Review Article

Approaches Used to Recognise and Decipher Ancient Inscriptions: A Review

H.M.S.C.R. Heenkenda* and T.G.I. Fernando

Department of Computer Science, University of Sri Jayewardenepura, Nugegoda, Sri Lanka

Date Received: 12-11-2020 Date Accepted: 02-12-2020

Abstract

Inscriptions play a vital role in historical studies. In order to boost tourism and academic necessities, archaeological experts, epigraphers and researchers recognised and deciphered a great number of inscriptions using numerous approaches. Due to the technological revolution and inefficiencies of manual methods, humans tend to use automated systems. Hence, computational archaeology plays an important role in the current era. Even though different types of research are conducted in this domain, it still poses a big challenge and needs more accurate and efficient methods. This paper presents a review of manual and computational approaches used to recognise and decipher ancient inscriptions.

Keywords: ancient inscriptions, computational archaeology, decipher, scripts

1. Introduction

Epigraphy includes deciphering the text of inscriptions and analysing the information on them (Khan and Vaidya, 2019). Inscription or epigraph is a historical or religious record cut, painted or written on hard surfaces such as stones, bricks, metals, rocks, or copper plates. Inscriptions can be classified based on the surfaces they engraved, the language used in the inscriptions, the age of the inscriptions and the geographical regions where the inscriptions are located. Different languages were used to write inscriptions such as Sanskrit, Sinhala, Mayan, Latin, Aramaic, old Chinese, Mycenaean Greek, Egyptian, Sumerian, Italian, Japanese, Russian, Iranian, Arabic, Kannada, Tamil, Telugu, Brahmi, Pali, Devanagari and Hoysala. Some inscriptions are meaningful only as graphics rather than textual signifiers. Incidental inscriptions defined as those not seriously meant for preservation such as wall scrawling of the graffiti on cheap materials like bricks and potsherds.

Inscriptions are an important source of information of past civilisations, economies, cultural events, names and designations of royal officers, trades and industries in ancient countries, religious activities and beliefs practised all over the world. In fact, inscriptions demonstrate the evolution of languages and development of characters over the past centuries. Hence, archaeological departments in respective countries have taken measures to preserve epigraphy. Deciphering the contents of inscriptions is a great challenge in the field of archaeology.

1.1 Inscriptions in the Sri Lankan context

Currently, 4,000 stone inscriptions have been discovered from Sri Lanka (Bandara and Warnajith, 2012). Around 1,500 of them are recorded. Sri Lankan inscriptions were written in Sinhala, Tamil, Brahmi, Chinese, Arabic, Persian, Sanskrit and Pali. These inscriptions contain the records of donations made to temples, maintenance of religious places, the establishment of tanks and directions for civilians.

**Correspondence: sankani.ruchirani@gmail.com*
Tel: +94 717744264

2. Materials and Methods

2.1 Manual methods to recognise and decipher ancient inscriptions

(a) Eye copy

Epigraphers translate the text on inscriptions by observing the inscription at the site it is located with the naked eye. Translated handwritten copy is known as eye copy. This method has some barriers like inconvenient to reach the inscription due to height (especially for cave inscriptions) and other natural factors like brightness, condition of inscription surface. Specially, it depends on human factors such as eyesight, specialised knowledge to understand ancient scripts, experience, etc. Hence, a great effort and time need in manual procedure.

(b) Estampages (rubbed copy)

Major and popular copying methods of inscriptions are called estampages (rubbed copies) (Dias et al., 2016). Initially, the inscription is cleaned by a brush. Then apply some water on the inscription and lay a dimai paper (shiny side lay on the inscription). Again, spray some water on the paper and need to press the paper until shapes of letters engrave on the paper (Figure 1). Next, apply roneo ink on paper. Then, it needs to rub the surface of the paper with a piece of cloth that is covered from ink. After several minutes, remove the paper from the stone. Archeology experts can read the estampage and note down essential information related to the inscription on the backside of the estampage.

Figure 1. During the rubbing of the inscription (Image credit: Department of Archeology, University of Sri Jayewardenepura).

(c) Tube view

Visualising the inscription via a tube is called a tube view. This method is used to prevent unnecessary darkness (Dias et al., 2016). Ultimately, it needs to be read manually.

(d) Mirror view

A mirror is directed to bright sunlight diagonally across the face of a gravestone. Then inscriptions are more visible and forwarded to eye copy method (Dias et al., 2016).

(e) Laser view

This method directly aligns lasers to the inscription. As a result, characters are enlarged and appeared on a digital screen. Then it can be forwarded to the eye copy method. Digitised images can be stored for future reference (Dias et al., 2016). This is not a cost effective method.

3. Significance of Computerised Approaches for Translating Ancient Inscriptions

Manual procedures are inefficient and time consuming. Archeological experts and epigraphers utilise a great amount of time to investigate inscriptions. Innovative computerised techniques can recognise, decipher and analyse inscriptions in a more accurate, and reliable manner. Hence, usability and attraction of archeological repositories can be enhanced by using computerised approaches.

4. Existing Computational Approaches to Recognise and Decipher Ancient Inscriptions

4.1 Optical character recognition (OCR)

Many researchers have utilised OCR to translate inscriptions. Publications focused on OCR are summarised below.

Bandara et al. (2012) implemented a computational system to establish the accurate alphabet fonts of early Brahmi scripts of Sri Lankan inscriptions. The research is conducted under four major stages. Initially, characters of early Brahmi scripts are produced from photographic data. Next, ancient characters are identified. Based on the results, a database is created. Finally, a web-based interface has been developed for scholars over KISSEL (Knowledge Integration Servers System for E-Learning) network which was developed by a research group of Ibaraki University to share the knowledge and experience among KISSEL users. In order to digitise a sample image of Brahmi characters, they used an optical image scanner with the resolution of 300 pixels per inch. Then the scanned image is saved in JPEG format. Adobe Photoshop CS3 was used for the primary image processing. Further processing is done using the MATLAB software. Black and white pixels are represented as 0 and 1 respectively. Next, it finds the center of the mass of the letter. Then, the Majority algorithm is introduced in order to find the ideal shapes of the alphabet of early Brahmi scripts. In the Majority algorithm, the values of identical pixels (i,j) of sample images are examined and the majority value (either 1 or 0) is assigned to the (i,j)th pixel of the font image. Due to the different shapes of letters in the same period, several sample letters have been examined and finally they have found the ideal shapes of Brahmi characters. In order to compare the produced fonts with the original image, a correlation function has been used in this research.

Dehiya and Verma (2014) focused on script identification that consists of Oriya, Telugu and Kannada letters based on OCR. Preprocessing has been done using two major steps: binarisation and line segmentation. Grayscale images are converted into 0 and 1. Projection method has been used for line segmentation that is the most efficient method for binary images as it simply divides the images into two categories. To classify scripts, horizontal and vertical projection profile based extraction techniques have been used. The features such as the number of vertical lines equal to a threshold height, upper profile component, frequency count in top segment, horizontal stroke like component in top segment and tick mark component have been extracted from the OCR images. For feature classification, k-nearest neighbor (KNN) classification approach has been used. The system achieved an accuracy of 98.64%.

Dias et al. (2016) developed a web-based system that recognises inscription characters and stores in an inscription database. Further, they showed a 2D map of tracking inscription sites. It can track geographical locations of inscriptions. Scanned images of inscriptions are stored in the database. This database supports OCR. In another database, a Geological Information System (GIS) of inscriptions is stored. Further, Sinhala interpretation of inscription text is stored in those databases. As input, inscription image or digitised estampage image is used after the optical scanning process has been completed. They deployed a web-based system that includes OCR and GIS modules. Archeological experts may use the OCR module and public users may use the GIS module. Binarisation, boundary detection, segmentation and thinning involved in the preprocessing stage. On arrival in the binarisation phase, Ostu's method

(Yousefi, 2011) is applied. After that, thinning, feature extraction and classification are performed. As many researches utilise clustering, the K-mean algorithm is used for clustering.

Gautum and Chai (2017) presented Brahmi script recognition system using OCR. A geometric method is used for feature extraction. Images are scanned by optical scanners. Cropping, thresholding and thinning are involved in the preprocessing stage. In the segmentation process, images are divided into lines and further divided into isolated characters from each identified line. The features of Brahmi characters are extracted using six entities of the geometric method such as corner point, bifurcation point, ending point, intersect point, circle and semi-circle.

Rajakumar and Bharathi (2012) revealed a method to recognise ancient Tamil characters from temple wall inscriptions. He recognised ancient Tamil characters by using Scale-Invariant Feature Transform (SIFT). Simultaneously, a novel method was introduced based on bags of keypoints representation. The method achieved a recognition accuracy rate of 84%.

Mara et al. (2014) presented OCR to read ancient Chinese texts on stones. An automated system was developed to process the ancient Chinese inscriptions (Sutras). The removal of noise from the image is done by Fast Fourier Transform (FFT) smooth low-pass filter. Multiscale Vessel Enhancement (MVE) filter is used to distinguish non-linear structures of the noise and the linear structures of the characters. COCR2: A Small Experimental Chinese OCR, Winton TH-OCR V9.0, Readiris pro 11 Asian OCR programs were applied for MVE filtered images.

Rajakumar and Bharathi (2011) developed a methodology to identify the century of the characters written in ancient Tamil inscriptions and convert them into current century’s form using MATLAB. The 3D contour-let method is adopted instead of 2D wavelet transforms to recognise ancient Tamil characters on inscriptions. A clustering mechanism is used to recognise characters from an input image. Noise presented in the image is removed using fuzzy median filters. Finally, neural networks have been used to train images and compare them with current Tamil letters.

Tomar et al. (2015) discussed various image pre-processing techniques and dimensionality reduction for feature extraction and classification. Filtering techniques such as mean filter, median filter, wiener filter, conventional adaptive median filter, decision based median filter, bilateral filter, butterworth low pass filter and gaussian low pass filter are available to denoise images. Among them, median filtering is best for the salt and pepper noise but in some cases non linear bilateral filter proves best for enhancing the quality, it smoothes but with sharp edges. However, all methods own certain limitations. Global thresholding is the simplest thresholding method. Dimensionality reduction can be performed using Principal Component Analysis (PCA), Kernel based Principle Components Analysis (KPCA) and Independent Components Analysis (ICA). This survey shows that PCA performs optimal solutions.

Bhat and Achar (2016) identified the period of Kannada scripts using advanced recognition algorithms. This algorithm includes image acquisition, noise removal, segmentation of character sets for feature extraction, classification and recognition of segmented characters. MATLAB was used to complete the experiment. The system achieved an accuracy of 80%.

Sreedevi et al. (2013) conducted research on NGFICA (Natural Gradient-based Flexible Independent Component Analysis) based digitisation of historic inscription images (Sreedevi et al., 2013). The paper addressed the issues encountered during digitisation and preservation of inscriptions such as perspective distortion and minimal distinction between foreground and background. A method has been introduced to minimise the dependency between foreground and background of inscription images. This novel system improved the accuracy of the OCR system by 65.3%.

Pires et al. (2015) presented the use of 3D scanning datasets for revealing invisible information from archaeological sites in Latin inscriptions from the Roman sanctuary. The paper presented techniques that have been recently used such as raking light photography, 3D scanning and PTM (Polynomial Texture Maps).

4.2 Geological information system (GIS)

The software developer (Arnold, 2008) established a website of Buddhist stone scriptures. He used OCR to recognise and GIS for present inscription information. Hence, the website displays textual and visual data of inscriptions.

Schwing measured important Buddhist stone inscription sites in Shandong and stored them in a database. The database contains geometrical information of inscriptions in 2D or 3D manner (Schwing, 2008). The field measurements are made with GPS for global positioning.

Laniga et al. (2015) proposed a system for the integration of the Buddhist inscriptions into a Spatial Data Infrastructure (SDI) using Open Geospatial Consortium (OGC). This system can list features such as prompts description, catalog with metadata, and search module to query the inscription database, 360° panoramas with 2D, 3D multimedia maps, annotated photographic pictures and GIS functionalities.

4.3 Google handwritten character recognition tool

Google's optical character recognition software recognises 248 languages. The technology extracts text from books, manuscripts, images etc. It preserves basic text formatting.

4.4 Supervised machine learning techniques

Publications that concentrate on supervised learning more frequently adopt pattern recognition.

(a) Support vector machine (SVM)

Rajkumar and Bharathi (2012) presented a system to recognise seventh century Tamil characters from walls of temples. Pre-processing operations performed to enhance the quality of images. Further, k-means algorithm and SVM classifier applied to get better overall classification accuracy. SVM classifier treated the bag of key points as the feature vector, to identify which character to assign to the image.

Subashini et al. (2012) used SVM to predict the period of Tamil epigraphical script. Classification was done using an SVM. The system successfully differentiated between four different centuries of characters. The SVM provided an iterative training algorithm that minimises an error function to construct an optimal hyperplane. Multiclass SVM turned out to be a very efficient method in the process of prediction. Overall accuracy achieved as 88.45%.

(b) Fuzzy logic

Soumya and Kumar (2014) implemented OCR and fuzzy logic to recognise Kannada text from ancient epigraphs. The system is tested for 100 epigraph images of ancient Kannada documents. Segmentation of input images, extracting statistical features of the character, classification and recognition of the segmented characters were carried out as the major phases of the system. The system computed mean, variance, standard deviation, skewness, kurtosis, entropy, homogeneity, correlation from the segmented characters of the given input image. Fuzzy based classifier which employs if-then rules for analysis used for classification. Mamdani fuzzy model adopted on Gaussian based computations. The system computed the model and classified characters of Brahmi and Hoysala script. After, retrieve the model values of the extracted features. The algorithm is as follows:

Input to the module is extracted ten features from the feature extraction module.
Output is recognised as a Kannada character in modern form.
Begin
Step 1: Find the model value
Step 2: Compare the model value with classes described in the dataset
Step 3: Compare model value in particular class applying if-then rules
Step 4: If (found nearest match) Retrieve the entire row in the dataset and display the result of modern Kannada character
Step 5: else display "result not found"
End

4.5 Unsupervised machine learning techniques

(a) Clustering techniques

Preethi and Mamatha (2016) implemented the optical character recogniser for the epigraphical script images. An artificial neural network, a clustering algorithm, K-nearest neighbor algorithm, Bayes classifiers and a support vector machine have been used as classifiers.

Rajkumar and Bharathi (2011) adopted clustering techniques for century identification and recognition of ancient Tamil characters. The characters from the input image are recognised through a clustering mechanism. Top down and bottom up segmentation have been applied.

4.6 Deep learning (DL)

(a) Convolution neural network (CNN)

Clanuwat (2018) discussed classification on early and modern Japanese characters and identified similarities of Kanji based on CNN. Two separate convolutional variational autoencoders were used. One on the Kuzushiji-Kanji dataset, and also a second on a pixel version of KanjiVG dataset rendered to 64x64 pixel resolution for consistency. Both datasets compressed into their own respective 64-dimensional latent space. Mixture Density Network (MDN) and Recurrent Neural Network (RNN) used to train the data. MDN is used with two hidden layers to model the density function. Sketch-RNN decoder modeled to generate Modern Kanji.

Guruviah et al. (2015) focused on improving OCR techniques with a combination of CNN for recognising ancient Tamil script that was used between 7^{th} and 12^{th} centuries. The input image is converted to gray-scale and the binarisation process is applied. Ostu's thresholding algorithm (Yousefi, 2011) is implemented to perform the binarisation that involves removing foreground text from the noisy background. Next, the binarised image is sliced into equivalent letter blocks each contacting Tamil characters. Then a set of cropped images are fed into CNN for image classification and detection. Transfer learning with data augmentation is used to train CNN to classify Tamil letters using Keras and TensorFlow. gTTS Python library used for the conversion of digitised text to an audio output. gTTS is a tool to provide an interface with Google Translates' text to speech API which supports automatic retrieval of a wide number of languages. Computed accuracy is recorded as 77.7%.

Liu and Gao (2018) investigated the recognition of oracle bone inscriptions (OBIs). 3,000 years ago, on cattle bones or turtle shells old Chinese characters were embossed and they are called OBIs. With the boom of deep learning in the current era, Liu and Gao (2018) used CNN to recognise OBIs. 8-layers are used to build the neural network. Manually scanned rubbed OBI images passed through a cascade of convolutional layers, where only filters of size 3×3 are employed. All convolutional strides are fixed to one. Gradually increased the number of feature maps from 64 (layer-1) to 512 (layer-4 and layer-5) after each max-pooling operation. The pooling operation is widely used to obtain shift invariance, which is important to OBI recognition. After five convolutional layers and four max-pooling layers, two fully connected layers, layer-6 and layer-7, are followed. The Stochastic Gradient Descent (SGD) technique is employed to train the CNN based model. Although the overall system obtained a high accuracy, it has certain limitations such as some OBIs cannot identify correctly.

(b) Artificial neural network (ANN)

Araokar (2015) has described an approach to recognise visual characters using ANN. In the paper, the author has embossed knowledge in pattern recognition and neural networking.

4.7 Other techniques

Mahalakshmi and Sharavanan (2013) discussed recognising and translating ancient Tamil inscriptions using LabView. As the initial step, characters are segmented from the stone inscriptions. Then the segmented images were enhanced. Followed by removal of the noise phase, recognition and translation of ancient characters were carried out. Ancient images from inscriptions were retrieved and recognised with the help of the segmentation process. Particle Swarm Optimisation (PSO) and DPSO (Darwinian PSO) algorithms were used for the segmentation process. Classification processes were carried out by the Fisher's linear discriminant algorithm. To enhance edges of letters, a contourlet transform method was used. Simulations were done using MATLAB and LabView.

Sigh (2016) explored a model that enabled an abstract representation of text written in Brahmi scripts. Stepped Distance Function and Dynamic Time Warping (DTW) algorithm utilised for the model.

Text Encoding Initiative (TEI) and eXtensible Markup Language (XML) were used by Bausi and Liuzzo to provide a clear XML representation of inscription information (Bausi and Liuzzo, 2018).

Papadaki et al. (2015) scanned damaged Greek inscriptions for revealing weathered letters using structured light scanning and Structure from Motion (SfM) technique.

5. Discussion

Inscriptions are foremost resources to provide information such as historical study of a civilisation, economy, cultural events, names and designations of royal officers, trade and industries, religious activities and beliefs throughout the world. Even though we are in a technological era, people have identified the value of inscriptions. So far, archeological departments of respective countries have established rules and regulations for the excavations and preservation of ancient inscriptions. Simultaneously, archeological experts, epigraphers and computational experts are developing systems to recognise and decipher inscriptions in an accurate manner. Inscriptions make a red carpet to boom tourism in many countries like Sri Lanka. In the name of socio-economic enhancements recognising and deciphering inscriptions are essential in this era.

Based on publications and discussions with archeological experts revealed that there are two major manual methods to translate inscriptions; (i) eye copy and (ii) estampage. These manual procedures own more barriers and disadvantages such as archeological expert necessity to visit the inscription site especially for eye copy method, inconvenient to reach the inscription due to height and other natural factors. Further, it completely depends on the human factors such as eyesight, specialised knowledge and experience to understand ancient scripts etc. In fact, an estampage procedure needs more equipment and materials.

Due to great effort, time consumption and lack of consistency of manual procedures, researchers have introduced computational methods. Among them, a great deal of systems have been established based on OCR. ZS Thinning algorithm, LW Thinning algorithm, WHF Thinning algorithm, Enhanced Parallel Thinning algorithm, Arabic Parallel Thinning algorithm and Matching algorithm can be used for the thinning phase. Quadratic Integral Ratio (QIR) algorithm and Ostu's algorithm utilised for thresholding. For classification supervised machine learning techniques, unsupervised machine learning and deep learning techniques have been mainly applied. According to these research, it can be seen that the quality of the original image is directly proportional to the final output. Hence, the results of partially damaged or erased inscriptions are substantially poor.

There are relatively few studies carried out to recognise inscriptions based on GIS. GPS for global positioning, 360° panoramas with 2D, 3D multimedia maps, annotated photographic pictures were used in GIS related studies. They connected geometrical information of inscription sites and images of inscriptions as 2D or 3D to the database.

Google's optical character recognition software recognises 248 languages. The tool preserves basic text formatting. Other text formatting and structuring elements get lost. Hence, it uplifts the faulty rate of output.

Currently, the vast majority of studies on recognising and deciphering epigraphs focus on machine learning approaches. When compared to other techniques machine learning approaches gain high accuracy. Especially neural networks related techniques obtain reliable output. However, existing methods failed to address the efficient and optimal solutions to recognise and decipher inscriptions.

Some inscriptions were damaged due to various reasons such as aging, civil war and natural disasters. Existing systems are not focused to predict text on partially damaged inscriptions.

The following taxonomy table shows comparative analysis on approaches that used to recognise and decipher inscriptions.

Table 1: Taxonomy table-recognising and deciphering inscriptions.

Title of Research	Hypotheses/Research Question/Research Objective	Technologies Used
Creation of precise alphabet fonts of early Brahmi script from photographic data of ancient Sri Lankan inscriptions (Bandara et al., 2012)	Producing accurate characters of early Brahmi scripts from photographic data	OCR has been used to identify Brahmi characters, Adobe Photoshop CS3 and MATLAB are used for image processing, to find the ideal shapes of Brahmi characters, Majority algorithm is used, to compare produced font images and original images used correlation function is used.
Script identification for tri-lingual image document (Dehiya and Verma, 2014)	Recognise scripts which contain Oriya, Telugu and Kannada letters	OCR is used to identify letters, for feature classification K-nearest neighbor (KNN) is used

Title of Research	Hypotheses/Research Question/Research Objective	Technologies Used
Thinning: A Preprocessing Technique for an OCR system for the Brahmi Script (Devi, 2006[a])	To present various algorithms may use in thinning as preprocessing technique	ZS Thinning algorithm, LW Thinning algorithm, WHF Thinning algorithm, Enhanced Parallel Thinning algorithm, Arabic Parallel Thinning algorithm and Matching algorithm
Thresholding: A pixel level image processing methodology preprocessing technique for OCR system for the Brahmi script (Devi, 2006[b])	To present various algorithms that may use in thresholding as a preprocessing technique	Quadratic Integral Ratio (QIR) algorithm and OTSU algorithm
Deep learning for classical Japanese literature (Clanuwat, 2018)	To classify early and modern Japanese characters and Identifying similarities of Kanji	Convolution Neural Network
A Roman to Devanagari Back-Transliteration algorithm based on Harvard-Kyoto convention (Nair and Sadasivan, 2019)	To recover the original Devanagari text which was transliterated into Roman text	UTF-16 encoding option, Hash Map and algorithm of Back-transliterate
A computational phonetic model for Indian language scripts (Singh, 2006)	To create computational phonetic model to enable an abstract representation of text written in Brahmi scripts	Stepped Distance Function, Dynamic Time Warping (DTW) algorithm
Inscriptions from Ethiopia: Encoding Inscriptions in Beta Masaheft (Bausi and Liuzzo, 2018)	To provide a clear XML representation of inscription information	Text Encoding Initiative (TEI), eXtensible Markup Language (XML)
The Design and implementation of AIDA: Ancient Inscription Database and Analytics System (Revez et al., 2019)	To store information related to ancient Minoan inscriptions	MYSQL, HTML, CSS, PHP
Exacting old Persian Cuneiform font out of noisy images (handwritten or inscription)(Mousavi and Lyashenko, 2017)	To recognise Persian Cuneiform characters from handwritten scripts and inscriptions	OCR, supervised and unsupervised machine learning algorithms for pattern recognition
Automatic writer identification of ancient inscription (Panagopoulos et al., 2009)	To classify Greek inscriptions according to author	Image extraction, pairwise comparison algorithm
Recognition of ancient Kannada epigraphs using fuzzy based approach (Soumya and Kumar, 2014)	To recognise Kannada text from ancient epigraphs	OCR, fuzzy logy
A review on automation of ancient epigraphical images (Preethi and Mamatha, 2016)	To prepare a review report on automated systems that use OCR for recognising epigraphical script images	OCR for identification and for classification Artificial Neural Network(ANN), clustering algorithm, K-nearest neighbor(KNN), Bayes classifiers and Support Vector Machine (SVM)

Title of Research	Hypotheses/Research Question/Research Objective	Technologies Used
A novel approach to OCR using image recognition based classification for ancient Tamil inscriptions in temples (Guruviah et al., 2015)	Improved OCR techniques to recognise ancient Tamil scripts	OCR, CNN
An architecture for an inscription recognition system for Sinhala epigraphy (Dias et al., 2016)	To recognise characters of inscriptions, store in a database and facilitate a 2D map for tracking inscription site	OCR (Ostu method for binarisation, OpenCV- Python)
Century identification and recognition of ancient Tamil character recognition (Rajakumar and Bharathi, 2011)	To recognise Tamil characters and century of stone inscriptions	Contour-let transform, clustering mechanism to recognise, fuzzy median filter for noise removal of image, artificial neural network for training and comparing data
Ancient Tamil script recognition from stone inscriptions using Slant Removal method (Rajakumar and Bharathi, 2012)	Improve offline ancient Tamil text lines recognition of stone inscriptions	Slant angle estimation technique for preprocessing, Cumulative Distribution Function (CDF) for core region detection
7th Century ancient Tamil character recognition from temple wall inscriptions (Rajakumar and Bharathi, 2012)	To recognise ancient Tamil characters of inscriptions which were written on temple wall	SIFT algorithm, K-means as clustering technique, support vector machine (SVM) as the classifier
Measurements and documentation of Buddhist stone inscriptions in China, Shandong (Schwing, 2007)	To measure, study and survey all Buddhist and stone inscriptions in China, Shandong region and create databases connected to geometrical information in 2D and 3D data	Geometric methods to get 2D and 3D geometrical data of stone inscription, GPS (Geological Positioning System), Orthorphotos and ortho-rubbings for copying text on inscriptions
Ancient Indian scripts image pre-processing and dimensionality reduction for feature extraction and classification: A survey (Tomar et al., 2015)	To conduct a survey on image pre-processing, segmentation, feature extraction and classification techniques via dimensionality reduction techniques	Filtering mechanisms such as ; (1)Spatial domain filters: mean filter, median filter, bilateral filter, (2) Frequency domain filters: Butterworth low pass filter, Gaussian low pass filter, Local thresholding: Niblack, Bernsen, Sauvola, Global thresholding: Otsu method Feature extraction and classification techniques: Principal component analysis, Kernel based principal component analysis, Independent component analysis
Interoperable integration of high precision 3D laser data and large scale geoanalysis in a SDI for Sutra inscriptions in Sichuan (China) (Lanig et al.)	To present concept for the integration of Buddhist inscriptions into a Spatial Data Infrastructure (SDI) using Open Geospatial Consortium (OGC) web services and link existing humanistic data to web based geographical information system	Terrestrial Laser Scanning (TLS), Open Geospatial Consortium (OGC)

Title of Research	Hypotheses/Research Question/Research Objective	Technologies Used
Optical character recognition for Brahmi script using Geometric method (Gautum and Chai, 2017)	To recognise Brahmi scripts	OCR, Geometric method for feature extraction
Oracle-Bone Inscription Recognition Based on Deep Convolutional Neural Network (Liu and Gao, 2018)	To recognise oracle bone inscriptions	Deep Convolutional Neural Network, Stochastic Gradient Descent technique to train CNN, KNN (k-nearest neighbors) and LR (Logistic Regression) as classifiers
Character recognition and period prediction of ancient Kannada Epigraphical scripts (Bhat and Achar, 2016)	To identify characters and era of stone inscriptions	OCR, Machine learning for classification
Accurate 3D scanning of damaged ancient Greek inscriptions for revealing weathered letters (Papadaki et al., 2015)	To reveal weathered letters from Greek inscriptions	Structured light scanning, Structure from Motion (SfM) technique
Information extraction and text mining of ancient Vattezhuthu characters in historical documents using image zoning (Vellingiriraj et al., 2016)	Develop a system to recognise Brahmi, Grantha and Vattezhuthu characters from palm manuscripts of Tamil ancient documents, analyse and translate to current Tamil digital text format	OCR, Zoning method for classification
A novel feature extraction and classification methodology for the recognition of historical documents (Vamvakas et al., 2009)	Reveal feature extraction and classification methods to recognise ancient documents	Structural feature extraction method, hierarchical classification scheme
A novel framework for 3D reconstruction and analysis of ancient inscriptions (Barmpoutis et al., 2009)	To perform 3D reconstruction and statistical techniques of ancient inscriptions	3D visualisation model for reconstruction on surface
GPU based optical character transcription for ancient inscription recognition (Mara et al., 2009)	Develop an automated system to transcript ancient Chinese inscriptions	Graphic Processing Units (GPU), Gaussian filter, FFT (Fast Fourier Transform) to remove noise of image, Compute Unified Device Architecture (CUDA) as image processing filter, Multi scale Vessel Enhancement (MVE) filter
Period prediction system for Tamil epigraphical scripts based on support vector machine (Subashini et al., 2011)	To predict time period of Tamil epigraphical scripts	Median filter technique to remove noise of images, Ostu's method for binarisation, vertical projection technique for segmentation, bilinear interpolation technique to scale segmented image, Support Vector Machine (SVM) for classification
A method for extracting text from stone inscriptions using character spotting (Aswatha et al., 2014)	To extract text from stone inscription images	Character spotting
Tamil character recognition from ancient epigraphical inscription using OCR and NLP (Manigandan, 2017)	To recognise various Tamil characters from inscriptions which were written between 9th and 12th centuries	OCR, NLP, contourlet transformation, Otsu method, SIFT algorithm, SVM classifier, Unicode mapping techniques

6. Conclusion

According to the literature survey, a great number of studies have been done to recognise and decipher inscriptions all over the world. As literature suggests, most research is based on OCR. Nevertheless, accuracy and reliability are comparatively low. Although researchers have utilised machine learning approaches, existing systems are not focused to predict text on partially damaged inscriptions. Researchers need to address those problems. Deep learning approaches with text embedding object detection algorithms like YOLO (You Only Look Once) may give better performance. Further, Recurrent Neural Network (RNN) may predict text which were erased or partially damaged inscriptions.

Acknowledgment

A sincere gratitude to Mr. Pathum Paranawitharana, Department of Archaeology, University of Sri Jayewardenepura, Sri Lanka and officials from Department of Archaeology, Sri Lanka for providing historical information.

References

Amato, G., Falchi, F. and Vadicamo, L., 2016. Visual recognition of ancient inscriptions using convolutional neural networks and fisher vectors. *Journal on Computing and Cultural Heritage*, 9:1-24.

Arnold, M., 2008. Buddhist stone scriptures from Shandong, China. *Electronic Visualisation and the Arts (EVA 2008)*, pp.1-8.

Aswatha, S.M., Talla, A.N., Mukhopadhyay, J. and Bhowmick, P., 2014. A method for extracting text from stone inscriptions using character spotting. In A*sian Conference on Computer Vision,* 598-611, Singapore. Springer.

Archaeology Glossary of Technical Terms, 2012. Official languages department of Sri Lanka.

Bandara, P.B.A.D.W., 2014. *Archaeological Study of Sri Lankan Brahmi Inscriptions by using Digital Image Analysis* (Doctoral dissertation).

Bandara, D., Warnajith, N., Minato, A. and Ozawa, S., 2012. Creation of precise alphabet fonts of early Brahmi script from photographic data of ancient Sri Lankan inscriptions. *Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition*, 3:33-39.

Bandara, R., 2008. Abilekana, Sri Lanka.

Bausai, A. and Liuzzo, P.M., 2018. Inscriptions from Ethiopia. Encoding Inscriptions in Beta Maṣāḥəft. 85-92.

Chandrakala H.T. and Thippeswamy, G., 2016. A comprehensive survey on ocr techniques for kannada script, *International Journal of Science and Research*, 5:35-39.

Clanuwat, T., Kitamoto, A., Yamamoto, K., Lamb, A. and Ha, D., 2018. Deep learning for classical Japanese literature, In: *32nd Conference on Neural Information Processing Systems*. Montreal, Canada.

Devi, H., 2006. Thresholding: A Pixel-Level image processing methodology preprocessing technique for an OCR system for the Brahmi script. *Ancient Asia*, 1.

Dhammananda, M., 2018. Sinhala Silalipimala (part 1 and 2), Sri Lanka.

Dias, M. ed., 1991. *Epigraphical Notes, Nos. 1-18*. Department of Archaeology, Sri Lanka.

Guruviah, V., Giridhar, L. and Dharani, A., 2015. A Novel Approach to OCR using Image Recognition based Classification for Ancient Tamil Inscriptions in Temples.

Giridhar, L., Dharani, A. and Guruviah, V., 2019. A novel approach to ocr using image recognition based classification for ancient tamil inscriptions in temples. *arXiv preprint arXiv:1907.04917*.

Hettiarchchi, K. and Dayananda, S., 2018. RajagalaSellipi, Department of Archeology of Sri Lanka.

Hillman, T., 2011. The inscription, translation and re-inscription of technology for mathematical learning. *Technology, Knowledge and Learning*, 16:103.

Inscription Analyser, http://kissel.base.ibaraki.ac.jp/intl/apps/inscription/, Accessed 17 January 2020.

Inscriptions of Ceylon, 2007. Department of Archaeology of Sri Lanka.

Karunarathne, K.G.N.D., Liyanage, K.V., Ruwanmini, D.A.S., Dias, G.K.A. and Nandasara, S.T., 2017.

Recognising ancient sinhala inscription characters using neural network technologies. *International Journal of Scientific Engineering and Applied Sciences*.

Khan, A. and Vaidya, U., 2019. Inscriptions: As a source of history.

Kivashilava, G., 2019. On the decipherment of linear inscriptions in the common colchian language: mathematics, international science and technology conference, Berlin, Germany, pp 856-864.

Lanig, S., Schilling, A., Auer, M., Höfle, B., Billen, N. and Zipf, A., 2011. Interoperable integration of high precision 3D laser data and large-scale geoanalysis in a SDI for Sutra inscriptions in Sichuan (China). *Geoinformatik*, 79-84.

Liu, G. and Gao, F., 2018. Oracle-bone inscription recognition based on deep convolutional neural networks. *Journal of Computers*, 13:1442-1450.

Mahalakshmi, M. and Sharavanan, M., 2013, April. Ancient tamil script and recognition and translation using LabVIEW. In: *2013 International Conference on Communication and Signal Processing* (pp. 1021-1026).

Manigandan, T., Vidhya, V., Dhanalakshmi, V. and Nirmala, B., 2017, August. Tamil character recognition from ancient epigraphical inscriptions using OCR and NLP. In: *International Conference on Energy, Communication, Data Analytics and Soft Computing*, pp.1008-1011.

Mara, H., Hering, J. and Kromker, S., 2009, September. GPU based optical character transcription for ancient inscription recognition. In: *2009 15th International Conference on Virtual Systems and Multimedia*, pp. 154-159. IEEE.

Mousavi, S.M.H. and Lyashenko, V., 2017, November. Extracting old persian cuneiform font out of noisy images (handwritten or inscription). In: *2017 10th Iranian Conference on Machine Vision and Image Processing (MVIP)*, pp. 241-246. IEEE.

Nair, J. and Sadasivan, A., 2019, March. A Roman to Devanagari Back-Transliteration Algorithm based on Harvard-Kyoto Convention. In: *2019 IEEE 5th International Conference for Convergence in Technology (I2CT)*, pp. 1-6. IEEE.

Nandasara, S.T. and Mikami, Y., 2015. Bridging the digital divide in Sri Lanka: Some challenges and opportunities in using Sinhala in ICT.

Ozawa, S., Dahanayaka, D.D.G.L., Warnajith, N. and Quarmal, S.B., 2017. Renovation of research and education in south asian countries by means of ICT.

Panagopoulos, M., Papaodysseus, C., Rousopoulos, P., Dafi, D. and Tracy, S., 2008. Automatic writer identification of ancient Greek inscriptions. *IEEE transactions on pattern analysis and machine intelligence*, 31:404-1414.

Panigrahi, S., Google's optical character recognition (OCR) software works for 248+ languages. opensource.com/life/15/9/open-source-extract-text-images. Accessed 6 February 2020.

Papadaki, A.I., Agrafiotis, P., Georgopoulos, A. and Prignitz, S., 2015, February. Accurate 3D scanning of damaged ancient Greek inscriptions for revealing weathered letters. In: *International Workshop on 3D Virtual Reconstruction and Visualisation of Complex Architectures, 3D-ARCH 2015*.

HR, M., 2016. A review on automation of ancient epigraphical images. *International Journal of Database Theory and Application*, *9*:143-150.

Priyanka, B., 2015. Lakdiwa sellipi asurion sakas karana lara daththa (1, 2, 3 A.C.), Sri Lanka.

Rajakumar, S. and Bharathi, V., 2012. Ancient tamil script recognition from stone inscriptions using slant removal method. In: *International Conference on Electrical, Electronics and Biomedical Engineering*. Penang, Malaysia, pp. 80-84.

Rajakumar, S. and Bharathi, V.D.S., 2012. 7th Century Ancient Tamil Character Recognition from Temple Wall Inscriptions. *Indian Journal of Computer Science and Engineering*, 3.

Rajakumar, S. and Bharathi, V.S., 2011. Century identification and recognition of ancient Tamil character recognition. *International Journal of Computer Applications*, 26:32-35.

Ranawalla, S., 2004. Sinhala sellipi wadan akaradiya, Department of Archaeology of Sri Lanka.

Revesz, P.Z. and Rashid, M.P., Tuyishime, Y., 2019. The design and implementation of AIDA: Ancient inscription database and analytics system, IDEAS'19, Athens, Greece.

Roth, W.M. and McGinn, M.K., 1998. Inscriptions: Toward a theory of representing as social practice. *Review of Educational Research*, 68:35-59.

Ruwanmini, D.A.S., Liyanage, K.V., Karunarathne, K.G.N.D., Dias, G.K.A. and Nandasara, S.T., 2016. An architecture for an inscription recognition system for sinhala epigraphy. *International Journal of Research-Granthaalayah*, 4:48-64.

Schwing, T., 2007. Measurements and documentation of buddhist stone inscriptions in China, Shandong. building information systems and GIS applications.

Sirimunasinghe, 2012. The script of the Sinhalese. https://sirimunasiha.wordpress.com/2012/01/02/the-script-of-sri-lanka, Accessed 20 December 2019.

Soumya, A. and Kumar, G.H., 2014, November. Recognition of ancient Kannada Epigraphs using fuzzy-based approach. In: *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 657-662. IEEE.

Subashini, P., Krishnaveni, M. and Sridevi, N., 2011, March. Period prediction system for tamil epigraphical scripts based on support vector machines. In: *International Conference on Information Systems for Indian Languages*, pp. 23-30. Springer, Berlin, Heidelberg.

Tomar, A., Choudhary, M. and Yerpude, A., 2015. Ancient Indian Scripts Image Pre-Processing and Dimensionality Reduction for Feature Extraction and Classification: A Survey.In: *International Journal of Computer Trends and Technology (IJCTT)*, 21:101-124.

Yousefi, J., 2011. Image binarisation using Otsu thresholding algorithm. *University of Guelph, Ontario, Canada*.

Frangi, A.F., Niessen, W.J., Vincken, K.L. and Viergever, M.A., 1998, October. Multiscale vessel enhancement filtering. In: *International conference on medical image computing and computer-assisted intervention*, pp. 130-137. Springer, Berlin, Heidelberg.