A Method for Extracting Text from Stone Inscriptions using Character spotting

Shashaank M. Aswatha, Ananth Nath Talla, Jayanta Mukhopadhyay, and Partha Bhowmick

Indian Institute of Technology Kharagpur, India

Abstract. A novel interactive technique for extraction of text characters from the images of stone inscriptions is introduced in this paper. It is designed particularly for on-site processing of inscription images acquired at various historic palaces, monuments, and temples. Its underlying principle is made of several robust character-analytic elements like HoG features, vowel diacritics, and location-bounded scan lines. Since the process involves character spotting and extraction of the inscribed information to editable text, it would subsequently help the archaeologists for epigraphy, transliteration, and translation of rock inscriptions, particularly for the ones having high degradations, noise, and a variety of styles according to the mason origin and reign. The spotted characters can also be used to create a database for ancient script analysis and related archaeological work. We have tested our method on various stone inscriptions collected from some of the heritage sites of Karnataka, India, and the results are quite promising. An Android application of the proposed work is also developed to aid the epigraphers in the study of inscriptions using a tablet or a mobile phone.

1 Introduction

Many of the stone inscriptions that are found across different regions of the world reveal the details of extravagance, lifestyle, economic condition, culture, and also of the administrative regulations followed by various rulers and dynasties particular to those regions. The information gained from these inscriptions can be corroborated with the information from other sources, in order to provide an insight into world's dynastic history, which otherwise lacks the completeness of contemporary historical records. Epigraphists use this information to identify the graphemes, clarify their meaning, classify their uses according to dates and cultural contexts, and to draw conclusions about the writing and its writers. Texts inscribed on stone are usually put up for public view to exhibit different cultures that prevailed during the period of inscription.

The inscriptions considered in our work are from Indian subcontinent. These Indic inscriptions have a composite mix of characters that evolved during the reign of several dynasties and kingdoms. They are usually found to be engraved on a variety of stones and other durable materials. Conventionally, they are studied offline by generating *estampages* of the inscription surface. For this, the surface of the stone inscription is first cleaned with water-soaked brush. Then, the stone surface is carpeted by a large piece of wet paper (or layers of paper), which is gently patted by a dabber made of some soft material. The dabber is smeared with Indian ink to get the impression of the surface. The paper is allowed to dry on the stone surface before taking it off. The ink impression (estampage) comes out as white letters (grooves of the characters) against black background on the paper. Epigraphers usually take several days to few weeks for reading, transliterating, and for translating these estampages.

With elapsing time, these inscriptions are gradually deteriorating to a undecipherable state. Although estampages are taken for many of them, it is very difficult to preserve these estampages, as they fade away very soon. Frequent generation of estampages would cause the inscription to degrade more, since it involves a physical activity on the inscription surface. Hence, with advancing technology, various imaging techniques have emerged for acquiring images with considerable originality and economical viability. However, extraction and processing of information and text from these images is a challenging problem due to various factors, such as uncontrolled illumination, multilingual text, low-contrast distinction between the groove and the surrounding surface, distortions due to perspective projection, and administrative constraints in using imaging devices at heritage sites.

2 Related Work and Our Contribution

There exists a series of research work on historical document processing, which deals mainly with preprocessing, word spotting, classification, and optical character recognition (OCR) [1, 2]. However, a little has been done so far to address the problem of extracting text and symbols from age-old inscriptions having historical importance. Moreover, the problem complexity and the related issues are often different in case of recognizing inscribed texts, as evident from the related literature [3–12]. Hence, for binarization of inscription images, an exclusive method is proposed recently in [12]. As the text in an inscription has a chiseled and engraved effect, which has often a degraded form owing to erosion through centuries, high-precision 3D measuring techniques are found to be useful, as shown in [3, 7]. Although these 3D techniques help in producing almost exact copies of the original inscriptions with an objective of processing them for better legibility, they are both expensive and time-consuming. To achieve a better processing speed, a GPU-based method for optical character transcription is proposed in [6], which is focused only on inscriptions written in a script that is highly structured in both horizontal and vertical directions. To identify the dating of a stone inscription by identifying its writer, other methodologies can be seen in [4, 5, 8, 9, 11]. In [5], enhancement of inscription images for recognizing the text using OCR is performed using *natural gradient based flexible ICA* (NGFICA). This is carried out by grouping the inscriptions by their inscribers [11]. A technique for identifying the consonants of a language from an inscrip-

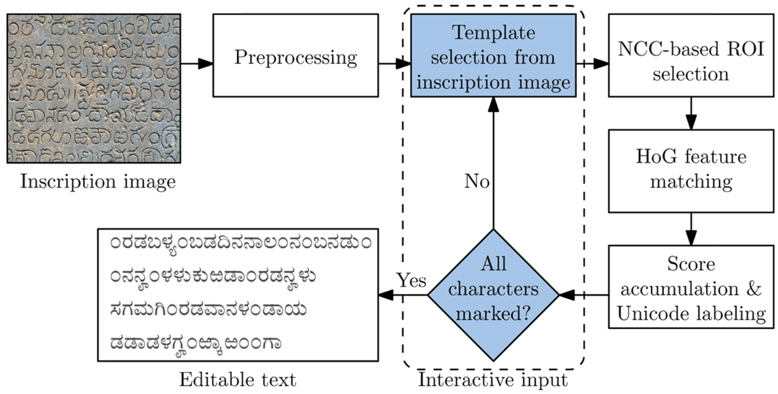

Fig. 1: The proposed character spotting algorithm. Note that the output editable text contains the modern Kannada characters equivalent to the ones found in the inscription image.

tion image by finding the feature vectors and classifying the characters based on SVM classifier, is also reported in [10].

This paper proposes an on-site semi-automatic method to extract text from the images of stone inscriptions using a novel technique of character spotting[1]. The proposed technique for character spotting uses *histogram of oriented gradients* (HoG) [13] as feature descriptor, and it does not use any classifier or training dataset. A labeling is made by the user through an interactive process, which is the input to subsequent processes. This requires a knowledgeable user, who is acquainted with the script, to convert the inscription to an editable text document, which can be used for further analysis. Since the marked and the spotted characters exist on the same surface of an inscription, the spotting process is robust to writing style and font, with a reasonable presumption that the entire inscription has a uniformity of script. The editable text can later be used for different analysis purposes like epigraphy, translation, and transliteration. The method accommodates human expertise in the process of delivering quality output.

The rest of the paper is organized as follows. The proposed method is explained in Section 3. We have tested our method on various stone inscriptions collected from the heritage sites of Karnataka, India. Some of these test results and our experimental setup are presented in Section 4. Finally, we conclude the paper in Section 5.

[1] Patent pending: System and method for converting substrate inscription into electronically editable format; patent filing reference no.: 452/KOL/2014, April 2014.

Fig. 2: An inscription image (cropped) with character spotting result (left: highlighted in yellow) corresponding to a template image (middle) having the HoG visualization as shown (right).

3 The Character Spotting Technique

The character spotting algorithm is shown in Figure 1. The image of an inscription is first preprocessed for contrast adjustment and denoising. Then the user starts by selecting a character in the inscription by specifying a rectangular box around the character (henceforth referred to as *template*) and selects the corresponding Unicode related to its equivalent modern character as input. Figure 2 shows an example of the *search* image and the *template* image. Iteratively, the first occurrence of each character is marked by the user, which is the spotted throughout the inscription. If any character is missed in the spotting process, it can be marked again in subsequent trials. The template is divided into equally sized N overlapping parts, whose scores of the *normalized cross correlation (NCC)* are accumulated at the center region of the original image portion corresponding to the template image fragments. For a template size equivalent to 100×100, $N = 8$ is observed to yield the desired result for our data.

The resulting correlation surface is thresholded to obtain the regions of interest (ROI) of possible character occurrences in the original image. All spotted characters from each of the iterations are marked using a distinct color. The above process is repeated until all the characters are marked. The process of character spotting involves manual interaction to choose the Unicodes of the characters that are being marked. The spotted characters and their locations are saved in a text file. The last stage of the character spotting process involves reading the spotted character positions and their Unicodes from the text file and generating an editable text document of the inscription in required script.

3.1 Feature Points and Thresholding

The normalized cross correlation is carried out in parts. Let I_s be the stone inscription image and I_t be the character template selected from I_s. Let I_{tp} be one of the overlapping parts of I_t. At each pixel location (x, y), $I_{st}(x, y)$ is

Fig. 3: Overlapping parts for the correlation corresponding to the template shown in Figure 2.

an image patch in I_s, centered at (x, y), confined under the character template fragment when I_{tp} is placed at that location. By performing correlation of all the parts, we compute the character centers in I_s, which are similar to the character in the template image I_t. The template image, which is divided into 8 overlapping parts, is shown in Figure 3. The normalized cross correlation at each of the pixel locations, for each of the parts is given by

$$\gamma(x,y) = \frac{\sum (I_{st}(x,y) - \overline{I_{st}(x,y)})(I_{tp} - \overline{I_{tp}})}{\sqrt{\sum (I_{st}(x,y) - \overline{I_{st}(x,y)})^2 (I_{tp} - \overline{I_{tp}})^2}},$$

where, $\overline{I_{tp}}$ is the constant-intensity image with the mean value of pixel intensities in I_{tp}, and $\overline{I_{st}(x,y)}$ is the similar one corresponding to $I_{st}(x,y)$. The maximum in the window of radius 4 is computed at each pixel location of the correlated image for each template fragment. The correlation values of all the parts are aggregated to get the correlation result for the entire template. The correlation result is then thresholded based on a threshold value, t_c. For the data considered in our work, the value of t_c has been empirically set to 30% of the number of fragments. After thresholding, we obtain points where the possible occurrences of the template exist. These points are the probable centers of the character in the template image. Figure 4 shows the correlation result and the thresholded correlation image.

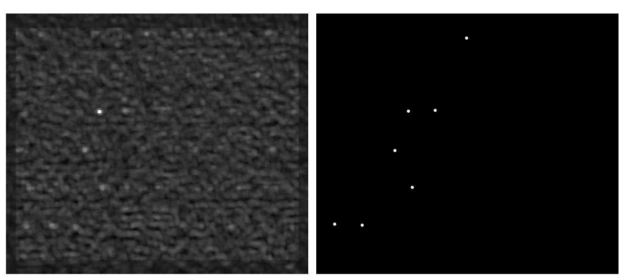

Fig. 4: Correlation result (left) and its thresholded image (right) corresponding to the inscription image and the template shown in Figure 2.

3.2 HoG Feature Description and Matching

HoG features capture the local edge information, with some tolerance to deformation, using the distribution of intensity gradients [13]. The local shape information gets captured even without precise information about the location of the edges. The HoG feature is computed by dividing the image region into small non-overlapping regions, called *cells*. A histogram of gradient directions or edge orientations is computed for each cell. Overlapping neighboring cell histograms are combined to form blocks, which are individually normalized to unit magnitude. A combined histogram entry from all the blocks is used as the feature vector describing the image region. HoG features are used in our work accounting to their relative invariance to local geometric and photometric transformations. The key points (character centers) that are identified by correlation results are considered to prune the search space for spotting the marked character. The character spotting is performed by describing the image regions of possible occurrences at each of the key points by patch descriptors. Similar to the correlation process, a neighborhood of the size of ROI is considered, which is described using HoG feature descriptor. To take into account the partial deterioration of the characters in stone inscriptions, the matching is performed by computing the scalar product of corresponding normalized feature vectors of each of the blocks. The mean value of all the block feature projection scores is accumulated to obtain a final score of the matching. If the template image is described by M blocks, then an identical image patch in the search image would have a maximum matching score of M. These mean scores for all the possible regions of occurrences are thresholded by a threshold value, t_f, to spot the selected character. The threshold is empirically set to 80% of the number of blocks in the template image. Figure 2(right) shows the HoG descriptor for a template image.

For computing HoG features, we have chosen a cell size of 8×8 pixels, the number of orientation bins as 9, and the block size to be of 32×32 pixels (that is, 4×4 cells). The input image is preprocessed using *non-local means (NLM)* denoising process [14]. It is found that application of Gaussian smoothing before applying the derivative mask did not improve the results, but yielded relatively poorer results in matching similar patches. Characters identified by the HoG matching, for each template, are displayed in a distinct color, as shown in Figure 2. For every spotting result, respective template size, spatial location of the spotted character center, matching score, and Unicode of the character in the template image are written to a text file.

This process of character spotting continues iteratively, until all the characters are spotted. Figure 5 shows an example of the search image (partial), after spotting 26 character templates.

3.3 Working with Vowel Diacritics

A *diacritic* is a glyph added to a letter. It is a modifier that changes the pronunciation of a consonant by associating either a consonant or a vowel to the base

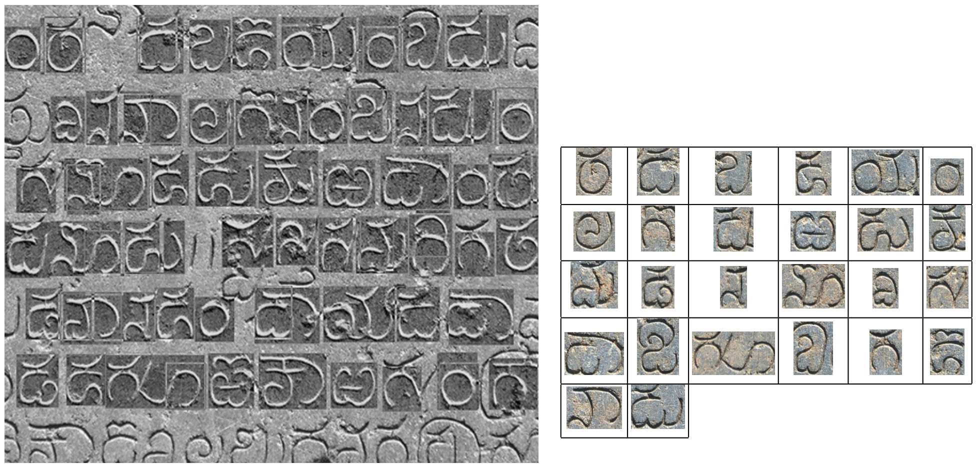

Fig. 5: Partial character spotting result for the inscription image shown in Figure 2 corresponding to 26 templates.

character. In general, diacritical marks appear above or below a consonant, or in some other position, such as within the letter or between two letters. However, in Kannada script, the diacritics can appear above, below, before, or after the consonants. Figure 6 shows four typical instances of vowel diacritics of consonants in Kannada.

To resolve the ambiguity in character diacritics, the matching scores and the areas associated with the character templates are considered. When multiple templates are found to match with the same character in the search image, the areas of the corresponding template images are used to decide the best. If the maximum area is at least 1.5 times the second maximum, then the one associated with the maximum area is considered and the rest is discarded. Otherwise, the template associated with the maximum score is considered. Figure 7 shows an example of spotting a Kannada consonant ‘na’ (template). The spotted character is a part of consonant’s vowel combination, ‘nu’, shown in blue rectangle. Here, matching scores are used to resolve the diacritic ambiguity.

3.4 Text Generation

The information about the spotted characters, which are obtained from the above process, contains the locations of the characters in the image. During the spotting process, the locations of the spotted characters along with the Unicodes

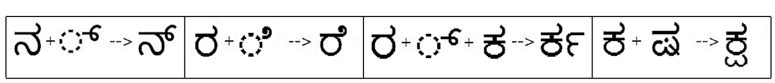

Fig. 6: Four typical cases of vowel diacritic of a consonant.

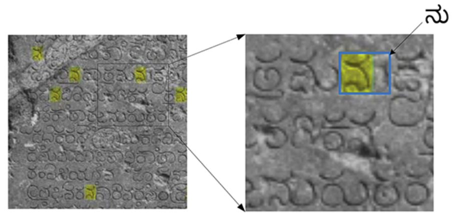

Fig. 7: An example of ambiguity during character spotting.

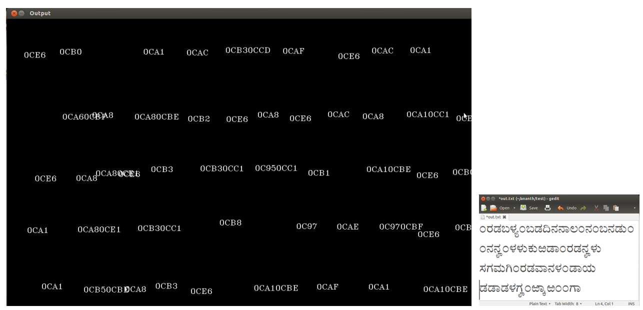

Fig. 8: An image with Unicodes at the locations of their respective characters (left) and the corresponding text image (right).

are stored in a text file. After completely spotting all the characters in the image (which is inspected manually), the text file contains all the locations and Unicodes of respective characters in the image. The entries of character centers in the text file are almost in a sorted order, in terms of the coordinates of the character centers in the search image. However, due to the complexity and non-uniformity involved in the script, the characters centers may not always lie exactly in the same line as in the search image. Figure 8 shows an image with the character centers along with their corresponding Unicodes. So, these character centers are sorted in row-major (i.e., left-to-right and top-to-bottom) order. For each character center (Figure 9), the upper boundary row index and the lower boundary row index, denoted by a and b respectively, are computed from the dimensions of the character. For the top-left character center $p(x, y)$, which serves as the *anchor point*, all the centers whose row index is less than b are considered, and the row indices of these centers are assigned the same index as p. The process is iterated with subsequent anchor points, as shown in Figure 9. After rearranging the character centers, the Unicodes corresponding to the sorted character centers are updated in the text file. Finally, these Unicodes are parsed to generate the inscription text.

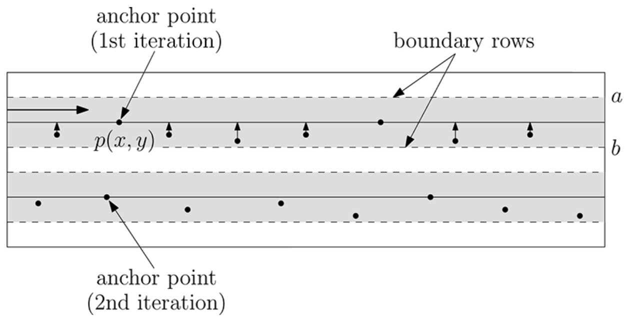

Fig. 9: Location-bounded scan for sorting the lines.

4 Experimental Results

The proposed character spotting method is tested on the images of different stone inscriptions collected from various parts of Karnataka, India. Stone inscriptions made by various dynasties that ruled over Karnataka display unique features in their style with reference to the type of stone, polishing, composition of text, inscribing on stone with color, engraving the text on stone, and also based on the position of erecting the stone in an appropriate place. Many of these inscriptions are deteriorated so badly that it is difficult to identify the meaningful data, particularly when the surface is broken or etched. Due to centuries of deterioration, majority of these ancient texts are in very poor condition, and many text portions are already missing. The damage has occurred to such an extent that either the fragments do not exist, or sections are no longer recognizable and beyond recovery. The performance result of the character spotting process on such images is also reported in this section.

4.1 Experimental Setup

A group of volunteers consisting of Sanskrit scholars, Kannada scholars, regional journalists, and archaeological students were identified to put the tool into use. These volunteers belong to different age groups and sects, and have knowledge of more than three regional dialects. A questionnaire form was given to each of the volunteers to identify the background of the volunteer. Feedbacks of the spotting tool by the volunteers are collected. It is found that the tool is indeed useful for extracting the text from stone inscriptions. Also, the volunteers have agreed that this tool will considerably reduce the effort and interpretation time in deciphering the inscriptions. Some of the inscriptions that were considered in our experiments, procured from various places in Karnataka, India, are mentioned below.

1. *Arasikere Temple, Karnataka:* Stone inscriptions found in Arasikere temples are characterized by artistic writing on huge slabs with small engravings. So, the images of such inscriptions were taken by us in slices. Figure 10

(a) Arasikere temple
(b) Chitradurga fort
(c) Virupaksha temple
(d) Shimoga wooden palace

Fig. 10: Some of the test images of stone inscriptions collected by us from different heritage sites of Karnataka, India.

shows the image slice of a stone inscription of size 7098×1114 on which the character spotting method was applied. This image has a non-uniform noise distribution owing to stone texture, which is difficult to separate from the foreground characters. The performance of character spotting on this image is shown in Table 1.

2. *Chitradurga Fort, Karnataka:* The image in Figure 10 is of an inscription in the fort of Chitradurga, Karnataka. The inscriptions found in this fort are made of black granite—a homogeneous rock that changes its chemical properties according to climatic conditions. The character spotting process is executed in the top segment of the inscription, and the corresponding result is shown in Table 1.
3. *Virupaksha Temple, Hampi, Karnataka:* Virupaksha temple in Hampi is a part of the group of monuments at the sites of Hampi, designated by UNESCO as one of the world heritage sites. Figure 10 shows one of the inscriptions found at the temple entrance. The performance result on the image of this inscription is shown in Table 1.
4. *Shimoga Wooden Palace, Karnataka:* Shimoga is a city in the central part of the state of Karnataka, India. The inscriptions found in the palace are not clear as there is a little difference between the characters and the background. Many of them are almost faded away due to constant exposure to climatic variations. Figure 10 shows an inscription found in the palace, and the performance of our algorithm on this is shown in Table 1.

4.2 Desktop Implementation

The spotting algorithm is implemented in C++ under Ubuntu 12.04 environment, using OpenCV [15] and Qt standard libraries. In all the experiments, inscriptions in Kannada script were considered. These images were acquired by high-resolution cameras (8 to 18 mega pixels) in daylight under auto settings without any external illumination. Some are sheltered indoor, while majority of them are kept in open grounds. The modern Kannada characters are displayed by using *Baraha* font library.

4.3 Performance Analysis

The performance of the spotting method is analyzed by estimating the measures of *sensitivity*, *specificity*, *positive predictive value* (PPV), and *negative predictive value* (NPV). For the spotting result of each character, we calculate the number of *true positives* (TP), *false positives* (FP), *true negatives* (TN), and *false positives* (FN), which correspond to correctly spotted, incorrectly spotted, correctly rejected, and incorrectly rejected characters, respectively. The ground truth for each of the inscriptions is obtained by the experts of Archeological Survey of India (ASI) and manual inspection.

Sensitivity is the ability of the method to correctly spot a character, given a similar character as input. Specificity is defined as the proportion of other characters (excluding the selected character) that are not correctly spotted. Precision is the fraction of the spotted characters that are same as the input character. The mathematical expressions of these parameters are as follows.

Table 1: Desktop performance (in %) and average CPU time (per spotting per template) for the inscription images shown in Figures 10.

Performance Measure	Fig 10a	Fig 10b	Fig 10c	Fig 10d
Sensitivity	86.05	73.82	68.27	52.50
Specificity	99.14	84.31	84.12	92.52
Precision	78.88	43.04	30.39	51.59
Accuracy	99.00	83.92	83.01	90.74
CPU time (sec.)	1.10	0.72	0.76	0.88

$$Sensitivity\ (True\ Positive\ Rate) = \frac{TP}{(TP+FN)}$$

$$Specificity\ (True\ Negative\ Rate) = \frac{TN}{(FP+TN)}$$

$$Precision = \frac{TP}{TP+FP}$$

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

Arasikere inscription 1.10 seconds Chitradurga inscription 0.72 seconds Viroopaksha inscription 0.76 seconds Shimoga inscription 0.88 seconds

It is observed from our experiments that, on varying the thresholds t_c and t_f (Sections 3.1 and 3.2), the results also vary to a great extent. If t_c and t_f are decreased from the chosen values, then the number of false positives increases. Since the search space also increases with decreasing thresholds, the time taken for execution increases. If t_c and t_f are increased beyond the chosen values, then the number of false negatives increases. In the latter case, since the search space decreases, the execution time also decreases. These thresholds are chosen empirically.

4.4 Android Implementation

The proposed method of character spotting presented and its related implementation with an appropriate interface caters the historical information to an archaeologist, who already has a database of stone inscription images taken by professional photographers. Wide adoption of mobile devices, especially smart phones with the app-store mobile application distribution model, supports an archaeological group to work out many problems on-site. Hence, we have also made an implementation of our method in mobile devices and tablets to initiate the interpretation process on-site, without necessarily requiring off-line workspace. The hierarchy of activity calls in our Android application is shown in Figure 11.

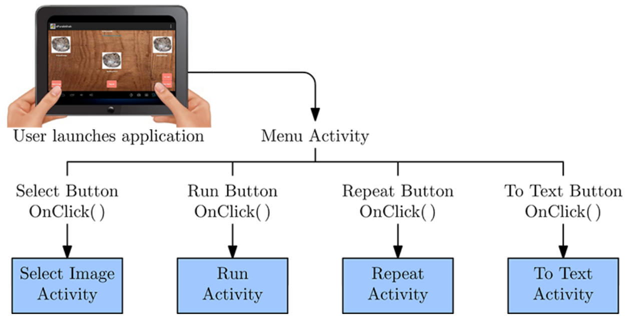

Fig. 11: Android application activity calls.

5 Conclusion

In this paper, we have proposed a useful interactive tool for epigraphers to read and archive ancient inscriptions in a convenient way. It can be used to replace the tedious task of obtaining estampages from stone inscriptions by ink-smeared manual dabbers, as followed in the conventional practice. The Android application developed by us can be used by the epigraphers and historians to analyze and interpret the inscriptions on-site. The character spotting results can be used to create a dataset of various characters of the concerned language, which would be helpful in studying the hierarchy of evolution of the script. This dataset can also be used to train a classifier for recognition process, which is in the purview of our future work.

Acknowledgement. This work is partially sponsored by Department of Science and Technology, Govt. of India through sanction number NRDMS/11/1586/2009.

References

1. Vamvakas, Gatos, Stamatopoulos, Perantonis: A complete optical character recognition methodology for historical documents. In: International Workshop on Document Analysis Systems. (2008) 525–532
2. Kim, S.K., Sin, B.K., Lee, S.W.: Character spotting using image-based stochastic models. In: International Conference on Document Analysis and Recognition. (2001) 60–63
3. Barmpoutis, A., Bozia, E., Wagman, R.: A novel framework for 3d reconstruction and analysis of ancient inscriptions. Machine Vision and Applications **21** (2010) 989–998
4. Galanopoulos, G., Papaodysseus, C., Arabadjis, D., Exarhos, M.: Exploiting 3d digital representations of ancient inscriptions to identify their writer. In: Advances in Visual Computing. (2012) 188–198
5. Shridevi, I.: Enhancement of inscription images. In: National Conference on Communications. (2013) 1–5
6. Mara, H., Hering, J., Kromker, S.: GPU based optical character transcription for ancient inscription recognition. In: International Conference on Virtual Systems and Multimedia. (2009) 154–159
7. Schmidt, N., Boochs, F., Schütze, R.: Capture and processing of high resolution 3D-data of Sutra inscriptions in China. In: Digital Heritage. (2010) 125–139
8. Papaodysseus, C., Rousopoulos, P., Arabadjis, D., Panopoulou, F., Panagopoulos, M.: Handwriting automatic classification: Application to ancient greek inscriptions. In: International Conference on Autonomous and Intelligent Systems. (2010) 1–6
9. Papaodysseus, C., Rousopoulos, P., Giannopoulos, F., Zannos, S., Arabadjis, D., Panagopoulos, M., Kalfa, E., Blackwell, C., Tracy, S.: Identifying the writer of ancient inscriptions and Byzantine codices. A novel approach. Computer Vision and Image Understanding **121** (2014) 57–73
10. Rajakumar, S.: Eighth century Tamil consonants recognition from stone inscriptions. In: International Conference on Recent Trends in Information Technology. (2012) 40–43
11. Rousopoulos, P., Panagopoulos, M., Papaodysseus, C., Panopoulou, F., Arabadjis, D., Tracy, S., Giannopoulos, F., Zannos, S.: A new approach for ancient inscriptions' writer identification. In: International Conference on Digital Signal Processing. (2011) 1–6
12. Shaus, A., Turkel, E., Piasetzky, E.: Binarization of first temple period inscriptions: Performance of existing algorithms and a new registration based scheme. In: International Conference on Frontiers in Handwriting Recognition. (2012) 645–650
13. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: International Conference on Computer Vision and Pattern Recognition. (2005) 886–893
14. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: International Conference on Computer Vision and Pattern Recognition. (2005) 60–65
15. OpenCVDocumentation: (OpenCV documentation) Accessed: 2013-11-10 `http:docs.opencv.orgmodulesimgprocdocobject_detection.html`.