Online Sumarians Cuneiform Detection Based on Symbol Structural Vector Algorithm

Kawther K. Ahmed*

Keywords: Sumarian Cuneiform, SSV, Detection

Abstract

The cuneiform images need many processes in order to know their contents and by using image enhancement to clarify the objects (symbols) founded in the image. The Vector used for classifying the symbol called symbol structural vector (SSV) it which is build from the information wedges in the symbol.

The experimental tests show insome numbersand various relevancy including various drawings in online method. The results are high accuracy in this research, and methods and algorithms programmed using a visual basic 6.0. In this research more than one method was applied to extract information from the digital images of cuneiform tablets, in order to identify most of signs of Sumerian cuneiform.

1.1. Introduction

Humans have been creating visual symbols to communicate since the beginning of humanity. However, the emergence of specific symbols representing sound, letters, or words in a spoken language occurred recently and fundamentally changed human society.

Hundreds of thousands of hardened clay cuneiform tablets rest in museums, libraries, and universities around the world. And more than half never been studied or even read by scholars. Copying, deciphering, and publishing their contents, and checking that work, is slow and difficult because among 6.5 billion people in the world maybe 300 can be read cuneiform.

Because of that, the access to the widely distributed tablets is expensive, time consuming, and difficult. For that on-line processing of cuneiform based on AI will be worked on numeric of Sumerian and that because its difficult to recognize and its so important .

1.2. Image Processing Techniques

The tablets images need further processes by the computer to know their contents; computer vision operates on mages that usually come in the form of arrays of pixel values. These values are invariably affected by noise, smoothing reduces the noise effecting on image. Because the nature of the cuneiform images a lot of attention should be paid on image enhancement. The reasons for that are:

* College of Education for women- Baghdad University.

there exists distortion on the original tablet (because of the age and nature), and due to the natural effect of the particular method of the image acquisition.

The object in an image in general and the cuneiform symbol specifically contain a great amount of redundant information, which hides the relevant detail, for example in the cuneiform symbol the only required parts of the image is a set of lines (direction and length of each line) to be able to recognize the symbol. Skeletonization is the process of eroding the boundaries of an object in the image until one pixel wide skeleton remains.

After that, the skeleton of a cuneiform symbol is to decompose it to its primitive'sfinding the straight lines which it is composed off; more than one technique are available to achieve that. The lines are then collected into the related groups to find the wedges and then identify the symbol.

1.3. Cuneiform Symbol Structure

The forms of the symbols consists of various combinations of different types and numbers of wedges, each wedge are upright, horizontal, diagonal or slopping. The writing system contained approximately 1,200 signs are known, but it is assumed that as many as 2,000 signs were in use. Only a few types of wedges were permissible for the composition of more complex signs.

a. Vertical
b. Horizontal
c. Oblique from upper left to lower right
d. Oblique from lower left to upper right
e. Wedging

At first, symbol were written from top to bottom; later, they were turned onto their sides and written from left to right. In later periods harder materials were also used. Five basic orientations are applied: horizontal, two diagonals, a hook and vertical stroke:

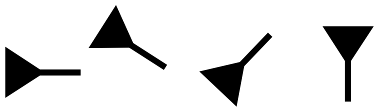


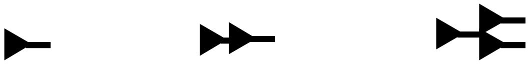
Figure (1) Symbols Structure

1.4. Order of cuneiform Symbol

The present day order of cuneiform signs such as used in sign lists is based on the (arbitrary) sequence show in figure (2), the sign to left is of higher order than the sign to it's right, this sequence is used in cuneiform dictionary Lab at. Also within the sequence starting with a horizontal stroke, this order applies to the remaining strokes.

With the signs starting with the sign precedes etc.

The different orthography of some of the signs makes this sequence not completely unique.



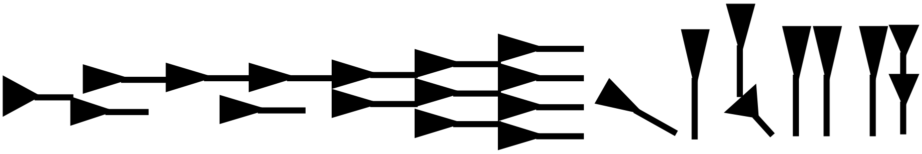


Figure (2) Ordering of the Cuneiform Signs

Another issue impacting cuneiform research and publication; is the fact that, unlike for most writing systems in the world, there is still no standard computer encoding for cuneiform text.

1.5. Classifications of Cuneiform Symbols

In order to build a computerized dictionary of the cuneiform symbols that might be used in case of automatic recognition of the symbols in tablets images, the dictionary should be indexed according to the shape and contents of the symbols itself, the number of cuneiform symbols is 595 with 1895 sound, the total number of wedges in all symbols are as show in table 1.

Table (1) Total Number of Symbols

Wedge type	Number
Horizontal	1600
Vertical	1344
Oblique	59
Style	919

1.6. Cuneiform Symbols Recognition Algorithms

The suggested method is built by making use of the fact that there are a limited number of cuneiform symbols. Each cuneiform signs according to the suggested method has its own finger print which is a sequence of integers calculated by the method making use of its image characteristics. The finger prints of all signs are stored in a signs database which is used in sign recognition.Applying the algorithm on all the cuneiform symbols image, and storing the results in a two dimensional array (B) will be the image database, each row in the array contain V1 and V2 of finger prints of a specific cuneiform symbol image. Identifying the symbol in a given cuneiform symbol image is done by the following algorithm.

Input: an image of cuneiform symbol, a two dimensional array of images finger prints
Output: a number represent the index of the finger prints in B

1. Produce the finger print of the given image using IFPG algorithm above, store V1 and V2 of the finger print in array A.
2. Calculate the total of the absolute difference between each entry in A and the corresponding entry in the first row in B. store the results in C[1].
3. Repeat step 2 for each row in B i.e., no of entry -1

$$C[i] = \sum_{j=0}^{noofentry-1} abs\,(B[i][j] - A[j]$$

   i: The row number, which contains the finger print of a cuneiform symbol image.
   j: An entry in finger print.
   Abs: an absolute difference function.
4. The index of entry of C which contains the smallest value is the number of image that best fit the given image.

The table in figure 3 shows an example of the implement of the suggested algorithm, the third entry in array c which contains the minimum value, signify that the third image is the most fit image to image in array A as mentioned above.
The calculation of the *c[2]* in the example is as follows

$$c[2] = abs(3-2) + abs(1-0) + abs(1-1) + abs(1-1) + abs(0-1)$$
$$= 1 + 1 + 0 + 0 + 1 = 3$$

Note that step 3 of calculating the finger print of the image (normalization step) minimizes the effect of changing the size of an image values in its finger print (i.e., the same finger prints (or almost the same) would be produced for two images of the same symbol with different sizes).

1	1	1	0	2

C
6
3
1
5
5
4

0	2	2	1	0
0	1	1	1	3
1	1	0	0	2
2	0	0	1	3
1	4	1	0	0
1	2	0	1	1

Figure (3). an example of implementation the algorithm

1.7. Building SSV (Symbol Structure Vector)

The algorithm to build the symbol structure vector (SSV) are stated below:

Algorithm - building the SSV vector for cuneiform symbol image

Input: image for cuneiform symbol Output: the SSV vector
1. Perform the image preparation process. 2. Perform the image skeletonization process image 3. Perform the final thinning 4. Find the image line skeleton feature points 5. Find the straight lines in the image skeleton. 6. Use the junction and bend points together with the length and slope of the line to recognize the wedges in the skeleton. 7. Find the sector number for each wedge.

The sector number is a normalized approximated location of the wedge and is calculated using the equation[1]:

$$S_n = 3 * int(3 * y/h) + in(3 * x/w) + 1 \ldots\ldots (1)$$

Where S_n is the sector number

w is the width of the image symbol

h is the height of the image symbol

x is horizontal coordinate of the wedge

y is vertical coordinate of the wedge

1.8. Results

In this work,the implementation was performed using Visual Basic 6.0. We've randomly picked images which have been used in designed application.

- **Online Detection**

By drawing a symbol of Sumarian numerical symbol will start detect and found the matched once of image that have been drawn with symbols found in database.

As shown below, for example number (3), the recognition will start implement an algorithm SSV to find a vector for images found in database and recognition one that drawn and will find the best result. A relevant or accuracy is48 (experimental result for drawn a "3" in Sumerian script in online) and maybe more or less as shown in figure 4., and this percent can be increase while detection of the drawn image be closer to one found in database.

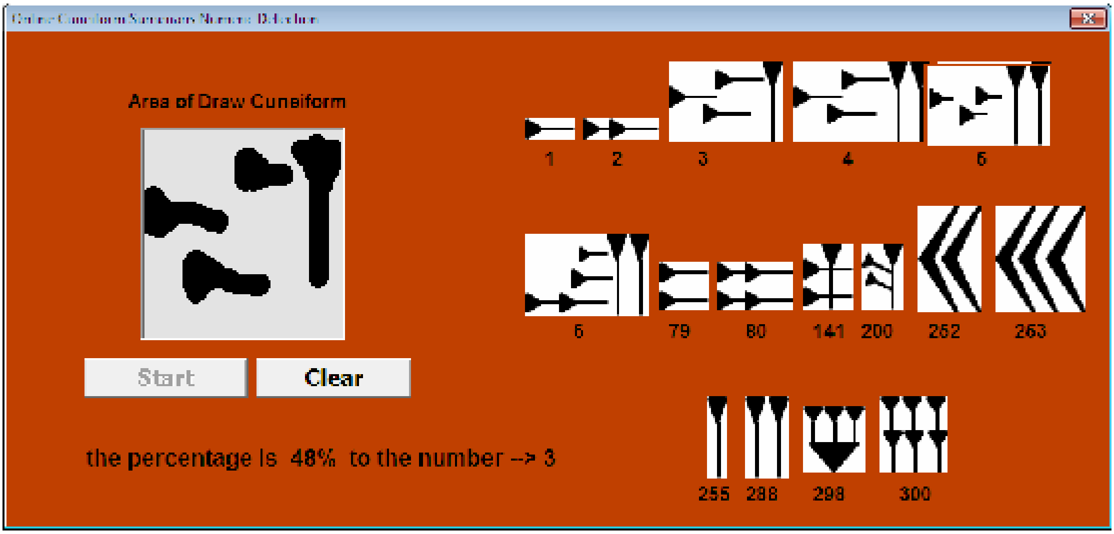


Figure 4, number (3) with accuracy 48%

Other example as shown in figure 5., it's a number (255), here will find the high relevance found, even when draw a number with more muddy appearance.

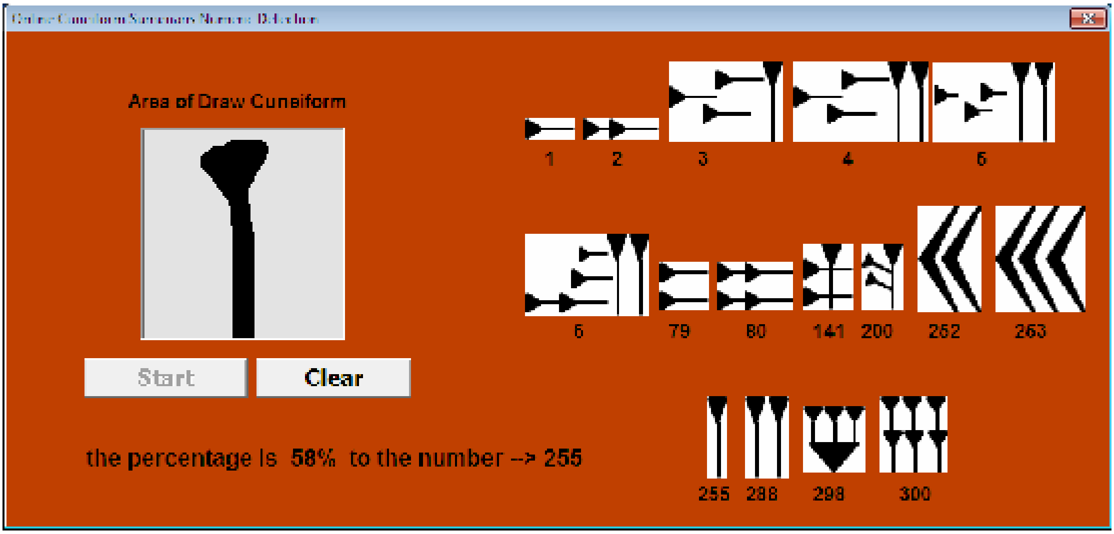


Figure 5, number (255) with accuracy 97%

When the symbol be more detailed and closer to original symbol, the relevancy will be increased as shown in figures 6 and 7.

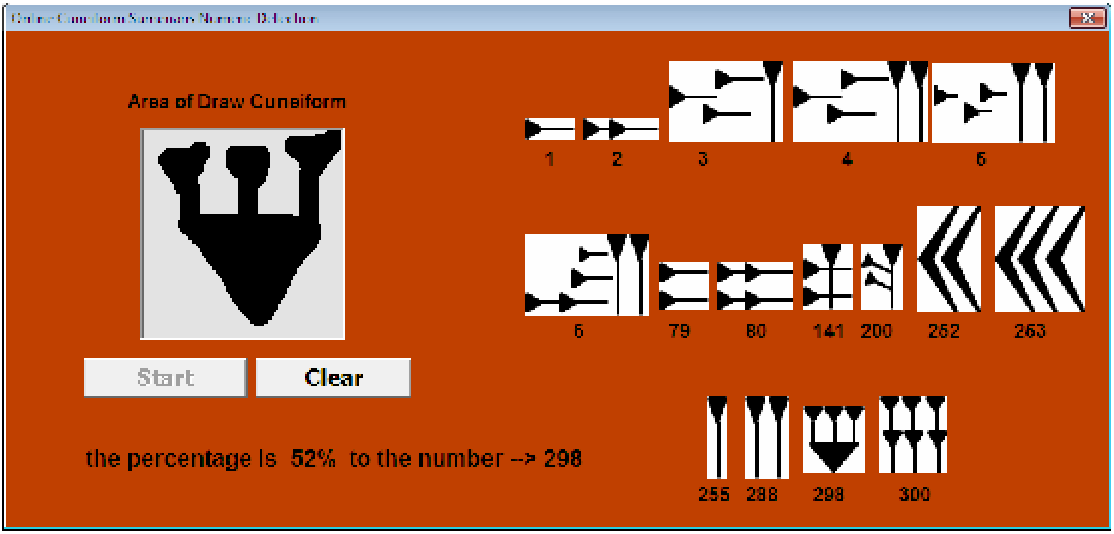


Figure 6, number (298) with accuracy 52%

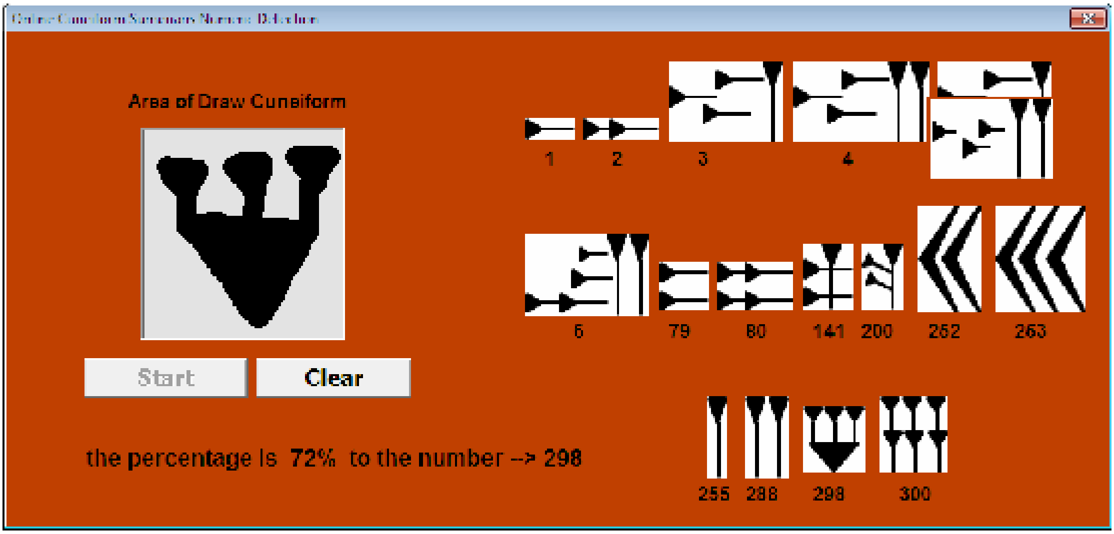


Figure 7, number (298) with accuracy 72%

And below shown in figures 8 and 9, the relevancy will be more, that because the drawn take more details for that the vector affected by the relations between the pixels densities of the regions.

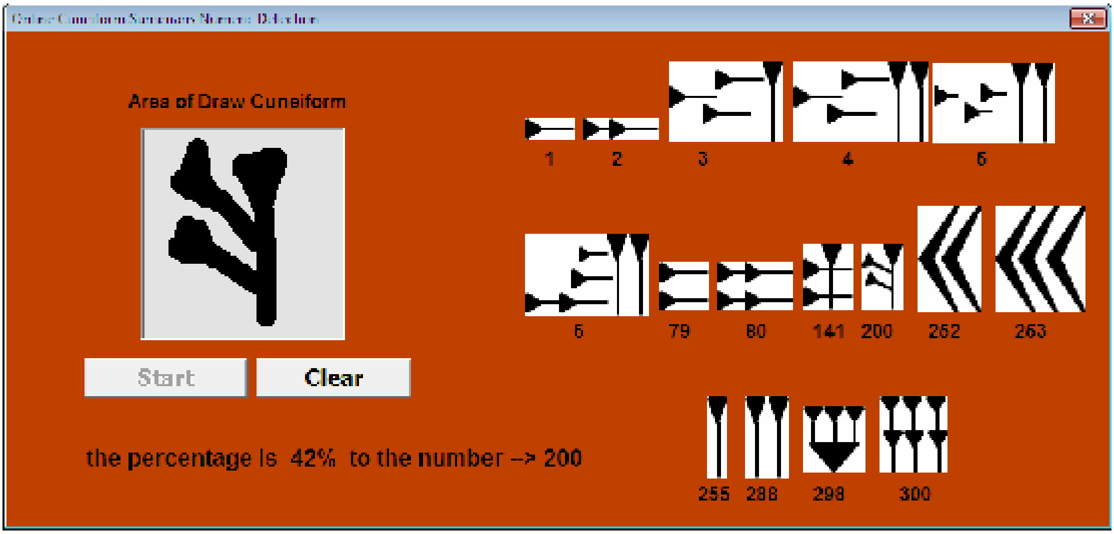


Figure 8, number (200) with accuracy 95%

Others tests taken into consideration as shown below which showed different numbers and different relevancy and more difficulties of draws in online method.

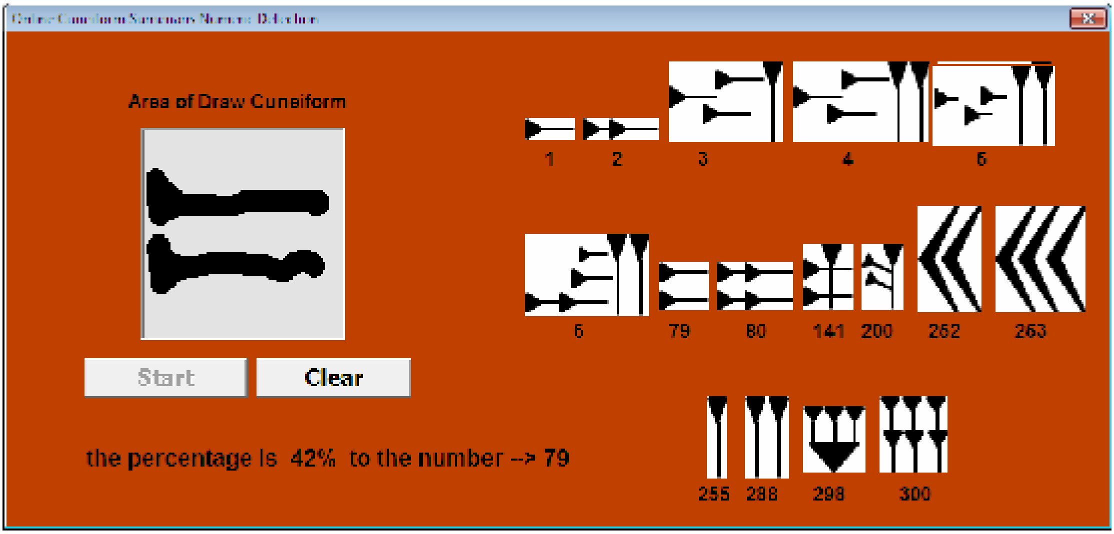


Figure 9, number (79) with accuracy 42%

Even if there is cut in edge will not be effect the recognition and keep finding the symbols only.

Conclusion

This research presents more than one method to extract information from the digital images of cuneiform tablets, its presents a methods Online and Offline, online its write or draw method which allowed to find out what matched depending on SSV vector. And in same time can load (offline) an tablet images and then extract a symbols in that tablet. and this proposed work is adopted to identify most of signs of Sumarian cuneiform.

As seen in the experimental results, the accuracy reached to 97% in recognition of Sumerian symbol. Even can increase the database by learning and saved each symbol stream in database and will be reach above accuracy when draw a number a symbol close to original drawn one.

References

1. Al-Ani H.; "Data Acquisition of Cuneiform Tablets Digital Images"; Ph.D. Thesis; University Of Technology: 2006.
2. Knrad Volk; "A Sumerian Reader"; Editrice Pontificio Istituto Biblico; (Book); 1996.
3. Gwendolyn Leick; "The Babylonians";Routledge; (Book); 2003.
4. "Journal of Semitic Studies"; Oxford Journals, Oxford University Press; Vol.1 Spring; 2009.
5. William W. Hallo; "The World's Oldest Literature, Studies in Sumerian Belles-Lettres"; Brill, Leiden-Boston; 2010.
6. Pearce, Laurie E.; "Cuneiform Cryptography: Numerical Substitutions For Syllabic And Logographic Signs"; 1982.

تميز الالوان المسمارية مباشرة بالاعتماد على خوارزمية هيكلية الرموز

م.م.كوثر كريم احمد
كلية التربية للبنات/ جامعة بغداد

الخلاصة:

ان الالواح المسمارية تحتاج الى العديد من المعالجات الاولية والصورية لغرض تمكن الحاسوب من معرفة محتوياتها، ومن الممكن اجراء عمليات معالجة صورية لغرض تحسين الصورة التي تحتوي على تشويهات او ضوضاء والتخلص منها لغرض توضيح العناصر ومنها الرموز المسمارية (كتابة، رموز، ارقام). في هذا البحث تم تعريف خوارزمية لإيجاد المستقيمات في الرموز وذلك بالاستفادة من نقاط خواص الموجودة فيه. واستخدم متجه هيكل الرمز (SSV) والذي يبني من معلومات المسامير في الرمز.

وأظهرت جميع الاختبارات ارقاماً مختلفة وعلاقات مختلفة باستخدامonline method والتي تعطي لنا دقة عالية باستخدام لغةVisual Basic 6.0،في هذا البحث نعرض اكثر من اسلوب واحد لاستخراج المعلومات من الصورة الرقمية من الجداول المسمارية،وذلك من اجدل تحديد معالم معظم الكتابات المسمارية السومرية.