A modified Thinning Algorithm for Handwritten Tamil Characters

Sobhana Mari S, G. Raju

***Abstract* —This works proposes a thinning algorithm suitable for offline handwritten Tamil character recognition. The proposed method is a modification of post processing step in Stentiford Thinning (ST) algorithm. In ST algorithm, for the removal of spurious line segments, a set of matrices are defined, which remove only vertical and horizontal line segment. We defined eight more matrices such that unwanted line segment in all the direction are removed. The visual quality of the thinned output given by the proposed algorithm is found to be better than given by a set of prominent thinning algorithms. Further we carried out character recognition experiments using character images thinned with the proposed algorithm. The results show that there is increasing recognition accuracy in comparison to the result obtained when thinning is performed with other prominent algorithms.**

***Index Terms*— HCR, MST, MLP, ST, thinning.**

I. INTRODUCTION

Thinning is a most important technique for achieving efficient handwritten character recognition (HCR) system. Offline handwritten characters have great variations of writing styles such as different size, thickness and orientation angle of the characters. Different handwritten characters have different thickness in style. So in HCR, if original handwritten characters are used as such, the accuracy of recognition seemed to be very low. Due to this, thinning of characters to adopted before applying recognition method to normalized characters. Thinning reduces the image components to their essential information so that further analysis and recognition become efficient. Thinning is basically a search and deletes process that removes only those boundary pixels whose deletion do not change connectivity of their neighbours locally and do not reduce the length of an already thinned curve.

Thinning technique extensively used in the area of pattern recognition, visual inspection, handwritten character recognition, finger print recognition etc. It reduces the amount of information in image patterns to the minimum needed for recognition. Basically, a HCR consists of three major steps, pre-processing, feature extraction and pattern classification. The preliminary step for recognizing handwritten character is the pre-processing, which involves operation on the digitized image intended to reduce noise and increase the ease of extracting structural features.

Manuscript received February 05, 2015.

Sobhana Mari S, School of Computer Sciences, Mahatma Gandhi University, Kottayam, Kerala, India, +91 9446135698.

Dr.G. Raju, School of Information Science and Technology, Kannur University, Kannur, Kerala, India, +91 9446106628.

Some of the common pre-processing operations performed include threshold, noise filtering, thinning and segmentation.

This work gives emphasize only on the pre-processing step of the HCR. The pre-processing step chosen is the thinning process as it is a fundamental pre-processing operation in image analysis. It is defined as the process of reducing the width of a line-like object from several pixels to a single pixel. The resultant image is called the skeleton. Skeletons have been a useful aid in HCR because most of the relevant information in characters is not related to the thickness of the character images. Using a thinned character images helps to simplify the classification and recognition process and also producing more accurate recognition. Therefore, the performance of the recognition system depends on the thinning algorithm used in the pre-processing step. A good algorithm for thinning is needed in order to reduce the errors in classification. This has become the motivation to do a research on finding a better thinning algorithm from the HCR perspective.

Over the years, many thinning algorithms have been proposed, and a comprehensive survey of these method is contained in Lam et al, 1992[1]. Some have obtained good results, but there are still some deficiencies. In the thinning algorithm research, there exist two main problems where the actual study focuses, namely the algorithm execution time and the resulting thinned image shape.[2][3][4][5]. Many of the more recent thinning algorithms were designed to improve the speed of the algorithm. Some only focused on the quality of the skeleton produced.

Most of the research in thinning algorithms focused on the implementation of a variation on an existing thinning method. The thinning algorithms produced may also generate good skeleton for some shapes but can produce poor skeletons for others. It is very difficult to develop a generalized thinning algorithm which can produce satisfactory results for all varieties of pattern shapes.

This works proposes a thinning algorithm suitable for offline handwritten Tamil character recognition. The proposed method is a modification of post processing step in Stentiford Thinning (ST) algorithm. In ST algorithm, for the removal of spurious line segments, a set of matrices are defined, which remove only vertical and horizontal line segment. We defined eight more matrices such that unwanted line segment in all the direction are removed. The visual quality of the thinned output given by the proposed algorithm is found to be better than given by a set of prominent thinning algorithms. We implemented following thinning algorithms namely improved parallel thinning [6][12], rotation invariant thinning [7][8], Zhang Suen thinning [14], NWG thinning [10][13], Parker thinning [9], Stentiford thinning [11] and also proposes modified stentiford thinning (MST) algorithms in Tamil handwritten character set. Further we carried out character recognition experiments using character images thinned with the proposed algorithm. The results show that there is increasing recognition accuracy in comparison to the result obtained when thinning is performed with other prominent algorithms.

The rest of the paper is organized as follows. Section II explains the review of thinning algorithms. The sentiford thinning algorithm is described in section III. The proposed MST algorithm is described in section IV. The different thinning algorisms used for comparisons are included in section V. In section VI, comparative analysis of performance of different thinning algorithm is described. Finally, section VII concludes the paper.

II. REVIEW OF THINNING ALGORITHMS

Pattern recognition and image processing applications frequently deal with raw inputs that contain lines of different thickness. In some cases, this variation in the thickness is an asset, enabling quicker recognition of the features in the input image. In other cases, the variation can be a liability, and can cause degradation in the accuracy and the speed of recognition. For example, in the case of handwritten characters, the degree of uniformity of the thickness of individual strokes directly impacts the probability of successful recognition; especially if neural network based recognition techniques are employed. For the latter category of applications, uniform thickness can be attained, prior to recognition stage, by first thinning the input pattern to a thickness of a single pixel. Thinning or skeletonization is a process by which a one-pixel-width representation (or the skeleton) of an object is obtained, by preserving the connectedness of the object and its end points. Thinning algorithms should also preserve topological and geometric properties of the original object as much as possible. This includes connectedness of components, no spurious endpoints, and no excessive erosion of the original object. Thinning algorithms can be classified as one of two broad categories (Fig 1):

1) Iterative thinning algorithms
2) Non-iterative thinning algorithms.

In general, iterative thinning algorithms perform pixel-by-pixel operations until a suitable skeleton is obtained. Non-iterative thinning methods use sequential pixel scan of an image. Iterative thinning algorithms delete successive layers on the edges of a pattern until a skeleton remains. Usually the pixels of an image are considered consecutively and a choice is made to either delete or keep the pixel. The criterion for this choice is a set of “rules” based on the pixels in the neighborhood around the current pixel. Usually the neighborhood is a 3*3 area around the pixel.

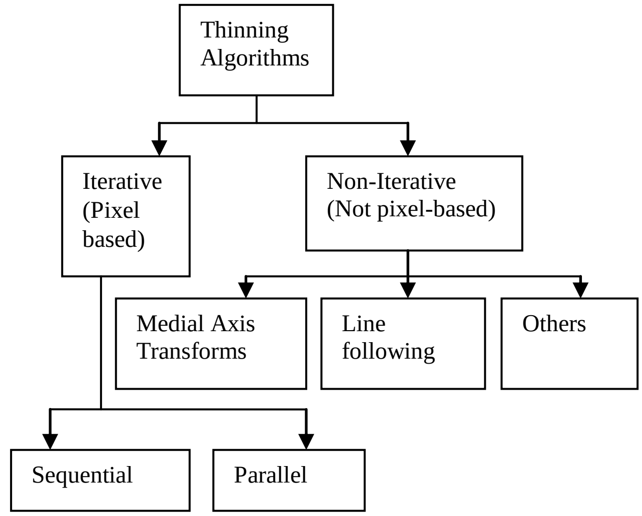


Fig 1 Classifications of common thinning algorithms

Iterative algorithms may be classified as either sequential or parallel. Sequential thinning algorithms [17] examine contour points of an object in a pre-determined order. Either raster scan or contour following accomplishes this. Raster scanning is essentially scanning an image until a contour point is found and then applying a set of rules. Contour following algorithms pre-compute the border pixels of connected objects and can provide a speed increase since every pixel of the image does not have to be examined. In parallel thinning algorithms [13], [15], [16], [18], [19], the decision for individual pixel deletion is based on the results of the previous iteration. Like sequential algorithms, parallel thinning usually considers a 3*3 neighborhood around the current pixel. A set of rules for deletion is applied based on pixels in the neighborhood. Fully parallel algorithms have trouble maintaining connectedness, so they are often broken into sub-iterations where only a subset of the pixels is considered for deletion.

Non-iterative thinning methods are not based on examining individual pixels. Some popular non-pixel based methods include medial axis transforms, distance transforms, and determination of centerlines by line following. In line following methods, midpoints of black spaces in the image are determined and then joined to form a skeleton. This is fast to compute but tends to produce noisy skeletons. It has been conjectured that human beings naturally perform thinning in a manner similar to this. Another method of centerline determination is by following contours of objects. By simultaneously following contours on either side of the object a continual centerline can be computed. The skeleton of the image is formed from these connected centerlines. Medial axis transforms often use gray-level images where pixel intensity represents distance to the boundary of the object. The pixel intensities are calculated using distance transforms. Over the years, many thinning algorithms have been proposed and comprehensive survey of this method is contained in [1]. Many thinning techniques [4][6][8][11] [14] reported have obtained fairly good results. But there are still performance limitations in the context of different scripts as well as applications.

Based on the review of related works, we have chosen the following algorithms in our study: Parallel Thinning [6], Rotation Invariant Thinning [8], Zhang Suen Thinning [14], NWG Thinning [13], Parker Thinning [9] and Stentiford Thinning [11] and proposes modified Stentiford thinning (MST) algorithms in Tamil handwritten character set and the results are compared as described below.

III. STENTIFORD THINNING ALGORITHM

Stentiford thinning algorithm [Stentiford & Mortimer, 1983] uses some pre-processing stages before a Zhang Suen thinning algorithm is applied. These pre-processing heuristics are specifically aimed at reducing the failure rate due to the defects in the data. The pre-processing steps such as hole removal, smoothing and acute angle emphasis are described below.

i) Hole removal: Hole removal is done using six patterns of bits or templates as shown in Fig.2. These holes are removed by either merging the hole with adjacent white areas by the removal of black elements or filling it.

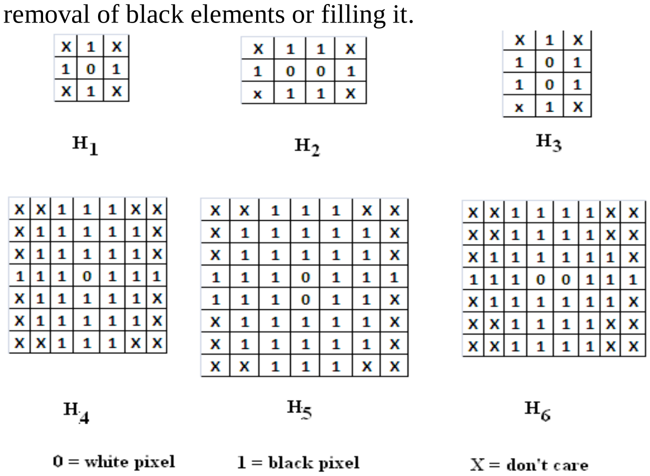


Fig 2 Shows patterns of hole removal [Stentiford & Mortimer, 1983]

ii) Smoothing: Smoothing is the removal of all black elements having less than three black neighbours and having connectivity 1. Stentiford thinning algorithm uses Yokoi's 8 connectivity number [24]. This has the effect of removing single elements projections and all isolated spots having one or two elements.

iii) Acute angle emphasis: The final pre-processing stage involves the detection of upward and downward acute angles (Di and Ui) between limbs by again scanning the character with the patterns of bits [9]. Five of these (Di) will fit the sharpest forms of downward pointing acute angles and (Ui) will fit the sharpest forms of upward pointing acute angles as shown in Fig.3. After a fit is found the central black element is deleted and the process is repeated twice more. Stentiford and mortimer described the summary of all preprocessing steps in [11].

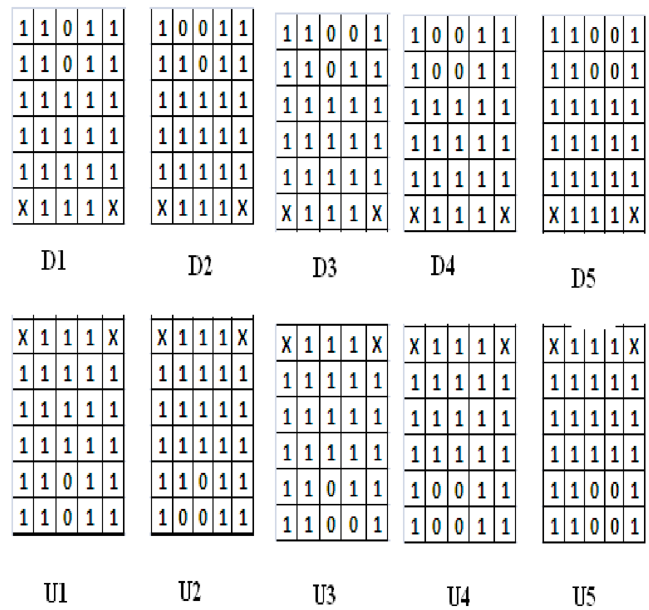


Fig 3 Templates used for the acute angle emphasis.

After pre-processing Zhang Suen thinning algorithm is applied. Four matrices Mi (where i =1, 2, 3 and 4) in Fig. 4 are scanned over the character and wherever a matrix fits the central black element is marked for deletion. Elements are not so marked if they are limb endpoints or if the 8-connectivity measure for that point is greater than one.

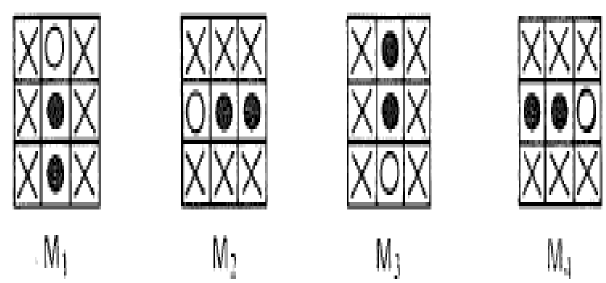


Fig. 4 Thinning matrices [Stentiford & Mortimer, 1983]

An endpoint is defined as a black element which is eight connected to only one other black element. Elements already marked are considered to be white for the purpose of subsequent end point or connectivity calculations. When all four matrices have been scanned in this way, all marked elements are deleted and the process repeated until no more erosion can take place.

The results of above process are sensitive both to the order of application of the matrix Mi and also the direction of scan. An ordering which minimize spurious tail production is given in a Table I. For example M3, the south edge eroding matrix is scanned from right to left moving upwards across the character.

Table I Matrix Ordering and Scan Direction

Order of Application	Direction of Single scan line	Direction of successive scan line
M1	Left to Right	Downwards
M2	Upwards	Left to Right
M3	Right to Left	Upwards
M4	Downwards	Right to Left

IV. MODIFIED STENTIFORD THINNING (MST) ALGORITHM

In this section, we propose a modification to the stentiford thinning algorithm. It also uses three steps such as pre-processing, thinning and post processing. The MST pre-processing steps are described given below:

A. Pre-processing

Hole removal is done using six patterns of bits or templates as shown in Fig.2. Smoothing is the removal of all black elements, having less than three black neighbours and connectivity 1. The Stentiford thinning algorithm uses Yokoi's eight connectivity number. But it does not preserve all the connectivity in a Tamil character image. So, the proposed MST thinning algorithm uses Zhang Suen eight connectivity numbers [12]. A 3x3 template window is used to define the pre-processing step as shown in Fig.5, where P1 refers to the current pixel of interest and P2 to P9 are considered as neighbouring pixels. This pre-processing step is applied only to pixels with value 1. A decision value is computed as the number of times the neighbourhood pixel value change from 0-1, when the neighbours are visited in the order P2, P3, P4, P5, P6, P7, P8, P9 and P2. If this decision value is 1 then the pixel under consideration is set to zero. For example, if the neighbourhood is visited in order P2, P3, P4, P5, P6, P7, P8, P9 and P2 results in a vector 0010000010, then the decision value is 2. This has the effect of removing single elements projections and all isolated spots having one or two elements. This step is followed by the detection of upward and downward acute angles (Fig 3) as described in section III.

P9	P2	P3
P8	P1	P4
P7	P6	P5

Fig 5 Designations of the nine pixels in a 3x3 window

B. Thinning

After pre-processing, Zhang Suen thinning algorithm is applied as described in Section V.

C. Post processing

In post processing, extra lines in thinned images are removed using endpoint removal matrices (Fig.6). The endpoint removal matrices *(Mi,Bi,Ci)* are scanned over the character and wherever a matrix fits, the central black pixel is marked for deletion. In stentiford thinning algorithm, only four matrices Mi are used to delete a central black pixel. But the proposed MST algorithm, we used two new matrices *Bi* and *Ci* in addition to *Mi* where i= 1 2 3 and 4. Because Mi scans matrices remove only left, right upward and downward extra lines. In order to remove slanting extra lines, the MST algorithm is used *Bi* and *Ci* scan matrices as shown in Fig.6. All matrices are scanned over the character and wherever a matrix fits, the central black pixel is marked for deletion. When all four matrices of *Mi*, *Bi* and *Ci* have been scanned in this way, all marked pixels are deleted and the process is repeated until no more erosion can take place. The pixel deletion processes are sensitive to both the order of application of the scan matrices (*Mi*, *Bi*, *Ci*) and the direction of scan (given in Table II).

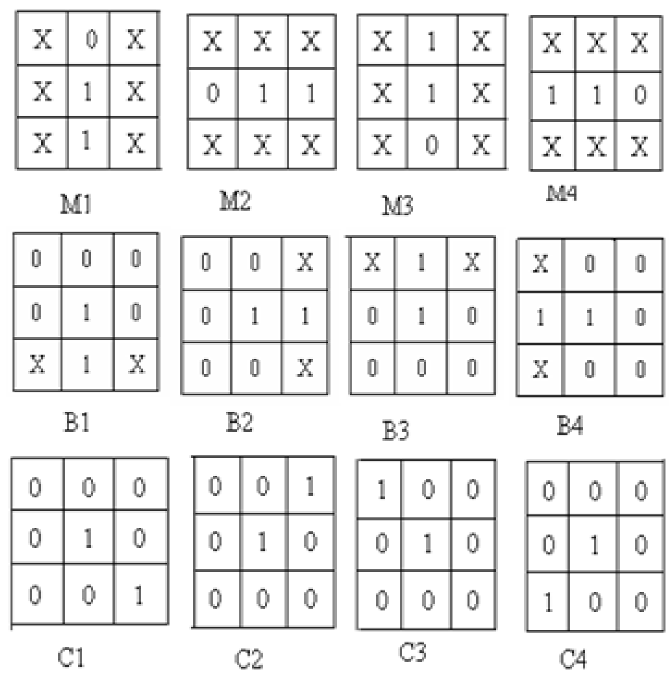


Fig.6 End point removal matrices

The post processing in the proposed MST algorithm can be summarized as follows:

Step1. Scan matrix M1 across character according to table II and identify next fit position
Step2. If the central element at a fit is not an endpoint and has connectivity value one then mark it for deletion.
Step3. Repeat Steps 1 and 2 for all fit positions
Step4. Repeat Step 1-3 for each scan of matrix M2, M3, M4.
Step5. Delete all marked elements.
Step6. If one or more elements are deleted in Step 4 then return to Step1.
Step7. Repeat Step 1-6 for scan matrices Bi and Ci .
Step8. Exit

Table II Matrix ordering and scanning directions of modified Thinning algorithm

Order of Application	Direction of Single scan line	Direction of successive scan line
M1,B1,C1	Left to Right	Downwards
M2,B2,C2	Upwards	Left to Right
M3,B3,C3	Right to Left	Upwards
M4,B4,C4	Downwards	Right to Left

Fig.7 illustrates the comparison of ST and MST algorithms. It is clear that the MST algorithm removes all spurious tails but ST algorithm removes only vertical and horizontal lines.

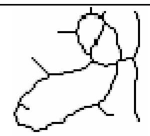
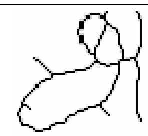
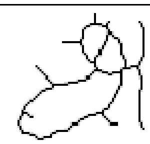
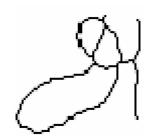
Thinning Algorithms	Post processing	
	inputs	Outputs
ST		
MST		

Fig.7 Comparison of ST and MST algorithm

V. DIFFERENT THINNING ALGORITHM USED FOR COMPARISON

A. Parallel Thinning Algorithm

The improved parallel thinning algorithm is fast and efficient for HCR. The deleting process take place based on some elimination rules [6]. Elimination rule is the kernel of a thinning algorithm which decides the performance of the thinning algorithm. All these rules are applied simultaneously to each pixel. In some peculiar cases these cannot be performed very well. The two pixel width or even pixel width in horizontal or vertical direction may be deleted, which would cause the loss of connectivity of object pattern. In order to keep up the connectivity, these pixels should not be deleted. So some new rules [7] are used to solve the problem of dis-connectivity. A drawback of this algorithm is that a large black block whose length and width are both greater than 1, is usually changed to a line or a dot [9]. In the field of handwritten character recognition this is not reasonable in some cases. In handwritten Tamil character recognition system, there are a lot of samples such as ஒ □□ எ which contains the black block. Fig 7 shows result of improved parallel thinning algorithm applied on a character எ. This may result in extra lines or loss of information or misidentification.

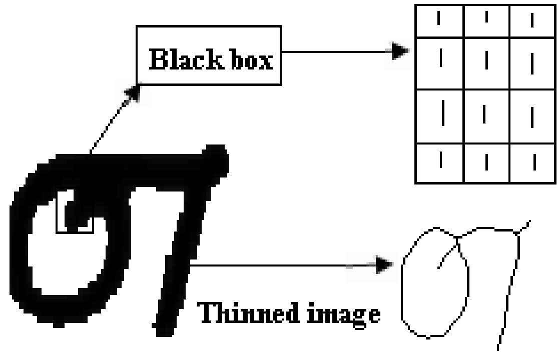


Fig 7 Result of improved parallel thinning algorithm

B. Rotation Invariant Rule Based Algorithm

This algorithm proposed by Ahmed and Ward is used to thin the character patterns to their central line. This algorithm proceeds by deriving a set of 20 rules [8] over the eight neighbours of the pixel, which is a candidate for deletion. It is iterative in nature. Each iteration it deletes every point that lies on the outer boundaries of the character pattern, as long as the width of the pattern is more than one pixel wide. The iterations are repeated until no further changes occur. If the resultant pattern, at some point, has width (measured in any direction) equal to one pixel then this pixel belongs to the central line of the symbol and will not be deleted. If the width at any point is two pixels, then the central line passes between these two pixels. This case is separately treated in the algorithm. The pixel is deleted if any one of the twenty conditions is satisfied. Drawback of the above procedure is that, when a part of a pattern is two pixels wide in the horizontal or vertical direction these two pixels may be deleted and this will disconnect these central lines. In order to solve this problem, we must first check whether the width of the pattern is two pixels or not in each iteration. Hence, the above algorithm is modified by incorporating these conditions [7]. By applying the above modified procedure into handwritten character, the characters are thinned to their central lines without extraneous branches. The resultant central lines are connected and are of one pixel width. In some cases, the algorithm may results in extra branches. A white pixel inside black boundary will turn to an extra hole after thinning as shown in Fig 8.

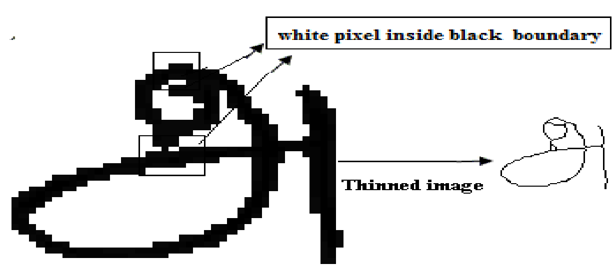


Fig 8 Results of rotation invariant algorithm applied on character அ

C. Nagendra Prasad-Wang-Gupta thinning algorithm

This algorithm [10] produces simpler and more elegant skeletons of handwritten characters at zero extra computational cost. The algorithm uses masks in order to select pixels to be turned off. The 8 closest neighbors are numbered following a clockwise walk around a pixel p, which starts at the middle of upper edge. The NWG algorithm is given below

```
Start
g = 1; h =1; Q0=Q ;( Initial settings)
While (h = 1)
h = 0; Q = Q0;
If (g = 1) then g =0;
        Else g=1;
End if
For (every pixel p belongs to Q)
If ((1 < b(p) <7) and (a(p) = 1 or c(p) = 1)) then
          If ((g = 0) and (e(p) =0)) then
             Erase p in Q0; h=1;
          End if
       End if
End for
End while
End
```

where b(p) is the number of neighbors of p which are on (pixels with value 1), a(p) is the number of off-to-on transitions when the neighbours are visited following a clockwise walk around p in the order of p(0), p(1),…..p(7),p(0) and the functions c(p), e(p) and f(p) are given by:

$c(p) = 1$ if $p(0)=p(1)=p(2)=p(5)=0$ and $p(4) = p(6) = 1$
1 if $p(2)=p(3)=p(4)=p(7)=0$ and $p(6) = p(0) = 1$
0 otherwise

$e(p)=(p(2)+p(4))*p(0)*p(6)$ and $f(p)=(p(6)+p(0))*p(4)*p(2)$

The skeleton obtained is connected and eliminates almost any redundant pixels which are not relevant in order to recognize the handwritten character. However, due to an asymmetry in the algorithm, some superfluous pixels are not removed. This may be corrected by simply changing the function c(p) in those iterations where g = 1 (this means odd iterations). That is, the condition a(p) = 1 or c(p) = 1 is replaced by

$a(p) = 1$ or $(1 - g)\cdot c(p) + g\cdot d(p) = 1$ with

$d(p) = 1$ if $p(1)=p(4)=p(5)=p(6)=0$ and $p(0) = p(2) = 1$
1 if $p(0)=p(3)=p(6)=p(7)=0$ and $p(2) = p(4) = 1$
0 otherwise

The result of NWG thinning algorithm usually produce more elegant skeleton in the sense that some redundant or confusing pixels are removed. It may result the loss of information in some cases as shown in Fig 9.

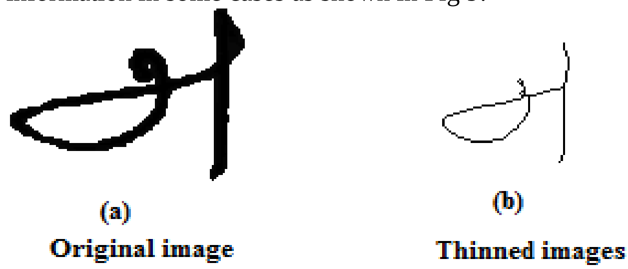


Fig 9 Results of NWG algorithm applied on character அ

D. Zhang and Suen thinning algorithm

This algorithm is described in [14], has been widely used as it is fast and simple to implement. This algorithm repeatedly deletes contour pixels respecting a number of conditions until a one-pixel wide 8-connected skeleton is obtained. In this algorithm, a 3 x 3 window is used. The 8 closest neighbors are numbered following a clockwise walk around a pixel p1, which starts at the middle of upper edge. Let NT (P1) be the number of zero (white) to non-zero (black) transitions in the ordered sequence <P2, P3…P9, P2>, and NZ (P1) be the number of nonzero neighbours of P1. Pixel P1 is deleted (set to zero). If $2 \le NZ(P1) \le 6$ and $NT(P1) = 1$ and $P2.P4.P8=0$ or $NT(P2) \neq 1$ and $P2.P4.P6=0$ or $NT(P4) \neq 1$.The process is repeated until there are no more changes in the image.

Although Zhang-Suen thinning algorithm has advantages of implementation speed and topological preservation, it still has other weaknesses. For example this algorithm deletes all 2X2 pixel image and some of the skeletons contain a defects such as necking and line fuzz as shown in Fig 10. Therefore, [Parker, 1997] has introduced an improved thinning algorithm to cater the problems posed by Zhang Suen thinning algorithm.

Fig 10 Results of Zhang Suen thinning algorithm applied on character image அ

E. Parker thinning algorithm

This algorithm merges three methods in a sequence order; Stentiford's pre-processing [11] scheme feeding images into Zhang-Suen's basic algorithm, followed with Holt's staircase removal as a post-processor. The classic thinning artefacts are necking, tailing and spurious projection, hairs or line fuzz. Necking happens when a narrow point at the intersection of two lines is stretched into a small line segment. Excess thinning done at where two lines meet at an acute angle may produces tails. Spurious projection refers to the creation of extra line segments joining a real skeletal segment. Basically, smoothing is done by making a pass over all pixels, deleting those having two or fewer black neighbours and having connectivity less than two. A procedure called acute angle emphasis [9] is used to deal with necking. Sometimes, when thinning is complete, there are still pixels that could be deleted. Principal among these are pixels that form a staircase; clearly half of the pixels in a staircase could be removed without affecting the shape or connectedness of the overall pattern. The result of Parker thinning algorithm usually produce more elegant skeleton. But some of the skeletons contain defects as shown in Fig11.

(a)
Original image

(b)
Thinned images

Fig 11 Results of parker thinning algorithm

VI.COMPARATIVE ANALYSIS OF PERFORMANCE OF DIFFERENT THINNING ALGORITHMS

Comparative analysis of the performance of different thinning algorithm is carried out by taking character recognition as an example. The details of the experiments are given the subsequent sections.

A. Feature extraction

In this work, the zero crossing method [21] is used for feature extraction. The feature extraction algorithm is as follows: Repeat step I and II for all preprocessed image stored in the database.

Step I a) The parallel thinning algorithm is applied on input character

b) The thinned image is divided into 9 equal sized blocks. For each of these blocks, the sum of black to white transition (zero crossing) of all pixels in each row and column is calculated and stored the 18 feature of the character and class id into a file.

Step II Same as step I(b) in rotation invariant thinning, Zhang Suen thinning, NWG thinning and Parker thinning algorithms

The output of the algorithm contains a set of five files, each containing feature vectors together with class id of the 10200 images.

B. Classifier

In this work, the MLP classifier [21] is used for classification. Multi layer feed-forward networks form an important class of neural networks [19]. Typically the network consists of a set of sensory units or input nodes that constitute the input layer, one or more hidden layers of neurons or computation nodes and an output layer. The performance of the MLP network will highly depend on the structure of the network and training algorithms [19] [20]. Multi layer Perceptron (MLP) neural networks with sufficiently many nonlinear units in a single hidden unit layer have been established as universal function approximates. The advantages of the MLP are:

- Hidden unit outputs (basis functions) change adaptively during training, making it unnecessary for the user to choose them beforehand.
- The number of free parameters in the MLP can be unambiguously increased in small increments by simply increasing the number of hidden units.
- The basic functions are bounded making overflow errors and round-off errors unlikely.

The MLP is a feed-forward network consisting of units arranged in layers with only forward connections to units in subsequent layers. The connections have weights associated with them. Each signal traveling along a link is multiplied by its weight. The input layer, being the first layer, has input units that distribute the inputs to units in subsequent layers. In this work, the network structure we used contains 18 input nodes, one hidden layer with 10 nodes and thirty four output nodes. MLP network is trained using feed forward Back propagation network algorithm.

C. Experimental Analysis

In the present study we considered only isolated Tamil characters of vowels and consonants. 300 handwritten pages containing the selected 34 characters are collected from different persons belonging to different age groups, qualification, and profession. The collected documents are scanned at 300 DPI. The resultant data base consists of 300 samples each of the 34 selected character classes. In the experiment, the 18 feature vector is given as input to the MLP network. The five thinning algorithms described above are tagged as Parallel Thinning (PT), Rotation Invariant Thinning (RIT), Zhang Suen Thinning (ZST), NWG Thinning (NWGT) Parker Thinning (PKT) Stentiford thinning (ST) and proposed modified stentiford thinning (MST) algorithms. For these methods, training and testing data sets are formed as follows:

The four pairs of training and testing sets are formed by randomly selecting 225 training samples per character and 75 test samples per character. In each training set, there are 7650 samples and remaining 2550 samples are used for testing. The seven thinning algorithms are applied to a database of Tamil handwritten character set and the results are compared based on skeleton produced, execution time and recognition rate. In order to evaluate the classification performance, feature extraction based on zero crossing and classification using

MLP are carried out on the thinned images using the different algorithms described above.

D. Analysis and Comparison of MST Algorithms

For evaluating the quality of thinning algorithms we consider connectivity preservation, the width of skeleton, shape of the character, classification performance and execution time. The seven thinning algorithms described above are tagged as Parallel Thinning (PT), Rotation Invariant Thinning (RIT), Zhang Suen Thinning (ZST), Nagendra prasad-Wang-Gupta Thinning (NWGT), Parker Thinning (PKT), Stentiford thinning (ST) and Modified Stentiford Thinning (MST). Table III shows the result of different thinning algorithms applied on five handwritten Tamil character images. Table IV shows the execution time of different thinning algorithms applied on five handwritten Tamil character images. Fig.12 shows the graphical representation of the execution time when applied to five handwritten Tamil characters. The proposed modified stentiford thinning algorithm gives better output and high recognition accuracy (90.9%) than the other six thinning algorithms. The execution time of the proposed algorithm is slightly greater than the other methods except NWGT. But the output of NWGT is relatively inferior to MST.

Table III Different thinning algorithms applied on five handwritten Tamil characters

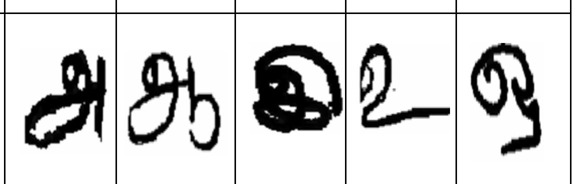
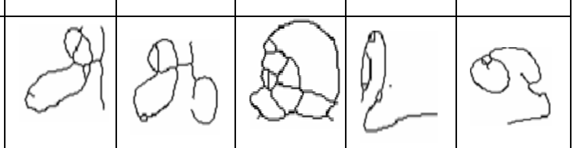
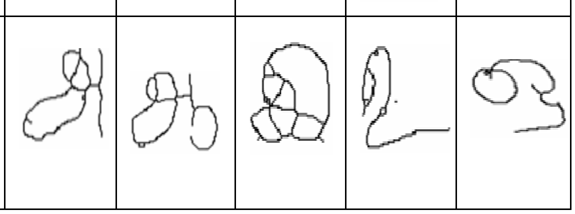
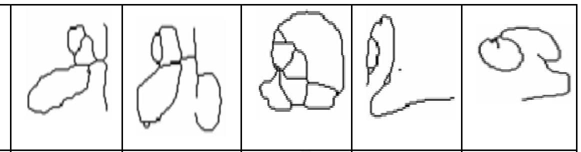
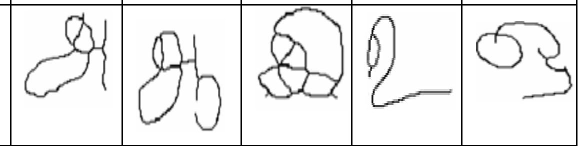
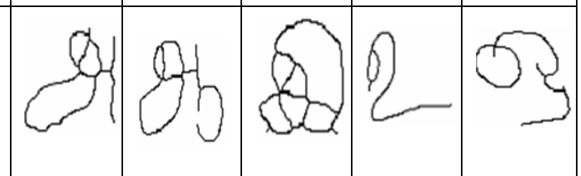
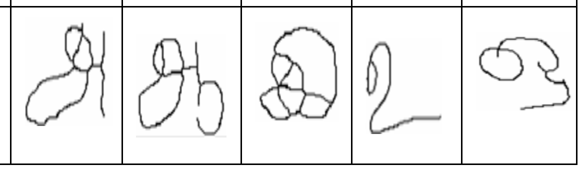
Methods \ Images	Sample 1	Sample 2	Sample 3	Sample 4	Sample 5
Original Images					
PT					
RIT					
ZST					
NWGT					
PKT					
ST					
MST					

Table IV Execution time of different thinning algorithms applied on five handwritten Tamil characters

Execution time in milliseconds					
Thinning methods	Sample1	Sample2	Sample3	Sample4	Sample5
PT	9.109	5.969	11.766	5.875	7.641
RIT	9.812	7.187	12.578	6.015	8.422
ZST	9.422	7.188	13.328	6.875	9.218
NWGT	2.5	2.156	1.641	1.454	2.719
PKT	10.14	7.562	14.019	7.383	8.891
ST	10.24	7.852	14.119	7.453	8.918
MST	10.28	7.902	14.12	7.543	8.924

E. Influence of Thinning Algorithms on Classification Performance

In order to evaluate the influence of the thinning on classification performance, experiments are conducted with thinned images obtained from the different thinning algorithms. Zero crossing method is selected on the feature and MLP on classifier. Features are extracted adopting Fixed Meshing (FM) strategies.

The four pairs of training and testing sets are formed by randomly selecting 225 training samples per character and 75 test samples per character. In each training set, there are 7650 samples and remaining 2550 samples are used for testing.

Table V shows the classification performance of different thinning algorithm applied on 10200 handwritten Tamil characters. The proposed thinning algorithm gives better performance (90.9%) than the other six thinning algorithms. Fig.13 shows the performance comparison of different thinning algorithm

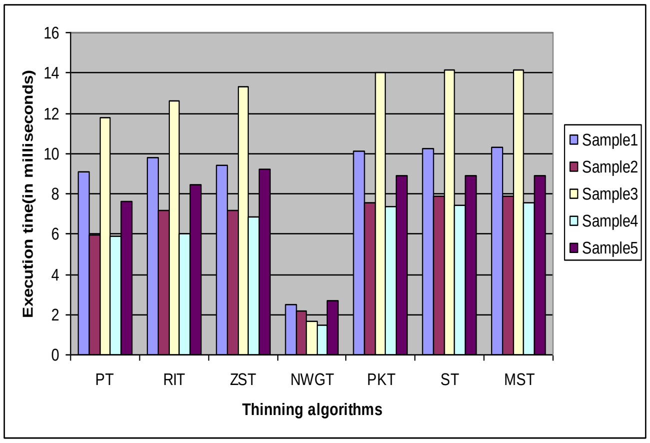


Fig.12 Execution time of five handwritten Tamil characters.

Table V **Recognition accuracy (%) based on different thinning algorithm**

Recognition accuracy (%)							
Data set	**PT**	**RIT**	**ZST**	**NWGT**	**PKT**	**ST**	**MST**
1	89.76	90.16	90.35	90.59	90.08	90.31	90.98
2	89.96	90.43	90.94	90.67	90.55	90.48	90.89
3	88.78	88.71	89.18	88.98	89.96	90.03	90.33
4	90.39	90.63	90.31	90.43	90.55	90.63	91.39
avg	**89.72**	**89.98**	**90.2**	**90.17**	**90.29**	**90.36**	**90.9**

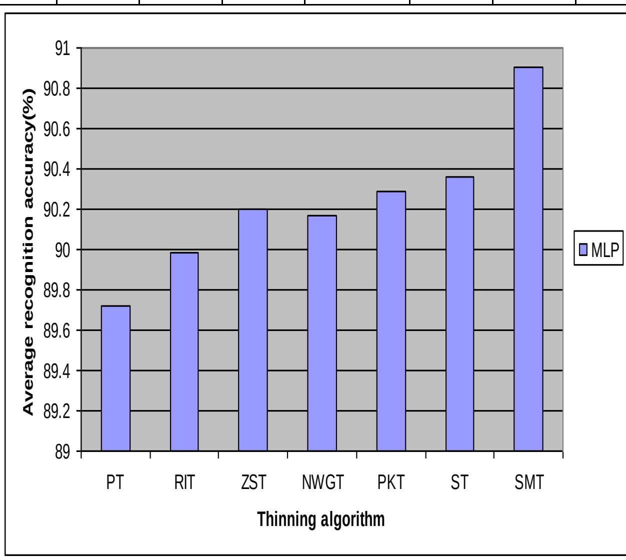


Fig. 13 Performance comparison of different thinning algorithm

VII. CONCLUSION

Thinning is a most important technique for achieving efficient HCR system. This work proposes a new thinning algorithm for offline handwritten Tamil character recognition. The proposed thinning algorithm is implemented and applied on a set of handwritten character images. The performance of the algorithm is analyzed in terms of image shapes, execution time and classification performance. In comparison with the performance of a set of six prominent thinning algorithms, the proposed method gives better performance.

REFERENCE

[1] Louisa Lam,Seong-Whan Lee, Ching Y.Suen,"Thinning Methodologies -A Comprehencsive survey" IEEE Transactions on pattern analysis and machine intelligence, vol. 14, no. 9, september 1992

[2] N.H.Han, C.W.La, and P.K.Rhee, "An Efficient Fully Parallel Thinning Algorithm," in Proc. IEEE Int.Conf.Document Analysis and Recognition, Vol.1,pp.137-141(1997).

[3] P. S. P. Wang and Y. Y. Zhang,"A Fast and Flexible Thinning Algorithm" , IEEE Transactions on computer society, vol. 38, no. 5, may 1989

[4] Ben K. Jang and Roland T, Chin"One-Pass Parallel Thinning: Analysis , Properties,and Quantitative Evaluation"IEEE Transactions on pattern analysis and machine intelligence, Vol.14, No.ll, November1992

[5] Antoine Manzanera,Thierry M. Bernard, Francoise Preteux and Bernard Longuet , "Ultra –fast Sekeleton based on an isotropic fully parallel algorithm" Discrete Geometry for Computer Imagery Lecture Notes in Computer Science Volume 1568, 1999, pp 313-324

[6] Lei Huang, Genxun Wan, Changping Liu ,"An Improved Parallel Thinning Algorithm" Institute of Automation, Chinese Academy of Sciences, Proceedings of the Seventh International Conference on Document Analysis and Recognition (ICDAR 2003), IEEE computer society

[7] Lajish V.L., "Adaptive Neuro-fuzzy Inference based Pattern Recognition Studies on Handwritten Character Images, Ph.D Thesis, Calicut University, 2006.

[8] Maher Ahmed and Rabab Ward, Fellow, IEEE, "A Rotation Invariant Rule-Based Thinning Algorithm for Character Recognition", IEEE Transactions on pattern analysis and machine intelligence, vol. 24, no. 12, december 2002

[9] Adeline anak and Engkamat ,"Enhancement of Parallel Thinning Algorithm for Handwritten Characters using Neural Network ,Ph.D thesis", Faculty of Computer Science & Information Systems University Teknologi Malaysia November-2005

[10] Rafael C. Carrasco and Mikel L. Forcada, "A note on the Nagendraprasad-Wang-Gupta thinning algorithm" , Pattern Recognition Letters, Volume 16 Issue 5, May 1995, Pages 539 - 541 ,Elsevier Science Inc. New York, NY, USA

[11] F.W.M.Stentiford and R.G.Mortimer," Some New Heuristics for Thinning Binary Hand printed Character for OCR" IEEE Transaction on systems , man, and cybernetics,vol.smc-3. no.1, January 1983

[12] Y.Y.Zhang and P.S.P.Wang, "Design of Parallel Thinning Algorithms", 9th international Conference on Pattern Recognition , College of computer Science , North eastern University, Boston, MA

[13] M.V.Nagendraprasad, Patrick S.P.Wang and Amar Gupta, Algorithms for Thinning and Rethickening Binary Digital Patterns" ,Massachusetts Institute of Technology, Cambridge, Digital signal processing 3,97-102(1993)

[14] H.Bunke, P.S.P Wang "Hand book of character recognition and Document Image analysis "PP. 1-47, world Scientific Publishing company

[15] R. W. Hall, "Fast parallel thinning algorithms: parallel speed and connectivity preservation," Commun. Assoc. Comput. Mach., vol. 32, no. 1, pp. 124-131, Jan. 1989.

[16] Z. Guo and R. W. Hall, "Parallel thinning with two-subiteration algorithms," Commun. Assoc. Comput. Mach., vol. 32, no. 3, pp. 359-373, Mar. 1989.

[17] C. J. Hilditch, "Linear skeletons from square cupboards," in Machine Intelligence (B. Meltzer and D. Michie, Eds.) New York: American Elsevier, 1969, pp. 403-420, vol. 4.

[18] N.H.Han, C.W.La, and P.K.Rhee, "An Efficient Fully Parallel Thinning Algorithm," in Proc. IEEE Int.Conf.Document Analysis and Recognition, Vol.1, pp.137-141(1997).

[19] H.H. Chen, M. T. Manry, and H. Chandrasekaran, "A neural network training algorithm utilizing multiple sets of linear equations," Neurocomputing, vol. 25, no.1-3, pp. 55-72, Apr 1999.

[20] C.Yu and M. T. Manry, "A modified hidden weight optimization algorithm for feed-forward neural networks," Thirty-Sixth Asilomar Conference on Signals, Systems & Computers, vol. 1, pp. 1034-1038, Nov. 2002.

[21] G.Raju and K.Revathy,"Wave packets in the Recognition of Isolated Handwritten Characters" Proceedings of the World Congress on Engineering 2007, Vo1.1, London,U.K.

Sobhana Mari S (MCA) is a research scholar, School of Computer Science, Mahatma Gandhi University, Kottayam, Kerala, India. She has presented over 8 papers in International and National conferences. Her area of interest includes Document image processing, neural networks, image processing and cryptography.

Dr. G. Raju (M Sc (Physics), MCA, M Tech (comp & IT), Ph D, B Ed) is Associate Professor & Head, School of Information Science and Technology, Kannur University, Kannur, Kerala, India. He has successfully guided 6 Ph D and 2 M Tech thesis. He has authored more than 40 journals papers and 60 conference papers. He has published 34 documents in Scopus; 57 citations; h-index 5. He is a member of various international and national conferences. He is a visiting faculty in reputed institutions – Central University of Kerala and Rajiv Gandhi institute of Technology, Kottayam. His area of interest includes Document Image Processing – Recognition of Handwritten Malayalam / Tamil documents, Data Mining -Rough sets in Data Mining, Fuzzy Clustering Algorithms, Text Mining and Web Mining and Image Processing –Medical Image processing (Segmentation, Fusion, classification, enhancement), Bio-metrics(Ear), medical image enhancement using soft computing techniques, Plant leaf classification.