Ensemble Classifier System for Offline Ancient Tamil Character Recognition

Merline Magrina M , Dr. Santhi M

[1]*ME / Communication Systems , Department of ECE , Saranathan College of Engineering*
Trichy, Tamilnadu
[2]*Professor, Head of the Department , Department of ECE , Saranathan College of Engineerimg*
Trichy, Tamilnadu

Abstract

Recognizing ancient Tamil characters enable archaeologists to reveal historical events in Chola period in 12th century with huge efforts of the archeological experts. The future research in the field of archeology will have negative impact due to inefficiency of the manual procedure. Optical Character Recognition (OCR) functionality is used to recognize ancient Tamil Inscription in stones. OCR module of application i.e. Ensemble Learning is mainly focused in this paper. The collected inscriptions are preprocessed and segmented. During pre-processing the color images are converted to gray images and they are enhanced. The resultant images are applied for noise removal process using various filters. The output is applied for post processing steps to remove further noises includes opening, closing, erosion and dilation. Segmentation process is performed using bounding box and the features like orientation points & magnitude points are detected using SIFT algorithm. Characters are classified using Ensemble learning classifier and the classified characters are matched using Unicode. The exact Modern Tamil Character is mapped and their Recognition rate & Segmentation Rates are calculated and compared.

Keywords—— *Image Processing, Character Recognition, Character Segmentation, OCR, SIFT, Ensemble learning method, Unicode.*

I. INTRODUCTION

Pattern Recognition (PR) is the process of recognizing patterns by using machine learning (ML) algorithm. PR is the classification of data based on the knowledge statistical information extracted from the patterns and their representations. In typical pattern recognition application, the raw data is processed and converted into a machine editable format. Pattern recognition involves classification & cluster of patterns.

- In classification process, an appropriate class label is assigned to a pattern based on an abstraction that is generated using a set of training patterns or domain knowledge. Classification is used in supervised learning.
- Clustering generated a partition of data which helps in decision making, the specific decision making activity of interest. Clustering is used in an unsupervised learning.

Features are represented as continuous, discrete or discrete binary variables. A feature is a function of one or more measurements computed so that it quantifies some significant characteristics of object. Learning is a phenomenon through which a system gets trained and becomes adaptable to give result in an accurate manner. It shows how well the system performs on the dataset provided it depends upon the algorithm used. Entire dataset is divided into 2 categories i.e. one which is used in training model (Training set) & one which is used in testing model (Testing set).

- **Training set** is used to build a model for extracting features. It consists of the set of images which are used to train the system by the features. Generally 80% of the dataset is taken as the training data.
- **Testing set** is used to test a model for recognition. It is a set of data which is used to verify whether the system is producing the correct output after being trained or not. Generally 20% of the dataset is taken for testing. Testing dataset is used to find the accuracy of the system.

Optical Character Recognition (OCR) is machine encoded text that is converted from images of typed, handwritten or printed text by the electronic conversion. OCR is a research field in pattern recognition, artificial intelligence & computer vision. Advanced systems are capable of producing a high degree of recognition accuracy for most fonts of common and with a support of variety of digital image file formats.

Epigraphy is a primary tool for archaeology in dealing with literate cultures. Epigrapher is the person, whose responsibilities are modernizing, translating & dating of inscriptions. They are engraved on pillars, stones, caves & rocks. Later they were inscribed on palm leaves, coins and copper plates and on temple walls. Inscriptions may be in Brahmi or Tamil-Brahmi scripts (Fig 1) Royal inscriptions [3] were also engraved on copper plates called Indian Copper Plate Inscriptions. The edicts of Ashoka contain Brahmi script as they are widely found in cave walls & later evolved as vattezhuthu alphabet (Fig 2) In 6th century,

the Pallava dynasty created a new script for Tamil called Grantha alphabet (Fig 3) which was evolved from the vattezhuthu script.
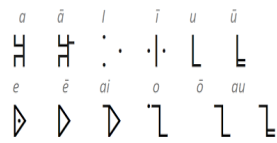

Fig 1: Tamil Brahmi script

Fig 2: Vattezhuthu script

Fig 3: Grantha script

The use of palm leaves (Fig 4) is a primary medium for writing that changes into scripts. The scribe had to be careful not the pierce the leaves with the stylus while writing because a leaf with a hole was more likely to tear & may decay. The use of the pulli to distinguish pure consonants became rare in stone inscriptions. The pure consonants are written as if the inherent vowels were present.

Fig 4: Palm script

BEAUTY OF 12ᵀᴴ CENTURY TAMIL CHARACTER[3]:

1. There is no pulli in the pure consonants present in the epigraphical stone inscriptions.
2. There is no separation between the words in the sentence present in the manuscript.
3. There is no difference between the vowel letters 'கி','கீ' in the stone inscriptions.
4. Both the letters 'மு' & 'மூ' and also 'ற' & 'ர' are written and identified as same in the stone inscriptions.

5. Yegara, Yegaara, Ogara, Oogaara, kuril, nedil differences were not followed in 12ᵗʰ century.
6. There is no full stop present at the end of the sentences present in the stone inscriptions.
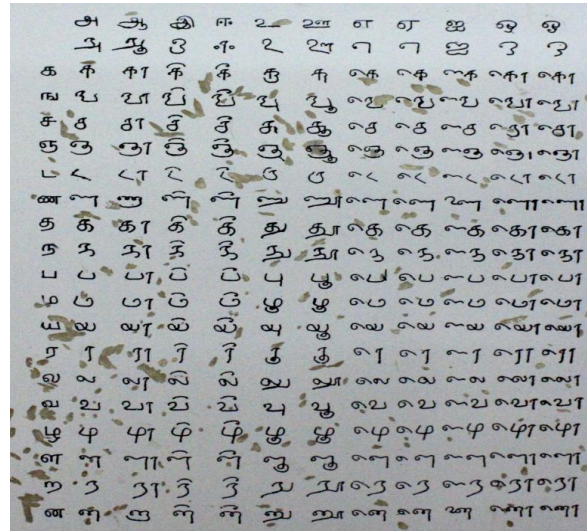7. If the same letter comes twice, the first character is considered as a pulli consonant.

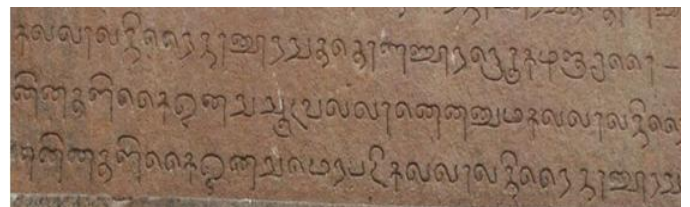**Fig 5: 12ᵗʰ century Tamil characters found in stone inscriptions [3]**

Fig 6: Sample stone inscription

Explanation of the sample inscription[3]:

கல்லால் நிரை நானூற்றுத் தொண்ணூற்றேழு கழஞ்சரை -

னின் தளிகை ஒன்று ஆடவல்லானென்னும் கல்லால் நிரை

னின் தளிகை ஒன்று மேற்படி கல்லால் நிரை நானூற்று .

Modern Tamil Script:

The forms of some of the letters were simplified in the 19ᵗʰ century to make the scripts easier to typeset. The earliest available text is Tholkaapiyam, a work describing the language of the classical period. There are several other famous works in Tamil like KambaRamayanam & Silapathigaram speaks the greatness of the Tamil Language. The two sentence poem "Thirukkural" explains the ideas to be followed in life of human. The modern Tamil has 12 vowels, 18 consonants (Fig 7) & they are combined to yield

216 characters and 1 special character counting a total of 247 characters.

Fig 7: Tamil vowels & consonants

1.1 Objectives of our work
- To recognize the characters using Offline Optical Character Recognition based on Ensemble Learning Classifier.
- To map the ancient Tamil characters of 12th century to modern Tamil characters using Unicode.
- To test the variety of inscriptions & find the performance characteristics under various noise environment.

II. LITERATURE SURVEY

OCR is implemented to identify the characters using image processing technique efficiently. To convert the inscription to machine editable format using Unicode values. This leads to more number of computations and but decreases the execution speed. The inscription images are evaluated and the performance & accuracy of the recognition is observed under different methodologies [6].

Tan Chiang Wei & A. Rahman [1] proposed the system to analyze the OCR performance for noisy English characters by Inception V3 network. The character recognition is complicated in nature for noisy images. To achieve the performance the testing samples & network is trained using Transfer learning method & the well trained characters are evaluated. The better recognition accuracy is achieved by DNN with the reduction error rate of 21.5% overall to existing OCR.

Recognition of ancient Tamil characters from the epigraphical inscriptions utilizes SVM classifier for classification and ANN with Back Propagation Model for recognition was proposed by K. DurgaDevi, Dr. UmaMaheswari [4] Technique is to capture the inscriptions directly from the stone using digital camera and subjected to OCR steps mainly for preserving data in digital format efficiently. SVM & ANN with BP model and KNN classification model is widely used for evaluating the

testing samples. The recognition rate of 66% accuracy is obtained.

SIFT (Scale Invariant Feature Transform) is an algorithm that extracts the feature data from an input image and comprises of characteristics that prevent image transformations such as image size and rotation in matching of feature points. This is followed by 4 steps namely pre-processing, feature point extraction process, matching descriptor process [5] by Hyun Bin Joo and Jae Wook Jeon.

V.Vaishali et al. [7] developed a system effectively for recognizing the Tamil inscriptions from the stone. Pre-processing, morphological operations & character matching is done using correlation matching and produce better mapping results. The main process deals with removing the noise from the inscription by pre-processing & using many filters for noise removal such as Weiner filter & Median filter. Weiner filters are used in adaptive filtering to remove GAUSSIAN noise. Median filters are best remove SALT & PEPPER noise that does not affect the smoothness & sharpness of the characters present in the stone inscriptions.

Rajasekar.M and Celine Kavitha proposed a method [8] to recognize the characters using ANN. The network consists of input layer, hidden layer & output layer. In pre-processing, the basic algorithm for segmentation of characters, normalizing the characters & De-skewing is applied. OCR is aimed at recognizing printed document. The input document is read pre-processed, feature extracted and recognized text is displayed in a picture. OCR eliminates the difficulty by making the data to be available in printed format. Then this is applied for different dataset for the testing process and the accuracy of the network is obtained.

Gunasekaran & Preethi [9] proposed a system of pre-processing the characters to remove noises completely present in it using filers. Segmenting the characters and features are extracted using transfer edge technique. To recognize those characters Artificial Neural Network is implemented. This approach can be improved by using different classifiers namely SVM, Fuzzy & Tree classifier.

Robi Polikar [10] proposed an ensemble system consists of 2 key components. (1) Need to build an ensemble that is diverse in nature. Some are bagging, boosting, AdaBoost, stacked generalization. (2) Need to combine the outputs of individual classifiers that the ensemble to make correct decisions which are amplified and incorrect ones are cancelled out. This helps to improve Machine Learning results by combining several models. Multi class classification algorithm called Ensemble learning [12] combines several machine learning (ML) techniques into a predictive model in order to reduce variance (bagging), bias (boosting) & improves prediction (stacking). The main idea is to improve the accuracy of supervised learning task. The challenges are how to construct ensemble & use

individual hypotheses of ensemble to produce classification. Increased accuracy is obtained using Bagging Technique for OCR as it uses weighted average voting method.

III. METHODOLOGY

The related work explains the way of digitizing Ancient Tamil inscriptions to Modern Tamil inscriptions. The proposed work concentrates on the recognition of 12th century Tamil characters from the stone inscriptions to 19th century Tamil characters for knowing the Chola period culture and history of their dynasty.

This proposed system for recognizing the ancient Tamil inscription characters uses OCR and Unicode Mapping. The architecture of this method is shown in the Fig 8. Inscription images of Chola period are collected from the TamilNadu Archaeological Department. Initially the collected dataset is classified into 80% for training samples & 20% for testing samples. The inscription images are pre-processed in order to remove noises. The pre-processed image is subjected to binarization process and segmented into characters. These images are applied for feature extraction & the features are extracted using Scale Invariant Feature Transform (SIFT) algorithm [4]. These features are used to classify the characters using Ensemble learning methods [10]. The identified character is mapped with Unicode character set and the characters are stored in the label set.
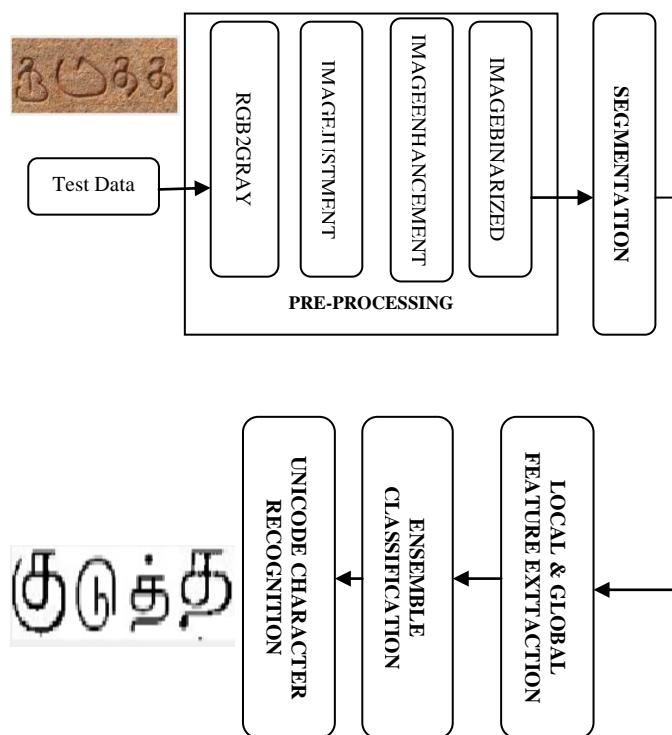

Fig 8: Block diagram of Optical Character Recognition

Pre-processing:

The input image is pre-processed to remove noises. The presence of noise reduces the efficiency of the character recognition systems. The image is resized process to a fixed size because all the collected images are of in different sizes. The resultant images are converted to gray scale images and they are enhanced. Then the image subjected to binarization process for the conversion to binary image. The binary image is applied to noise removal process by using median filter to remove "Salt and Pepper noise". Further the wiener filter is also applied to remove "Gaussian noises".

Steps involved in pre-processing:

1. **Rgb2gray conversion:** The function rgb2gray converts RGB images to grayscale images by eliminating the hue & saturation information while retaining the luminance. When we convert to gray scale image then we have stored only 8 bit values per pixel for the gray scale image value. These images will be mere 33% to the size of the original image.

2. **Image Enhancement**: The image enhancement is the process of digitally manipulating a stored image using software by changing various properties of an entire image. In Image enhancement technique **Image Adjustment** is that which adjusts the image intensity values. It maps the intensity values in gray image to new values such that 1% of data is saturated at high & low intensity values. This method increases the contrast of the output image.

3. **Binarization:** Binary image is of black and white. The colour used for the object in the image is the foreground colour in white while the rest of the image is the background colour in black. In Binarization technique **Imbinarize** [4] is that which creates a binary image from the enhanced image by replacing all values above a threshold with 1's and below a threshold with 0's.

4. **Noise removal:** Noise can be removed from the image using the Median filter [6] of size 5*5. In median filter the pixel values are replaced by median pixel values of neighbouring values. The median pixel value is calculated by first sorting all the pixel values from neighbourhood into numerical order & afterward supplanting the pixel being considered with the centre pixel value. In Noise removal technique **Median filter** is a non-linear digital filtering technique often used to remove noises from the image to improve the results of further processing steps.

Post-Processing:

To remove further noises present in the image, the morphological operations such as

Closing, Opening, Erosion and Dilation are used as a post processing steps.

1. **Closing:** It performs morphological closing on the gray scale image with the structural element 'disk' that returning opening image. The structural element (SE) must be a single element returned by structural (strl) function. The morphological close operation is erosion followed by a dilation using the same structuring element for both operations.

2. **Opening:** It performs morphological opening on the gray scale image with the structural element 'disk' that returning opening image. The structural element (SE) must be a single element returned by structural (strl) function.

3. **Erosion:** It performs morphological erosion on the gray scale image with the structural element 'disk' that returning opening image. The structural element (SE) must be a single element returned by structural (strl) function.

4. **Dilation:** It performs morphological dilation on the gray scale image with the structural element 'disk' that returning opening image. The structural element (SE) must be a single element returned by structural (strl) function.

Segmentation:

Image segmentation is a challenging task in image processing areas. Segmentation is a process of distinguishing lines, words & characters of an inscription image or printed document. It is a crucial step that extracts the meaningful regions for analysis. As the input image is a word, the segmentation process is **Character Segmentation**, the most crucial step for any OCR system. The key factor is deciding the accuracy of the system if there is a good segmentation of characters, the recognition accuracy will be so high.

$$Segmentation\ Rate = \frac{no\ of\ correctly\ segmented\ characters}{total\ no\ of\ characters} * 100$$

Feature Extraction:

Feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative & supervised learning in machine learning, pattern recognition & image processing,. Feature extraction is purely related to dimensionality reduction. It is important in the area of optical character recognition.

1. **Edge Detection technique**: **Canny Edge Detection** [2] operator uses a multi stage algorithm to detect a wide range of edges in images. It is a technique to extract useful structural information from different objects & dramatically reduce the amount of data to be processed. General criteria: 1. Detect the edges with low error rate. 2. Accurately localize the centre of the edge. 3. Noise in image should not create any false edges.

2. **Corner Detection Technique**: **Harris Corner Detector** [2] is a corner detector that is commonly used in computer vision algorithm that extracts corners and interfeatures of an image. It takes the differential of the corner score into account with the reference to directions directly instead of shifting patches for every 45 degree angles & more accurate in distinguishing edge & corner detections. The best corner points of 30 is chosen for the feature extraction.

3. **Region Properties:** Measure the properties of regions present in an image. The 9 region properties calculated for the features are

1) **Area:** It specifies the actual number of pixels in the region.

2) **Major Axis Length:** It specifies the length of the major axis of the ellipse that has the same normalized second central moment as the region.

3) **Minor Axis Length:** It specifies the length of the minor axis of the ellipse that has the same normalized second central moment as the region.

4) **Eccentricity:** It specifies the eccentricity of the ellipse that has the same second moment as the region. Eccentricity is the ratio of the distance between the foci of the ellipse & its major axis length. The eccentricity value is to be between 0 and 1.

5) **Orientation:** It specifies the angle between the x-axis and major axis of the ellipse that has the same second moment as the region. The value is between -90 to +90 degrees.

6) **Equivdiameter:** It specifies the diameter of a circle with the same area as the region. **Equivdiameter = 4*Area/pi**

7) **Solidity:** It specifies the proportion of pixels in the convex hull that are also in the region. **Solidity = Area/ConvexArea**

8) **Extent:** It specifies the ratio of pixelsin the region to pixels in the bounding box. **Extent = Area/ Area of the bounding box**

9) **Perimeter:** It specifies the distance around the boundary of the region.

Scale Invariant Feature Transform (SIFT) Algorithm

SIFT [5] [12] is the feature extraction algorithm to detect and describe local features in images. The key-points are first extracted from the set of reference image and stored in a database. A testing inscription sample is recognized by individually comparing each feature from the testing image to the built database and finding the candidate matching features. For any object in an image, interesting points on the image can be extracted to provide a "feature description" of

the object. The features that are extracted from the training image which can then be used to identify the object. Another important characteristic of these features is that the relative positions between them in the original scene should not change from one image to another. SIFT extracts large number of features from the images that reduces the errors caused by variations in the average error of all feature matching errors. The features that are extracted using SIFT is 40, i.e. 20 for orientation points & 20 for descriptor points. For an image sample the gradient magnitude & orientation are computed using pixel differences,

Magnitude…

$$m(x,y) = \sqrt{\left(L(x+1,y) - L(x-1,y)\right)^2 + \left(L(x,y+1) - L(x,y+1)\right)^2}$$

Orientation…

$$\theta(x,y) = tan^{-1}\left[\frac{L(x,y+1) - L(x,y-1)}{L(x+1,y) - L(x-1,y)}\right]$$

All together the total features extracted for each character is
79 i.e *9 - region properties, 30 - corner points, 20 – orientation points & 20 – descriptor points.*
OCR is successfully created using these features of the Training Samples.
Classification:

        Classification is a process of finding small parts of an image that match the template image. Template matching compares the picture element values within the picture element values within the sub-region of the picture. If the match is good then it would mention the particular objects that are present in the image. So template matching is the basic step in OCR technique to determine the similarities between the template and region with the same size in an image that show similarity. After feature extraction classification is undertaken by comparing the input character with the set of templates from each character class. After all templates have been compared with the observed character image, the character identity is assigned.

        Classification Method: **Ensemble Learning Method** [10] helps to improve machine learning (ML) results by combining several models. This model produce better predictive performance compared to a single model. They are multi class classifier algorithm that combines several machine learning techniques into one predictive model in order to decrease variance (**BAGGING**), Bias (**BOOSTING**) or improve predictions (**STACKING**).
Ensemble methods are classified into 2 groups
- SEQUENTIAL
- PARALLEL

Sequential: Ensemble methods where the base learners are generated sequentially (ADABOOST). The main idea of sequential methods is to exploit the

dependence between the base learners. The classification performance can be boosted by weighing previously mislabeled examples with higher weight.
Parallel: Ensemble methods where the base learners are generated in parallel (RANDOM FOREST). This is to exploit theindependence between the base learners since the error can be reduced dramatically by averaging.
- ✓ **Bagging**: Bagging stands for bootstrap aggregation. The variance of an estimate can be reduced by averaging multiple estimates. It uses voting for classification & averaging for regression. Computed the ensemble as

$$f(x) = \frac{1}{M} \sum_{m=1}^{M} f_m(x)$$

Accuracy for decision tree = 0.63 & KNN = 0.70
        In random forest, each tree in the ensemble is built from a sample drawn with replacement from the training set. A random subset of features is selected instead of using all the features thereby randomizing the tree. The bias value of the forest increases slightly, but due to averaging of less correlated trees, its variance decreases, resulting in an overall better model.
Algorithm
1. Let the number of training points be N and the number of features in the training data be D.
2. Choose L to be the number of individual models in the ensemble.
3. For each individual model I, choose d_I be the number of input variables I. It is common to have a single value of d_I for all the individual values.
4. For each individual model I create a training set by choosing d_I features from D with replacement and train the model.
- ✓ **Boosting**: **Boosting** is a machine learning ensemble meta-algorithm for primarily reducing bias, and also variance in supervised learning, and a family of machine learning algorithms that convert weak learners to strong ones. Predictions are combined through a weighted majority voting (classification) & weighted sum (regression) to produce final results.
Algorithm
1. Each classifier is trained using a random subset of overall training set.
2. After training at any level, Adaboost assigns weight to each training item.
3. Mis-categorized label is assigned higher weight so that it appears in the training subset of next classifier with higher probability.
4. Also weight is assigned based on the accuracy.

Recognition:

Pattern recognition is a branch of machine learning (ML) that emphasizes the recognition of data patterns in a given scenario. Machine learning can be either "supervised" where previously known patterns can be found in a given data or "unsupervised" where entirely new patterns are discovered.

Recognition Method: The **Unicode standard** reflects the basic principle which emphasizes that each character code has a numerical value. Unicode text is simple to parse and process and well defined. After classification the characters are recognized and a mapping table is created in which the Unicode's for the corresponding characters are mapped. Pattern matching algorithm is used to match the characters.

Ancient Tamil Character	Modern Tamil Character

Fig 9. Sample modern characters mapping

Character Recognition: The testing sample inscription is passed through various blocks of functions and finally compared with the recognition details from the mapping table. The corresponding Unicode's are accessed and printed using standard Unicode fonts so that the Character Recognition is achieved.

$$Recognition\ rate = \frac{no\ of\ correctly\ classified\ characters}{total\ no\ of\ characters} * 100$$

PERFORMANCE METRICS

✓ **Accuracy** is a measure of number of samples that are correctly classified.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

where
TP= number of characters which are correctly assigned to the given category
TN= number of characters which are correctly assigned not to belong to category
FP= number of characters which are incorrectly assigned to the category
FN= number of characters which are incorrectly not assigned to the category

IV. EXPERIMENTAL RESULTS

The original inscription image is collected from the dataset. It can be any document of different dimensions. This captured image (Fig 10) of dimension 401*147 in jpeg format is fed to pre-processing section.
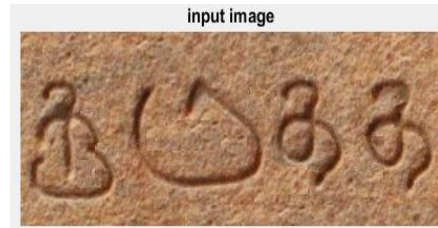

Fig 10: Input image

The binary image is subjected to noise removal using median filter where the pixel values are replaced by median pixel value of neighbouring pixel values as in Fig 11

Fig 11: Noise removed image

The next basic morphological operation is Dilation which adds pixels to boundaries of an image. The number of pixels in an image purely depends on the size & shape of the structuring element. To eliminate further interference and make the edge smooth Dilation operation is done as in Fig12.

Fig 12. Post procesed image

It is a process of distinguishing characters of an inscription image and extracts meaningful regions for analysis. Rectangular bounding box is used to bind each character into smallest possible rectangle as in Fig13.
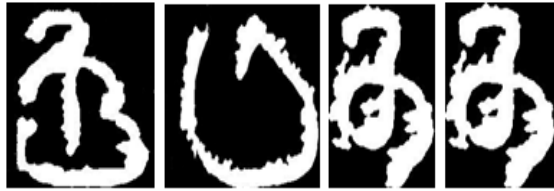
Fig13. Segmented Character images

The value of the points higher than the threshold is called as corner. Likewise the 30 corner points are detected as features in Fig.14.
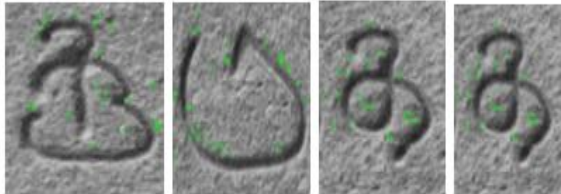

Fig14. Corner points detection

The exact characters are recognized to map the characters with corresponding Unicode values [53, 45, 40, 3]. The recognized modern Tamil characters are shown in Fig 15.
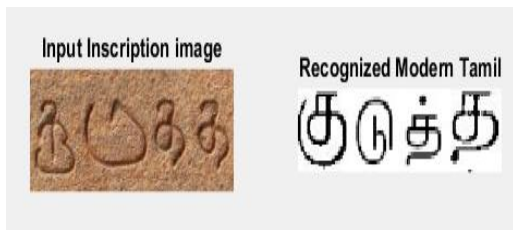

Fig15. Recognized Modern Tamil Character

Table 1: Performance of Ensemble methodology

Methodology & Type		Segmentation rate	Recognition rate
Bagging	**KNN**	*97.11%*	*96.52%*
	Discriminant	*97.68%*	*96.52%*
Boosting	AdaboostM2	8.5%	9.2%
	RUSBoost	8.9%	7.23%
	LPBoost	7.9%	7.7%
	TotalBoost	11.56%	10%

Table 2: Performance of noisy images under bagging method

Method	Type	With noise filtering		Adding noises & without noise filtering	
		SR	RR	SR	RR
Bagging	KNN	97.11%	96.52%	85.6%	43.25%
	Discriminant	97.68%	96.52%	85.6%	40.8%

Table 3: Performance noisy images with corresponding filters

Method	Type	Salt & Pepper noises& with median filter		Gaussian noise & with Weiner filter	
		SR	RR	SR	RR
Bagging	KNN	97.86%	96.52%	93.1%	92.5%
	Discriminant	96.8%	98.32%	92.8%	91.2%

V. CONCLUSION

Thus the proposed work focuses on the recognition of ancient Tamil characters from epigraphical inscriptions. The traditional methods of inscription identification are time consuming & laborious & imprints quality is poor to digitize. Hence an innovative technique is developed to identify the exact modern Tamil characters from ancient Tamil characters. For this the images are subjected to pre-processing by median filters, segmentation through bounding box, features are extracted by SIFT algorithm. The extracted features are applied for classification using Ensemble learning classifier. The modern Tamil character is mapped using Unicode. The performance metric calculated for segmentation & recognition rates of 97.11% and 96.52% are obtained respectively using Bagging methodology & it is found to be the best among the ensemble learning methods.

REFERENCES

[1] Tan Chaing Wei, Sheikh U.U and Ab-Al Hadi Rahman, "Improved OCR with Deep Neural Network", IEEE 14th International Colloquium on Signal Processing and its application (ICSPA 2018), Mar 2018, pp: 245-259.

[2] Beihai Tan, Chao Hu and Zepei Zhang, "Character Recognition based on Corner Detection", IEEE International Conference on Natural computation, Fuzzy System and Knowledge Discovery (ICNC-FSKD), Mar 2017, pg: 503-507.

[3] Bhavani.B (2017), "Tamilnadu Historical Documents (Epigraphical Inscriptions)".

[4] DurgaDevi. K and Uma Maheswari. P, "Insight on Character Recognition for calligraphy digitization", IEEE International Innovations in ICT for Agriculture and Rural Development (TIAR), Jun 2017, pp: 78-89.

[5] Hyun Bin Joo and Jae Wook Joen , "Feature point extraction based on improved SIFT algorithm", International Conference on Control Automation and System (ICCAS), Oct 2017, pp: 345-350.

[6] Karunarathne K.G.N.D, Liyanage K.V, "Recognizing Ancient Sinhala Inscription Characters using Neural Network Technologies", International Journal of Scientific

Engineering & Applied Sciences (IJSEAS), vol:3, issue:1, Jan 2017, pp: 37-49.

[7] Janani. G, Vaishali. V and Mohan, "Recognition and Analysis of Tamil Inscriptions & Mapping using Image Processing Technique", IEEE International Conference on Science and Technology Engineering & Management, Apr 2016, pp: 181-184.

[8] Rajasekar. M, Celine Kavitha and Anto Bennet.M , "Performance and analysis of Handwritten Tamil Character Recognition using Artificial Neural Network", International Journal of Recent Scientific Research (IJRST), vol: 7, Jan 2016, pp: 8611-8615.

[9] Preethi. N and Gunasekaran. T , "Language Specific Offline character recognition using Neural Network Classifier", International Journal of Advanced Research in

Electronics & Communication Engineering (IJARECE), vol: 4, Feb 2015,   pp: 218-221.

[10] Anitha Mary, Chacko and Dhanya, "Mutiple Classifier System for Offline Malayalam Character Recognition", International Conference on Information & Communication Technologies, vol: 46, Jan 2014, pp: 86-92.

[11] Robi Polikar, "Ensemble based systems in decision making", IEEE Circuits & Systems Magazine, vol:6, Sep 2010, pp:21-45.

[12] Warren Cheung and Ghassan Hamarneh, "n-SIFT: n dimensional scale invariant feature extraction", IEEE Transactions on Image Processing, vol: 18, Sep 2009, pp: 720-723.

.