write only with modern Tamil characters. Learning the evolution of Modern Tamil from ancient Tamil is a time consuming process therefore a recognition system helps to teach, understand and also to research the ancient cultures and heritages. To design a good recognition system this paper proposes feature extraction of acquired images.

Handwritten character recognition is one of the most difficult tasks in the pattern recognition system. There are lot of difficult things need in many image processing techniques to solve. The difficulties are, how to separate cursive characters into an individual character, how to recognize unlimited character fonts and written styles, and how to distinguish characters that have the same shape but different meaning such as character ‘o’ and number ‘0’. Many researchers try to apply many techniques for breaking through the complex problems of handwritten character recognition. There are many applications need to take advantage of the handwritten character recognition system namely, automatic reading machine, non-keyboard computer system, and automatic mailing classification system. The Tamil alphabet consists of 12 vowels, 18 consonant,combination of vowels and consonant 216, and one Ayutha letter.The structure of Tamil words is written in a four-line level style.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

II. FEATURE EXTRACTION

Feature extraction is defined as the problem of extracting from the raw data the information which is most relevant for classification purposes. Feature extraction methods are

- Template matching
- Deformable Template
- Unitary image Transform
- Graph description
- Projection histogram
- Zoning
- Zernike moments

A. Template matching

In gray scale Template matching a similarity or dissimilarity measure between each template T and character image Z is computed. In similarity measure the template t which has highest similarity measure is identified. In the case of dissimilarity measure, the lowest dissimilarity is identified. In binary template matching several similarity measure is identified.

B. Deformable templates

Each template is deformed in a number of small steps called local affine transform to match the candidate input pattern. The number and types of transformations before match is obtained.

C. Unitary image transform

In the transformed space, the pixels are ordered by their variance and pixel with the highestvariance are used as features. The unitary transform has apply to training set to obtain estimates of the variances of the pixels in the transformed space. Transforms have been utilized to extract features from images. A unitary transformis a specific type of linear transformation. An image transformation can be interpreted as a decomposition of the image data into a generalized two-dimensional spectrum . Each spectral component in the transform domain corresponds to the amount of energy of the spectral function within the original image. This type of generalized spectral analysis is useful in the investigation of specific decompositions that are best suited for particular classes of images.

D. Graph description

Extracts approximate strokes from skeletons. Extract graph descriptions from skeletons as features derive hierarchical attributed graphs to deal with variations in stroke lengths and connectedness due to variable writing style of different writers.

E. Projection histogram

Feature, image is scan along a line from one side to another side and number of foreground pixel on the line is counted. Thus it is also known as histogram projection count.The horizontal projection is the number of pixels with the features can be scale independent by using a fixed number of bins on each axis and dividing by the total number of print pixel in the character image. Here the background pixel is 0 and foreground pixel is 1. Similarly, vertical projection histogram can be calculated. This feature is used in several pre-processing steps of document Image segmentation where it is used for segmenting text lines, words and characters . Calculateboth horizontal and vertical projection histograms and combine them into a feature vector for training and testing. This measurement is not image size invariant but all the character data used for the recognition process have same size. The feature does not consider stroke width variation in handwritten characters.

F. Zoning

To extract the features an image is divided into some non-overlapping or overlapping zones. Then the number of foreground pixel is counted and the density is computed for each zone. Sometimes zoning is considered with other features. The character is divided into zones. The zoning method is used to compute the percentage of pixel values. Zoning is relatively scaling and slant invariant. The feature vector of the experiments is designed to contain the densities of 4 X 4 = 16 zones for each image.The character is divided into zones. The zoning method is used to compute the percentage of black pixels

G. Zernike moments

The Zernike moments are projections of the input image onto the space spanned by orthogonal function. A set of complex orthogonal polynomials are used.

International Journal of Recent Development in Engineering and Technology
Website: www.ijrdet.com (ISSN 2347 - 6435 (Online)), Volume 2, Special Issue 3, February 2014)
International Conference on Trends in Mechanical, Aeronautical, Computer, Civil, Electrical and Electronics Engineering (ICMACE14)

Zernike moment is overcome the shortcomings of information redundancy present in the geometric moments. Zernike moments are rotation invariant.

III. Proposed System

Combining zoning with Structural feature extraction method is used in proposed method. It hasthe potential to improve the accuracy of existing applications. The database created with the set of trained character images. Test image is compared with trained image to find the correct character. Average matrix is calculated .After pre-processing add all the training image matrix into single matrix of 64*64 and then divided it by total number of samples to get average of all training image called average matrix. In average matrix if 1 is present then it means in each and every sample image that particular pixel is present. After creating average matrix, subtract it from each and every sample character image matrix to get unique features, results in matrix. Use fuzzy logic as the classifier for recognition.

IV. Module Description

The proposed work is formulated as following modules:

- Cropping
- Noise Removal
- Feature Extraction
- Classification

Cropping

Cropping refers to the removal of the outer parts of an image to improve framing. It is used to removing unwanted areas from a character image. A crop made from the top and bottom of a image.

Noise Removal

Noise is the result of errors in the image acquisition process that result in pixel values that do not reflect the true intensities of the real scene. Noises result in distortions of final result so it has to be removed. The Median filtering Technique is used for removing noise.

Feature Extraction

The character image is divided into 3x3 zones.From each zone features are extracted to form the feature vector.

The goal of zoning is to obtain the local characteristics instead of global characteristics.

Classification

Classification is the problem of identifying to which of a set of categories a new observation belongs, on the basis of a training set of data containing observations.

V. Conclusion

Thus, in this work, an improved feature extraction method for character recognition has been proposed. Zoning with structural features are used for feature extraction. Feature extraction method that is best suited for one particular recognition application may not give optimum performance for the other application. This feature set captures close region, convexity of stroke from different direction, and flat region of thinned character images. The extracted features are represented as a shape-based graph and acts as an invariant feature set for discriminating similar type character images. The future work includes testing the efficiency of the algorithm with recognition system and ancient Tamil characters will be recognized accurately.

REFERENCES

[1] E. K. Vellingiriraj and P. Balasubramanie, “Recognition of Ancient Tamil Handwritten Characters in Palm Manuscripts Using Genetic Algorithm”, International Journal of Scientific Engineering and Technology May 2013

[2] Swapnil A. Vaidya, Balaji R. Bombade, “A Novel Approach of Handwritten Character Recognition using Positional Feature Extraction” , International Journal of Computer Science and Mobile Computing , June 2013

[3] Gaurav Y. Tawde ,Mrs. Jayashree M. Kundargi, “ An Overview of Feature Extraction Techniques in OCR for Indian Scripts Focused on Offline Handwriting”, International Journal of Engineering Research and Applications , January -February 2013.

[4] N.Sridevi and Dr.P.Subashini “Moment based Feature Extraction for classification of Handwritten Ancient Tamil scripts”, International Journal of Emerging trends in Engineering and Development November 2012

[5] Bilal Bataineh ,FarshidPirahanSiah, "Character Recognition Based on Global Feature Extraction", 2011 International Conference on Electrical Engineering and Informatics 17-19 July 2011

[6] M. ZahidHossain,m. AshrafulAmin,hong Yan, "Rapid Feature Extraction For Bangla Handwritten Digit Recognition", Proceedings of the 2011 International Conference on Machine Learning, July, 2011

[7] Soumen Bag and Gaurav Harit, " Topographic Feature Extraction For Bengali And Hindi Character Images", Signal & Image Processing : An International Journal, June 2011

[8] NiraseFathimaAbubacker and Indra Gandhi Raman, "An Approach for Structural Feature Extraction for Distorted Tamil Character Recognition", International Journal of Computer Applications, May 2011

[9] NiraseFathimaAbubacker and Indra Gandhi Raman, "An Approach for Structural Feature Extraction for Distorted Tamil Character Recognition", International Journal of Computer Applications, May 2011

[10] Aria Pezeshk and Richard L. Tutwiler, "Automatic Feature Extraction and Text Recognition From Scanned Topographic Maps", Ieee Transactions On Geoscience And Remote Sensing December 2011.