

Computer Analysis of Photographic Data of Sri Lankan Early Brahmi Inscriptions

Nalin Warnajith¹, Dammi Bandara^{1,2}, Sarkar Barbaq Quarmal¹, Masanori
Itaba¹, Atsushi Minato¹ and Satoru Ozawa¹

¹ (Graduate School of Science & Engineering, Ibaraki University, Hitachi 316-8511, Japan)

² (Department of Archaeology, Colombo 7, Sri Lanka)

Abstract: Ancient Sri Lankan documents inscribed on stone surface are usually contaminated with various kinds of noises such as scratches, cracks, voids, etc. In authors' previous paper, ideal shape of Brahmi letters has been studied by a digital method and some alphabet fonts of early Brahmi scripts have been created from the analysis of a lot of photographic data of ancient Sri Lankan inscriptions of the same period in the history. The present paper deals with a precise method of identification of ancient letters with the aid of the alphabet fonts, which leads to automatic reading of ancient inscriptions by computers.

Keywords: *alphabet fonts, archaeology, Brahmi inscriptions, image analysis*

I. INTRODUCTION

Written sources have been played main roles in studies of history of any countries. Ancient Sri Lankan people used stone surfaces for writing letters. The documents inscribed on stone surfaces are called "inscriptions" [1, 2]. The first work for a newly unearthed inscription is the identification of letters. The stone surface is usually damaged by various kinds of noises such as scratches, cracks, voids, etc. [3, 4]. There found different shapes for the same letter depending on the condition of the stones and the tool of inscribing [5]. The aim of this paper is to obtain a better method of computer-aided identification of ancient Sri Lankan letters found in inscriptions. In authors' preliminary work [6], alphabet fonts of early Brahmi letters have been created by using majority algorithm (see Section III of Ref. 6). The present paper deals with (1) refinement of the majority algorithm to obtain more ideal shape of alphabet fonts which enable better identification of early Brahmi letters of inscriptions, and (2) proposal of a modified correlation function method for comparisons between the shape of the alphabet fonts and that of the sample letters. This work will prepare for the development of digital repository of Sri Lankan ancient documents, which has a function of automatic reading of ancient letters.

II. GENERALIZED MAJORITY ALGORITHM

The detailed procedures of obtaining alphabet fonts of early Brahmi letters from a lot of inscriptions are described in authors' previous paper [6], where the majority algorithm has been adopted to find out the ideal shape of the alphabet of early Brahmi script. The rough story is the following. Firstly, we correct N sample images of one alphabet of Brahmi script from the estampage documents (see Section III of Ref. 6). The image is composed of 300×300 pixels and is expressed by a two-dimensional array data. Each (i, j) element of the array has a value either 0 or 1 depending on the pixel is black or white. The method of the majority algorithm examines the value of the (i, j) element of N array data of the samples, and the value of the (i, j) element of the alphabet font is determined by the majority of the values of (i, j) element in N samples. This process is mathematically expressed by the following equation.

$$F(i, j) = \begin{cases} 1, & \text{if } \sum_{k=1}^N f_k(i, j) / N > 0.5 \\ 0, & \text{if } \sum_{k=1}^N f_k(i, j) / N \leq 0.5 \end{cases} \quad (1)$$

Here, $F(i, j)$ is the array data of the alphabet font, $f_k(i, j)$ is the array data of k -th sample image, (i, j) is the pixel position ($0 \leq i \leq 299$, $0 \leq j \leq 299$), N is the number of the samples. In order to get better font shape, we generalize above method. We introduce a parameter, γ which represents the degree of matching. The generalized algorithm is expressed by the following.

$$F_\gamma(i, j) = \begin{cases} 1, & \text{if } \sum_{k=1}^N f_k(i, j) / N > \gamma \\ 0, & \text{if } \sum_{k=1}^N f_k(i, j) / N \leq \gamma \end{cases} \quad (2)$$

Here, $F_\gamma(i, j)$ is the font array data which is produced by $\gamma \times 100$ percent matching.

By changing the value of γ , we can produce various fonts for one alphabet of Brahmi scripts. Table 1 show examples of the fonts for a Brahmi script which corresponds to the Latin letter "A". Here, we used fifty samples, and isolated major noises on the estampages were removed prior to the analysis with the aid of the image processing software, Adobe Photoshop CS3. It is seen from Table 1 that the thickness of the font image is decreasing for larger value of γ and the font is broken in the case $\gamma > 0.65$.

Table 1: Font images with different γ values

γ -value	Font image	Total number of white pixels
0.40		20,380
0.45		18,699
0.50		16,323
0.55		14,718
0.60		12,313
0.65		10,655
0.70		8,192
0.75		6,462
0.80		3,588

Now, let us examine which font in Table 1 is best suited for our study of the letter identification. So, we compare the font image with estampage data. As an example, we use here a photographic data of an inscription found in Nācciyārmalai area in Trincomale district, Eastern province, Sri Lanka [4]. It includes 13 letters as shown in Fig. 1. The letters are numbered. Hereafter, the letters are called by the number. The analysis of the comparison was carried out by using correlation function method (see Section IV of Ref. 6).

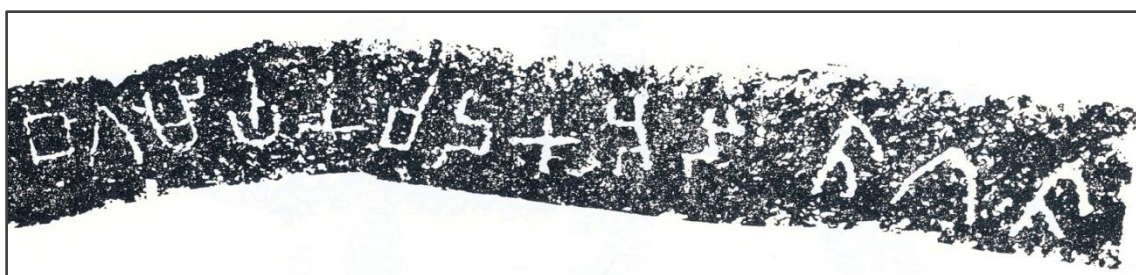


Figure 1. Photograph of original estampage data of inscription found in Nācciyārmalai area in Trincomale district, Eastern province, Sri Lanka

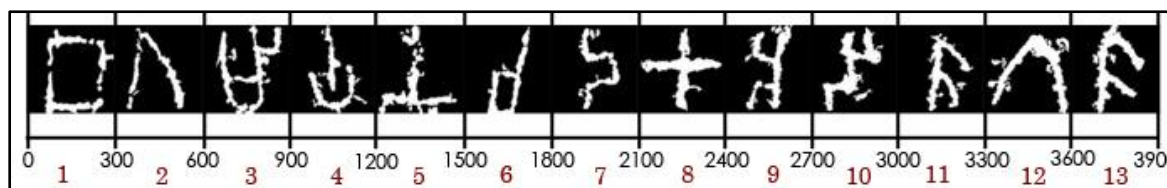


Figure 2. Formatted inscription image of the data of Fig. 1. The horizontal scale is in pixel numbers along the original image. The red letters are the letter numbers.

From the original estampage data of inscription, we firstly created a formatted inscription image (see Fig. 2), where the letters are arranged horizontally in x -direction with an equal spacing and major noises are

eliminated. And then, the correlation function, $P(x)$ in horizontal direction between one font in Table 1 and the letters in the formatted inscription image was calculated, where the correlation function is defined as the following.

$$P(x) = \frac{\sum_{i,j} G(i-x, j)F(i, j)}{\sum_{i,j} F(i, j)} \times 100 \quad (3)$$

Here, $F(i, j)$ is the array data of the best font and $G(i, j)$ is that of the letter in the formatted inscription image. The calculation was repeated for all the fonts in Table 1. Figure 3 shows the correlation function curves obtained for the different fonts whose γ -values are between $0.4 \leq \gamma \leq 0.8$. It is seen from the figure that the correlation function has sharp peaks at positions where the font is best correlated to the letters. The height of the peak takes a maximum value when the font matches to the correct letter. Therefore, we can identify the best matched letter to the font by analyzing the peak height in the correlation function. In order to study the degree of discrimination of the correct letter from wrong letters, let us introduce a parameter which is defined by

$$R_\gamma = \frac{1}{N} \sum_n \frac{H_\gamma}{h_{(\gamma,n)}} \quad (4)$$

Here, $h_{(\gamma,n)}$ is the peak value of the correlation between the font with a γ -value and the n -th letter in the formatted image, and H_γ is the maximum peak value of the correlation between the font and the correct letter. It is, in the case of Fig. 2, the letter, No. 9, which is located in the range, $2400 \leq x \leq 2699$. The parameter R_γ was calculated for different values of γ . Fig. 4 shows the result of the calculation. It is seen from the figure that the discrimination is the best in the case where the font of $\gamma = 0.55$ is adopted. Thus, the best font in Table 1 is decided. It should be noted, however that the best γ -value depends on the shape of letters and we must repeat the above procedure for all alphabets of early Brahmi letters.

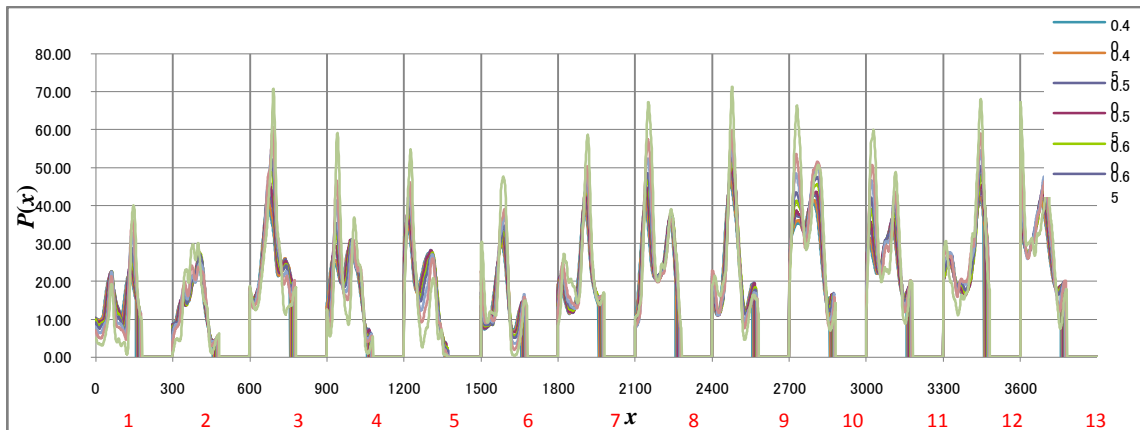


Figure 3. Correlation between the shape of sample letters in the formatted inscription and that of the fonts with different γ -values. The correlation functions are shown by different colored lines depending on the γ -values which are shown at the top-right corner of the figure. The red letters in x axis show the letter numbers in Fig.2.

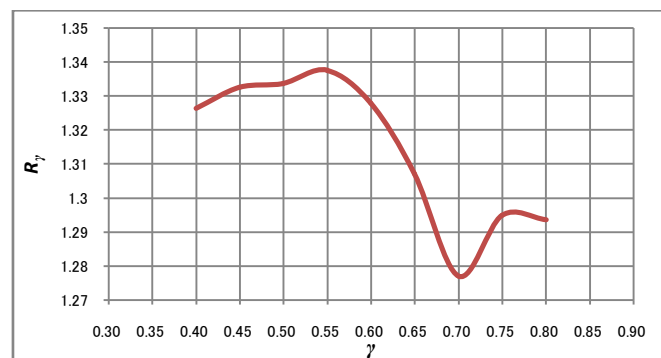


Figure 4. Dependence of R_γ on γ

III. MODIFIED CORRELATION FUNCTION METHOD

In the previous section, we noted the peak values in the correlation function and assumed that the letter which gives the highest peak is the correct letter to be identified. This method works in most of the cases of identification experiments. However, in some cases, the method fails to find out the correct letter. Let us show such an unsuccessful example. The original estampage data is given in Fig. 5 [4] and the formatted image is given in Fig. 6. Figure 7 shows the curve of the correlation between the best font found in the previous section and the letters in Fig. 6. The first two letters, No. 1 and No. 2 are the correct ones to be identified. The peak heights for the letters, No. 1 and No. 2 are less than those of the letters, No. 8 and No. 9. So, the peak method fails to find out the correct letters, and we need to have a better method of Brahmi letter identification which successfully works in this example.

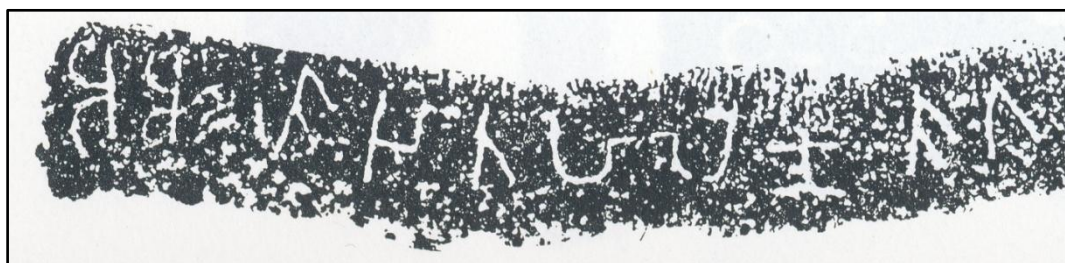


Figure 5. Photograph of original estampage data of inscription found in Dirñbul ñgala: N ñmalpokuñ a area in Trincomale district, Eastern province, Sri Lanka

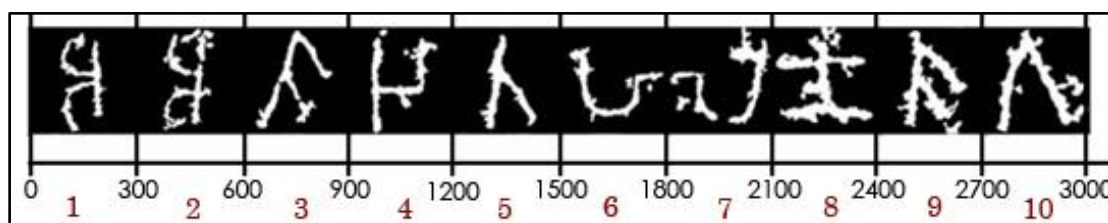


Figure 6. Formatted inscription image of the data in Fig. 5. the horizontal scale is in pixel numbers along the original image. The red letters are the letter numbers.

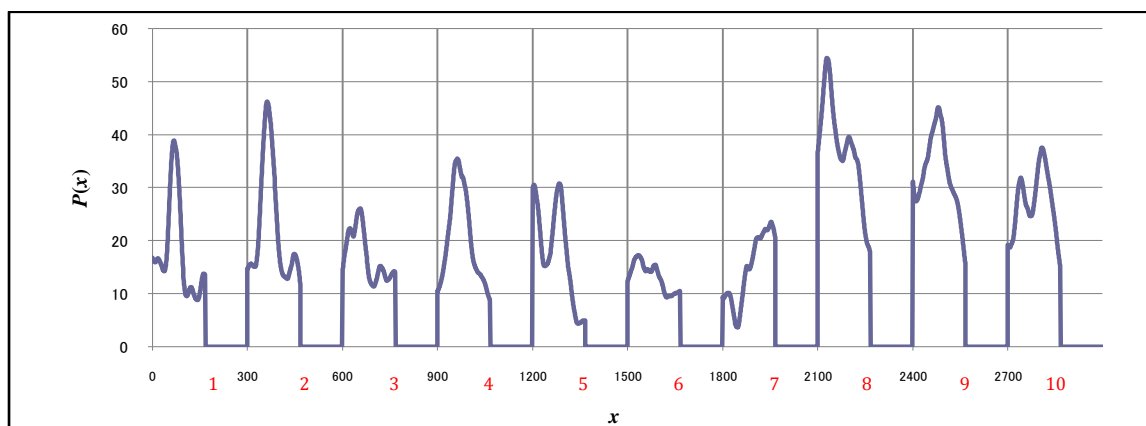


Figure 7. Correlation between the shape of sample letters in the formatted inscription and that of the best font with $\gamma = 0.55$. The red letters in x-axis show the letter numbers in Fig.6.

The point of the better method is to note the shape of correlation function rather than the peak heights. The processes of the sophisticated indentation method of Brahmi letters is the following.

(1) The auto-correlation function of the font of the letter to be identified is calculated, which is defined by

$$A_{font}(x) = \frac{\sum_{i,j} F(i-x, j)F(i, j)}{\sum_{i,j} F(i, j)} \times 100 \quad (5)$$

Here, $F(i, j)$ is the array data of the font. The scale of auto-correlation function is adjusted so that its peak value is 100.

(2) The each curve of correlation between the font and one letter in the formatted image is scaled by the peak height of the auto-correlation function so that they have the same peak height, 100.

$$P_0(x) = \frac{100}{P_{\max}} \times P(x) \tag{6}$$

(3) Then, the sum of the square of the difference between $A_{font}(x)$ and $P_0(x)$ for the x-range of one letter are calculated.

$$D = \sum_x [A_{font}(x) - P_0(x)]^2 \tag{7}$$

Here, D is a measure of the difference between the font shape and the letter shape. Actually, if the letter is the font itself, P_0 is the auto-correlation of the font, and therefore D becomes zero.

(4) The value of D is calculated for all specimen letters in the formatted image. The letter for smaller D value is the better matched letter with the font.

Now, let us show that the above algorithm of identification of letters surely works for the specimen given in Fig. 6. Firstly, we calculated the auto correlation function of the best font found in the previous section, namely the font of $\gamma = 0.55$ in Table 1. The auto-correlation function curve is shown in Fig. 8.

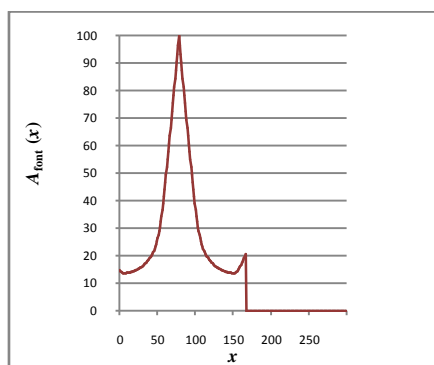


Figure 8. Auto-correlation function of Brahmi font for the Latin letter "A".

The normalized curve of the correlation between the font and each letter in the formatted image is presented in Fig. 9, together with the auto-correlation curve.

The D-value calculated for each letter in the specimen is shown in the bar graph, Fig. 10. It is clearly seen from the figure that the D-values for the first two letters, No.1 and No.2 in the specimen are much less than those of the other letters. Thus, we have succeeded to identify the font letter in the specimen.

This sophisticated identification method was applied also for the specimen of Fig. 2, where the peak method was also successful. The D-value calculated for the specimen is shown in the bar graph, Fig. 11. The discrimination of the correct letter from the wrong letter is much better than the previous peak method for the letters, No. 3, 8, 10, 12 and 13. However, the discrimination efficiency for the letter, No.6 is a little poorer than the other letters. In such case, the combined use between the peak method and the sophisticated method may be recommended.

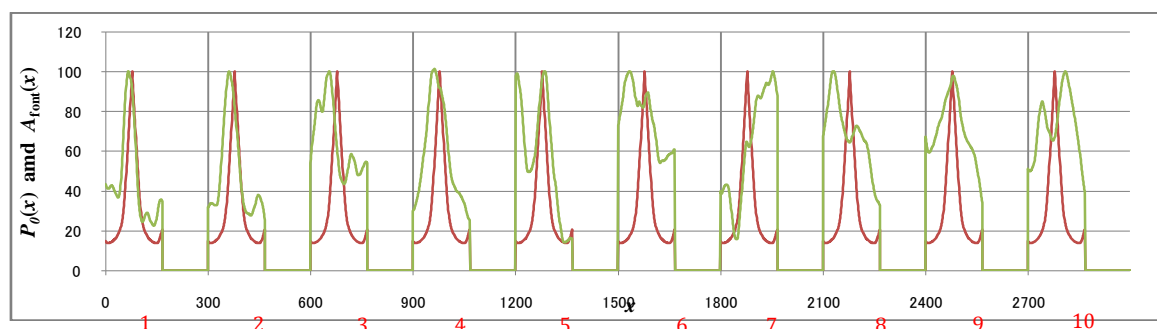


Figure 9. Modified correlation function, $P_0(x)$ for each letter (green curves) and auto-correlation function, $A_{font}(x)$ of the font (red curves). The red numbers are the letter numbers in Fig.6

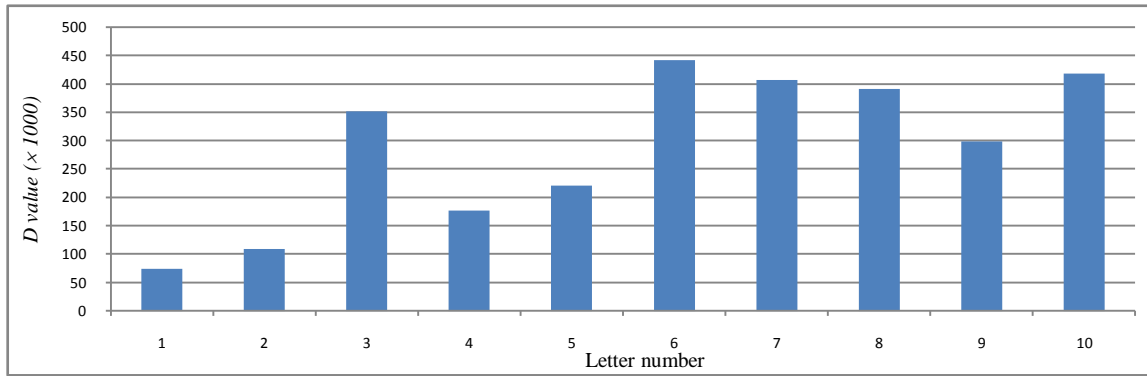


Figure 10. D-value for each letter in formatted inscription image, Fig.6

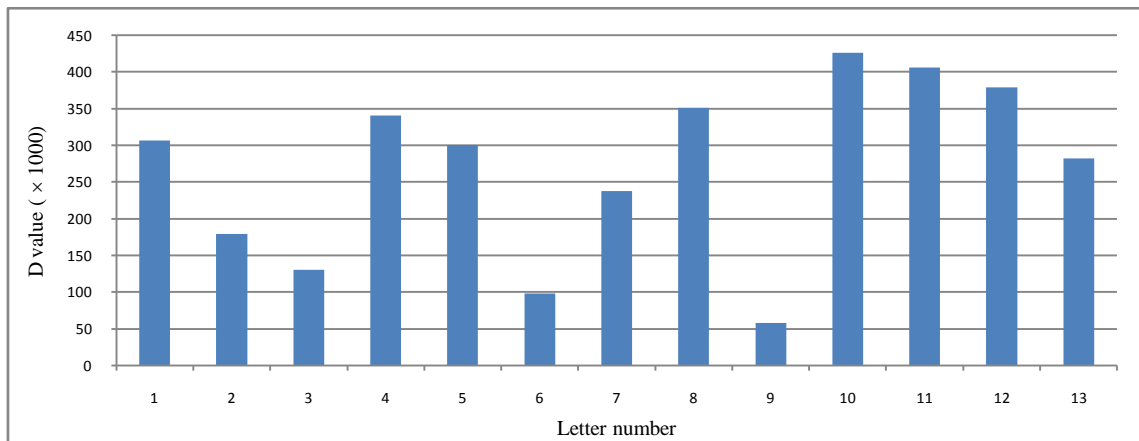


Figure 11. D-value for each letter in formatted inscription image, Fig.2

IV. CONCLUSION

This study is concerned with automatic identification of early Brahmi inscriptions by computer. The shape of the letters has been changed in time. Even in a same period, slightly different shape has been found for one letter. In this study, algorithm of the automatic indentation of ancient letters has been established. Namely, the modified correlation function method has been proposed which is much more sophisticated than the previous correlation peak method. The author's next work is producing a digital repository of Sri Lankan inscriptions which has a function of automatic reading of letters by computers. The design of the digital repository will be reported elsewhere.

V. ACKNOWLEDGEMENTS

This work was supported by Department of Archaeology, Colombo, Sri Lanka. The first and second authors equally contributed to this work. This work was also supported by the Grants in Aid 20500825, the Grand in Aid 24650124 and the Grants in Aid 24300278, Ministry of Education, Culture Sports, Science and Technology, Japan

REFERENCES

- [1] M. Dias, Lakdiwa Sellipiwalin heliwana Sinhala Bhashawe Prathyartha namayange vikashanaya, pp.1, Department of Archaeology, Colombo Sri Lanka,1996.
- [2] Rev. K. Amarawansa, Lakdiwa Sellipi,pp 10, MD Gunasena, 1969
- [3] L.S. Perera, Lanka Ithihasaye Muulashra, Journal of Sri Lankan History, vol1, pp 62,Vidyalankara university Sri Lanka.
- [4] S. Paranavitana, Inscriptions of Ceylon, Department of Archaeology, Colombo, Sri lanka, 1970
- [5] A.S Hettiarachchi,. Investigation of 2nd, 3rd and 4th century Inscriptions, pp 78, Department of Archaeology, Colombo, Sri Lanka, 1990.
- [6] Dammi Bandara, Nalin Warnajith, Atsushi Minato and Satoru Ozawa, Creation of precise alphabet fonts of early Brahmi script from photographic data of ancient Sri Lankan inscriptions, pp33-39,Vol3 No 3 ; Canadian Journal on Artificial Intelligence, Machine Learning and Pattern Recognition