Advances in the Development of an Algorithm for Parametric Identification of Egyptian Hieroglyphs Using Artificial Vision

Rafael Bolívar León*, César Augusto Peña, and Gonzalo Guillermo Moreno

Faculty of Engineering and Architecture, University of Pamplona, Colombia;
Email: cesarapc@unipamplona.edu.co (C.A.P.), gmoren@unipamplona.edu.co (G.G.M.)
*Correspondence: rbolivarl@unipamplona.edu.co (R.B.L.)

***Abstract*—This article presents the development of an algorithm for identifying Egyptian hieroglyphs written on papyri. For the development of the algorithm, the implementation of parametric artificial vision techniques allowed the reduction of computational power required. A study of the main morphological characteristics used in artificial vision was carried out, some relevant ones were selected, and others were adapted to be normalized and quantified quickly. It was shown that the established characteristics allow the differentiation and identification of the hieroglyphs of the ancient Egyptian alphabet. The developed algorithm has the advantage that it allows to differentiate characters, regardless of their initial size.**

***Keywords*—hieroglyphs, artificial vision, parametric identification, optical character recognition, alphabet**

I. Introduction

Egyptian culture has been characterized as a topic of global interest. The significant advances that developed millennia ago have captivated the interest of many scientists. Several constructions indicate that their authors used very advanced tools for their time. Many researchers have spent their lives unraveling the secrets of ancient Egyptian culture [1–6]. A primary tool in interpreting Egyptian culture is the hieroglyphs written on obelisks, monuments, caves, stones, and papyri.

This article proposes a preliminary step for the semantic interpretation of writings from the ancient Egyptian culture to identify the most relevant symbols known as the ancient Egyptian alphabet. Several works have been presented that contribute to interpreting these symbols. The researches [7, 8] present the application of deep learning techniques for hieroglyph recognition; Barucci *et al.* [7] present augmented reality and machine learning as identification tools for these symbols. This study developed a specifically dedicated Convolutional Neural Network (CNN) named Glyphnet, tailoring its complexity to its classification task. Moustafa *et al.* [8] show Deep learning techniques, such as EfficientNet, MobileNet, and ShuffleNet. This study has been tested on two hieroglyph datasets. This paper describes a flutter-based mobile application named Scriba. This application provides as an advantage an exact translation of hieroglyphs. Arrivault *et al.* [9] presented a Fuzzy Hierarchical Attributed Graph (FHAG) approach for handwritten Egyptian hieroglyphs. This paper combines techniques in hierarchical modeling, and fuzzy grammar definitions seem natural; Ly *et al.* [10] illustrates the use of graphics incorporating the skeletonization of hieroglyphs. The application of a technique based on obtaining a Histogram of Oriented Gradients is presented in [11]. This algorithm offers high accuracy as an advantage but requires post-processing, which can be a disadvantage. Additionally, several relevant works can be found, among which can be highlighted [12–14]. The goal of the research [12] is to discover the linear order of a ancient two-dimensional script, Hieroglyphic Luwian. This paper records a complete decipherment process, including encoding, modeling, parameter learning, optimization, and evaluation. This algorithm has an elevated computational cost. Iglesias-Franjo and Vilares [13] manage hieroglyphic texts of Ancient Egyptians, considering two peculiarities: the lexical and encoding level for its application in Egyptology and Digital Heritage. Pinilla-Buitrago *et al.* [14] presented a method to extract segments from hieroglyphs by encoding the extracted segments through local descriptors. This study proposes to use them under the technique named the Bag of Visual Words (BoVW) model.

Previous works make a significant contribution to the field of study. However, most of them employ techniques that require significant computational resources. There are some advances in this regard, as is the case of the Scriba application presented in [8].

In this article, an algorithm with a low computational cost is proposed. This algorithm allows the selection of identification characteristics. This allows the user to determine the degree of precision in the identification according to the computational power with which it counts. This algorithm is an advance for the development of real-time applications for both mobile devices and portable augmented reality systems. Other contributions proposed in this article are the normalization of the

Manuscript received February 9, 2023; revised April 20, 2023; accepted April 28, 2023; published August 11, 2023.

characteristics that describe the symbols to a range [0, 1] and the creation of a new descriptor called the difference index. These contributions allow the parametric identification of symbols independently of their size.

This article is distributed as follows: the second section illustrates the methodology used to segment and identifies hieroglyphs corresponding to the ancient Egyptian alphabet. Section three presents the results with their corresponding discussion, and conclusions and bibliography are presented.

II. Methodology

The interpretation of Egyptian hieroglyphs is a reasonably broad and highly complex problem. Hieroglyphs are usually found in different materials, such as stone, ceramics, wood, and papyrus, among others. Visual identification will be significantly different depending on the material used to elaborate them. This problem is because each material provides a distinct colors and textures. On the other hand, many hieroglyphs can change their horizontal orientation to show a different meaning in what they want to express. In this article, the subdivision of the problem is proposed, and initially, the identification of well-defined hieroglyphs written on papyri in good condition is taken as an objective. Given the great variety of symbols, the identification of 24 is taken, known as the hieroglyphs of the ancient Egyptian alphabet (See Fig. 1).

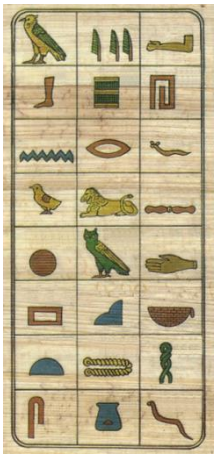


Figure 1. Image of the ancient Egyptian hieroglyphics alphabet.

As a methodological framework, the development of an algorithm using artificial vision techniques is proposed for the identification of symbols. This algorithm must segment each of the symbols and generate a quantification through parametric characteristics that allow the differentiation between them.

For the development of the algorithm, three stages were proposed. The first stage involves identifying the background corresponding to the papyrus. The second corresponds to the segmentation of the symbols correcting the empty spaces. Finally, in the third stage, the identification characteristics of each symbol are selected.

In order to carry out the first stage of the algorithm, a frequency diagram or histogram of the three channels of an image was made [15]. The processed image corresponds to the ancient Egyptian hieroglyphic alphabet. The three channels correspond to the red, green, and blue color components. This histogram is shown in Fig. 2. In each channel, it is possible to identify the superposition of distributions with a Gaussian trend. The one with the highest height is supposed to correspond to the bottom of the papyrus.

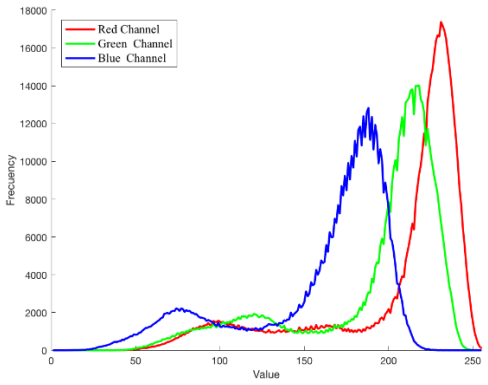


Figure 2. Histogram of the papyrus with hieroglyphs.

By taking the histogram data of each image channel as a reference, a Threshold can be performed on each color component to eliminate the pixels corresponding to the background. This Threshold is illustrated in Fig. 3. For the red channel, the pixels located in the range 170 to 255 were removed. For the green channel, the pixels greater than 160 and less than 255 were removed. Last, for the blue channel, the pixels range from 120 to 255. These ranges are variable parameters that depend on the conditions of the papyrus on which the hieroglyphs were written. Notice that in the red channel, several pixels corresponding to hieroglyphs were confused with the color of the papyrus (see the birds, the lion). The same effect can be seen in the green channel, although attenuated; however, there is a higher percentage of noise in the image. Finally, greater effectiveness was achieved in the blue channel, although it could have been better.

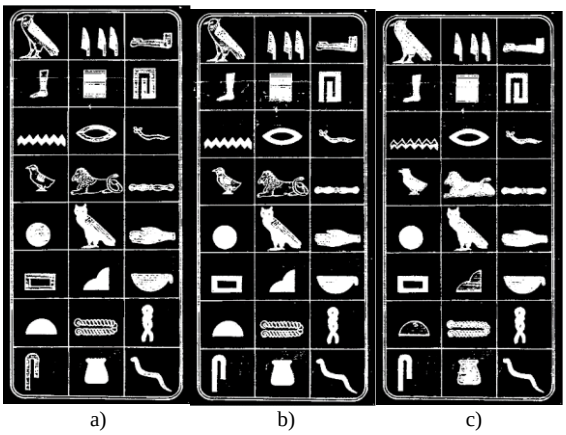
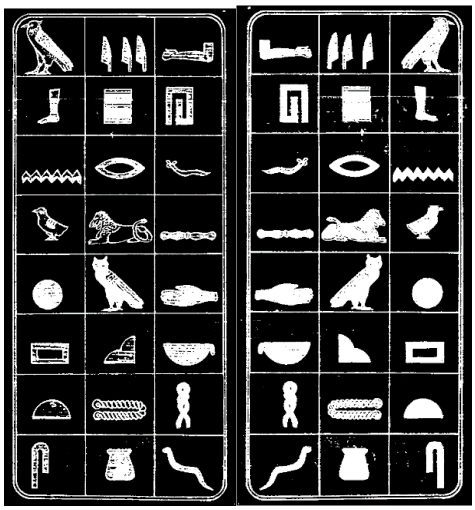


Figure 3. Threshold of image channels using RGB color space. a) Channel R; b) Channel G; c) Channel B.

To increase the recognition of the background, one could think of expanding the ranges of the color channels; however, as the symbol recognition improves, the presence of noise also increases. In this order of ideas, it is proposed to work with the predefined ranges for the channels and use logical operations for their combination [16]. For example, Fig. 4 shows the logic operations: intersection (AND) and union (OR) of the three channels simultaneously. The AND operation results correspond to an intersection of the pixels; therefore, a reasonably high restriction is obtained. Therefore, some pixels (that correspond to the interior of the symbols) are identified as background due to color similarity. The logical OR operation is equivalent to the union of the pixels in the three channels. In this case, good recognition of the pixels of the hieroglyphs can be seen, although with the presence of noise (seen the second row of hieroglyphs).



a) b)

Figure 4. Application of logical operations on the channels of the image. a) AND operation; b) OR operation.

Another alternative is proposed to increase the efficiency of the background recognition algorithm. This alternative was based on the Euclidean error between the most characteristic background color and each pixel in the image. This coordinate system is based on the RGB ortho-normal space. Eq. (1) presents the estimation of this error.

$$\text{Error}_{pixel} = \sqrt[2]{\left(R_{ij} - R_0\right)^2 + \left(G_{ij} - G_0\right)^2 + \left(B_{ij} - B_0\right)^2} \qquad (1)$$

where the Error variable is the Euclidean error of the pixel to the target color, R is the red color component, G is the green color component, B is the blue color component, i is the vertical position of the pixel in the image, j is the horizontal position of the pixel in the image, and subscript 0 denotes the values corresponding to the target color.

Fig. 5 shows the results of applying the Euclidean error equation to each image pixel. The maximum error found was 314.89, and the minimum was zero. Pixels with minor errors correspond to those most similar to the background.

Subsequently, the conversion of the error image to a binary image is carried out, applying a threshold, where the values greater than 110 will become one and the lesser ones zero.

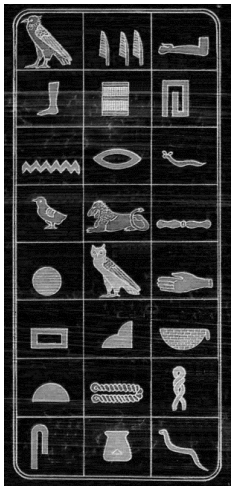
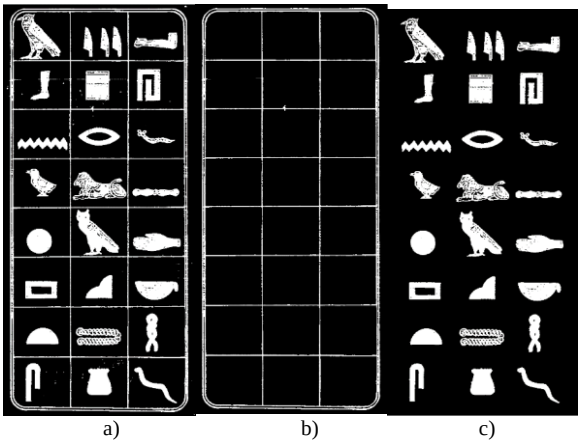


Figure 5. Error to a target color.

The result of this thresholding is presented in Fig. 6a. Additionally, the removal of the sets with less than 10,000 pixels was applied to obtain the image in Fig. 6b. In this image, significant elements, such as dividing lines, that do not correspond to hieroglyphs can be observed. In many papyri of an artistic nature, prominent figures appear that are not part of the symbols. The objective is to generate an image with the elements of large dimensions that are not required to be identified and remove them from the first binary image shown in Fig. 6a. The subtraction of these two images is illustrated in Fig. 6c, where the symbols corresponding to the hieroglyphs can be seen. It is noteworthy that in this image, sets of pixels with less than 100 elements were also eliminated to attenuate the image's noise.

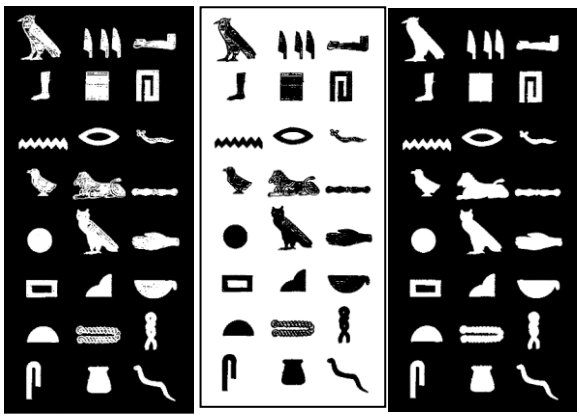


a) b) c)

Figure 6. Binary image of the error a) Initial; b) Elements with more than 100 pixels; c) Elements with a size according to the symbols.

In the algorithm's second stage, the filling problem of the hieroglyphs' internal holes was solved for their

subsequent identification. Note in Fig. 7a) that there are hieroglyphs with predefined holes that are part of the symbol. One case is the rectangle, where the hole cannot be eliminated. On the contrary, in the first hieroglyph exists noise in the background identification. This problem can be solved by making the following steps: first, the binary image was inverted, as shown in Fig. 7b), the sets with less than 500 pixels were eliminated, and finally, the image was inverted again, as shown in Fig. 7c). This set of instructions avoids the elimination of large holes that are part of the symbol and leads to the elimination of the holes corresponding to noise. The results obtained using this technique are much higher than those obtained in Fig. 4.



a) b) c)

Figure 7. Elimination of internal noise without elimination of significant holes. a) Initial image; b) Inverted image; c) Image without holes.

In the third stage of the algorithm, the identification characteristics of each of the hieroglyphs are selected and obtained. The most common features that are often used in computer vision for black-and-white regions are Area, Centroid, Bounding Box, Subarray Index, Major Axis Length, Minor Axis Length, Eccentricity, Orientation, Convex Hull (a matrix that specifies the smallest convex polygon that can contain the region), Convex Image, Convex Area, Filled Image, Filled Area, Euler Number, Extrema (a matrix that specifies the extrema points in the region), Equivalent Diameter, Solidity, Extent, Pixel Index List, Pixel List and Perimeter [17–19].

The most relevant features must be selected to make the algorithm more efficient at a computational level. Additionally, it must be taken into account that, as far as possible, the characteristics must be chosen or adapted so that they are as general as possible. For instance, the magnitude of the hieroglyphs will depend on the size of the papyrus where it was written. This feature can be compared to the font size in modern texts. Therefore, a new normalized index was generated. This index calculates all areas of hieroglyphs present in the papyrus and divides them over the area of the largest hieroglyph. This feature will allow having values greater than zero and less than one that describe the area of the hieroglyph to be analyzed concerning the entire alphabet. A proportion was calculated between each perimeter hieroglyph and the alphabet's most significant magnitude.

Another selected feature is derived from two parameters of the Bounding Box item. This feature corresponds to the division of the height over the width of the hieroglyph. This measure may contain values more significant than one. However, it will be invariant to the general size of the symbol. This feature is because a proportion is followed regardless of size. In Fig. 8, the Bounding Box of the hieroglyphs of the ancient Egyptian alphabet is illustrated. Additionally, a number is observed, which will be taken as a reference for identification in this document.

Eccentricity allows us to determine how circular a hieroglyph is, and its value ranges between zero and one. This feature will also be invariant to hieroglyph size. In order to follow the same normalization guidelines, the following characteristics were selected: Euler Number, Orientation, and Fill. The Euler Number is a scalar that describes the number of objects in the region minus the number of holes in those objects. The orientation was taken as an index ranging from zero to one. The orientation is measured to the region's greater axis concerning the horizontal (greater axis of an ellipse containing all the pixels in the region). A value of 0 corresponds to an orientation of −90º, and a value of 1 corresponds to 90º.

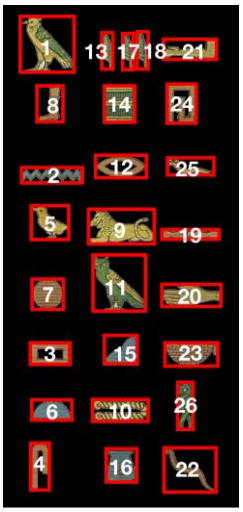


Figure 8. Hieroglyphs characterization.

Another index, called Fill, was established as a ratio between the number of pixels the hieroglyph has over the total area of the Bounding Box. This feature will also be invariant to the size of the hieroglyphs. In this way, seven standardized characteristics are completed that describe each of the hieroglyphs of the ancient Egyptian alphabet. Table I shows the numerical values of these seven characteristics for the alphabet symbols.

A new "difference index" was generated to improve the quantitative description of hieroglyphs through features. This index was obtained by converting the images (containing each of the hieroglyphs) to a size of nine pixels as a preliminary step. This conversion consists of dividing the image horizontally and vertically into three sections of equal size. Subsequently, the pixels with logical value 1 of each section are added and divided over

the total area of the section. Finally, the two 3×3 images of the two hieroglyphs to be compared are subtracted, and the absolute value of all the resulting pixels is summed. This operation can be defined by employing Eq. (2).

$$\text{Difference index} = \sum_{r=1}^{3}\sum_{c=1}^{3}|Image_1(pixel\ r,c) - Image_2(pixel\ r,c)|\ /9 \quad (2)$$

where r is the row number, c is the column number, and images 1 and 2 are the images of the same size to be compared.

Note that the 3×3 size images provide relevant information about the percentage of pixels with logical value in one of the original images.

TABLE I. NORMALIZED CHARACTERISTICS OF THE SYMBOLS OF THE ANCIENT EGYPTIAN ALPHABET

Symbol	Area	Eccentricity	H/W	Perimeter	Euler Number	Orientation	Fill
1	0.989	0.845	0.009	0.864	1	0.180	0.437
2	0.412	0.984	0.109	0.723	1	0.500	0.584
3	0.453	0.822	0.348	0.445	0	0.504	0.705
4	0.402	0.914	0.967	0.624	1	0.947	0.613
5	0.427	0.866	0.219	0.477	1	0.233	0.447
6	0.463	0.844	0.403	0.377	1	0.502	0.758
7	0.554	0.162	0.358	0.360	1	0.367	0.774
8	0.292	0.928	0.127	0.437	1	0.924	0.414
9	1.000	0.900	0.127	0.850	1	0.459	0.582
10	0.670	0.920	0.290	0.918	1	0.496	0.729
11	0.966	0.901	0.195	0.792	1	0.215	0.435
12	0.426	0.891	0.121	0.429	0	0.501	0.498
13	0.196	0.957	0.103	0.313	1	0.042	0.646
14	0.763	0.586	0.109	0.481	1	0.985	0.913
15	0.429	0.811	0.414	0.402	1	0.705	0,612
16	0.629	0.579	0.596	0.453	1	0.023	0.822
17	0.186	0.959	0.111	0.302	1	0.043	0.647
18	0.208	0.958	0.100	0.318	1	0.048	0.670
19	0.316	0.991	0.160	0.484	1	0.499	0.757
20	0.675	0.940	0.193	0.566	1	0.506	0.690
21	0.442	0.965	0.027	0.537	1	0.521	0.548
22	0.303	0.982	0.348	0.564	1	0.294	0.174
23	0.685	0.894	0.261	0.504	1	0.507	0.738
24	0.628	0.657	0.111	1.000	1	0.008	0.733
25	0.200	0.974	0.135	0.432	1	0.418	0.336
26	0.394	0.958	1.000	0.519	1	0.999	0.720

Additionally, converting all the images to the same size allows the possibility of comparing them, which is done through the subtraction indicated in Eq. (2). The absolute value allows calculating and estimating the magnitude of the error regardless of the sign, and dividing the sums by nine allows obtaining a range of results between zero and one.

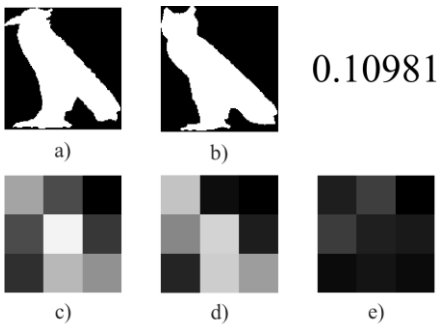


Figure 9. Example 1 - Hieroglyph Comparison. a) B/W image of the first hieroglyph; b) B/W image of the Second hieroglyph; c) 3×3 image of the first hieroglyph; d) 3×3 image of the Second hieroglyph; e) Absolute value of the subtraction of the 3×3 images.

Fig. 9 shows the application of the algorithm to compare two very similar symbols; In this case, it corresponds to two birds. The reduced images to a size of 3×3 can be observed in Fig. 9c and 9d. The resulting absolute value obtained from the subtraction of the 3×3 images are shown in Fig. 9e. By applying Eq. (2) in this case results in the value of 0.10981. This index allows us to know that the images are similar. However, they have a different percentage greater than 10%. This value is significant in the process of identifying hieroglyphs.

Fig. 10 shows a second example of comparing two similar symbols. In this case, the central pixel of the two 3×3 images are similar. However, a significant difference is noted in the others, allowing a difference of 26.598%.

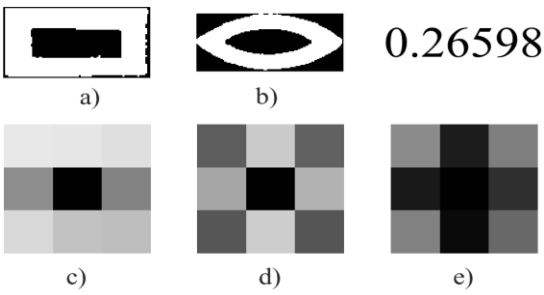


Figure 10. Example 2 - Hieroglyph Comparison. a) B/W image of the first hieroglyph; b) B/W image of the Second hieroglyph; c) 3×3 image of the first hieroglyph; d) 3×3 image of the Second hieroglyph; e) Absolute value of the subtraction of the 3×3 images.

Fig. 11 illustrates the comparison of two entirely different symbols, achieving a difference of 40.123%. In

the 3×3 images, the difference in the distribution of the pixels is noticeable. Finally, Fig. 12 shows an example of comparing two images with very different height and width proportions, which shows the applicability of the proposed algorithm.

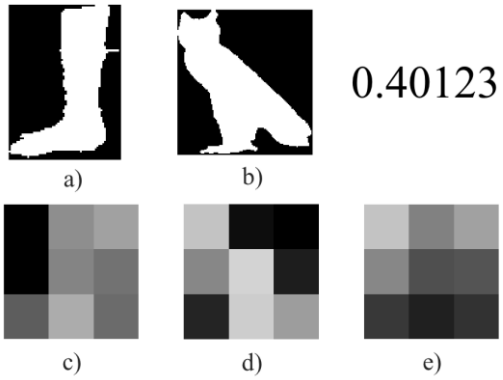


Figure 11. Example 3 - Hieroglyph Comparison. a) B/W image of the first hieroglyph; b) B/W image of the Second hieroglyph; c) 3×3 image of the first hieroglyph; d) 3×3 image of the Second hieroglyph; e) Absolute value of the subtraction of the 3×3 images.

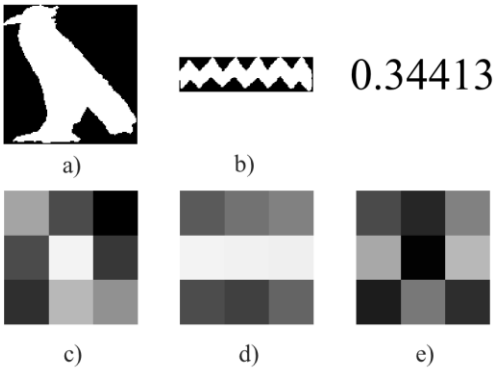


Figure 12. Example 4 - Hieroglyph Comparison. a) B/W image of the first hieroglyph; b) B/W image of the Second hieroglyph; c) 3×3 image of the first hieroglyph; d) 3×3 image of the Second hieroglyph; e) Absolute value of the subtraction of the 3×3 images.

With the difference index, eight normalized characteristics were completed that allow the identification of hieroglyphs. For this process, the error between the hieroglyphs characteristics detected concerning the predefined database is calculated, as expressed in Eq. (3). It should be noted that if the images were similar, the error value would be low; otherwise, its value would increase. This procedure is applied between the hieroglyph image to be identified and all the images included in the database. The smallest error will indicate the hieroglyph to which it most closely resembles.

$$\text{Error} = \sqrt[2]{\sum_{i=1}^{n}(CHI_i - CHD_i)^2} \tag{3}$$

where CHI is a characteristic of the hieroglyph to identify, CHD is a characteristic of the database hieroglyph, n is the total number of characteristics to consider, and *i* is the type of characteristic. 1) Area, 2) Eccentricity, 3) H/W, 4) Perimeter, 5) Euler Number, 6) Orientation, 7) Fill, 8) Difference Index.

According to the described methodology, the pseudocode of the proposed algorithm is shown as following. It is noteworthy that the reduction in the number of characteristics that it is desired to use allows a reduction of computational cost (to the extent that the user considers it pertinent).

Algorithm 1. Parametric Identification of Egyptian Hieroglyphs

Step 1: Identification of the image background

1: Start
2: Read image (RGB)
4: For $i \leftarrow 1$ up to the number of RGB rows
5: For $j \leftarrow 1$ up to the number of RGB columns
6: $\text{Error}_{pixel} = \sqrt[2]{(R_{ij} - R_0)^2 + (G_{ij} - G_0)^2 + (B_{ij} - B_0)^2}$
7: If $\text{Error}_{pixel} < 110$, then
The pixel of row i and column j of the
binary image (BW) takes a value of 0
indicating that it belongs to the background.
8: Otherwise:
The pixel takes the value of 1, indicating the
possibility of being a symbol

Stage 2. Segmentation of symbols and filling of empty spaces

9: Elimination of the sets with a size greater than 10000 pixels from the BW binary image
10: Noise reduction in the image, eliminating groups with less than 100 pixels
11: Filling the internal holes of the hieroglyphs

Stage 3. Selection and obtaining of the identification characteristics of each one of the symbols

12: Selection of the characteristics (CHI) with which you want to execute the algorithm (Area, Eccentricity, H/W, Perimeter, Euler Number, Orientation, Fill, difference index)
13: Characteristic normalization (Conversion to a range [0-1])
14: Calculation of the error between the characteristics of the hieroglyphics to be identified and the database of the ancient Egyptian alphabet. Error $= \sqrt[2]{\sum_{i=1}^{n}(CHI_i - CHD_i)^2}$
15: Identification of the hieroglyph by selecting the symbol with the least error.
16: End

III. Results

To analyze the effectiveness of the new index, called the difference index, its application was proposed to compare all the hieroglyphs of the ancient Egyptian alphabet in the database. Tables II and III present the results of all the comparisons. It can be observed that the lowest error values correspond to the comparison of symbols 13, 17, and 18. This error occurs because they are the same symbol written repeatedly; these combined symbols form a single hieroglyph, which does not produce a confusion problem. The other values demonstrate a clear differentiation between the symbols.

As shown in the previous section, the error calculation identifies the hieroglyphs by comparing preselected characteristics 1) Area, 2) Eccentricity, 3) H/W, 4) Perimeter, 5) Euler Number, 6) Orientation, 7) Fill, 8) Difference Index. The number of these features can vary according to the computational power that the user wants to use. Similarly, the effectiveness will vary depending on the number of features it wants to employ. To measure the algorithm's efficiency, it takes the calculation of the average of the minimum value of the errors when comparing all the hieroglyphs of the alphabet (database).

TABLE II. DIFFERENCE INDEX OF THE SYMBOLS OF THE ANCIENT EGYPTIAN ALPHABET—PART I

Symbol	1	2	3	4	5	6	7	8	9	10	11	12	13
1	0.00												
2	0.34	0.00											
3	0.46	0.51	0.00										
4	0.55	0.30	0.33	0.00									
5	0.09	0.31	0.46	0.51	0.00								
6	0.38	0.31	0.32	0.41	0.38	0.00							
7	0.36	0.20	0.38	0.35	0.33	0.18	0.00						
8	0.32	0.33	0.39	0.44	0.33	0.41	0.37	0.00					
9	0.22	0.32	0.36	0.43	0.22	0.24	0.29	0.40	0.00				
10	0.36	0.35	0.18	0.28	0.35	0.21	0.22	0.31	0.25	0.00			
11	0.11	0.40	0.44	0.56	0.15	0.41	0.40	0.40	0.18	0.33	0.00		
12	0.38	0.29	0.27	0.36	0.33	0.28	0.26	0.28	0.33	0.25	0.39	0.00	
13	0.28	0.22	0.42	0.40	0.28	0.19	0.22	0.42	0.25	0.29	0.33	0.30	0.00
14	0.46	0.36	0.23	0.30	0.45	0.18	0.17	0.49	0.32	0.18	0.46	0.40	0.30
15	0.43	0.41	0.41	0.49	0.45	0.32	0.40	0.29	0.32	0.31	0.44	0.46	0.46
16	0.37	0.33	0.27	0.35	0.36	0.09	0.15	0.43	0.22	0.17	0.39	0.31	0.21
17	0.29	0.22	0.43	0.39	0.28	0.21	0.21	0.43	0.26	0.30	0.33	0.32	0.02
18	0.27	0.23	0.42	0.37	0.27	0.22	0.22	0.44	0.25	0.29	0.30	0.33	0.05
19	0.32	0.17	0.37	0.28	0.33	0.23	0.14	0.37	0.23	0.21	0.38	0.33	0.22
20	0.37	0.21	0.42	0.35	0.31	0.17	0.13	0.40	0.27	0.28	0.41	0.25	0.22
21	0.40	0.25	0.44	0.41	0.41	0.29	0.30	0.29	0.23	0.31	0.36	0.36	0.30
22	0.26	0.41	0.60	0.51	0.27	0.59	0.58	0.38	0.40	0.54	0.26	0.42	0.45
23	0.39	0.30	0.30	0.32	0.33	0.28	0.19	0.40	0.36	0.25	0.40	0.29	0.28
24	0.34	0.30	0.21	0.29	0.34	0.25	0.25	0.32	0.20	0.12	0.32	0.29	0.28
25	0.16	0.43	0.41	0.53	0.22	0.47	0.45	0.36	0.25	0.38	0.11	0.31	0.37
26	0.32	0.29	0.24	0.34	0.33	0.20	0.20	0.34	0.23	0.18	0.37	0.33	0.21

TABLE III. DIFFERENCE INDEX OF THE SYMBOLS OF THE ANCIENT EGYPTIAN ALPHABET—PART II

Symbol	14	15	16	17	18	19	20	21	22	23	24	25	26
14	0.00												
15	0.37	0.00											
16	0.12	0.35	0.00										
17	0.30	0.48	0.22	0.00									
18	0.27	0.49	0.22	0.07	0.00								
19	0.20	0.37	0.19	0.23	0.19	0.00							
20	0.26	0.43	0.22	0.22	0.23	0.17	0.00						
21	0.38	0.21	0.33	0.31	0.32	0.25	0.33	0.00					
22	0.72	0.54	0.63	0.44	0.47	0.56	0.52	0.46	0.00				
23	0.18	0.48	0.24	0.29	0.28	0.27	0.24	0.42	0.59	0.00			
24	0.19	0.27	0.21	0.29	0.26	0.19	0.30	0.24	0.54	0.23	0.00		
25	0.56	0.48	0.47	0.38	0.38	0.43	0.45	0.37	0.20	0.49	0.39	0.00	
26	0.20	0.40	0.16	0.22	0.23	0.18	0.26	0.32	0.52	0.24	0.21	0.43	0.00

In other words, one hieroglyph is taken and compared to all the others; the minimum error value is noted. The second hieroglyph is taken and compared with the others until all the symbols of the alphabet are compared. Finally, the average of the previously calculated minimum values is estimated, and this data is taken as a reference.

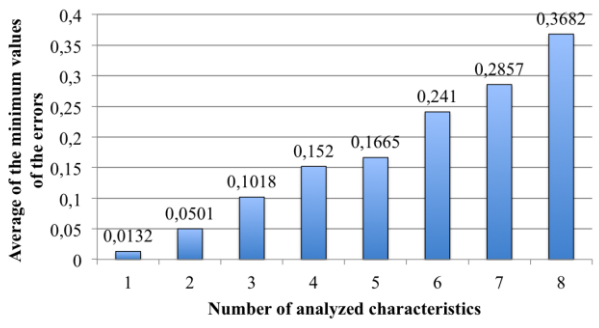


Figure 13. Average of the minimum values of the errors versus the number of analyzed characteristics.

For example, when using only the area characteristic, the average of the minimum error was 0.0132. Considering the first two characteristics (Area and Eccentricity), the average of the minimum error was 0.0501. This procedure is repeated analogously, obtaining the values presented in Fig. 13. It is noteworthy that the algorithm showed excellent results when using the eight preselected characteristics. The most relevant characteristics were: H/W, Perimeter, Orientation, and Difference Index. However, using the eight allows an index of 0.3682, which is very high considering that the minimum errors correspond to the most similar images.

IV. CONCLUSION

The progress to date shows that work is being done in the right direction. A selection of features that well describe the hieroglyphics of the ancient Egyptian alphabet has been achieved.

Normalizing the characteristics is a significant advance because it allows the application of the algorithm in

papyri with larger or smaller symbols, that is, that have a different font size.

The development and incorporation of the different indexes between two images allowed the rapid comparison of two images that may initially have different sizes. This index significantly increased the efficiency of the hieroglyph identification algorithm.

In future work, the incorporation of other hieroglyphs is proposed to expand the algorithm's database. In the same way, the development of an algorithm that combines the parametric techniques presented with artificial intelligence techniques for the recognition of hieroglyphs that are damaged or in poor condition is proposed. Finally, the semantic interpretation of the hieroglyphs is proposed to understand the message written in these papyri.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

RBL conducted the research; CAP implemented the artificial vision algorithm and performed the verification tests; GGM took the database of the images and carried out the documentation and its analysis. All authors had participated in the writing of the paper. All authors had approved the final version.

REFERENCES

[1] C. D. Hollings and R. B. Parkinson, "Two letters from Otto Neugebauer to Thomas Eric Peet on ancient Egyptian mathematics," *Historia Mathematica*, vol. 52, pp. 66–98, 2020.

[2] A. M. Metwaly, *et al.*, "Traditional ancient Egyptian medicine: A review," *Saudi Journal of Biological Sciences*, vol. 28, no. 10, pp. 5823–5832, 2021.

[3] S. Meyer, *et al.*, "Multidisciplinary studies of heavily fragmented and commingled ancient Egyptian human remains found in KV 40 (Valley of the Kings, Luxor, Egypt): A pragmatic workflow and first results," *Journal of Archaeological Science: Reports*, vol. 29, 102069, 2020.

[4] J. Russell, *et al.*, "An investigation of the pharmacological applications used for the ancient Egyptian systemic model 'RA-IB' compared with modern traditional Chinese medicine," *Journal of Ethnopharmacology*, vol. 265, 2021.

[5] M. Vandenbeusch, R. Stacey, and D. Antoine, "Rediscovering nestawedjat: Embalming residue analyses reunite the mummified remains of an ancient Egyptian woman with her coffins," *Journal of Archaeological Science: Reports*, vol. 40, 103186, 2021.

[6] G. Verly, F. W. Rademakers, C. Somaglino, P. Tallet, L. Delvaux, and P. Degryse, "The chaîne opératoire of middle kingdom smelting batteries and the problem of fuel: Excavation, experimental and analytical studies on ancient Egyptian metallurgy," *Journal of Archaeological Science: Reports*, vol. 37, 102708, 2021.

[7] A. Barucci, C. Cucci, M. Franci, M. Loschiavo, and F. Argenti, "A deep learning approach to ancient Egyptian hieroglyphs classification," *IEEE Access*, vol. 9, pp. 123438–123447, 2021.

[8] R. Moustafa, *et al.*, "Hieroglyphs language translator using deep learning techniques (Scriba)," in *Proc. 2022 2nd International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*, 2022, pp. 125–132.

[9] D. Arrivault, N. Richard, and P. Bouyer, "A fuzzy hierarchical attributed graph approach for handwritten Egyptian hieroglyphs description and matching," in *Proc. Eighth International Conference on Document Analysis and Recognition (ICDAR'05)*, 2005, vol. 2, pp. 898–902.

[10] T. T. Ly, T. Batabyal, J. Thompson, T. Harris, D. S. Weller, and S. T. Acton, "Hieroglyph: Hierarchical glia graph skeletonization and matching," in *Proc. 2020 54th Asilomar Conference on Signals, Systems, and Computers*, 2020, pp. 251–255.

[11] R. Elnabawy, R. Elias, and M. Salem, "Image based hieroglyphic character recognition," in *Proc. 2018 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS)*, 2018, pp. 32–39.

[12] S. Lin and K. Knight, "Discovering the linear writing order of a two-dimensional ancient hieroglyphic script," *Artificial Intelligence*, vol. 170, no. 4–5, pp. 409–421, 2006.

[13] E. Iglesias-Franjo and J. Vilares, "TIR over Egyptian hieroglyphs," in *Proc. 2016 27th International Workshop on Database and Expert Systems Applications (DEXA)*, 2016, pp. 198–203.

[14] L. A. Pinilla-Buitrago, J. F. Martínez-Trinidad, and J. A. Carrasco-Ochoa, "Encoding hieroglyph segments to represent hieroglyphs following the bag of visual word model for retrieval," *Expert Systems with Applications*, vol. 201, 116983, 2022.

[15] D. Vijayalakshmi and M. K. Nath, "A novel multilevel framework based contrast enhancement for uniform and non-uniform background images using a suitable histogram equalization," *Digital Signal Processing: A Review Journal*, vol. 127, 103532, 2022.

[16] C. Lezcano, J. L. Vázquez Noguera, D. P. Pinto-Roa, M. García-Torres, C. Gaona, and P. E. Gardel-Sotomayor, "A multi-objective approach for designing optimized operation sequence on binary image processing," *Heliyon*, vol. 6, no. 4, p. e03670, 2020.

[17] A. Cabarcos, C. Paz, R. Pérez-Orozco, and J. Vence, "An image-processing algorithm for morphological characterisation of soot agglomerates from TEM micrographs: Development and functional description," *Powder Technology*, vol. 401, 2022.

[18] H. Jia, W. Cai, H. Huang, and Y. Xia, "Learning multi-scale synergic discriminative features for prostate image segmentation," *Pattern Recognition*, vol. 126, 108556, 2022.

[19] X. Liu, M. Pedersen, and R. Wang, "Survey of natural image enhancement techniques: Classification, evaluation, challenges, and perspectives," *Digital Signal Processing: A Review Journal*, vol. 127, 103547, 2022.