AN ARCHITECTURE FOR AN INSCRIPTION RECOGNITION SYSTEM FOR SINHALA EPIGRAPHY

D.A.S. Ruwanmini [1], K.V. Liyanage [2], K.G.N.D. Karunarathne [3], G.K.A. Dias [*4], S.T. Nandasara [5]

[1, 2, 3, *4, 5] University of Colombo School of Computing, Sri Lanka

Abstract

Sinhala Inscriptions are used as one of the major sources of getting information about ancient Sri Lanka. Revealing the Information from these inscriptions becomes a huge challenge for archeologists. This research paper focused on Sinhala character recognition in ancient Sri Lankan inscription. Our intention is to ease this process by developing a web based application that enable recognition of inscription characters through scanned images and store them in an inscription database. Using this system people can track geographical location of inscriptions. Epigraphist could be able to easily obtain Sinhala interpretation of Sri Lankan inscriptions via the optical character recognition feature in our system. Our work on this research project provides benefits to researchers in archaeology field, epigraphists and general public who are interested in this subject. Inscription site tracking module will present a map that user can go around easily by tracking the locations of inscriptions. This paper presents the Architecture for this Sinhala Epigraphy system.

Keywords: Epigraphy; Sinhala Inscriptions; Optical character recognition (OCR); Archaeology.

1. Introduction

This research project focused on ancient Sinhala character recognition in ancient Sri Lankan inscription. Inscriptions are used as one of the major sources of getting information about ancient Sri Lanka. It is a huge challenge to reveal the Information from these inscriptions. Main purpose of this project is to make an effort to provide a better solution to overcome some existing problems in inscription reading.

1.1.Research Background

There are many Inscriptions found in ancient cities of Anuradhapura and Polonnaruwa in Sri Lanka. Thonigala inscription, Galpotha (Stone Book) inscription (see Fig. 1.1), Mirror wall are some of them.

Figure 1.1: Polonnaruwa Galpotha (Stone Book) Inscription.

These inscriptions are very important as they are the major sources of getting information about ancient Sri Lanka. These inscriptions provide valuable information about the time, place and situation connected with the inscription. Inscriptions are considered as a reliable source for the knowledge of political, economic, cultural and social history of ancient times. Also it provides information on early human settlements and the Sinhala kings[1]. Inscriptions also show the evolution of the languages over the centuries. They show the development of each letter and structure. Inscriptions from the 9th century onwards had lengthy sentences and uses flowery language. Inscription from the 9th to 13th centuries carry conjoint consonants which came in through Mahayana Buddhism. [1]

Revealing the content of these inscriptions can be highly valuable to investigate the history of ancient Sri Lanka. Currently the content of each of these inscriptions are translated to modern Sinhala language manually by an archaeology expert who has specialized knowledge to understand the ancient scripts. So the Inscriptions letters read through human eye with great effort and this manual procedure would be time consuming and generate some uncertain outcomes due to lack of consistency.

1.2.Problem Statement

Epigraphy is the study of interpretation of ancient inscriptions which are engraved on a rock, stone or caves at the beginning of a piece of writing[2] [3]. Although the Inscriptions were used as a one of the information source to recognize expansion of Sinhala language, recognizing content in this inscription become a huge challenge due to the cause of various reasons. Some of characters used in inscription are partially or fully erased due to the reasons of war, weather conditions etc. In such cases where inscriptions are damaged and are incomplete, recognizing the characters would not be easy as the isolated character in the inscription cannot be clearly viewed.

In addition to the above difficulties in reading inscriptions, there are some other problems in reading inscriptions. Lack of specialized knowledge and lack of resources can be identified as some major issues. In the archaeology field there is a demand for scholars who have experienced knowledge of reading inscriptions. Therefore currently the existing archaeology experts have to put considerable amount of effort to read inscriptions. Further they also recognize the characters in these inscriptions manually. This manual procedure is very much inefficient and time consuming. It is often find that archaeology experts take long time period to investigate and experiment these inscriptions.

In general content of inscriptions represent as several lines of characters. With this situation it was really difficult to identify each word or sentence separately. It is also found that some sculptors engrave inscription in diametrically opposite way in ancient times. Due to these points it was really difficult to recognize isolated words or sentences.

Obtaining the meaning of each recognized word or sentence becomes a huge challenge. From the ancient times different nations has migrated to our country. They have changed the characters structure and the meaning of words for their convenience. As a result of this the meaning of a word can be different from time to time or from region to region. Therefore the meaning of the recognized word in inscription doesn't exactly match with the current version of Sinhala. There is also no way of finding the meaning of the inscriptions for researchers and visitors who are visiting these inscription sites. These inscriptions may have already translated by experts.

There is limited inscription information one can find in the internet. Based on our survey, application to refer Sinhala inscription information over the different centuries could not be found. Hence it will be very useful if a web based information system is publicly available to get inscription information related to Sri Lanka via the World Wide Web easily.

1.3. Research Purpose

The purpose of this research is to make an effort to provide a better solution to overcome some existing problems in inscription reading. This project has great value to the society by discovering historical records from inscription reading.

Main focus of the project is to recognize each isolated character set in these inscriptions using a proper character recognition process and map them into modern Unicode characters. Further as an IT solution we hope to provide an attractive system with the features to track the inscription by using geographical location, to search inscriptions using an inscriptions database module to convert historical inscriptions in all the sites with meaningful descriptions in Sinhala according to inscription content. Visitors to these inscription sites would be provided with inscription details without difficulty.

1.4. Objectives of the Research Project

The following are the objectives of our research.
- Recognize the ancient characters and digitize them.

- Map ancient character identified with modern Sinhala character
- Obtain Sinhala interpretation of Sri Lankan inscriptions via the character recognition feature
- Track geographical location of inscriptions

Design and develop a user-friendly software solution to overcome some existing problems of researchers in archaeology field, epigraphists and general public who are interested in this subject.

1.5.Research Methodology

As mentioned above section evaluating feasible OCR solutions will be taken as the initial step of this research project. Before that it is essential to prepare proper dataset. In that sense research methodology is broken down into three steps as mentioned below.

Step 1: The first step refer to creating a database for training using a set of ancient characters.
Step 2: Try out different feasible OCR solutions and evaluate results of each. Each OCR solution will follow the character recognition procedure of preprocessing, feature extraction and character recognition. Here we will mainly focused to find better OCR technique which has high recognition rate.
Step 3: Finally, in the third step, best resulted OCR solution is incorporated into the system.

2. Literature Review

2.1.Introduction of the Review

This review undertakes the responsibility to share the knowledge about some different and important Inscriptions' Character Recognition areas to the readers through a systematic search of research papers, academic journals, articles and websites.

2.2.Review of Literature

2.2.1. Manual Inscription Reading Methods

According to article done at RootWeb.Ancestry.com about Alternative Gravestone Reading Methods (2015) [5]; Hand Rubbing is possible on a uniformly colored stone surface, to lightly brush the surface with the palm of hand, which raises a light dust and leaves the recessed inscription as a dark color. Another method is by using a mirror to direct bright sunlight diagonally across the face of a gravestone, so that it can easily cast shadow in indentations which will makes inscriptions much more visible and easy to read. Another approach is using a viewing tube which is held against the stone to prevent light entering and then tilt the end of the tube touching the stone slightly so that a little light enters and then views the inscription through the tube.

However, currently researchers pay much attention to computerized character recognition task to achieve more efficiency and convenience.

2.2.2. *Optical Character Recognition (OCR)*

Publications focus on OCR is summarized below.

P.Niranjan (2015) [6], P.Nikhi, V.Jayakumar, S.kolkure (2015) [7] and Yasser Alginahi (2004) [8] have discussed in their research papers about Optical character Recognition process and technologies.

Rajakumar S. (2012) [9] conducted a research to recognize 7th century ancient Tamil characters from temple wall inscriptions. In this paper he attempt to recognize ancient Tamil characters by using Scale-Invariant Feature Transform (SIFT) features and present a new and efficient approach based on bag-of key points representation. This system achieves a maximum recognition accuracy of 84% using SIFT features.

Hubert Mara, Jan Hering and Susanne Kromker (2014) [10] presented Optical Character Recognition (OCR) to read ancient Chinese texts carved into stones. They have developed an automated system for processing of ancient Chinese inscriptions (sutras).

S.Rajakumar, V.Subbiah Bharathi (2011) [11] presented a methodology for Century Identification and Recognition of Ancient Tamil Characters. In this paper researchers have addressed the issue of Tamil character recognition, such as difficulties to understand Tamil characters. These characters are different from current century's Tamil characters. Therefore in this research paper presented a way of ancient Tamil character recognition in inscription using MATLAB with contourlet transform method. A Neural network has been used to train the image and compare the data with the current century's character.

Abhishek Tomar, Minu Choudhary, Amit Yerpude (2015) [12] conducted a survey research focused on Indian inscription character recognition areas by understanding the script and the languages in the images and carried out prehistoric inscription analysis.

H. K. Anasuya Devi (2006) [13] has conducted a research regarding preprocessing algorithms of an OCR system for Brahmi script reading. The researcher has presented various algorithms which were used in low level processing stage of image analysis for Optical Character Recognition.

Sachin S Bhat, H.V. Balachandra Achar (2015) [14] have conducted a research with the purpose of period identification of various ancient Kannada scripts using advanced recognition algorithms. They have proposed an algorithm including basic components of the images processing namely image acquisition, noise removal, segmentation of character sets for feature extraction, classification and recognition of segmented characters. The main specialty of this research is that they have developed a system to predict the era by examining a few characters in Kannada inscription. Their experiments have been done using MATLAB and they have mentioned that they can achieve 80% of accuracy and efficiency of the final results.

Ayatullah Faruk Mollah, Nabamita Majumder, Subhadip Basu and Mita Nasipuri (2011) [15] have conducted a research to design a complete Optical Character Recognition system for

camera captured image/graphics embedded textual documents for handheld devices such as s high-end cell-phones, Personal Digital Assistants (PDA), smart phones, iPhones, iPods, etc.

Shashank Araokar (2015) [16] has described a simplistic approach for recognition of visual characters using artificial neural networks. Basically in this paper researcher has shared knowledge in pattern recognition, neural networking and related disciplines.

Indu Sreedevi, Rishi Pandey, N.Jayanthi, Geetanjali Bhola and Santanu Chaudhury (2013) [17] have conducted a research on NGFICA (Natural gradient-based flexible Independent Component Analysis) based Digitization of Historic Inscription Images. This paper addresses the problems encountered during digitization and preservation of inscriptions such as perspective distortion and minimal distinction between foreground and background. In general inscriptions possess neither standard size nor shape nor color difference between the foreground and background. This paper proposed a method to enhance the minimal difference between text and the background of the inscription image. For minimizing the dependency among the foreground and background of historical inscription images, they have used NGFICA for obtaining independent components of the images. The proposed method improves word and character recognition accuracies of the OCR system by 65.3%.

Hugo Pires a, José Martínez Rubio b, Artzai Elorza Arana (2015) [18] have presented recent developments regarding the use of 3D scanning data sets for revealing invisible information from archaeological sites. For this study they have considered the most frequently used techniques like raking light photography, 3D scanning and PTM (Polynomial Texture Maps). They have used a Latin inscription from the Roman sanctuary and a panel with engravings from the rock art site of Monte Faro.

2.2.3. Handwritten Character Recognition (HCR)

Research articles focused on HCR is summarized below.

Theingi Htike and Yadana Thein (2013) [19] have done research to recognize Myanmar handwritten characters based on Competitive Neural Trees (CNeT). In this research work thirty three alphabet handwritten Myanmar characters are scanned and the training datasets are 660 and testing datasets are 330. Recognition accuracy rate 97% is obtained with testing data 330.

G.Vamvakas, B.Gatos, N. Stamatopoulos, and S.J.Perantonis [20] in their work presented an OCR methodology for recognizing historical documents, either printed or handwritten without any knowledge of the font. This methodology consists of three steps; the first two steps refer to creating a database for training using a set of documents, while the third one refers to recognition of new document images. They presented segmentation-free approach for recognizing old Greek handwritten documents especially from the early ages of the Byzantine Empire.

Luke V Rasmussen, Peggy L Peissig, Catherine A McCarty, Justin Starren (2012) [21] have done research to develop optical character recognition pipeline for handwritten form fields from an electronic health record (EHR). This research work was done under the two main phases. First one is focused on designing forms to capture hand-printed data specifically for OCR

processing. Second one has utilized custom-developed OCR engines to perform the handwriting recognition.

2.2.4. *Google Handwritten Character Recognition Tool [32]*

Google's Optical Character Recognition (OCR) software now works for over 248 world languages (including all the major South Asian languages including Sinhala). The technology extracts text from images, scans of printed text, and even handwriting, which means text can be extracted from old books, manuscripts, or images. When processing a document, it attempt to preserve basic text formatting such as bold and italic text, font size and type, and line breaks. However, detecting these elements is difficult and user may not always succeed. Other text formatting and structuring elements such as bulleted and numbered lists, tables, text columns, and footnotes or endnotes are likely to get lost.

2.2.5. *Recognition of Inscriptions with GIS Technology*

Matthias Arnold (2008) [22] has introduced a website of the Buddhist Stone Scriptures. It was to display textual and visual data and character recognition of the inscription together with GIS information in a XML database.

Thorsten Schwing (2008) [23] has shown the task of his project to study ,survey and measure all important Buddhist stone inscription sites in Shandong and document those information in a database connected to geometrical information in 2D or 3D.

Sandra Laniga, Arne Schillinga, Michael Auera, Bernhard Hoflea, Nicolas Billenb, Alexander Zipfa (2015) [24] have presented concept for the integration of the Buddhist inscriptions into a Spatial Data Infrastructure (SDI) using Open Geospatial Consortium (OGC). In this research paper they have implemented functionalities to display the textual description of all inscription sites, an inscription catalog with metadata about the texts, a reading tool to explore the inscriptions and a search module to query the inscription database. Also multimedia map combines geographic 2D/3D visualization with 360°-panoramas, annotated photographic pictures and GIS functionality.

3. Design of the Solution

3.1.Introduction

In section 1 and 2 we have briefly explained about the research problem, significance of the study, objectives, scope, research methodology and the gap in the literature. As a summary this chapter presents a detailed description of the design overview of the solution to the research.

Scanned photographs of original inscription or impression of an inscription made on inked paper which is called estampage have been used as main resources in this research project. Therefore we put considerable effort to gather these types of data as the   initial step in this research.

The collection and organization of data is the integral and critical part of the research process. Collecting data was highly challengeable task due to the cause of some major difficulties. The whole research project is based on ancient Sinhala character inscriptions and most of those inscriptions placed in garrigues or places which are very difficult to visit. Collecting inscriptions which have better surface level clearness become very difficult task due to the cause that some parts of these inscriptions are erased or damaged. Giving Sinhala interpretation based on these partially damaged inscriptions will be meaningless. Therefore to capture Photographs from better intelligible inscription or estampages of inscriptions is recognized as one of the major problems. Further the data which needed for this type of research is not publicly available. In that situation we have to deeply refer many archaeology scholarly books [25] [26] [27] [28] [29] [30] with great effort. Accessibility to inscriptions to capture or take photographs is also pretty much difficult due to the cause of government restrictions and inscription protection rules.

3.2.Data Collection Methods

The following data collection methods were used during the research to capture the data.
1) Interview with archeology experts
2) Field visits & Observation
3) Refer related archeological books
4) literature search

3.2.1.   Interview with Archeology Experts

Conducting interviews with experts is one of major data gathering techniques. Expert's perceptions, their knowledge and the opinion are essential in order to carry out the project successfully. We gathered qualitative data by conducting interviews with archaeology experts. Discussed in depth with a large number of field officers, knowledge experts which encouraged us to gather variety of information regarding this research. Face to face interviews were conducted with some field officers in Postgraduate Institute of Archaeology, Professors in Universities in order to gather information about current methods used in inscription character reading, the methods used to obtain estampages of inscriptions, inscription sites details in Polonnaruwa, expansion of ancient Sinhala characters, and problems they had faced as field officers when they find new inscriptions. When studying the expansion of ancient Sinhala characters, we have recognized that there are huge variations in each character. Good guidance was obtained from them for our further works. Subsequently opinions of different knowledge experts were carefully investigated to obtain better understanding about this research project. It became more challengeable task to compromise different ideas and come up with a better conclusion.

3.2.2.   Field Visits & Observation

Going into the "field" for observation is one of the effective way of gathering reliable evidence. Authors visited Polonnaruwa ancient ruins in order to gather original photos of selected inscriptions. A photographer with DSLR camera was hired to get quality images for this purpose. Also gathered more digital inscription images by visiting Archaeology Department of Sri Lanka and the National museum in Colombo. Direct observation is   essential to gain the understanding

of the research problem. From these visits we identified how difficult it is to select inscriptions that are needed for our project.

3.2.3.   *Archaeological Scholarly Books Referred*

On the other hand we gathered quantitative data and qualitative data both by referring archaeology scholarly books such as "Ancient Inscriptions of Ceylon"[26] , "Epigraphia of Zeylanica" [25] , "Seegiri Graffiti" [27] , "Sinhala Akshara Rupa Wikasaya saha Bharatheeya Abashaya" [28], "Shila Lekana Sangrahaya"[29] and "Sinhala Warna Malawe Vikashaya" [30]. These archaeological books encouraged us to
- gather inscription sites information which are engraved in different places,
- collect some estampage images of inscriptions,
- Gather details about Inscriptions sites' locations and so on.

3.3.The Data Sources Collected

Following Figure [3.1] shows one of the original image data used for our research project and Figure [3.2] shows the estampages 'image of Nissankamalla's Inscription found in Polonnaruwa.

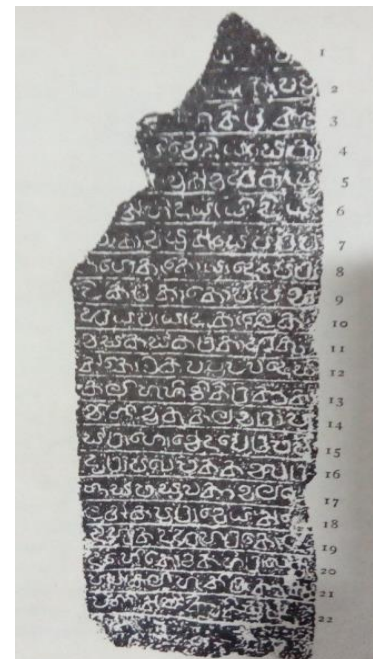

Figure 3.1: Original photograph of Galpotha inscription

Figure 3.2: Estampageage of Nissankamalla's Inscription found in Polonnaruwa

The identification of the ancient character set in selected time period and identification of each character variations and styles become more predominant part in this research.

3.3.1. Properties of Ancient Sinhala Inscription Characters

There are many Character variations found in Sinhala inscriptions and each character has at least five variations. When inscriptions are found closer to the current year, character variations are increased due to the cause of population diversity. Some character variations found are shown in figure 3.3. [25]
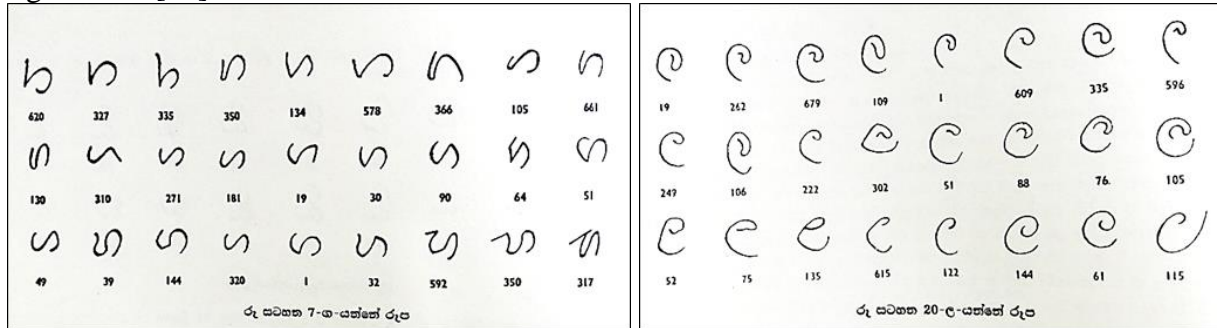

Figure 3.3: Letter variations of 'ග' ('ga') and 'ළ' (la) characters

The identification of character set is a predominant condition in this research project as shown above. The following figure 3.4 summarizes the analyzed characters after referring to the Epigraphia of Ceylanika volume VI and "Shilalekana Sangrahaya".
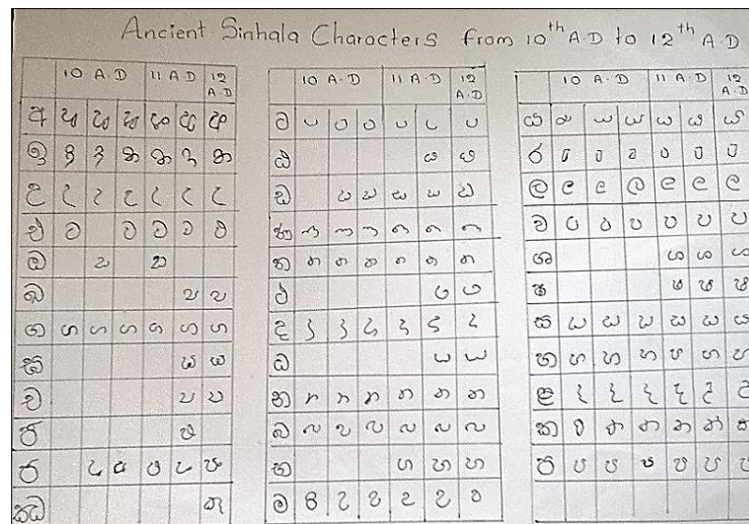

Figure 3.4: Manually prepared chart found in Epigraphia of Ceylanika [25] and Sigiri Graffiti [27]

3.3.2. The database

All the data will be entered in to the database for storing the data in a useful manner. In our research we are mainly creating two databases. One database to support OCR functions. In this database details of inscription characters will be stored with details about commonly used words in most inscriptions with its transcript data. The other database is for storing GIS information of each inscription. Descriptions about inscriptions and their Sinhala interpretations are also stored in the database.

3.4.Design of the Final Proposed System

Final solution will be responsive web application which includes two modules; OCR module and GIS module. OCR module and GIS module will be developed as separate functions in one platform. In this research we target two main user types. They are archeology experts and public users. OCR module will be developed especially for the archaeology experts. Any public user can access the GIS module.

Archaeology experts/researches can upload the inscription image and he can view the modern letters that are recognized by the OCR module. Public User can search inscription sites by using search catalog and find inscription's location. Public user can also view descriptions for each inscription. Sinhala interpretation of inscription and story of the inscription could be viewed from the inscription site. Overall System Architecture is given in Figure 3.5.
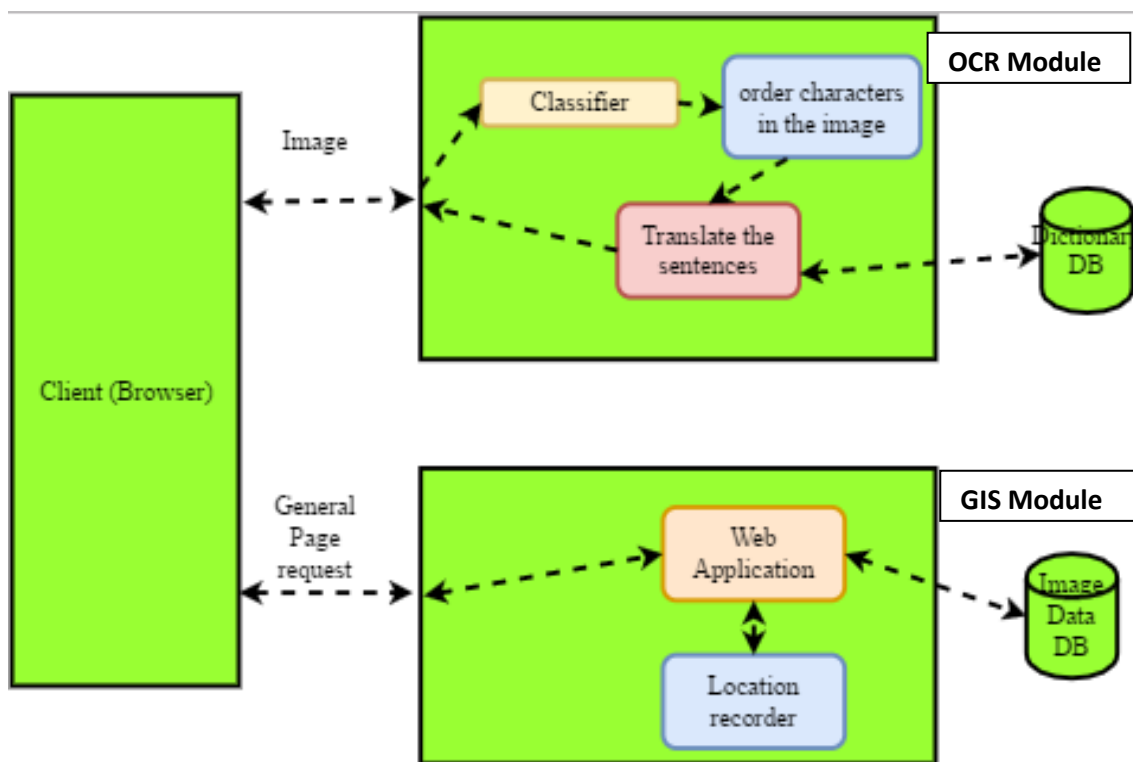

Figure 3.5: Overall System Architecture

3.4.1. OCR Design approach

The Optical Character Recognition process can be divided into three major steps:
1) Preprocessing
2) Character Recognition
3) Post processing

In this section we describe each step of character recognition process separately.

Acquisition of Image
Input scanned image is the pre-condition for the preprocessing process. Firstly original inscription image or digitalized estampage image of input data is optically scanned. The scanned image can be any document of different dimensions. This scanned input image is fed to pre-processing section for further analysis.

3.4.1.1. Preprocessing

Preprocessing the input images facilitate to increase the recognition accuracy. OCR loads an image from a given source and performs different algorithms to clean the image. These algorithms apply techniques such as blurring, threshold to the image to reduce the noise. In this process it will enhance the visual appearance and the quality of ancient Sinhala characters in inscription images. Some different actions performed during pre-processing are listed below. [31][33]

1) Binarization
2) Boundary Detection
3) Segmentation
4) Thinning

Binarization
Binarization plays an important role in pre-processing. Binarization is a technique by which the gray scale images are converted to binary images. Basically separation of background and actual image area referred as foreground of a scanned image. When converting a grayscale image into binary image by using thresholding, Otsu method (Otsu's method, named after Nobuyuki Otsu) was used to perform thresholding. Otsu's method automatically performs histogram shape-based image thresholding for the reduction of a gray-level image to a binary image. The algorithm assumes that the image for thresholding contains two classes of pixels either foreground or background and then calculates the optimum threshold separating those two classes so that their intra-class variance is minimal. [33][34]

Boundary Detection
The binarized image is now applicable for boundary detection with operation of dilation and erosion using Opencv - Python. In this operation the boundaries of scanned image are detected. It is necessary to detect the boundaries so as to select an individual character. Therefore Dilation and Erosion are operations which increase or decrease objects in size and can be very useful during the preprocessing stage. Erosion makes an object smaller by removing or eroding away the pixels on its edges whereas dilation makes an object larger by adding pixels around its edges. [31]

Segmentation
Segmentation is done in order to improve the analysis of an image. This will be done by separating the pixels of an image according to semantic content and facilitating the manipulation and visualization of the data with a computer. Character detection is one of the main part of the segmentation. Here all the characters' with middle zone consider as the the character segmentation portion. This is because when considering the inscriptions it may contain for eg.

[Dias et. al., Vol.4 (Iss.12): December, 2016]
ICV (Index Copernicus Value) 2015: 71.21

ISSN- 2350-0530(O), ISSN- 2394-3629(P)
IF: 4.321 (CosmosImpactFactor), 2.532 (I2OR)
InfoBase Index IBI Factor 3.86

"කි", "කා" , of that individual character (ක) with upper zone and lower zone. Main purpose of character segmentation is detecting individual character with middle zone. [33]
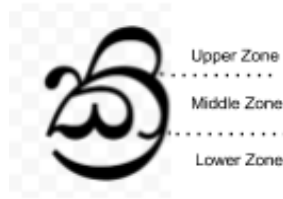
Thinning

Thinning or skeletonization is used to clean and visualize the skeleton of scanned input image. This process deletes the dark points in the image which means remove selected foreground pixels from binary image that approximate the center skeletons of the regions. Thinning used to infer shape in the original image. Consider the "අ" Character.

It consists basically of two strokes or line-link pieces (two horizontal) and two curves connected in a certain manner. The thickness of the strokes and curves are irrelevant to the recognition problem. What is important is the topology of how the strokes and curves are connected together. In such situations it is convenient to simplify the input as much as possible in order to make the topological analysis as simple as possible. In that sense it is very important to obtain the skeleton of the character inside the image. [33]

3.4.1.2. *Character Recognition*

After preprocessing of image document, OCR technology can be used to read and translate characters. Mainly Character Recognition consist of the basic stages to compare the characters in the scanned image to the characters in the learned set involving,

- The procedure of extraction and isolation of each character from an image
- Determined the properties of extracted character.
- Comparison of the properties of extracted character and learned character.

As mentioned above these basic stages are named as feature extraction and classification. [33]

Feature extraction

In this stage a set of features are extracted from the character. More features extracted for the character increases the probability of its recognition. Extracting the text data from images is important for reading, editing and analyzing the text context contained in the images. Computers cannot recognize the text data directly in images. Thus the design of computer program called "Optical Character Recognition" which can recognize text in images play a vital role. [33]

Classification

The extracted feature data must go through the process of classification. This step classifies the extracted individual character. Several approaches can be used to classify the characters such as clustering, K-mean algorithm, etc. [33]

3.4.1.3.   Post Processing

After recognition process, post processing act as the final stage. This is needed because sometimes the recognized character does not match with the original character. In some other cases character could not be recognized from the input image. In post processing stage it mainly involves error checking. [33]

Steps involved with the Neural network application to be developed which describes the above mentioned overall character recognition procedure is given below.
- First of all Inscription scanned image is uploaded to the system.
- Selecting the training technique provided
- Train the image using the selected technique.
- Extract characters.

3.5.Inscription Sites Tracking

This module will present a map that user can go around easily by tracking the locations of inscriptions. The 2D map provides detailed information about the different inscription sites. This is a kind of a regional geographical map which shows where the inscriptions are located in selected era of our research. This component will guide the archaeology researchers who visit ancient ruins. Given figure 3.6 shows the main interface of the designed prototype of the responsive web application to be developed.

Figure 3.6: Designed Interface of Inscription Site Tracking

The designed solution is to provide a web application which included sections on space, time and inscriptions. The section on "space" defines the location of the site, describes its characteristic topography, the grouping of inscriptions and elucidates on the place name. In the "time" section a site is placed in the context of inscriptions history and story in general, discussing its dating, or its relations to monasteries and historic figures. In the section on "inscriptions" the inscribed texts are divided in groups based on content and shape. Finally, they are analyzed with focus on art historical characteristics, like engraving techniques and style of calligraphy. It will also

provide some catalogs to access for searching inscriptions. They can load further metadata, like detailed information on the inscriptions measurements, location, and condition. In simple , when a researcher or visitor wants to visit the places the web application should provide road map guidelines to reach there and search for nearby inscriptions and provide related information of that particular place.

4. Conclusion

This research paper provides a clear overview of the areas going to touch through this entire research project. Based on the entire review we can conclude that Optical Character Recognition systems are well developed in modern society. OCR plays an important role of character recognitions in different sources of handwritten notes, printed documents and Inscription content readings. When comparing what the number of researches have done in handwritten character recognition, and inscription character recognition it can be concluded that still there is lot of scope for Inscription character recognition research. According to our review we could found that OCR techniques heavily used in India for Tamil character recognition. Technological wise artificial neural networking, image processing techniques significantly contribute to develop OCR systems.

In this paper we have provided the summary of completed design tasks. Our objectives were met to a particular level for the current investigations by capturing the necessary details using interviews with archeology experts and referring archaeology scholarly books. All image processing experiments done by using digitalized photographs of estampage page of inscription and original photographs of inscriptions.   We have concluded that digitalized estampage page photographs will be more convenient and accurate.

Before the implementation of the proposed system for Sinhala inscription OCR, we will test various OCR modules which are based on different techniques. Every logical step will be implemented and tested as separate modules and examined the outputs for each. Each module's results explicitly analyze and compared to find the best resulted OCR solution. Best resulted OCR solution will be included as a module for the proposed system with the GIS module.

References

[1] Pieris.K ,"Inscriptions of ancient and medieval Sri Lanka", Sri Lanka's National Newspaper,1918.[online].Available:http://archives.dailynews.lk/2009/09/18/fea30.asp
[2] Wikipedia,"Stone Inscriptions in Sri Lanka", March,2016.[Online]. Available: https://en.wikipedia.org/wiki/Stone_inscriptions_in_Sri_Lanka
[3] Wikipedia, "Epigraphy", March, 2016 [online]. Available: https://en.wikipedia.org/wiki/Epigraphy
[4] Wikipedia, "Geographic information systems in geospatial intelligence", July, 2014. [Online]. Available: https://en.wikipedia.org/wiki/Geographic_information_systems_in_geospatial_intelligence
[5] Root Web [2015] Alternative Gravestone Reading Methods [ONLINE] Available at: http://www.rootsweb.ancestry.com/~iamcpcrp/pcri/reading0.htm
[6] P.Niranjan [2015] Literature Review of Segmentation Problems in Nepali Optical Character Recognition   [ONLINE]   Available   at:

https://www.researchgate.net/publication/291348878_Literature_Review_of_Segmentation_Problems_in_Nepali_Optical_Character_Recognition

[7]   P.Nikhi,V.Jayakumar,S.kolkure [2015] OPTICAL CHARACTER RECOGNITION: AN ENCOMPASSING REVIEW [ONLINE] Available at: http://esatjournals.net/ijret/2015v04/i01/IJRET20150401062.pdf

[8]   Yasser Alginahi [2015] Preprocessing Techniques in Character Recognition [ONLINE] Available at: http://cdn.intechopen.com/pdfs/11405.pdf

[9]   Rajakumar S. [2015] 7th Century Ancient Tamil Character Recognition from Temple Wall Inscriptions [ONLINE] Available at: http://www.ijcse.com/docs/INDJCSE12-03-05-055.pdf

[10] Hubert Mara, Jan Hering and Susanne Kromker [2014] GPU Based Optical Character Transcription for Ancient Inscription Recognition [ONLINE] Available at: https://www.researchgate.net/publication/224610728_GPU_Based_Optical_Character_Transcription_for_Ancient_Inscription_Recognition

[11] S.Rajakumar, V.Subbiah Bharathi [2011] Century Identification and Recognition of Ancient Tamil Character Recognition [ONLINE] Available at: http://www.ijcaonline.org/volume26/number4/pxc3874237.pdf

[12] Abhishek Tomar , Minu Choudhary , Amit Yerpude [2015] Ancient Indian Scripts Image Pre-Processing and Dimensionality Reduction for Feature Extraction and Classification: A Survey [ONLINE] Available at: http://www.ijcttjournal.org/2015/Volume21/number-2/IJCTT-V21P116.pdf

[13] H. K. Anasuya Devi [2006] Thinning: A Preprocessing Technique for an OCR System for the Brahmi Script [ONLINE] Available at: http://www.ancient-asia-journal.com/articles/10.5334/aa.06114/

[14] Sachin S Bhat, H.V. Balachandra Achar [2015] Character recognition and Period prediction of ancient Kannada Epigraphical scripts [ONLINE] Available at: http://www.ijarcce.com/upload/2016/si/nCORETech-16/nCORETech%2024.pdf

[15] Ayatullah Faruk Mollah , Nabamita Majumder , Subhadip Basu and Mita Nasipuri [2011] Design of an Optical Character Recognition System for Camerabased Handheld Devices [ONLINE] Available at: https://arxiv.org/ftp/arxiv/papers/1109/1109.3317.pdf

[16] Shashank Araokar [2015] Visual Character Recognition using Artificial Neural Networks [ONLINE] Available at: https://arxiv.org/ftp/cs/papers/0505/0505016.pdf

[17] Indu Sreedevi, Rishi Pandey, N.Jayanthi, Geetanjali Bhola [2013] NGFICA Based Digitization of Historic Inscription Images [ONLINE] Available at: https://www.researchgate.net/publication/258397837_NGFICA_Based_Digitization_of_Historic_Inscription_Images

[18] Hugo Pires a, José Martínez Rubio b, Artzai Elorza Arana [2015] Techniques for revealing 3D hidden archeological features Morphological residual models as virtual-polynomial texture maps [ONLINE] Available at: https://www.researchgate.net/publication/276507424_Techniques_for_revealing_3D_hidden_archeological_features_Morphological_residual_models_as_virtual-polynomial_texture_maps

[19] Theingi Htike, Yadana Thein [2013] Handwritten Character Recognition Using Competitive Neural Trees [ONLINE] Available at: http://www.ijetch.org/papers/574-ST0017.pdf

[20] G.Vamvakas, B.Gatos, N. Stamatopoulos, and S.J.Perantonis [2008] A Complete Optical Character Recognition Methodology for Historical Documents ONLINE] Available at: http://users.iit.demokritos.gr/~bgat/3337a525.pdf

[21] Luke V Rasmussen,Peggy L Peissig, Catherine A McCarty, Justin Starren [2012] Development of an optical character recognition pipeline for handwritten form fields from an electronic health record [ONLINE] Available at :http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3392858/

[22] Matthias Arnold [2008] Buddhist Stone Scriptures From Shandong, China [Online] Available at: https://www.researchgate.net/publication/228410775_Buddhist_Stone_Scriptures_From_Shandong_China [Accessed 6th June 2016]

[23]   Thorsten Schwing [2008] Measurements and Documentation of Buddhist Stone Inscriptions in China, Shandong [ONLINE] Available at: https://www.fig.net/resources/proceedings/fig_proceedings/fig2007/papers/ts_5f/ts05f_02_schwing_1247.pdf

[24]   Sandra Laniga, Arne Schillinga, Michael Auera, Bernhard Hoflea, Nicolas Billenb, Alexander Zipfa [2015]Interoperable integration of high precision 3D laser data and large scale geoanalysis in a SDI for Sutra inscriptions in Sichuan (China) [Online] Available at : https://www.researchgate.net/publication/266182349_Interoperable_integration_of_high_precision_3D_laser_data_and_large_scale_geoanalysis_in_a_SDI_for_Sutra_inscriptions_in_Sichuan_China

[25]   Paranavitana.S," Polonnaruwa Pillar Inscription ", Epigraphy of Ceylonica, Volume iv ,Archaeology Department of Sri Lanka, pp 127 - 250.[1950]

[26]   Dr.Muller.E, Ancient inscription of Ceylon,Volume i, London,Trubner,pp 10-232, [1883]

[27]   Paranavitana.S, Sigiri Graffiti, Volume 1, Archaeology Department of Sri Lanka, [1956]

[28]   Tero O.pangnasara, Sinhala Akshara Rupa Wikasaya saha Bharatheeya Abashaya, Volume 1. [2005]

[29]   N.Mudiyanse , Sinhala Akshara Malawe Prabavaya saha Vikashaya" .

[30]   Tero O.pangnasara,,Sinhala Basha Vikashaya saha Shila Lekana Vimarshana, PP 10-160, [2010]

[31]   OpenCv , OpenCv python tutorials , [online].Available at : http://docs.opencv.org/3.0-beta/doc/py_tutorials/py_tutorials.html

[32]   Google's Optical Character Recognition (OCR) software https://opensource.com/life/15/9/open-source-extract-text-images

[33]   Optical Character Recognition https://en.wikipedia.org/wiki/Optical_character_recognition

[34]   M. Sezgin & B. Sankur (2004). "Survey over image thresholding techniques and quantitative performance evaluation". Journal of Electronic Imaging. 13 (1): 146–165.

*Corresponding author.
E-mail address: gkad@ucsc.cmb.ac.lk