DOI:10.24996/ijs.2018.59.3A.20

ISSN: 0067-2904

# Cuneiform Tablets Image Preprocessing Proposed Algorithms Techniques for Pattern Recognition

**Ali Adel Saeid*, Abdual Munim S. Rahma, Muhsen J. Abdul Hussien**
Computer Science Department, University Of Technology, Baghdad, Iraq

**Abstract**
Cuneiform symbols recognition represents a complicated task in pattern recognition and image analysis as a result of problems that related to cuneiform symbols like distortion and unwanted objects that associated with applying Binrizetion process like ***spots and writing lines***. This paper aims to present new ***proposed algorithms*** to solve these problems for reaching uniform results about cuneiform symbols recognition that related to (select appropriate Binerized method, erased writing lines and spots) based on statistical Skewness measure, image morphology and distance transform concepts. The experiment results show that our proposed algorithms have excellent result and can be adopted as a preprocessing technique.

**Keyword s**: cuneiform, distance transform , image Binariztion , image preprocessing

# الخوارزميات المعالجات الأولية المقترحة لتمييز انماط صور الألواح المسمارية

**علي عادل سعيد، عبد المنعم صالح رحمة، عبد المحسن جابر عبد الحسين**
قسم علوم الحاسبات ، الجامعة التكنولوجية ، بغداد، العراق

**الخلاصة**
تمثل عملية تميز الرموز المسمارية عملية معقدة في مجال تمييز الأنماط وتحليل الصور الرقمية ،وذلك لتعلق الأمر بالتشوهات المرافقة لهذة الصور بصيغتها ثنائية اللون (ابيض،اسود) كالبقع وخطوط الكتابة المسمارية . يهدف هذا البحث الى اقتراح خوارزميات لحل هذة المشاكل وذلك باستخدام معامل التفلطح الأحصائي ،مورفولوجية الصور وتحويل المسافة. من خلال النتائج المتميزة المتحققة يمكن اعتماد هذة الخورزميات لدعم المعلجات الأولية للعمليات المتعلقة بتميز الرموز المسمارية .

## 1. Introduction

Cuneiform writing is one of the oldest writing systems that was born in Mesopotamia, in the third millennium BC in the city of Urk [1], [2]. The letters of this language consist of a set of cuneiform symbols that take either horizontal, diagonal or vertical directions [3], which have significant implications, see Figure-1. The cuneiform language differs from the ancient Egyptian hieroglyphs that the cuneiform character expresses another meaning when it is in the context of the sentence in relation to the other letters. There are approximately one hundred thousand of cuneiform tablets are hold in various museums of the world and the large proportion of them in the Iraqi Museum in Baghdad, which is related to the civilization of Babylonian and Assyrian [4], There are many obstacles to the interpretation of cuneiform language including the difficulty of language and the lack of interpreters [1]. This requires the use of information technology to solve this problem with the difficulty of writing style in the form of three-dimensional, in addition the distortions problem that associated with cuneiform tablets. There are many researches that have gone towards this goal with multiple faces,

*Email: zeenali736@gmail.com

therefore with recognition face. Hilal Yousif [1] proposed recognition method for cuneiform symbols depending on intensity curve features about the cuneiform symbols, and with same context. Fahimeh [5] presented recognition method of Persian cuneiform characters by neural network with back propagation model [6]. Raed presented a method for extract cuneiform symbols from clay tablets based on selected wavelate bases algorithm and with retrieving process. Leonard [3] proposed retrieving method by initially applying Bag-of-Features technique for word spotting and constructing the retrieving quarry by Hidden Markov probability Model, therefore, dealing with aspects of cuneiform writing like (recognition, extraction,..) faced with many obstacles like the spots results from distortion effects. This is evident after selecting a suitable Binriztion techniques as seen in Figure-2(b) which creates a critical factor in recognition process.

**Figure 1-** Assyrian cuneiform symbols

( a) (b)

**Figure 2-** a):cuneiform image character , b) spots problems

The other problem is concern with the **(writing line)** that associated with cuneiform writing, which depends on the writing of some cuneiform texts or not. Therefore, a difference in this pattern affects negatively the process of recognition state depending on crated uniform image feature vector as seen Figure-3.

Figure 3- same cuneiform character with different features as present or absence of writing lines

However this research proposed a new method for solving mentioned problems that have not been addressed previously related to cuneiform image for recognition task by suitable proposed algorithms of choosing a proper binriztion method and get a uniform image symbols **that free from spots and lines,** these achieved by using (statistical Skewness metrics , image distance transform and connection component) techniques.

## 2. Image Binarization

Image Binarization is an important initial step in many image and document recognition that leads to convert the gray image colors tone to binary color (0 or 1), that reflects all image features to foreground and background sides, as seen in the following (eq1). However there are many metrics adopted to achieve this task like entropy, clustering and thresholding. With last metric, choosing the optimal and general thresholding technique among several ways is difficult because each of them leads to various results according to the selected dataset [7]. The various thresholding techniques with their features are described as follows.

$$G(X,Y)=\begin{cases}1 \ \ if \ \ f(x,y) > t \\ 0 \ \ if \ \ f(x,y) \le t\end{cases} \tag{1}$$

### 2.1 Otsu's Method

Otsu's method is global thresholding method that converts gray image to binary image [8]. It's linear discriminant statistical method that separates the image features to homogeneity of two colors bands, the first one related to foreground (objects, symbols) and other background [9]. Otsu's thresholding method starting with iterative histogram procedure separates the image colors as two colors intervals (I0=dark, I1= light),[7]. The color density with first is I0={0,1,2,3…,I}, and second is I1={I+1,i+2,..,k-1}. Therefore the global thresholding value is computed by the following formula [10],[9].

$$\boldsymbol{\delta}^2_w = w_b(I) * \boldsymbol{\delta}^2_b(I) + w_f(I) * \boldsymbol{\delta}^2_f(I) \quad ...(2)$$

Where

$$w_b(I)=\sum_{j=1}^{I} \boldsymbol{p}(\boldsymbol{j}). \tag{3}$$

$$w_f(I)=\sum_{j=I+1}^{I} \boldsymbol{p}(\boldsymbol{j}). \tag{4}$$

$$\mu_b(I)= \sum_{j=1}^{I} \boldsymbol{j} * \boldsymbol{p}(\boldsymbol{j})/w_b(I). \tag{5}$$

$$\mu_f(I)= \sum_{j=I+1}^{l} \boldsymbol{j} * \boldsymbol{p}(\boldsymbol{j})/ w_f(I) \tag{6}$$

$$\boldsymbol{\delta}^2_b(I)= \sum_{j=1}^{vI} (\boldsymbol{j} - \mu\mathbf{b}(\mathbf{I}))\ \mathbf{2}/w_b(I) \tag{7}$$

$$\boldsymbol{\delta}^2_f(I)= \sum_{j=I+1}^{l} (\boldsymbol{j} - \mu\mathbf{f}(\mathbf{v}))\ \mathbf{2}/ w_f(I) \tag{8}$$

with repeating the mention equations to compute (eq2) of each new color histogram and then choosing the optimal global thresholding value that achieve reaching the weighted minimum variance value among different computed values.

### 2.2 Niblack Method:

Niblack method is a local thresholding technique that is applied locally to determine thresholding that associated with each image's rectangle window by computing the mean and variance color values to reach the threshold value as with following form [11]

$$T=M+k\sigma \tag{9}$$

Where k is a constant value [0, 1], (m, σ) represent the mean and stander deviation.

### 2.3 Sauvola's Method

Sauvola's method is proposed to repair the drawback that related to Niblack method especially with noise or the spots that result after applying these methods. However this thresholding method has adopted the same parameters required to be applied with the previous form (9) for (**niblack method**), as follows [11].

$$T=m(1-k(1-\frac{\sigma}{r}) ) \tag{10}$$

Where k and r are set to be 0.5, 128 respectively

### 2.4 Iterative threshold

Iterative threshold method is a global thresholding method, and the iterative procedure for this method as in the following steps [10][12].

**Step1**: determines the initial thresholding value T where it is equal the mean value of color density.

**Step2**: with T separates the image colors to two pixels region and computes mean value for each region (M1, M2).

**Step3**: computes the new threshold value with the following T= (M1+M2)/2.

**Step4**; reapetes the steps starting from 2 to3 until the successive value of threshold value approaching to match.

## 3. Image connected-component labeling*:*

### 3.1introduction

Image connected-component labeling (CCL) is important operation in pattern recognition and computer vision [13]. It requires multiple application like fingerprint recognition, target recognition [14]. However, in this technique, each connected segments in binary image will have distinguish label that discriminates it from other segments, see Figure-4. CCL methods can be classified to four types as Hybrid algorithms, Two-scan algorithms, Multi-scan algorithms and tracing-type algorithms [14]. This research adopted the third type based on the image morphology technique depending on the dilation principle.

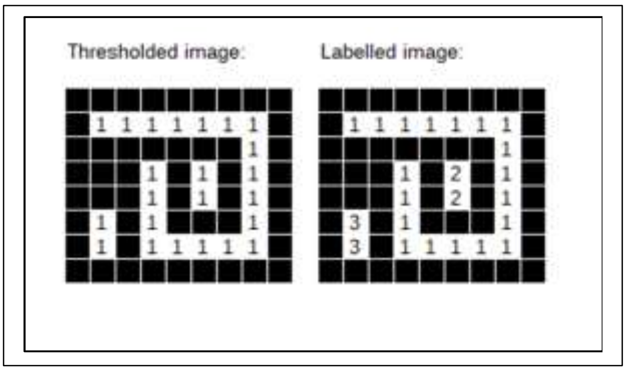



**Figure 4-**each connected-component has a distinguish label.

### 3.2 Extraction of Connected Components

To reach image labeling based on third strategy (malty scan), depending on image morphology technique (dilation process), the initial state starting with initializing the stricture elements B and scan the binary image to locate the first foreground pixel *p* that represents a seed point to generate reconstructed matrix $\mathbf{X_K}$. However with iterative process, it is applied by the following form with structure element.

let $X_0$=p where K=0,1,2…n

$$X_K= (X_{K-1+}B)\cap A \quad K=1,2,3.... \tag{11}$$

The termination state will be satisfied after reaching the condition, where $X_K=X_{K-1.}$ and this regenerated procedure is repeated with each connected component to create a **distinguish label** value.

## 4. Image distance transform

Image distance transform (DT) plays an essential role in many applications like pattern recognition, computer vision, robotics and image matching particularly for binary image matching with using suitable features crated by matching approaches [14],[15]. Distance transform (DT) is a conversion process that is applied on binary image to produce gray-level image in which each pixel of represents real value corresponds the minimum distance between object pixel (*ob*) and background pixel (*Bg*) That can be defended as in the following form[16],[1].

Where I(x,y) ε {Ob,Bg}.

$$I_d(x,y)=\begin{cases} 0 & I(x,y)\varepsilon\{Bg\} \\ \min I(x,y)\,\varepsilon\{Bg\}\left||x-x0,y-y0|\right| & I(x,y)\varepsilon\{Ob\} \end{cases} \tag{12}$$

There are many approaches adopted by distance transform algorithms as follows [17].

1- Morphological Approach.

2- Two Pass Algorithms Approach.

3- Vector Propagation Approach.

4- Boundary Propagation Approach.

5- Sampled Functions Approach

**Distance Transforms with Sampled Functions [18].**

Distance transform by sample function represents generalized approach for distance transform of binary image on grid (rows, columns) instead of binary value as depended. Therefore with samples functions, the basic intuition for computing the image distance transform depending on appearance or loss feature with each pixels defined as cost feature related with each pixel [10]. Let (£={1,2,3,…, n}) uniform **1D** is one dimensional grid, F: £→R where F is a function of grid defined as sample function , then the distance transform defined by sampled function is demonstrated as follow .

$$D_f(p) = \min_{q\epsilon£}((p-q)^2 + f(q)) \quad (12)$$

About every point q϶£, there is a restriction where distance transform (F) is bordered from above by parabola presenting the rooted position (q, f(q)). Therefore distance transform is realize by lower envelop of these parabolas, Figure-5 and its value corresponds the high of lower envelop.

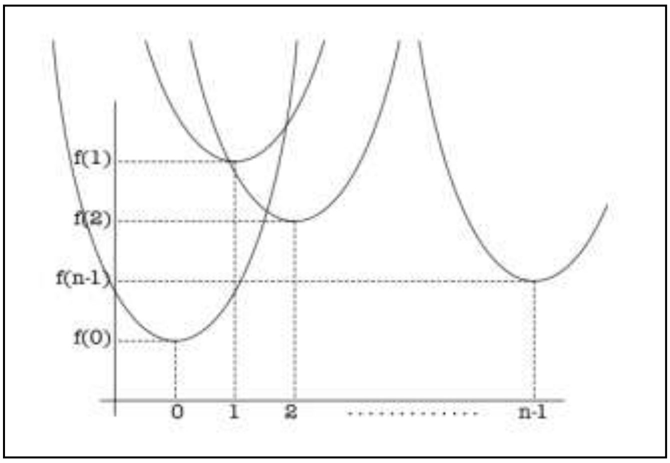



**Figure 5-** the lower envelop about n parabolas

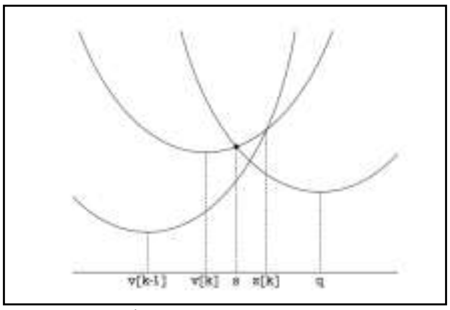
For computing image distance transform, the following two steps must be implemented

1- calculate the lower envelope of n parabolas.

2- solve the mention equation (12) by substituting the lower envelope's height at grid position. Where the two parabolas determine the distance transform that are intersect at single point therefore the intersection (s) position between two parabolas defined by grid positions (r,q) as follows

$$S=\frac{(f(r)+r2)-(f(q)+q2)}{2r-2q} \quad (13)$$

The lower envelop is calculated by sequentially calculating the first q ordered parabolas related to their horizontal positions. Where parabola is considered from q and find the intersection position with another parabola in v[k]. Therefore there are two states that are satisfied. First, Figure-6 if the intersection position is after z[k] then the lower envelope must be adjusted as with the following algorithm steps (6). The second opposite state Figure-7 considers the deleted state of th k parabola that lead to K parabola from, v[k] is not contained in new lower envelop steps



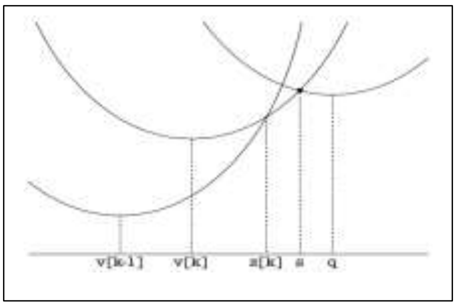




**Figure 7-** state 2.

The following one dimension distance transform algorithm is represented by sample function.

| Algorithm(1) : cuneiform image thresholding<br>Input: gray image<br>Output: Binary image | Step7: |
|---|---|
| Step1: k=0.<br>Step2: v[0] =0.<br>Step3: z[0] =-∞.<br>Step4: z[1] =+∞.<br>Step5: for q= 1 to n-1<br>Step6: s=(( f (q)+ q2) -( f (v[k])+ v[k]2))/(2q - 2v[k]). | 7.1 if s<z[k] Then k=k-1.<br>7.2 Go to 6<br>7.3 Else<br>7.4 K=k+1.<br>7.5 V[q]=q.<br>7.6 Z[k]=s.<br>7.7 Z[k+1]= +∞.<br>Step8: k=0.<br>Step9: for q= 0 to n-1<br>Step 10 while z[k+1] <q<br>K=k+1.<br>$D_f$(q)= (q - v[k])2+ f (v[k]). |

Therefore to implement image distance transform, ***the previous algorithm will be applied on each row in image and reapplied it again on each column's matrix output from initial applied***.

**5. The proposed algorithms**

This section presents the proposed algorithms that related to solving the cuneiforms tablet problems distributed in three sections initially with Binrized proposed algorithm and the proposed algorithms of erosion spots and cuneiform writing lines

**5.1 The proposed thresholding method**

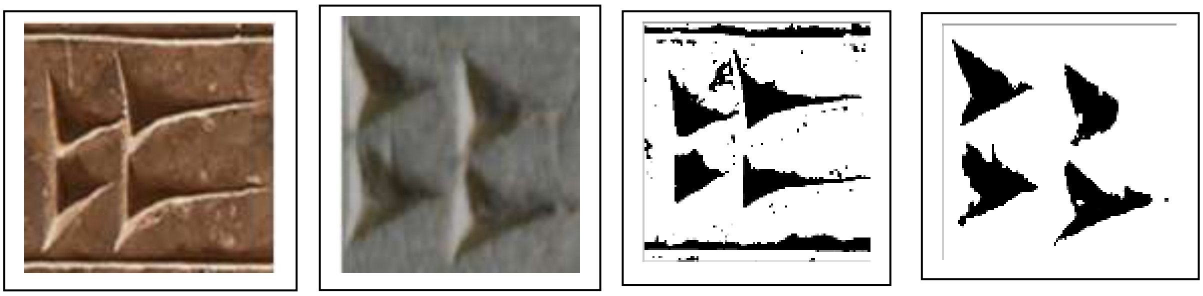
This research proposed a new Binarizaed method with thresholding metric. However, the threshold value is computed depending on (Sauvola, Niblack) method where the selection of each of them is based on computed value of **statistical Skewness metric** as illustrated in following proposed algorithm.

| Algorithm(2) : cuneiform image thresholding<br>Input: gray image<br>Output: Binary image | $I_{Skewness}$= ($I_{Mea}$- $I_{mode}$)\ $I_{SD}$;<br>Step7 : if $I_{Skewness}$≥0 then K = $I_{Mea}$ +$h_1$* $I_{SD}$<br>Else . k= $I_{Mea}$ *(1-$h_2$*(1-( $I_{SD}$ /128))) |
|---|---|
| Step1: read the Gray image grayimage (x,y).<br>Step2: convert the 2D array to 1D array gayimage(:) .<br>Step 3: find the $I_{Mea.\ where\ a=}$mean(grayimage (:).<br>Step4: find the $I_{SD}$.<br>Where $_{SD}$ =standard division(grayimage(:)).<br>Step5: find $I_{mode.}$. $I_{mode}$= mode(grayimage(:)).<br>Step6: compute I $_{Skewness.}$ | where K=threshold value , h1= 0.7 , h2=0.3<br>step 8: applied global thresolding process.<br><br>if grayimage(x,y)≥K<br>grayimage(x,y)=1<br>Else grayimage(x,y) = 0;<br>Step 8: $I_b$= grayimage(x,y).<br>Step 9: return(binary image($I_b$) ). |

And through the above proposed algorithm, the computed thresholding value from Niblack depends where the **Skewness metric** value is positive, conversely the Sauvola method depends where a Skewness is negative.

**5.2. The elimination of rejected elements**

After the binary cuneiform tablets were created by the previous stage, the rustle may contain undesirable elements like spots and lines. However the presence this element without any elimination process, will lead to get a wrong results, in other words these elements will subscript negatively with advanced recognition state. Therefor before reaching to classification state all cuneiform symbols must submitted to filtration task. For illustrating the mentioned content, the browser can recognize the difference between the features that extracted from each binary image below Figure-8 where each of them represents the same character and number of conform symbols, (with different lines and spots), after they submitted to biniriztion process where they are different in features related to number of content elements in each binary image. Therefore for reaching **to stander and uniform form of cuneiform binary image free of any lines or spots,** the next sections review the suggested methods by suitable algorithms distributed with following as two sides.



(a) (b) (c) (d)

**Figure 8-**writing line problem. a, b) : cuneiform images. c ,d) the output of binary images with difference features (number of spots and liens).

### 5.2.1 Removing Spots

The presence of spots in the binary image, as a result of the nature of the material written on it, whether they are stone or clay, this will get the images with different features, as have seen in figure (2.b). However for eliminating the spots from cuneiform binary image based on proposed algorithm satisfy the concept of (**Extraction of Connected Components**).

### 5.2.2 The proposed algorithm to eliminate spots

The proposed algorithm for eliminated spots is based on (CCL) concept as illustrated by previous, with iterated implementation to extract each connected components segments in binary cuneiform image, as a separated labels for each one of them. Therefore to eliminates spots, each separated connected density labels of pixels must be computed and subject to division by total number of pixels in image. However each

CCL in image will stay or removed depending on applied comparison process between the predefined threshold value with corresponding it's divided value which explained in advance.

**Algorithm (3): spots elimination.**
**Input: cuneiform binary image .**
**Output: Binary spots off cuneiform image** .

Step1: read the cuneiform binary image c.
Step2: initialize the label counter (LC) ,LC=1.
~~Step3: initialize the stricture element B.~~
Step4: initialize threshold T=α.
Step5: initialize the index matrix (Indx_mat).
Step6: initialize the labeled matrix (Lab_Mat).
Step7: repeat
Step8: locate the first foreground pixel p and it's location p(x,y).
Step9:k=0;
Step10:intilize the connection component matrix $X_K(0,0)$..
Step11 :set $X_K(x,y)$ =p(x,y).
Step12: **repeat**
Y=Xk
Applied the dilation process on $X_k$ and interest the result with original Image ***IB*** as following formula.

$X_{k+1}$= dilation(B,$X_K$ )∩ ***IB*** .
K=K+1.
**Until** ( Y==$X_{K+1}$)
Step13: save the index of matrix Y in Indx_mat where Y =1.
Step14:set Y=0;
Step15: for each foreground element in $I_b$ correspond Indx_mat , Set $I_{b=0}$;
Step16 :for each element in Lab_Mat correspond Indx_mat , set Lab_Mat =LC .
Step17 : LC=LC+1.
Step18: until(sum ($I_b$)==0).
Step19: initialize the count label matrix(CL)
Step20: applied histogram operation on Lab_Mat to determine the count for each ( pixel label) and save the result in (CL) .
Step21: divide each element in CL by the total number of pixel CL=CL/(Row*Column ).
Step22: for each element in CL with index (i) if (CL(i)>=threshold)
CL(I) =1
Else CL(I) =0.
Step 23: for each element of Lab_Mat update it's value with correspond value of element matric CL where label value= rank row of CL.
Step24:cuniform_spots_off= Lab_ Mat.
Step25:return(cuniform_spots_off).

### 5.2.2 Removing writing lines

This section review the proposed algorithms for removeing unwanted lines in binary cuneiform image to guarantee and satisfy uniform results about cuneiform symbols. However removing the lines symbols from cuneiform binary image is applied after removing spots, where the proposed algorithms basically adopted **distance transform technique** to solve this problem, specifically by sample functions approach.

#### 5.2.2.1 The proposed algorithm to eliminate writing lines

This section will demonstrate the proposed algorithms with their features.

1- Statistical lines remove method.

2- MSE lines remove proposed method.

#### 5.2.2.2: Statistical lines remove proposed Algorithm

This proposed algorithm is related to remove lines depending on the statistical counts of each distance pixels, where the counts of amount of distance pixels values (after applied distance transform) of line symbol is different from the same distance pixels values in cuneiform symbols as seen in the following pseudcode of the statistical line remove proposed Algorithm.

**Algorithm (4) . - Statistical lines remove method. Input: spots off binary cuneiform image.**
**Output: lines off binary cuneiform image.**

Step1: read cuneiform spotoff binary image ($I_{SO}$)
Step2: applied CCL algorithm (1) to determine the number of connected labels α.
Step3: initialize the number of arrays $MAT_i$ equal number of labels α each one like the size (mxn) of $I_{SO}$ .
Step4: separate each connected label(i) in $I_{SO}$ with independent matrix MATi. Where i=1… α.
Step5: for each i set **f**= MATi.
\\ applied distance transform on each MATi
Step6 : for each row of **f** applied the distance sampling procedure with m iterations as the following and save the result about each row matrix (cal)
6.1: k=0.
6.2 v[0] =0.
6.3 z[0] =-∞.
65.4: z[1] =+∞.
6.5: for q= 1 to n-1
6.6: s=(( f (q)+ q2) -( f (v[k])+ v[k]2))/(2q - 2v[k]).
6.7 if s<z[k]
Then k=k-1.
Go to 6.6
Else
6.8: k=k+1.
6.9:V[q]=q.
6.10 : Z[k]=s.
6.11:Z[k+1]= +∞.
6.12: k=0.
6.13: for q= 0 to n-1
6.14 while z[k+1] <q
6.15 K=k+1.
6.16 $D_f$(q)= (q - v[k])2+ f (v[k]).
Step7: applied transpose process on each row in cal \and save the result in transposed matrix (TM).
Step8: set f= TM.
Step9: got to (step6) applied distance sample procedure to recompute cal matrix.
Step10: set distance matrix (Dm) =cal.
Step11: compute the number of distance pixels for each value less than 7 **<take7 as threshold value >**
Step12: compute the number of distance pixels (all distance pixels ) DT .
Step13: compute distance ratio DR where RD = DC/DT.
Step 14: if RD > threshold value then **set MATi=0**;
Step15 :next i. < **with each connected object >**
Step 16: compute lines off matrix ($I_{lo}$) where $I_{lo=}$ $MAT_1$+$MAT_2$+…+MATi.
Step 17:return $I_{lo.}$

#### 5.2.2.3 MSE lines remove proposed method.

The second proposed method for removing cuneiforms liens adopts MSE as decision cratering of unwanted liens, where the MSE is computed **between** the first result matrix which came from applied distance transform sample function on each row of binary image, and second matrix (image distance transform) resulted after applying the previous transform on each column's matrix (output from starting transform). However the basic principle of this algorithm depends on concept that the MSE computed values (as mention) have a distanced verities difference value between vertical and horizontal cuneiform symbols (where the symbols shape limit to line). **That means MSE values for**

**cuneiform horizontal linas with previous proposed concept has low value compared with it's length.**

**Algorithm (5) . MSE lines remove proposed method. .**
**Input: spots off binary cuneiform image.**
**Output: lines off binary cuneiform**

Step1:read cuneiform spotoff binary image ($I_{SO}$)
Step2: apply image labeling algorithm (1) to determine the number of connected labels α.
Step3: initialize the number of arrays $MAT_i$ equal number of labels α each one like the size (mxn) of $I_{SO}$ .
Step4: separate each connected label(i) in $I_{SO}$ with independent matrix MATi.
Where i=1… α.
Step5: for each i set f= MATi.
< **applied distance transform on each MATi>**
Step6 : for each row of f applied the distance sampling procedure with m iterations as the following and save the result about each row in calcul
ate matrix (cal)
6.1: k=0.
6.2: v[0] =0.
6.3 z[0] =-∞.
6.4: z[1] =+∞.
6.5: for q= 1 to n-1
6.6: s=(( f (q)+ q2) -( f (v[k])+ v[k]2))/(2q - 2v[k]).
6.7: if s<z[k]
Then k=k-1.
Go to 6.6
Else
6.8:K=k+1.
6.9:V[q]=q.
6.10:Z[k]=s.
6.11:Z[k+1]= +∞.
6.12: k=0.
6.13: for q= 0 to n-1
6.14 while z[k+1] <q
6.15 K=k+1.
6.16.D $_f$(q)= (q - v[k])2+ f (v[k]).
Step7: applied transpose process on each row in cal and save the result in transposed matrix (Tm).
Step8: set f= TM.
Step9: got to (step6) applied distance sample procedure to recompute cal matrix.
Step10: set distance matrix (Dm,) =cal.
Step11: compute mean square error mse= MSE( Dm, Tm,)
Step12: compute the length of cuneiform symbols cun_l .
Step13: if mse < cun_l set MATi=0
Step14 :next i.
Step 15: compute lines off matrix ($I_{lo}$) where $I_{lo=}$ $MAT_1+MAT_2+…+MATi$.
Step 16:return $I_{lo.}$

**6. Results and discussion**:
**a.Thresholding**

This section reviews the results that are achieved after applying the proposed algorithm to test and demonstrate the proposed algorithms efficiency where the test data consist of 85 images taken from the Iraqi Museum from the Assyrian Hall. Therefore, each proposed algorithm illustrated will be tested with correspond there problems as follows:

when applying the thresolding techniques by previous (Otsu's, Niblack, Sauvola and Iterative) methods the browser can see the drawback result that related especially to first and last one as seen in Figure-9.

(a) (b) (c) (d) (e)

**Figure 9-** a) original cuneiform image, b,c) the binrizeed images result from Otsu's and Iterative *method* , d,e*)* binrizeed images result from *Niblack, Sauvola method.*

Generally, this problem about thresholding technique is appeared clearly where the cuneiform image's histogram is distributed has getting closer near the dark colors tone as seen in Figure-10

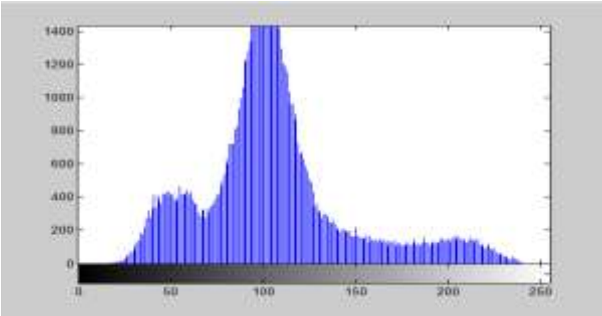


**Figure 10**-image color histogram closed to dark area.

where the binrized clear results came from (Niblack, *Sauvola*). However to select the best one of them through by proposed algorithm (2) depending on statistical **Skewness metric** then the binrized binary image resulted from Niblack method is selected related to its positive Skewness value subjected to the algorithm's condition where the values of (Skewness, mean, standard division, mode) are (0.0917, 113.2, 45.9, 109) respectively

**b. The elimination of rejected elements.**

After the binary cuneiform tablets were created by the previous stage, the results may contain undesirable elements like spots and lines. However the presence of this element without any elimination process, will lead to get a wrong results, in other words, these elements will subscript negatively with advanced recognition state. Therefor before reaching to classification state, all cuneiform symbols must be submitted to filtration task. Therefore this experiment section has distributed into two sections as follows

**1. Removing Spots:**

The presence of spots in the binary image, as a result of the nature of the material written on it, whether they are stone or clay, thus will get the images with different features. However for eliminating the spots from cuneiform binary image based on proposed **algorithm (3)** that satisfy the concept of (**Extraction of Connected Components**). Each component in cuneiform image (symbols, lines, spots) their connected pixels will have a **distinguish** value (label), with any size of density pixel. However it can be seen that the results after applying the proposed algorithm on cuneiform binary image are shown in Figure-11(b)

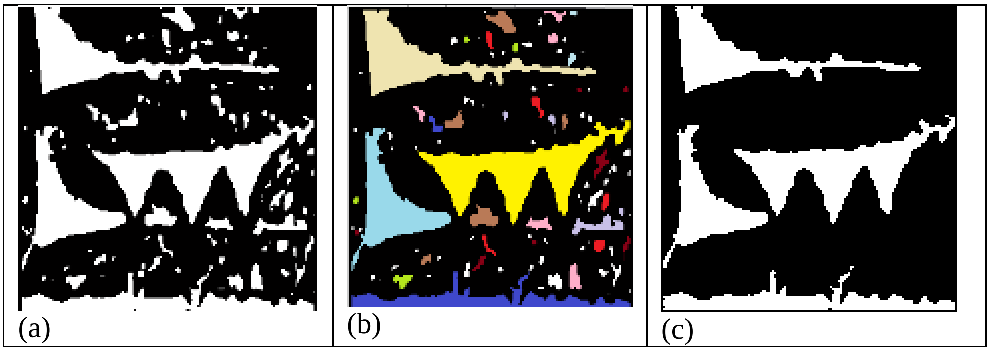



**Figure11**-image labeling . a) binary image , b) image labeled with distinguish color labels

Then the results of proposed algorithm can be seen with the following Figure 11(c) where the predefined threshold value is equal to (0.04), c) binary cuneiform image pure clear from any spots and writing line

Generally when applying the previous procedure as with proposed algorithm with predefining the same threshold equal (0.04). That leads to reach to unwanted results as seen in Figure bellow.

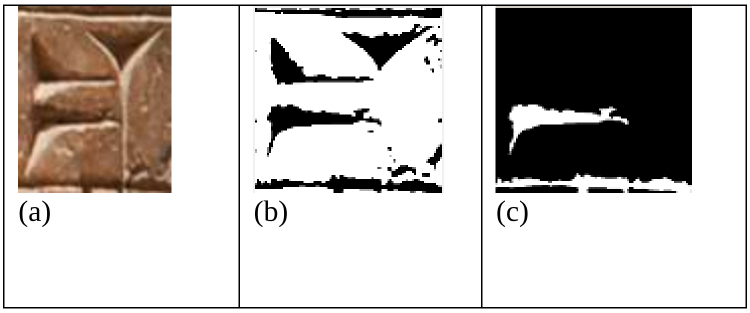



**Figure 12-** threshold problem, a) cuneiform image , b) binary image content spots and lines , c) unwanted result.

As seen with the above figure the threshold value isn't appropriate and it has an effect on eliminating some cuneiform symbols. Therefore the value of threshold must be less than 0.04. However the solving problem of erasing the writing line problem is solved with next proposed algorithm unwanted.

**2.Removing lines**

This section deals with elimination process of writing lines to get lastly a uniform binary cuneiform image for advance recognition task. Therefore the first and second line removed proposed algorithm applied to test. when applying the first statistical algorithm (4) on the previous spot off cuneiform binary image can see the output results as pure binary cuneiform image content just cuneiform symbols where the value of threshold distance equal (0.7) Figure-13(c).

The problem that reflated with this proposed algorithm (4) is the selected value of threshold Figure -13(b). However the next second proposed algorithm decision criterion depends on comparing between the value of MSE for each symbol and its length of density pixels as seen with Figure-(14) , Table-(1).

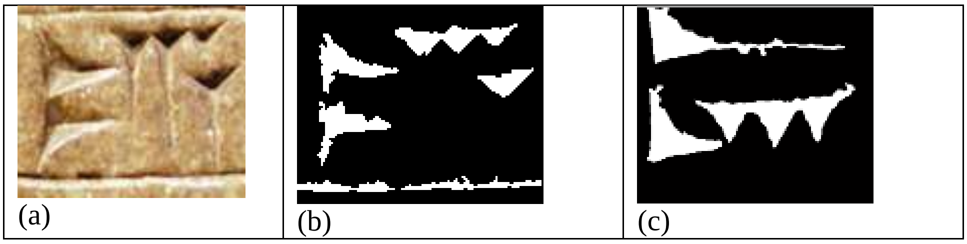



**Figure 13-**the drawback of statistical algorithm. b) the writing line is remain ,c) line off binary cuneiform image.

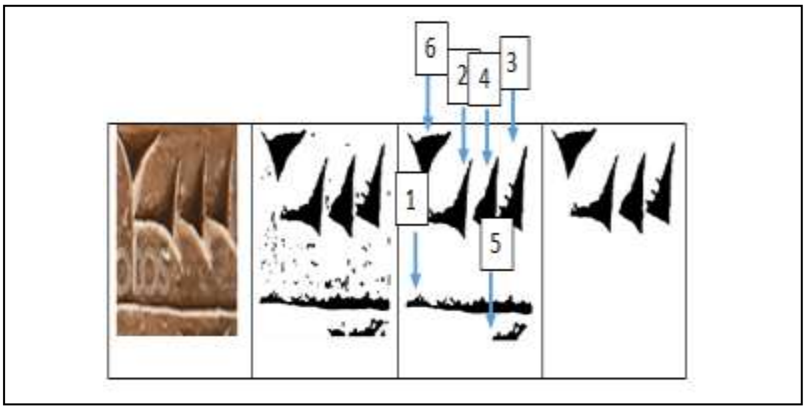



**Figure 14-**Where the values of MSE and length of each cuneiform symbols was illustrate with following tables.

**Table 1-**MSE values for each uniform symbols according to the previous algorithm with their length

| cuneiform symbols NO | MSE | cuneiform symbols Length |
|---|---|---|
| **1** | **130.56** | **436.374** |
| 2 | 823.336 | 117.11 |
| 3 | 971.371 | 156.8 |
| 4 | 1373.3 | 138.102 |
| **5** | **12.62** | **110.89** |
| 6 | 11494.4 | 149.14 |

As seen in previous tables the values of MSE all cuneiforms symbols are greater than the values of length pixels except the first and fifth symbols (line symbols) that will be eliminated from image.

**7. Conclusions**

In this research a new proposed algorithms that related to binrizetion and remove the unwanted object, and keeping them will lead to negative results with advance analysis recognition statute, therefore the aim of this research is to achieve the pure binary cuneiform image of the binrized task through by thresholding. This paper presents the drown back of (uots, iterative ) methods when the color histogram is close to dark interval and the computed thresholding value selected between niblack and salvu methods depending on value of Skewness metric. However about the spots problem this paper proposed a new algorithm eliminates these spots based on CCL concepts applied by image morphology. The last problem about writing lines erosion is solved by image distance transform with sample function was employed with two proposed algorithms.

**References**


1. Hilal, Y. Abdul., M.R. Haithem. Alani. **2006.** "*Cunform Symbols Recognition Using Intensity Curve". The international Arab journal information Technology*, **3**(3), July.
2. Jonathan Cohen, Donald Duncan, Dean Snyder, Jerrold Cooper, Subodh Kumar. **2004**. iClay: Digitizing Cuneiform "the 5th International conference on Virtual Reality, Pages 135-143, Archaeology and Cultural Heritage VAST (2004).
3. Leonard Rothacker, Denis Fisseler, Gerfrid G.W. Müller. **2015.** "Retrieving Cuneiform Structures in a Segmentation-free Word Spotting Framework" ACM 2015 Article.
4. Sean Eron Anderson Marc Levoy. **2002**. Unwrapping and Visualizing Cuneiform Tablets. *IEEE Computer Graphics and Applications*, **22**(6): 82-88.
5. Fahimeh Mostofi, Adnan Khashman. **2014.** "Intelligent Recognition of Ancient Persian Cuneiform Characters". Intelligent Systems Research Center, 1Department of Computer Engineering. August 2014.
6. Raed Majeed, Zou Beiji, Hiyam Hatem, and Jumana Waleed. **2015**. "Ancient Cuneiform Text Extraction Based on AutomaticWavelet Selection", *International Journal of Multimedia and Ubiquitous Engineering*, **10**(6): 253-264.
7. Puneet. Naresh Kumar Garg. **2013**. Binarization Techniques used for Grey Scale Images, *International Journal of Computer Applications* (0975–8887)Volume 71– No.1, June 2013.5
8. Patil Priyanka Vijay, Patil Priyanka Vijay. **2016.** Gray Scale Image Segmentation using OTSU Thresholding Optimal Approach. *Journal for Research*, **02**(05): 2395-7549.
9. Jamileh Yousefi. **2011.** "Image *Binarization using Otsu Thresholding Algorithm"*. University of Guelph, Ontario, Canada. April 18.
10. Sheenam Bansal, Raman Maini, **2013.** "A Comparative Analysis of Iterative and Ostu's Thresholding Techniques", *International Journal of Computer Applications* (0975 – 8887)Volume 66– No.12, March 2013.3
11. Rashmi SainiDocument, **2015**. "Image Binarization Techniques,Developments and Related Issues",: A Review. *International Journal of Computer Applications* (0975–8887) Volume 116 – No. 7, April 2015 4.
12. Rafael C. Gonzalez. **2001**. *Digital Image Processing*, Second Edition. Prentice Hall 2001.
13. Lifeng He, Yuyan Chao, Kenji Suzuki, Kesheng, Wu. **2009.** "Fast connected-component labeling". *Pattern Recognition*, **42** (2009): 1977 – 1987.

**14.** Phaisarn Sutheebanjard, Wichian Premchaiswadi. **2011**. "Efficient scan mask techniques for connected components labeling algorithm", *Sutheebanjard and Premchaiswadi EURASIP Journal on Image and Video Processing*, **12**(-606).

**15.** Muhammad H. Alsuwaiyel, Marina Gavrilova, **2000.** "On the Distance Transform of Binary Images ". The 2000 international conference on imaging science system and Technology. Vol. I. Las Vegas, Nev. p 83–6.

**16.** Donald Graeme Bailey**. 2004.** " An Efficient Euclidean Distance Transform" , Conference Paper *in* Lecture Notes in Computer Science · December 2004.

**17.** Donald G Bailey. **2012**. "Accelerating the Distance Transform, "International Conference on Image and Vision Computing New Zealand. 2012

**18.** Pedro, F., Felzenszwalb. **2012.** "Distance Transforms of Sampled Functions", *Theory Of Computing*, **8**(2012): 415–428.