ISSN 2229-5518

# Image Glazing for Thinning of Ancient Tamil Characters

R. Angelin Jeniffer, G. Bhuvaneswari

**Abstract**— The Conversion of Ancient Tamil characters to modern text is a necessary automation in image processing. These ancient Tamil characters are images taken from Ancient Tamil Stone inscriptions. Recognising the ancient Tamil characters 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 a method called Image Glazing for Thinning of acquired images. Thinning is the process of peeling thick layers of boundaries from an image to create a one pixel width representation of the image. Thinning not only reduces the storage needs of extracted features but also increases the recognition rate of the whole system. The proposed work produces fine-tuned thin images for each and every input character.

**Index Terms**— Thinning, Ancient Tamil Characters, Pattern Matching, Skeletonization, Stone Inscriptions, Filtering Operations.

## 1 Introduction

TAMIL is one of the oldest southern languages of India. The evolution of Tamil started from 3rd century BC. The topology of characters used to represent Tamil language then, differ diversely from those used now. The beginning of evolution of Tamil comprises the period between the 3rd century BC and the 6th century AD, Medieval Tamil existed between the 6th century AD and the 12th century AD, and Modern Tamil, from the 12th century down to the present day. Figure 1.1 shows the evolution of vowels in Tamil script over time which has been inscribed on a stone. The ancient Tamil characters that is, the early and medieval Tamil can be mostly found only in stone inscriptions and palm leaves. Only epigraphists can read those stone inscriptions. To extend the readability and to preserve the ancient historical values, we need a good recognition system that can convert the ancient text to modern text. This paper proposes methods to develop such a recognition system.


Fig 1.1 Evolution of Vowels in Tamil Script

## 2 Overview Of Image Thinning

In computer vision, and machine recognition of patterns the need for reducing the amount of information to be processed to the minimum is necessary. Averaging operation over a square window with a high threshold results in thinning of input image. The thinned characters are used for recognition. In addition, the reduction of an image to its essentials can eliminate some contour distortions while retaining significant topological and geometric properties. The thin-line representations of patterns would be more suitable for extraction of critical features such as minutiae's, connectivity between curves, and connections among the components. The result of thinning should be a skeleton that clearly indicates the original structure of input character in topology. For the thinning algorithm to be really effective, it should ideally reduce data, retain all important features subjective to human perception and eliminate local noise without affecting the topological characteristics of original pattern. To accomplish all that using the simplest and fastest algorithm is the challenge involved.

This paper proposes image processing technique to recognize the ancient Tamil characters from images of stone inscriptions efficiently. Fig 2.1 shows the overall process as system architecture diagram. The character identification process involves two main stages: a preprocessing step and a recognition step. The preprocessing step dedicates to acquiring the image from stone inscription and then processing the acquired image for recognition. The recognition step involves training the classifier and then testing a new input based on the trained data. In order to enhance the recognition rate thinned images are used for training and testing by the classifier.

## 3 Literature Survey

A study of various thinning methods in use and those that has been proposed was made to come up with this novel approach. The following are few of the prominent thinning algorithms which contributed to the novel approach proposed in this paper.


- *R.Angelin Jeniffer is currently pursuing masters degree program in Computer Science and engineering in DMI college of Engineering, India, E-mail: angelinjeniffr@gmail.com*
- *G. Bhuvaneswari, is an Assistant Professor with the Department of Computer Science and Engineering in DMI College of Engineering Chennai-602103, and a Research Scholar at Anna University, TN, India, E-mail-bhuvankerani@gmail.com*

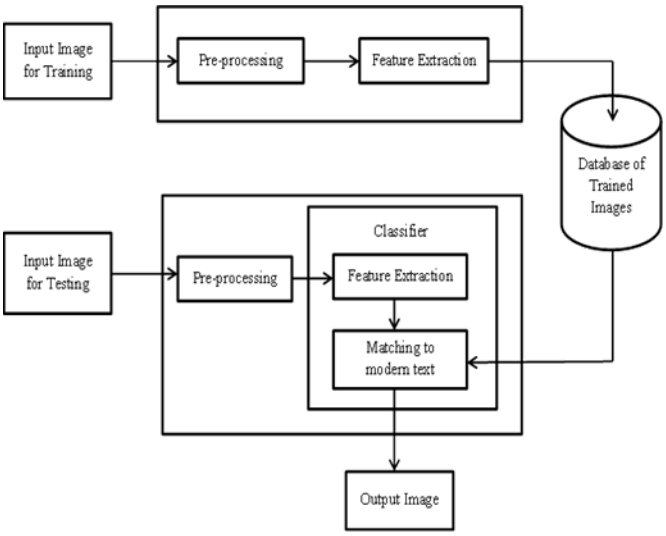


Fig 2.1: System Architecture Diagram

The general agreement for any thinning methodology to be effective are; they should, Preserve topological and geometric properties, Preserve isotropy, Must be reconstruct able, Must have high processing speed[1].

Improved ZS Thinning algorithm proposed in Improved Zhang-Suen Thinning Algorithm in Binary Line Drawing Applications utilizes the median filter for image processing, then refines the skeleton to a single pixel width and removes redundant segments[3].

Sequential Order independent Thinning Algorithm proposed in On Order–Independent Sequential Thinning makes use of Sequential algorithms that produce identical skeletons for any visiting orders and requires certain conditions for the order-independence of template based sequential algorithms [4].

An Order-Independent two-pass Parallel Algorithm proposed in On Order–Independent Sequential Thinning works as follows. First, thinning rules are formulated based on weighted values of input image and then rule based skeletonization is performed. Pass 1 uses 23 rule masks procedure and pass 2 uses 4 rule masks procedure [5].

## 4 PREPROCESSING

The central role of the preprocessing step involves preparing the acquired image for recognition

### 4.1 Background Subtraction

An image of Ancient Tamil characters is initially acquired from stone inscription. The acquired image is fed into image processing software and initially digitized. As the first process the acquired image is subtracted from its back ground. The image's regions of interest here are the Tamil characters in its foreground. The stone inscription's image has its background to be a stone, stones are rough surfaces and contain more noise while processing and in attempt for digitalizing its images. Thus Background subtraction is a prerequisite step. The stone background is subtracted by retaining only the characters inscribed on it, which is the ROI (Region of Interest). Otsu's method of global thresholding is used to do so. In Otsu method, pixels are divided so that the sum of spread of foreground and background pixels is at its minimum.

### 4.2 Noise Removal

Noise removal, in the background subtracted stone image involves removing noise and identifying regions likely to contain the background pixels which were not subtracted during background removal. Noise may also distort the character's structural perception. Cleaning an image infused with noise is thus an important area in Tamil character recognition system. The major contribution of noise is from the rough surfaces, holes and dents present on the stones. Most general approach used to remove these noises is to use thresholding. Thresholding the background subtracted image is best method to remove noise from stone inscriptions.

### 4.3 Image Glazing

The Noise removed image gives a rough edge perception for the characters. Such characters when thinned results in a weaker skeleton which does not preserve topology of character. Therefore Image glazing is what refines the skeleton that is formed by iterative thinning. The process of image glazing involves the following steps:

(i) Growing the regions of rough edges to give a proper shape to define the character. This also reduces the disconnected components.

(ii) To refine the boundaries of the image and to avoid noise to be mistaken for image we apply a linear Gaussian filter that removes boundary noises from the connected image pattern.

### 4.4 Thinning

Thinning is an image processing operation in which binary valued image regions are reduced to lines that approximate the center skeletons of the regions. This process is performed iteratively on every pixel.

The classic and most efficient method for thinning is Zhang- Suen Thinning algorithm. This Algorithm requires that the pattern to be thinned has a proper shape, if otherwise; the skeleton produced may contain lots of branch points and unwanted minutiae's. This is due to the complicated nature of the processing as well as various evaluation measures that should be considered. This paper makes use of that classic Zhang-Suen method[2] for thinning by meeting all the prerequisites by using other pre-processing steps. It is fast and simple to be implemented. It comprises of two sub-iterations. Figure 4.1 shows the structural element used by this algorithm[2].

In the first Iteration, a pixel P(i , j) is marked for deletion if the following conditions are satisfied:

1. If the Sum of pixels in a 3x3 structural element excluding the pixel P(i,j) is 1.

2. If the sum of Black pixels in the neighbourhood of P(i,j) is greater than 2 but less than 6.

3. If the product of pixels in P2, P4 and P6 equals zero.

4. If the product of pixels in P4, P6 and P8 equals zero

In the second sub-iteration the conditions in step 1 and 2 remain same but the steps in 3 and 4 are as follows,

3. If the product of pixels in P2, P4 and P8 equals zero.

4. If the product of pixels in P2, P6 and P8 equals zero.

| | | |
|---|---|---|
| P9 (i-1,j-1) | P2 (i-1,j) | P3 (i-1,j+1) |
| P8 (i,j-1) | P1 (i,j) | P4 (i,j+1) |
| P7 (i+1,j-1) | P6 (i+1,j) | P5 (i+1,j+1) |

Fig 4.1: Structural element

At the end, pixels that satisfy these conditions will be deleted. At the end of either sub-iteration if there are no pixels to be deleted, then the algorithm stops and produces the skeleton of image.

## 5 Recognition System

### 5.1 Feature Extraction

Universe of discourse is defined as the shortest matrix that fits the entire character skeleton. The Universe of discourse is selected for feature extraction based on the positions of different line segments in the character image. After the selection of universe of discourse, the image is divided into equal sized matrices, and feature extraction is done on individual matrix. Each matrix is called a zone. This gives more information about unique details of character skeleton. To extract different features in a particular zone, the entire skeleton image in that zone should be traversed. Each zone is to the process of extracting features and this is continues until features from all the zones in the skeleton are extracted. When all the pixels are visited, the traversal stops. Every zone creates a feature vector corresponding to it. The different features taken are, total number of horizontal lines, total number of vertical lines, the number of Right diagonal lines, the number of Left diagonal lines, the Length of all horizontal lines, the Length of all vertical lines, the Length of all right diagonal lines, the Length of all left diagonal lines, Area of the Skeleton. Line type normalization is done using,

Pixel value = 1 - ((number of lines/10) x 2)

Normalized length is found using,

Length = (Total Pixels in line type)/ (Total zone pixels)

### 5.2 Pattern Matching

Pattern Matching is the assignment of a label to a given input value; it attempts to assign each input value to one of a given set of classes in trained data set. Pattern Matching attempts to correlate all possible inputs to the most likely trained data taking into account their statistical variation. Here the correlation is done using the feature vectors of training image and that of the testing image. That is, supervised learning. Supervised learning assumes that the set of training data which has been provided consists of a set of instances that have been properly labelled by hand with the correct output. The matching set then matches the testing data with the closely related label

## 6 Results and Discussions

This Paper was implemented using Matlab 11b. The Skeleton obtained by using Image Glazing and without using Image glazing is shown in Figure 6.1 (b) and (c) respectively.

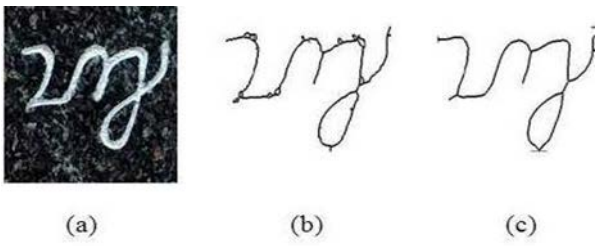


Fig 6.1: Result obtained for thinning of Ancient Tamil character using Image Glazing.

The quality of these images were calculated by comparing with the ground truth image by using the quality metrics given in Table 6.2 and the Skeleton obtained by Image glazing was found to be better in quality than the previous one.

| Image Quality Metrics | Ground Truth Vs. Zhang-Suen Thinning | Ground Truth Vs. Glazed Thinning |
|---|---|---|
| Mean Square Error | 1.9817e+003 | 1.6130e+003 |
| Peak Signal to Noise Ratio | 15.1604 | 21.0544 |
| MNormalized Cross-Correlation | 0.9815 | 0.9866 |
| Average Difference | 1.7384 | 0.4424 |
| Structural Content | 1.0060 | 1.0015 |
| Maximum Difference | 255 | 255 |
| Normalized Absolute Error | 0.0380 | 0.0327 |

Table 6.2: Image Quality Metrics

Also the recognition rate increased considerably when image glazed skeletons were used for feature extraction. Out of 20 character inputs 14 characters were recognized when skeletons obtained without image glazing were used. That is, its recognition rate was 70%. When the same sets of inputs were passed through the recognition system which used image glazing for thinning, it produced a recognition rate of 95% that is, 19 characters were correctly recognized.

## 7 Conclusion

Image Glazing is introduced in this paper. It produces a fine tuned image for thinning the characters. The effectiveness of this approach has been subjected to human perception and it has proved to be efficient when compared to the traditional thinning method. Thus the overall recognition rate was also increased.

## REFERENCES

[1] Louisa Lam, Seong-Whan, Ching Y.Suen,"Thinning Methodologies – A Comprehensive Survey " in IEEE Transaction on Pattern Analysis and Machine Intelligence(Sep 1992)

[2] T. Y. Zhang and C. Y. Suen, "A Fast Parallel Algorithm For Thinning Digital Patterns" in Research Contributions Image Processing and Computer Vision (Mar 1984)

[3] Wei Chen, Lichun Sui, Zhengchao Xu,Yu Lang ,"Improved Zhang-Suen Thinning Algorithm in Binary Line Drawing Applications" in International Conference on System and Informatics (May 2012)

[4] Péter Kardos, Kálmán Palágyi ,"On Order–Independent Sequential Thinning" in 3rd IEEE International Conference on Cognitive Infocommunications (Dec 2012)

[5] A. Jagna,"An Order-Independent two-pass Parallel Algorithm for Binary Image Thinning" in International Journal of Applied Information Systems (May 2012)

[6] Miss. G. V. Padole, Dr. S. B. pokle, "New Iterative Algorithm for Thinning Binary Images" in Third International Conference on Emerging Trends in Engineering and Technology (Nov 2010)

[7] Peter Tarabek, "A Robust Parallel Thinning Algorithm for Pattern Recognition" in 7th IEEE International Symposium on Applied Computational Intelligence and Informatics (May 2012)

[8] Ms. Aarti Desai, Dr. Latesh Malik, "A Modified Approach to Thinning of Devanagri Characters" in IEEE International Conference on Image Processing (Apr 2011)

[9] Jun-Sik Kwon, "Improved Parallel Thinning Algorithm To Obtain Unit-Width Skeleton" in The International Journal of Multimedia & Its Applications (IJMA) (Apr 2013)

[10] Digital Image Processing by Rafael C. Gonzalez.