# Characterization and Recognition of Stone Inscription

Puneeth P[1], Shashwathi Nayaka M P[2], Sridevi M N[3,] Varsha S[4]

*Maharaja Institute of Technology Mysore*

*(E-mail: puneeth0789@gmail.com, nayaka.shashwathi17@gmail.com , sridevimn11298@gmail.com, varshapintoo2013@gmail.com )*

*Abstract*—This electronic document is a "live" template and already defines the components of your paper [title, text, heads, etc.] in its style sheet. *CRITICAL: Do Not Use Symbols, Special Characters, or Math in Paper Title or Abstract. (Abstract)*

*Keywords—component; formatting; style; styling; insert (key words)*

## I. INTRODUCTION

Kannada language, one of the 5 ancient Dravidian languages has a history of more than 2000 years. In this period Karnataka was mainly ruled by many dynasties starting from Kadambas, Rashtrakutas, Gangas, Chalukyas, Hoysala and Vijayanagara Empire. Rich heritage of the empires have been carried over generation through the manuscripts and historic writings. The first written record in Kannada can be traced back in the time of Ashoka"s Brahmagiri edict (3rd century B.C) whereas Halmidi inscription(4th century A.D) of Kadamba"s is treated as the first stone inscription in Kannada. These inscriptions generally found on stones, temple walls, coins, palm leaves, pillars and copper plates. Analysis of these inscriptions is very important to through light and to understand the history, culture, socio-economic status, administration and literature of that period. Epigraphy is the art of identifying the inscriptions on rocks, plates, palm leaves etc. And Paleography is the study of ancient inscriptions and the practice of deciphering and reading of historical manuscripts. Paleographer must have the knowledge of the language, corresponding text of that period and various styles of handwriting and customs which where in use. Paleography is the study of ancient handwriting and the practice of deciphering and reading historical manuscripts. The paleographer must have the knowledge of the language of the text and the historical usages of various styles of handwriting, common writing customs, and scribal abbreviations. Prediction of the period of a given ancient script is a follow-on member of an ancient script recognition system and can be used as a component of the OCR system for ancient scripts. This knowledge can be used by archaeologists, historians and paleographers for further explorations. Any language can be written in any script. Presence or absence of a script for any language is not an impendent or a hurdle for the language. There were three different varieties of ancient scripts from which the current 20+ different scripts of present day in India have been evolved. They were Indus valley script, Brahmi and Kharoshti scripts. Scripts of all modern languages have been developed from one of the scripts over centuries.

Kannada script has been used to write in Kannada language. Kannada which belongs to the Dravidian family is the official language of the state Karnataka and one of the most ancient languages of India with a long historical heritage. Their earliest records date back to around second century BCE and Kannada script has developed gradually from the ancient script of Brahmi while it has undergone several modifications, twins and turns. Linguists have marked the evolution of Kannada script in three different phases. First, the „Halegannada" started with the Halmidi inscription of 4th century in the Kadamba"s era covering the time period of almost 8 hundred years. Though first written records of Kannada was found in the era of Maurya"s which dated back to 3rd century B.C is generally treated as Pre-ancient Kannada. „Halegannada" gradually developed into „Nadugannada" with different changes and evolutions in script under the rule of many dynasties from Gangas to Rashtrakutas to Chalukyas. „Nadugannada" in the era of Hoysalas, Vijayanagara and Mysore dynasties over five centuries has undergone many changes so far which is now categorised as „Hosagannada".

## II. LITERATURE SURVEY

Preserving the heritage of ancient civilizations furthers the education and understanding of world cultures that may help to promote world peace. Without digital preservation of important cultural sites, future generations may never know the achievements of past civilizations, our heritage. As pollution, wars, and natural disasters take their toll on these ancient structures, it is increasingly important to accomplish preservation of those historical monuments for cultural heritage of a nation.

Optical Character Recognitions with high accuracy are reported in literature for non- Indian languages like English, which have very minimal character set and also are less complex structurally. Character Recognition of Indian scripts is relatively complex because of large character set and the presence of compound characters. Not much work is carried out on recognition of Indian languages. Though there are few conventional OCR systems available for present Kannada script, almost nil work has been reported in the field of recognition of ancient scripts, which is much more complex as the characters have transformed to present form over years.

S.Rajkumar[1] has worked on recognition of eighth century Tamil consonants from stone inscriptions using SVM classifiers. He has proposed a neural network approach for the recognition of Tamil characters. Based on global texture analysis character segmentation and recognition processes are carried out. Same author[2] has published the work on 8th century character identification of Tamil characters based on clustering methods. Contour let-based method on offline text-independent ancient Tamil handwriting identification is carried out here. Rajitkumar[3] has worked on the template matching method for recognition of stone inscripted characters based on correlation analysis. H S Mohana[4] has proposed a technique for era identification of Ganga and Hoysala phase Kannada stone inscriptions characters using advanced recognition algorithm. This method is normally implemented by first picking template and then it call the search image, then by simply comparing the template over each point in the search image and it calculate the sum of products between the coefficient. Based on this calculated product value it recognizes the character. In this present paper, we extended the work for various Kannada dynasties of different centuries. Experimental results have shown good accuracy in recognizing the stone inscription characters of Kadamba, Hoysala, Chalukya and Vijayanagara time frames with better results obtained.

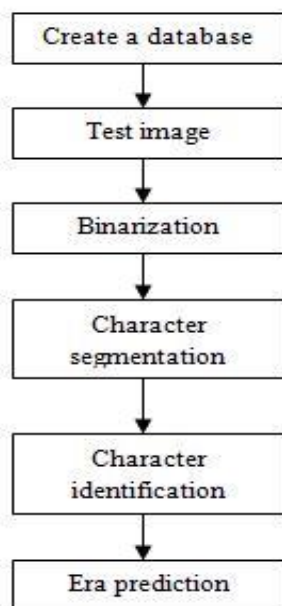### III.   PROPOSED METHODOLOGY



Fig 1. Proposed algorithm

Proposed method consists of the following steps:

A.   Create a database of Kadamba, Chalukya, Hoysala and Vijayanagara phases alphabets and corresponding current Kannada alphabets in jpg or *bmp* format. We have considered these four periods as transition of Kannada characters from one form to another form can be clearly

distinguishable in these. Sample test image is shown in fig.2

B.   Test image: Capture the image of a stone inscription using a digital camera where characters are visible clearly. If the characters in the inscriptions are too small, crop a particular part of the image and consider it as a test image file.

C.   Binarization: Binarization is the process of converting a gray scale image into binary image by selecting a global threshold that separates background from foreground. Separating foreground object from the background object is called thresholding. Typically two peaks comprise the histogram gray scale values of a document image, a high peak analogous to the white background and a small peak corresponding to the background. Image after binarization is shown is fig 3.

### IV.   EXPECTED RESULT

A.   Character segmentation: Connected component method is used for line and character segmentation. When a black pixel is encountered, it identifies the complete character through connected components. This character is segmented and centroid is computed. Similarly all the characters are segmented and their centroids are computed. Computed values can be used to find Euclidian distance between the centroids. Segmented characters using connected components method is shown in fig 4.

V. Character identification:

(i)   First, we have created the database of various stone inscriptions of different era like Kadamba, Chalukya, Hoysala and Vijayanagara alphabets images in jpeg/bmp format. Then resize all the images into a fixed format so that calculation of the characteristics if these images becomes easy.

(ii)   Secondly, to match the characters of the captured image with the characters of the database image, we need to calculate the mean and variance values. Mean value returns the arithmetic mean values of the elements along different dimensions of an array. For a data set, the terms arithmetic mean, mathematical expectation, and sometimes average are used synonymously to refer to a central value of a discrete set of numbers i.e. sum of total values divided by the number of values of an M*N matrix.

Variance value is calculated of each row or column of the input image. The variance of an M-by-N matrix is the square of the standard deviation which is given by:

$$y = \sigma^2 = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}|u_{ij}|^2 - \frac{\left|\sum_{i=1}^{M}\sum_{j=1}^{N}u_{ij}\right|^2}{M*N}}{M*N-1}$$

(iii)    Thirdly, Sum of absolute differences algorithm measures the similarity between image blocks. It takes the absolute difference between each pixel in the original block and the corresponding pixel in the block being used for comparison. Mean, variance, absolute differences algorithms are used to match the character of inscription image with the images of the database. Characters from the database which matches to the segmented characters depending on the similarity of mean, variance and absolute difference values is shown in fig 5.
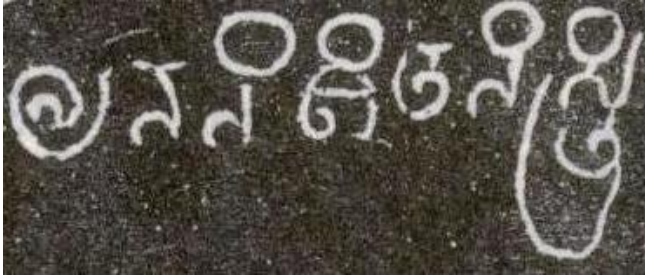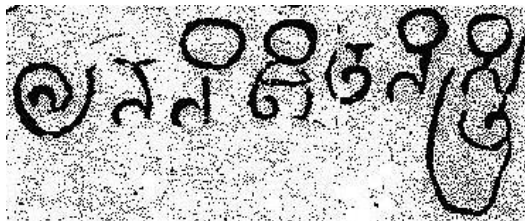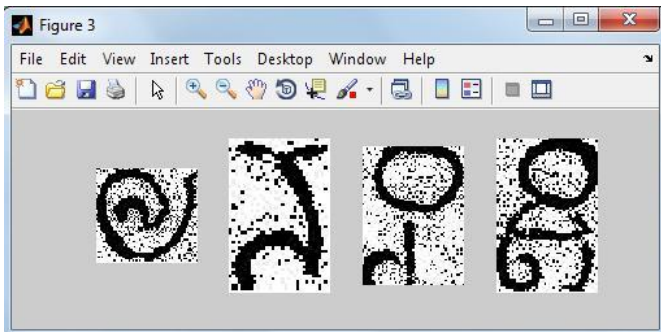


Fig 2. Test image



Fig 3. Dilated image
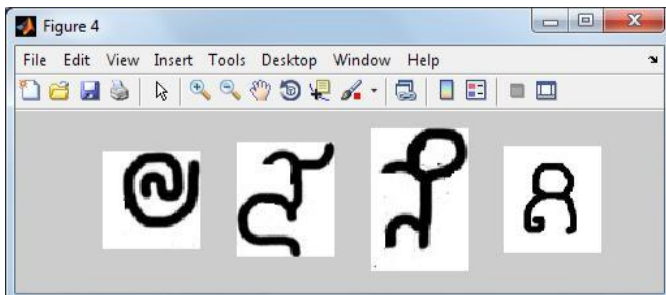


Fig 4. Character segmentation

Fig 5. Matched characters from database
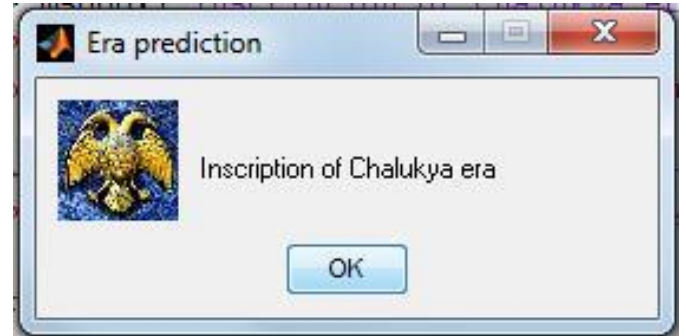


Fig 6. Era prediction

## VI.  CONCLUSION

An easy and efficient recognition system for identification of different time periods of Kannada stone inscriptions characters were introduced in this paper. This provides the good platform for identifying the stone inscriptions and their era. In this work, a simple digital camera is being used to capture and digitize the epigraphs. Captured image is taken as the test image. A database of Kannada characters of different time periods have been generated in jpg format and resized to a same shape. Test image is cropped and binarized. Each character is segmented separately using connected component method. Mean, variance and standard deviation values of each character is calculated for both segmented test image as well as the character images in the database. Similarity between the image blocks of test image and database is measured using absolute differences algorithm. Matching characters of segmented test image will be replaced with the characters of database to predict the era of characters. Experiments were performed using MATLAB R2013a software tool and tested with chosen characters with datasets of four different time periods. Result achieved is efficient and accuracy is around 80%.

### REFERENCES

[1] Read and Recognition of old Kannada Stone Inscriptions Characters using Novel Algorithm Rajithkumar B K, Dr. H.S. Mohana, Uday J, Bhavana M B, Anusha L S.IEEE-2015

[2] Information Extraction and Text Mining of Ancient Vattezhuthu Characters in Historical Documents Using Image ZoningE.K.Vellingiriraj,Dr.M.Balamurugan,Dr.P.Balasubraman ie.IEEE-2016.

[3] Grantha Script Recognition from Ancient Palm Leaves Using Histogram of Orientation Shape Context.Amrutha Raj V, Jyothi R L, Anilkumar A.IEEE-2017.

[4] A Framework for Recognition of Handwritten South Dravidian Tulu Script.Dr. Antony P J, Savitha C K.IEEE-2016.

[5] Tamil Character Recognition from Ancient Epigraphical Inscription using OCR and NLP.Manigandan T, Dr. V.Vidhya, Dr.Dhanalakshmi V, Nirmala B.IEEE-2017.

[6] Era Identification and Recognition of Ganga and Hoysala Phase Kannada Stone Inscriptions Characters using Advance Recognition Algorithm.Dr. H S Mohana, Mr. Rajithkumar B K.IEEE 2014.

[7] An Efficient Positional Algorithm for Recognition of Ancient Stone Inscription Characters. Mrs.G.Bhuvaneswari, Dr.V.Subbiah Bharathi. IEEE 2015

[8] Line and Word Segmentation of Kannnada Handwritten text documents using Projection Profile Technique.Banumathi K L, Jagadeesh Chandra A P.IEEE 2016.

[9] Handwritten Character segmentation for kannada scripts.C naveena, V N Manjunath aradhya.IEEE 2012.

[10] Recognition and Analysis of Tamil Inscription And Mapping using Image Processing Techniques.G. Janani, V. Vishalini, Dr .P. Mohan Kumar.IEEE 2016.

[11] Adapting moments for Handwritten Kannada Kagunitha Recognition.Leena Ragha, M sasikumar.IEEE 2010.

[12] Handwritten Kannada Character Recognition using Wavelet Transform and Structural Features.Saleem Pasha, M.C.Padma.IEEE 2015.

[13] Feature Extraction Of Handwritten Kannada Characters Using Curvlets and Principal Component Analysis. M.C.Padma, Saleem Pasha.IEEE 2015.

[14] SVM classifier for the prediction of ERA of an epigraphical script.Soumya A, G Hemanth Kumar.IJP2P 2011.

[15] "KANNADA LIPI VIKASA" DR. Manjunatha ,Yuvasadhane Publications, Bengaluru.