VOL. 5, NO. 3, MARCH 2010 ISSN 1819-6608
ARPN Journal of Engineering and Applied Sciences


www.arpnjournals.com

CLASSIFICATION AND IDENTIFICATION OF TELUGU HANDWRITTEN CHARACTERS EXTRACTED FROM PALM LEAVES USING DECISION TREE APPROACH

Panyam Narahari Sastry[1], Ramakrishnan Krishnan[2] and Bhagavatula Venkata Sanker Ram[3]
[1]Department of Electronics and Communication Engineering, CBIT, Hyderabad, India
[2]Indian Institute of Space Science Technology, Department of Space, Trivandram, India
[3]Department of Electrical and Electronics Engineering, JNTUH College of Engineering, Hyderabad, India
E-Mail: ananditahari@yahoo.com

ABSTRACT

Research in character recognition is very popular for various application potentials in banks, post offices, defense organizations, reading aid for the blind, library automation, language processing and multi-media design. Even though Epigraphical work dealing with stone inscriptions have been analyzed, these have been done largely manually and also on 2D traces. A large collection of these are available in the classical Indian languages like Sanskrit, Tamil, Pali etc as well as in more modern languages like Telugu. These characters on the palm leaf have the additional properties like depth, an added feature which can be gainfully exploited in character recognition. In this paper, we explore how these 3D features can be extracted and how they can be used in the recognition and classification process. This paper describes a system to identify and classify Telugu (a south Indian language) characters extracted from the palm leaves, using Decision Tree approach. The decision tree is developed using SEE5 algorithm, which is an improvement from the predecessor ID3 and C4.5 algorithm. The identification accuracy obtained is 93.10% using this method.

Keywords: palm leaf, Telugu handwritten characters, decision tree, SEE5 algorithm, 3D features.

1. INTRODUCTION

Character recognition is one of the oldest applications of automatic pattern recognition. Research in character recognition is very popular for various application potentials in banks, post offices, defense organizations, reading aid for the blind, library automation, language processing and multi-media design.

1.1 Offline and on-line handwritten character recognition systems

Off-line handwriting recognition is the task of recognizing the image of a handwritten text, in contrast to on-line recognition where the dynamic characteristics of the writing are available as well (as is the case with the pen input of handheld devices). They do not carry temporal or dynamic information such as the number and order of pen-on and pen-off movements, the direction and speed of writing and in some cases the pressure applied while writing a character. They are not real time recognizers. Off-line conversion usually requires costly and imperfect pre-processing techniques prior to feature extraction and recognition stages.

1.2 Difficulties encountered in handwriting recognition

1.2.1 Styles of Handwritten data

Handwritten data is also characterized by the writing style of users. Actually, the difficulties encountered in handwriting recognition lie in the freedom the user takes when he writes. The more freedom the user takes, the more difficult the recognition of the data will be.

1.2.2 Irregular writing

Irregular handwriting aggravates ambiguities and makes it harder to group symbols and to distinguish relations among them. A cause of this is due to inexperienced users, because they normally take excessive freedom with the location and alignment of handwritten symbols. Other kinds of irregular writing arise during the correction, deletion, and insertion of symbols. Something like this could generate a very complex expression, which could not be recognized even by humans.

1.3 Background on palm leaves

Palm leaves have been a popular writing medium for over two thousand years [1] in South East Asia. Use of palm leaves for recording literary and scientific texts have been reported from about the fifth Century B.C, with the oldest existing documents dating from the second century A.D. The manuscripts are typically created by using a metallic stylus to etch letters into the dried leaf and enhancing the contrast and legibility of the script by applying lampblack or turmeric mixed with aromatic oils chosen for their insect repellent qualities.

A survey by the Institute of Asian Studies, Chennai, indicates that there are still about a hundred thousand palm leaf manuscripts surviving in South Indian repositories alone with many more scattered across India, Nepal, Thailand Cambodia and other south east Asian countries. These manuscripts contain religious texts and treatises on a host of subjects such as astronomy, astrology, architecture, law, medicine and music.

Most of these palm leaves are nearing the end of their natural lifetime or are facing destruction from elements such as dampness, fungus, ants and cockroaches. This has spurred many new preservation projects to

protect these valuable historical documents. Conversion of the palm leaf contents into machine readable form is the only sure way of preserving the contents of the palm leaves.

1.4 Characteristics of Telugu

India is a multi-lingual multi-script country [2] and Telugu is one of the prominent scripts in India and Asia with more than 62 million speakers. There are 18 vowels, 36 Consonants, and three dual symbols in this language. The progress of character recognition in Asian and particularly Indian scripts is in a relatively nascent stage as compared to English, which is in a mature stage of development due to the following reasons;

a) Compared to English, Indian languages have more basic characters and also composite characters.
b) Handwritten character recognition is more complex, than the printed character recognition
c) Because of the sheer, number of English speakers, OCR in English is highly developed. With smaller number of people, languages like Telugu have not attracted equivalent efforts.
d) Palm leaf character recognition is even more complex, than handwritten character recognition. This is due to the different palm leaf characteristics, like softness etc and also due to aging and deterioration of palm leaf. Some of the Telugu characters (Vowels) are shown below in the Table-1 along with their pronunciation details.

Table-1. Pronunciation of Telugu vowels.

Vowels:							
అ	ఆ	ఇ	ఈ	ఉ	ఊ	ఋ	ౠ
a	Aa	i	ee	u	oo	aRu	aRoo
ఎ	ఏ	ఐ	ఒ	ఓ	ఔ	అం	అః
e	ae	ai	o	oa	ow	am	aha

1.5 Existing methods of handwritten character recognition systems and their limitations

Hidden Markov Model [3] has been popularly used for off-line recognition of cursive handwriting with the reported recognition rate over 98% on the word level. The algorithm is applied to the skeleton graph of a word that extracts the edges in a particular order representing information about the location of the edge relative to the four reference lines, its curvature and the degree of the nodes incident to the considered edge. Creative use of fuzzy logic [4] for unconstrained handwritten character recognition in which, they used box (subdivided binary image) approach by using vector distances between pixels for feature extraction. The recognition rate obtained is 87%.

Recognition of online handwritten Tamil characters with two different approaches using the subspace method [5] and elastic matching technique using dynamic programming principles has also been proposed. The first method evaluates effective recognition capability while the second one is important for studying the effectiveness of a scheme in real time applications. A unique method [6] visualizing Telugu characters as being composed of circular segments of different radii was used to extract feature vectors of individual basic characters. These were isolated characters and the recognition scores ranged from 78% to 95%. Another important method for handwritten character recognition is Principal Component Analysis (PCA) [7] followed by Euclidian Distance to measure the similarity between any two images.

1.6 Decision tree and See5 / C5 algorithm

A decision tree is a classification scheme which generates a tree and a set of rules, representing the model of different classes, from a given data set. The set of records available for developing classification methods is generally divided into two disjoint subsets a training set and a test set. The training set is used to derive the classifier, while the test set is used to measure the accuracy of the classifier. The accuracy of the classifier is determined by the percentage of the test examples that are classified.

The attributes of the records are categorized into two types. Attributes whose domain is numerical are called numerical attributes, and the attributes whose domain is not numerical are called the categorical attributes. There is one distinguished attribute called the class label. The goal of classification is to build a concise model that can be used to predict the class of the records whose class labels are not known.

The major strengths of the decision tree are:

a) Ability to generate understandable rules.
b) Ability to handle both numerical and categorical attributes.
c) Ability to provide a clear indication of which fields are most important for prediction or classification.
d) Ability to perform classification without requiring much computation.

The weaknesses are:

a) Decision trees are less appropriate for estimation tasks where the goal is to predict the value of a continuous attribute.
b) Decision trees are prone to errors in classification problems with many class and relatively small number of training examples.
c) The process of growing a decision tree is computationally expensive. At each node each candidate splitting field is examined before its best split can be found.

C5/See5 rules and its previous version C4.5 (Quinlan, 1993) are the extension of Quinlan.s famous Iterative Dichotomizer 3 (ID3) algorithm (Quinlan 1979).

The See5 / C5 algorithm automatically learns and applies mark-up rules. The advantages of C5 algorithm are that it is very fast, it is not sensitive to missing features, it can deal with large number of features and it is incremental.

There are two basic steps in the technique viz building the tree and applying the tree to the database. It has been successfully used in expert systems in capturing knowledge. It mainly uses inductive methods to the given values of attributes of an unknown object to determine appropriate classification according to decision tree rules.

The main component of the ID3 algorithm is the classification-learning algorithm. ID3 uses fixed set of examples to build a decision tree. The resulting tree is used to classify future samples. ID3 uses a statistical property called Gain which helps to select the attribute as a decision node. Gain gives the measure of how good a given attribute separates training examples into target classes.

For any handwritten character there are several pixels. For each of the pixels there are three attributes / features. Each set of attributes along with the class name forms a Record pertaining to a pixel of a character. If more than 50 % of the records are matched it is said that the test character is identified successfully.

Information gain is defined in terms of Entropy, which is used to measure how informative is the node of a decision tree. Entropy is a formula to calculate the homogeneity of a sample. If the entropy is zero “0” it is completely homogeneous sample. When the entropy is one “1” the sample is non homogeneous. For training set S entropy is defined as the measure of the average amount of information required in identifying the set S. Mathematically Entropy can be expressed as

$$Entropy(s) = \sum_{i=1}^{c} p_i \log_2 p_i \quad (1)$$

1.7 Approaches to overcome the limitations

Depending upon the similarity between the characters of Telugu, in the first stage, the most similar letters can be grouped using 2 dimensional correlation techniques. In the next step, SEE 5 algorithm is used to classify and differentiate the most similar letters using the third dimension information. The third dimension is the amount of depth at each pixel for any character. This amount of depth is directly proportional to the amount of pressure applied by the scriber at that pixel point of each character. A 3D method of character recognition is a new area of research, which is still not explored. Hence, in this research work, three dimensional data is used to identify and classify the Telugu characters scribed on the palm leaves using decision tree approach.

2. METHODOLOGY AND SYSTEM DESIGN

2.1 Image acquisition

Character recognition in general involves scanning a document and storing it in the computer system, which is used as an input image to the character recognition problem. However the input to the system is entirely different in the proposed work. In the proposed method, any basic Telugu character from the given palm leaf is chosen and pixels along the boundary of the character are identified. The (x,y) co-ordinates for each pixel is measured using digital measuroscope (XY co-ordinate measuring instrument). Dial Indicator Plunger Assembly, having an accuracy of 0.01 microns least count, is used to measure the depth information at each pixel, which varies from 10 microns to 150 microns for various pixels along the contour. After obtaining the (x, y, z) co-ordinates, by using the Microsoft excel, the character is obtained on to the computer to store and process further for identification of the characters. Minimum Euclidean distance concept is used on these patterns to compare with the available patterns in the database of the computer. The flow chart of the data acquisition process is shown in Figure-1 and can be described as follows:

Step 1: Identifying the pixel positions for the character selected ranging from 13 to 30 pixels based on the character. The left most pixel is chosen as the origin O (0, 0) coordinate and with reference to this point all the other pixel coordinates are measured using Measuroscope.
Step 2: The x and y coordinates are measured using Nikon Measurescope Model No. 20 with Nikon SC-102 as the display.
Step 3: Z coordinate is measured using a plunger type dial gauge indicator and SYLVAC 50 digital reader.

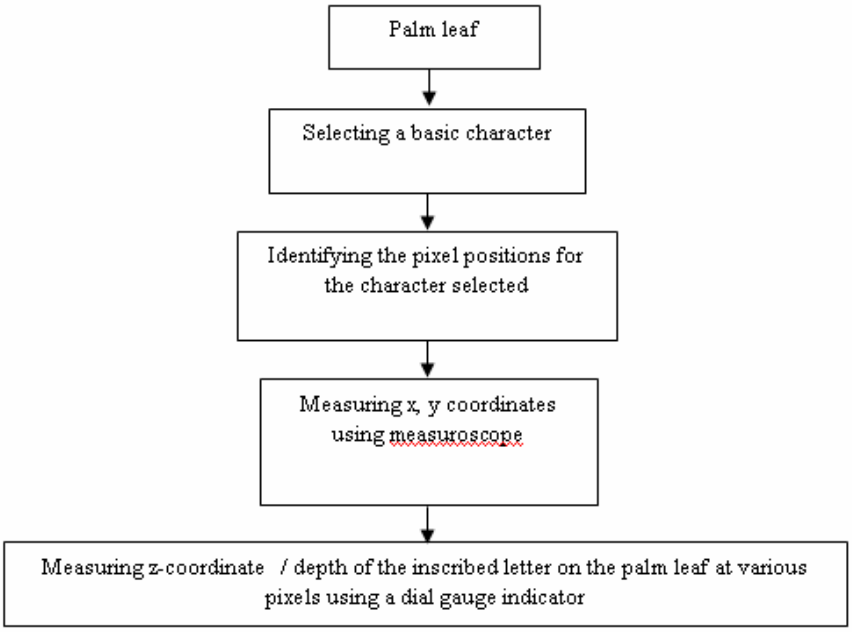


Figure-1. Procedure for image acquisition.

Palm leaf showing the holes of the Folio, which helps to store the leaves between the wooden boards is shown in Figure-2.

Figure-2. Palm leaf showing the holes of the folio, which helps to store the leaves between the wooden boards.

2.2 Correlation co-efficient for printed characters

Initially for reference set, the printed Telugu characters are considered for grouping, using two-dimensional correlation coefficient. The most similar letters are grouped together which had correlation coefficient of 0.75 and above.

The two-dimensional correlation coefficient between two matrices A and B can be obtained provided A and B are matrices or vectors of the same size.

$$r = \frac{\sum_m \sum_n \left(A_{mn} - \overline{A}\right)\left(B_{mn} - \overline{B}\right)}{\sqrt{\left(\sum_m \sum_n \left(A_{mn} - \overline{A}\right)^2\right)\left(\sum_m \sum_n \left(B_{mn} - \overline{B}\right)^2\right)}} \quad (2)$$

Where $\overline{A}$ = mean2 (A), and $\overline{B}$ = mean2

2.3 Extraction of characters from palm leaves

As discussed in the earlier sections of this paper, using image acquisition method, all the Telugu characters are extracted for the recognition purpose. The proposed method is experimentally evaluated with a data set extracted from palm leaves pertaining to Telugu script.

The pixel points for every character are selected according to the catastrophe theory, a branch of singularity theory in mathematics, which uses the concept of critical points (CP). The critical points are the simple minima (first derivative vanishes, second derivatives positive) and the simple maxima (first derivative vanishes, second derivative negative).

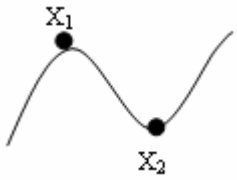


Figure-3. A smooth function with two critical points (CP): a simple maximum (X1) and simple minimum (X2).

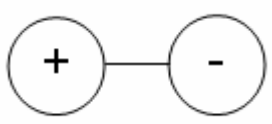


Figure-4. A graph representing the function in terms of its CPs.

A smooth function f(x) has a simple minimum (at x_1) and a maximum (at x_2) inside a domain D as shown in the

Figure-3. The positive is the maximum and the negative is the minimum as shown in Figure-4 which can be mathematically represented as follows:

$$\frac{df}{dx} = 0, \quad \frac{d^2f}{dx^2} < 0, \quad x=x_1 \tag{3}$$

$$\frac{df}{dx} = 0, \quad \frac{d^2f}{dx^2} > 0, \quad x=x_2 \tag{4}$$

After selecting the pixel points of a character, the left most pixel is chosen as the Origin O (0, 0) and with reference to this point all the other pixel coordinates are measured using measuroscope as discussed in the methodology. This is shown in Figure-5.

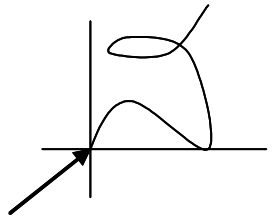


Figure-5. Selection of origin as reference among all the pixels.

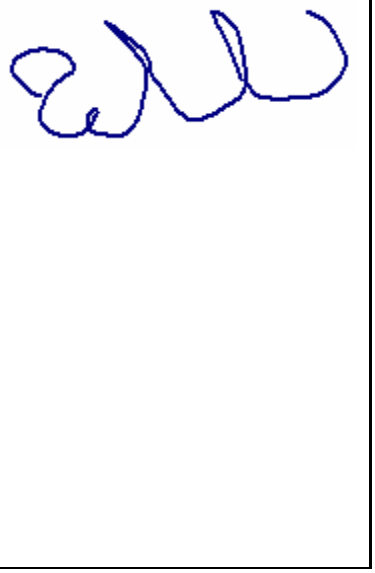
Using Microsoft Excel program images, for all the characters are generated by using the X and Y coordinates. For example consider two alphabets Ae and Ru characters shown in Tables 2 and 3.

Table-2. X and Y coordinate of Ae.

S. No.	x	y
1	1.44	1
2	1.48	1.24
3	1.08	1.5
4	0.65	1.49
5	0.48	1.29
6	0.55	0.993
7	0.9	0.926
8	1.42	0.91
9	0.81	0.91
10	1.4	0.85
11	1.87	0.84
12	1.93	1.19
13	1.85	1.65
14	1.43	2.16
15	0.75	2.62

Table-3. X and Y coordinate of Ru.

S. No.	x	y	S. No	x	y
1	0.81	1.08	21	2.68	1.03
2	0.39	1.19	22	2.6	1.51
3	0.39	1.47	23	2.09	2.11
4	0.63	1.69	24	2.45	1.78
5	1.12	1.85	25	2.88	1.37
6	1.64	1.77	26	3.34	0.91
7	1.8	1.57	27	3.89	0.74
8	1.63	1.34	28	4.47	0.8
9	1.26	1.27	29	4.73	0.97
10	0.9	1.18	30	4.66	1.32
11	0.78	0.95	31	4.05	1.85
12	1.01	0.6	32	4.64	1.42
13	1.38	0.55	33	5.06	1.14
14	1.73	0.56	34	5.63	0.99
15	1.79	0.68	35	6.45	0.93
16	1.86	0.84	36	6.94	0.98
17	1.83	0.64	37	7.09	1.17
18	2.06	0.51	38	7.01	1.49
19	2.41	0.57	39	6.25	1.86
20	2.67	0.64			



The above procedure is followed for obtaining the images of all the characters on to the computer. Also each character is obtained from multiple positions and the images are found to be highly reproducible. Using the minimum rectangle method, all the characters are made both scale invariant, rotational invariant. After this stage, these images are considered to be ready for use in various algorithms.

As can be appreciated, the data acquisition method used has many advantages compared to the scanned document. Problems such as skew and noise are eliminated thus avoiding major steps of pre-processing.

2.4 Implementation procedure of SEE5

a) The values of X, Y, Z (numerical values) are known as the attributes of the pixel.
b) The Record is obtained by separating each attribute by a comma and followed by the class label.
c) The test and the data set for each character is made by concatenating the records of each character.

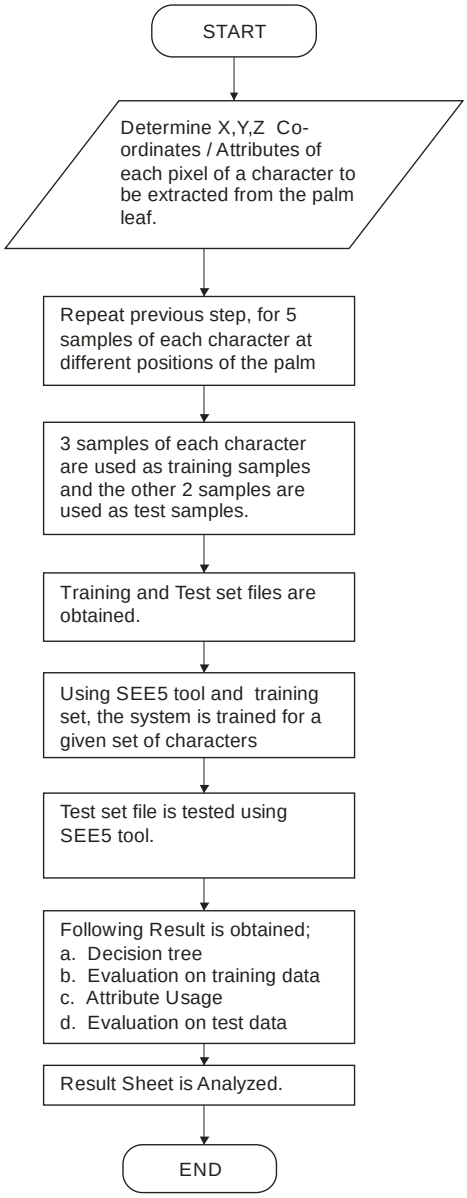


Figure-8. Flow chart showing general steps of application of SEE5.

Flow chart showing general steps of application is described by Figure-8. The most similar characters of Telugu are grouped into one group namely Group-1. These are (a, aa, ala, bra, la, ta, tha, khaa).For each pixel in a character, there are 3 attributes which are numerical and a Class name. Here the Co-ordinates of the pixel are X, Y where as Z is the attribute showing the depth of the indentation at that pixel which is proportional to the pressure applied by the Scriber at that point. For example if a pixel of a character **"**a**"** has 1.091, 0.159, 24 as the values for X,Y and Z, then the Record for that pixel is 1.091, 0.159, 24, a. The Training and test data set for each character is made by concatenating these records of each character.

3. SOFTWARE AND PROGRAMS USED IN THIS WORK

All the experiments are carried on a PC machine with P4 3GHz CPU and 512MB RAM memory under Matlab 7.0 platform. After the data acquisition, the (X, Y) data is used to plot the Telugu character and generate the same on the computer using MS excel. Further in the preprocessing stage. Photoshop software is used to normalize and the concept of minimum rectangle fit is used to get all the Telugu characters into the size of 50x50 pixels size.

4. RESULTS AND DISCUSSIONS

4.1 Establishing the reliability of the data acquisition method used

For establishing the reliability of the data acquisition method used in this work Printed Telugu characters are considered initially. By using the correlation concept, a correlation matrix between each of the Test Telugu character is generated [8] with every other Telugu character and tabulated the correlation coefficient in XY plane, using the method described in the recognition model earlier.

It is observed that the characters which are most similar had correlation coefficient of 0.75 and above. From the above generated matrix, all the characters are grouped successfully into six different sub groups as shown in the Figure-6.

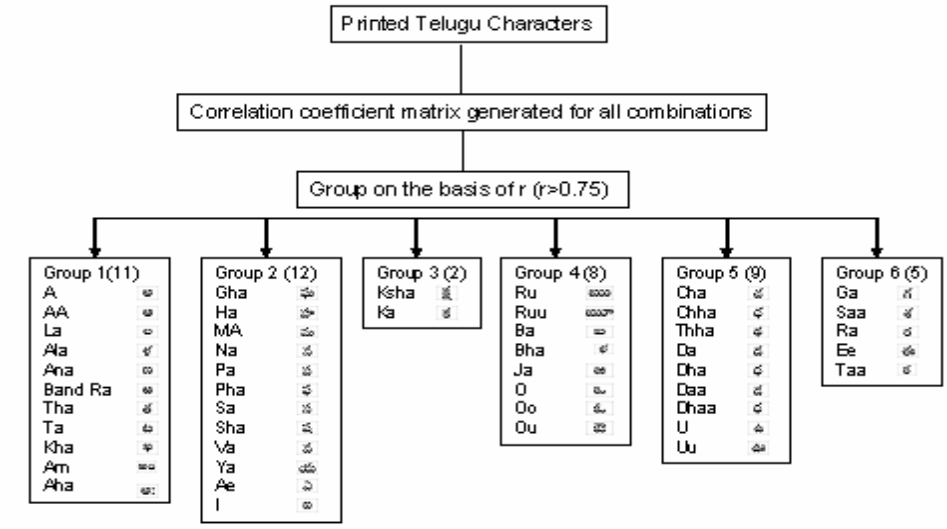


Figure-6. Sub-grouping of printed Telugu characters based on correlation coefficient.

4.2 Confusion of similar characters in XY plane

Considering the grouping results obtained for the printed characters, as discussed earlier in this section, it is very clear that there is lot of similarity and hence the confusion for the characters on XY plane within the same group. This is true even in the case of handwritten characters, and hence the result for character recognition obtained in this method is very low.

Three letters of a single group having a correlation coefficient of 0.75 and above showed lot of confusion in the normal XY plane, where as in the YZ plane, the patterns are completely different from each other. Images in Figure-7 are examples of the characters pertaining to a single subgroup with maximum confusion.

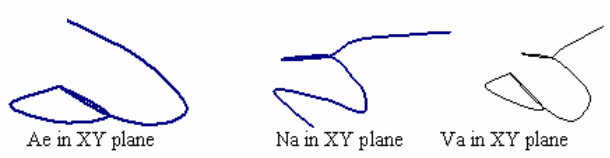


Figure-7. Characters in the XY plane.

5. RESULT SHEET

(As observed after running the SEE 5 Algorithm)

Group1 (a,aa,ala,bra,la,ta,tha, khaa)

Using the SEE5 tool after training the system with the training data, the test data is used to test the validity of classification. After running the SEE5 tool, the evaluation result obtained is as follows:

5.1 Evaluation on training data

```
(a)  (b)  (c)  (d)  (e)  (f)  (g)  (h) <-classified as
----  ----  ----  ----  ----  ----  ----  ----
 45                                       (a): class a
  1    46                  1             (b): class aa
            32    1                       (c): class ala
                  42                      (d): class bra
                        51                (e): class khaa
  1                     34    1           (f): class la
             2                  34        (g): class ta
                                 1    53   (h): class tha
```

5.2 Attribute usage

```
100% 3.z
 93% 2.y
 87% 1.x
```

5.3 Evaluation on test data

```
(a)  (b)  (c)  (d)  (e)  (f)  (g)  (h) <-classified as
----  ----  ----  ----  ----  ----  ----  ----
 14                                  1    (a): class a
      16                                  (b): class aa
  1          7    2    1                  (c): class ala
             1   13                       (d): class bra
                       17                 (e): class khaa
                            11    1       (f): class la
            1                1    8    2  (g): class ta
  1              1                1   15  (h): class tha
Time: 0.1 secs
```

Table-4 shows the number of records for each Telugu character and the percentage of attribute usage for classification.

Table-4. Summary of results from SEE 5 tool.

Groups/ character	Number of records			No. of nodes in decision tree	Attribute usage for classification		
	Tested	matched	% accuracy		X	Y	Z
Group I							
A	15	14	94	61	87	93	100
Aa	16	16	100				
ala	11	7	64				
bra	14	13	93				
Kha	17	17	100				
La	12	11	92				
Ta	12	8	67				
Tha	18	15	84				
Group II							
Ae	11	9	82	66	97	100	90
Bha	20	18	90				
Bhaa	19	16	85				
Na	10	4	**40**				
Pa	12	6	50				
Sa	12	7	59				
Shaa	14	12	86				
Va	13	5	**39**				
Ya	17	15	89				
Group III							
ka	20	20	100	8	100	71	28
ksha	18	18	100				
Group IV							
Cha	12	12	100	17	59	100	100
Da	14	12	86				
Dha	16	14	88				
Group V							
Ba	11	8	73	19	79	79	100
Bha	15	14	94				
Ja	20	17	85				
Ru	26	25	97				
Group VI							
Ee	20	20	100	9	84	65	100
Ga	12	12	100				
Saa	16	10	63				

The characters Va and Na have got less than 50% of records matched and hence they are the characters not identified by the system.

The above matrix is showing the number of Records of a Test character matched with the Training characters respectively. If more than 50 % of the records are matched, then we say that, the test character is identified successfully.

Here all the characters of the test character file are matched to the database / trained characters. The Attribute Usage is found to be 100 % for Z, whereas 93% for Y attribute and 87 % for X attribute. This tells that, Z i.e. the depth of indentation at any pixel, which is proportional to the pen pressure applied at that pixel point, is an important attribute for classification and identification.

For this classification and character recognition, it takes very small amount of time i.e. 0.1 seconds.

A sample set of 29 different Telugu characters is tested and 27 out of 29 of them were correctly recognized. Hence 27/29 = 93.10% of overall accuracy is obtained.

Figure-9 is a bar graph which gives the percentage of Accuracy group wise including the overall Accuracy.

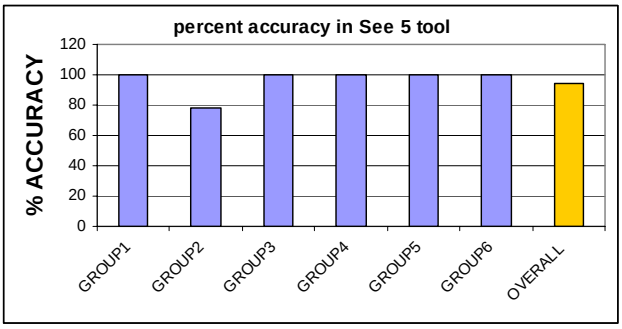


Figure-9. Group wise accuracy in see 5 tool.

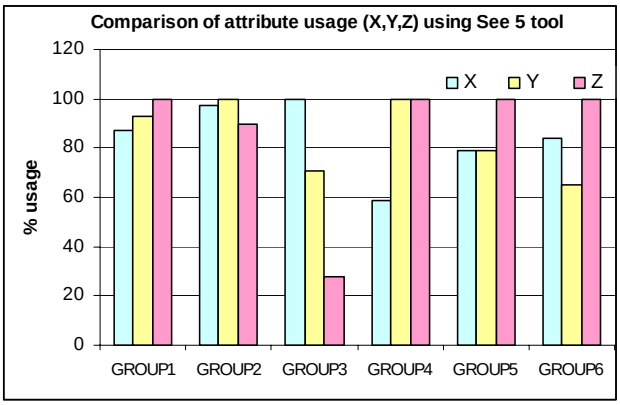


Figure-10. Percentage usage of X, Y and Z attributes for classification.

The graph shown in Figure-10 gives the attribute usage of X, Y and Z for Classification. Group3 shows less percentage of Z attribute usage, since the number of characters in this group is only two, and the X attribute is sufficient for recognition.

6. CONCLUSIONS

SPC (Scribed palm leaf character**)** recognition has not been reported in Literature. The Telugu Handwritten characters do not have high recognition rates for small sample sizes.

This paper describes a novel method to recognize and classify Telugu (a south Indian language) characters, written on the palm leaves. Hand written characters on the palm leaves were extracted by measuring X, Y and Z coordinates of pixels of each character reproducibly. The complexity of the characters in the language is reduced by dividing all the characters into 6 subgroups using printed characters and their correlation coefficient values**.** Z dimension is the depth at that particular pixel which is directly proportional to the pen / stylus pressure on the palm leaf.

Further, in the next step, considering the Co-ordinates of X, Y, Z of each pixel of Telugu handwritten characters (extracted from palm leaves) and SEE 5 algorithm, the system is trained to identify and classify another set of characters pertaining to these trained classes. Even the most similar Telugu Characters within the same subgroup are successfully recognized and classified.

The classification accuracy obtained is 100%. A tree structure along with the various nodes showing the various Telugu characters as leaves is obtained. The overall recognition accuracy obtained ranged from 93.10% using SEE5 algorithm.

Although some of the letters like Va, Ma, Pa and Ya, are very similar to each other, the proposed method is able to identify them distinctly by application of 3D features in the recognition process.

All the characters of the test character file are tested with the database / trained characters. The Attribute Usage is found to be 100 % for Z, whereas 93% for Y attribute and 87 % for X attribute. This tells that Z, i.e., the depth of indentation at any pixel is proportional to the pen pressure applied at that pixel point is an important attribute for classification and identification

For this classification and character recognition, it takes very small amount of time i.e. 0.1 Seconds.

This method is applied only to the basic Telugu characters and can be extended for samayukt akshars (combination of two or more basic characters) also in future studies. The method of data collection can be improved in future by automated process of measurement like using a laser technique instead of manual data collection mainly for the Z dimension (depth information).

ACKNOWLEDGEMENTS

The author whole heartedly acknowledges the co-operation extended by Sri S. Anand, Finance Officer, RSVP (Rashtriya Sanskrit Vidyapeeth), Tirupati in procuring the palm leaves from Oriental Research Institute, Tirupati, A.P, India. Further, the author expresses sincere gratitude to Dr. Vally Maya who has actively participated in the technical discussions and rendered appropriate suggestions at every stage in the work. It would not be out of place to express thanks to Sri A. V. Narasimha Rao, who assisted me in getting the manuscript into an appropriate shape.

REFERENCES

[1] Shi Zhixin, Setlur Srirangaraj and Govindaraju Venu. 2005. Digital Image Enhancement Using Normalization Techniques and their Application to Palm Leaf Manuscripts. CEDAR. Center for Excellence for Document Analysis and Recognition. New York. U.S.A

[2] Ashwin T. V and Sastry P.S. 2002 A Font and Size-independent OCR system For Printed Kannada documents Using Support Vector Machines. Sadhana. Vol. 27, Part 1. 35-58.

[3] Bunke H, Roth. M, Schukat-talamazzini. E.G. 1995. Off line Cursive Handwriting recognition using Hidden Markov Models. Pattern Recognition. Pergamon. pp. 1399-1413.

[4] Gader Paul D, Keller James M, Krishnapuram Raghu, Chiang Jung-Hsien and Mohamed Magdi. A. 1997. Neural Methods in Handwriting Recognition. Research Feature. IEEE. pp. 79-85.

[5] Joshi Niranjan, Sita G, Ramakrishnan A.G and Madhvanath Sriganesh. 2004 Tamil Handwriting recognition using subspace and DTW based Classifiers. Springer. Vol. 3316/2004. pp. 806-813.

[6] Rao P.V.S and Ajitha T.M 1995. Telugu Script Recognition-A feature Based Approach. ICDAR. IEEE. pp. 323-326

[7] Aradhya Manjunath. V.N, Hemantha Kumar. G and Noushat.S. 2007. Multilingual OCR system for South Indian Scripts and English Documents: An Approach based on Fourier Transform and Principle component Analysis. Engineering Applications on Artificial Intelligence. Elsevier.

[8] Panyam Narahari Sastry, Ramakrishnan Krishnan, Bhagavatula Venkata Sanker Ram. November 2008. Telugu Character Recognition on Palm Leaves-A three dimensional Approach. Technology Spectrum. Volume 2, No. 3, 19-26.