

Sumerian Character Extraction by Using Discrete Wavelet Transform and Split Region Methods

Moahaimen Talib

Computer Science Dept.

College of Science

Al-Mustansiriyah University

Baghdad, Iraq

moahaimen@gmail.com

Jamila Harbi S

Computer Science Department.

Al-Mustansiriyah University

Baghdad, Iraq

dr.jameelahharbi@gmail.com

Abstract— this paper proposed a new method to extract characters from Sumerian Texts in Sumerian cuneiform tablets from the Ur III period. The work was confronted by the challenges posed by the fact that Sumerian is not a well understood language and it is not similar to any other ancient or modern language, so we offered a new method for extracting characters from Sumerian tablets, it has an accurate results and better time consuming than other methods, taking many tablet images and applying preprocessing methods to enhance and segment the image and then discrete wavelet transformation and we extract characters for each tablet image by split region algorithm, this work will be very helpful to Cuneiforms and scholars in their field.

Keywords— Sumerian Texts, discrete wavelet transformation, Sumerian tablets, Features extract.

1. INTRODUCTION

Sumerians an advanced people at their time which their origin still a theory till now, these great men establish a marvelous civilization in Mesopotamia (The Land between the Rivers), they have great contributions in the Humanity progress along all times, due to their advanced technology at their time they needed a way to record their achievements, victories, their Gods and Heroes, all of that must be recorded, so the invent the first character their first hand written that ever made by human, this system of writing they invented is known as cuneiform texts which are written on tablets mostly clay tablets, along a 30 centuries this system is used and invested in other civilization like Acadian Eelam, Babylonian, Assyrian and other great civilization[1].

At most Mesopotamian tablets have two axes, the horizontal for the numerical information, and the vertical axis, for the data are assigned to different individuals or spaces [2]. The character that is shown in multiple places of tablet texts is useful to understand the Sumerian words and the style of writing [3].

In the research are mainly used Meaning of Sumerian texts and the way of their readings from the Cuneiform Digital Library Initiative (CDLI) [4], Pennsylvania Sumerian Dictionary (PSD) [5].

In the next section we will discuss the main methods used to read and understand the cuneiform

texts off course with their advantage and their disadvantages

2. METHODS AND MATERIALS

Reading cuneiform symbols is an important issue for learning from cuneiform tablets. Cuneiformists have used two main strategies to represent and archive the clay tablets:

A. Hand-Drawn Copies (Autographs)

The usual registration of writes in tablets and sealing is done with pencil, and other tools. Trained conformists accurately take every single detail and make a measurement and also make a copy of all details from the original drawings, the success of the method will be achieved by those how are a domain expert in epigraphy of the script and/or iconography [6]. Disadvantages that the autographs that are written by hand is very exhausting, it take more time, and errors might happen and the scholar need to have an access to the tablets directly. The whole operation will be slow, sensitive, high resources, tedious, and, at last, unproductive, because the Cuneiformists must travel for long distance from its home to read a tablet because they kept in certain places such museums in London, Iraq, Pennsylvania etc. and also the Cuneiformists must travel again to compare other readings and there is a wide variety in reading and drawings of texts between Scholar's [7].

B. 3D Methods

3D method scans the tablet from all its sides because the tablets and cuneiform letters are in 3D (Saving Digital information of the clay tablets using 3D-Scanning) as shown in Figure 1, its good method to get the texts, in the later years, some solutions are introduces to obtain the full 3D model of tablets[8]. But it also has disadvantages such time consuming hardware problem software of scanning 3D tablets hardware is also so complex and need professional technical support some the scanning need a technician with some domain expert, but the main target of the scanning is to read the texts not all other properties of the tablet the Scholar also need to travel long distance to the depots or the museums where the 3D are stored in it [9] [10].



Figure1: 3D impressions in the clay, and the streams of the writing extending on the tablet sides.

3. WAVELET ANALYSIS

Wavelet analysis gives a windowing technique with variables region spaces. It allows the use of high length time intervals that have more accurate low frequency information is needed and regions with less spaces that have the information high frequency is needed [11].

The wavelet main best feature is it can do analysis locally that is "localized area can be analyzed by a signal that is much larger". Wavelet analysis has the ability of aspects can be revealed from data that other techniques of analyzing signals couldn't, aspects like self-similarity, trends, discontinuities in higher derivatives, and breakdown points. In addition, because it can give a good scene of data that is different from those presented by ordinary techniques, wavelet analysis can be used as a method of de-noising signal or compressing it with almost no degradation noticed [12].

Wavelet analysis can be implemented to higher-dimensional data such as the two-dimensional data images; and, in principle, the wavelet transform has taken a very wide acceptance in signal processing and image compression. Wavelets can be taken from a single mother wavelet by dilations and shifting [8]. The discrete wavelet transform (DWT) also known as a method that is a high efficient and flexible for decomposing the sub band of a signal [13].

Below is the two-dimensional scaling function.

$$\left. \begin{aligned} \psi^H(x, y) &= \psi(x)\varphi(y) \\ \psi^V(x, y) &= \varphi(x)\psi(y) \\ \psi^D(x, y) &= \psi(x)\psi(y) \end{aligned} \right\} \dots \dots \dots (1)$$

Where $\psi^H(x, y)$, $\psi^V(x, y)$, And $\psi^D(x, y)$ are called horizontal, vertical, diagonal wavelets, and one separable 2-D scaling function see Figure 2 [12].

The high-scale and low-frequency parts of the signal are known as approximations. The low-scale, high-frequency parts are known as details. The most primary level of the filtering process is shown in figure2:

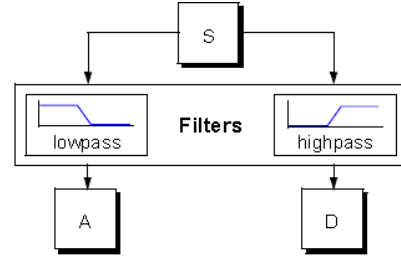


Figure 2: The process of filtering

4. WAVELET RECONSTRUCTION

DWT may have a certain job that it usually used for analyzing, or decomposing, signals and images. But how these parts can be regrouped into the signal of origin with losing no information is something to be considered. This operation known as synthesis or reconstruction. The mathematical manipulation that effects synthesis is called the inverse discrete wavelet transforms (IDWT) [13].

5. REGION SPLITTING TECHNIQUE

Region splitting technique is one of the Region-Based Segmentation (RBS); the main target of other techniques of segmentation is splitting the image to regions, but RBS techniques are based on finding the regions directly. An image is split initially into four disjoint quadrants, separated regions and then merges and/or split the regions in try to satisfy the terms of segmentation

6. PROPOSED SUMERIAN CHARACTER EXTRACTION METHODS

The benefit of the image processing is we take an input image in our case is the tablet image which may has several problems such as being broken in some parts or some characters is destroyed by several factors, sometimes the tablets are stolen or missed and only their images are exist.

In this paper we enhanced an algorithm to help solving many problems that the tablets as shown in figure3:

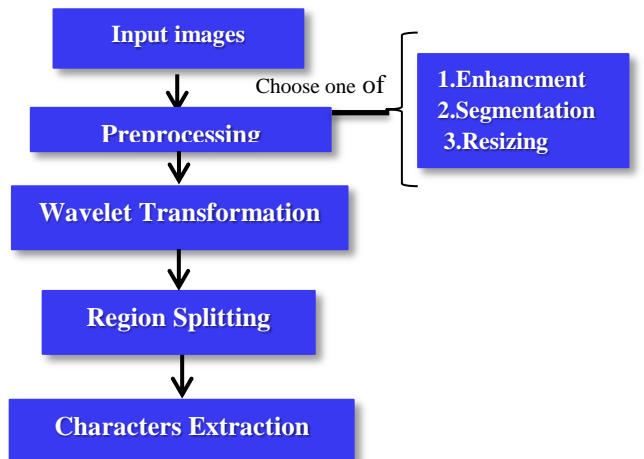


Figure 3: The Stages Involved in the Proposed Method

7. RESULTS

Our offered algorithm was tested on Matlab 2015b; we took 20 images from CDLI Cuneiform library at Cornell University as shown in figure 4. Figure5 illustrated the results of our proposed system.

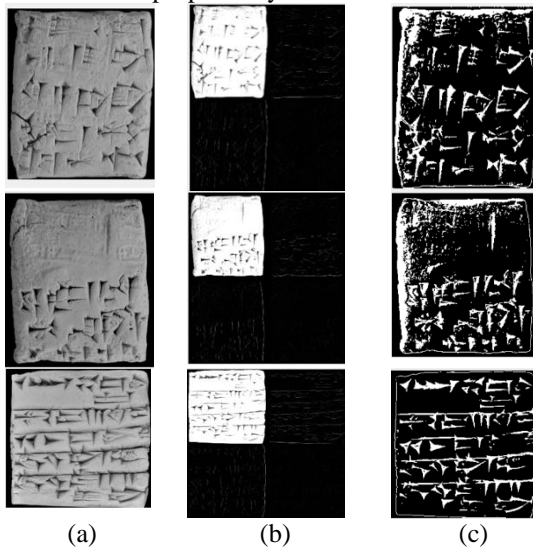


Figure 4: The output of our method:

a) Original image, b) Wavelet image, c) some extracted characters

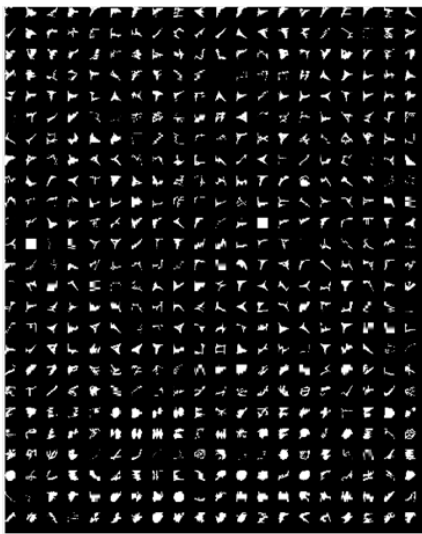


Figure 5: Characters Extracted from all 20 Tablets by using our proposed system

8. DISCUSSION

Our proposed system obviously from figure5 succeeds to extract Sumerian characters separable from Sumerian texts that are written in the clay tablet. Also, we are introduced the first research in this field and we are proposed a new method. The extract characters will be used in other stage of our proposed system to recognition and identification.

9. CONCLUSION

Sumerian language due it is the earliest known written language it's really hard to be read and understand it is not similar to any other language whether the language was ancient or modern, in this research we provide scholars and cuneiforms with the best help they can have

its better to the cuneiforms to take an image from a university website library and use our method to extract characters from Sumerian texts.

10. REFERENCE

- [1] N.Samual Kramer, "The Sumerians their History, Culture, and Character," Chicago University Press, Chicago, 1963.
- [2] R.Eleanor , "Tables and Tabular Formatting in Sumer, Babylonia and Assyria, 2500-50 BCE," Oxford University Press, Oxford,2003.
- [3] S. Gabriella, "Two Old Babylonian Model Contracts," Cuneiform Digital Library Journal, vol.24, 2014.
- [4] ePSD, "The Electronic Pennsylvania Sumerian Dictionary Project," psd.museum.upenn.edu/epsd/index.html,(2017).
- [5] CDLI, "The Cuneiform Digital Library Initiative," cdli.ucla.edu, (2017).
- [6] H. Hendrik, W. Geert, "New Visualization Techniques for Cuneiform Texts and Sealings," Journal of Akkadica, vol.132, pp. 163-178, 2011.
- [7] V. Carlo, V. Luc, V. Karl, V. Jay Rompay and W. Patrick, "Digitizing The Cuneiform Tablets From Beydar," in Images and Artifacts of the Ancient World, Bowman, A. and M. Brady, London, pp. 85-96,2005.
- [8] K. Subodh, S .Dean, D. Donald and C. Jerry, "Digital Preservation of Ancient Cuneiform Tablets Using 3D-Scanning," In Proceedings of the IEEE Fourth International Conference on 3-D Digital Imaging and Modeling, 2003.
- [9] V. Daniel, Hahn, C. Kevin Baldwin, and D. Donald Duncan "Non-Laser-Based Scanner for Three-Dimensional Digitization of Historical Artifacts," Journal Appl. Opt, vol. 46, pp. 2838-2850, 2007.
- [10] J. Kantel, P. Damerow, S. Köhler and C. Tsouparopoulou, "3D-Scans Von Keilschrifttafeln - ein Werkstattbericht," Assmann, W., Hausmann-Jamin, vol.26, DV-Treffen der Max-Planck-Instituut, atenverarbeitung, pp. 41-62,2010.
- [11] R. Majeed, B. Zou Beiji, H. Hiyam and W. Jumana, "Ancient Cuneiform Text Extraction Based on Automatic Wavelet Selection," International Journal of Multimedia and Ubiquitous Engineering, vol.10, pp.253-264, 2015.
- [12] H.Tzu-Heng Lee, "Wavelet Analysis for Image Processing," Institute of Communication Engineering, National Taiwan University, Taipei, Taiwan, 2017.
- [13] C. Rafael Gonzalez Gonzalez, E. Richard Woods and L. Steven Eddins, "Digital Image Processing Using Matlab,"

2004.

Biography

Moahaimen Talib was born at 1979 in Baghdad, graduated from Al-Rafiedien University College. Computer Sci, Dept.2002. Recently, he is in research year of MSc.

Prof.Dr.Jamila was born at 1966 in Baghdad. She was graduated from Baghdad Uni. Ph.D. in 2001, from Al-Mustansiriyah Uni. M.Sc. in 1996 and B.Sc. in1989. Her interested in theoretical fields and analysis of; image processing, data compression, digital signal processing, Pattern Recognition, and multimedia systems.
<https://scholar.google.com/citations>.