Searching the Past: An Improved Shape Descriptor to Retrieve Maya Hieroglyphs

Edgar Roman-Rangel [1, 2]
eroman@idiap.ch

Carlos Pallan Gayol [3]
pallan.carlos@gmail.com

Jean-Marc Odobez [1, 2]
odobez@idiap.ch

Daniel Gatica-Perez [1, 2]
gatica@idiap.ch

[1] Idiap Research Institute. Martigny, Switzerland.
[2] École Polytechnique Fédérale de Lausanne, EPFL. Lausanne, Switzerland.
[3] National Anthropology and History Institute of Mexico, INAH. Mexico City, Mexico.

ABSTRACT

Archaeologists often spend significant time looking at traditional printed catalogs to identify and classify historical images. Our collaborative efforts between archaeologists and multimedia researchers seek to develop a tool to retrieve two specific types of ancient Maya visual information: hieroglyphs and iconographic elements. Towards that goal we present two contributions in this paper. The first one is the introduction and analysis of a new dataset of 3400+ Maya hieroglyphs, whose compilation involved manual search, annotation and segmentation by experts. This dataset presents several challenges for visual description and automatic retrieval as it is rich in complex visual details. The second and main contribution is the in-depth analysis of the Histogram Of Orientation Shape Context (HOOSC), and more precisely, the development of 4 improvements that were designed to handle the visual complexity of Maya hieroglyphs: open contours, mixture of thick and thin lines, hatches, large instance variability, and a variety of internal details.Experiments demonstrate that the adequate combination of our improvements to retrieve Maya hieroglyphs, provides results with roughly 20% more precision compared to the original HOOSC descriptor. Complementary results with the MPEG-7 shape dataset validate (or not) the proposed improvements, showing that the design of appropriate descriptors depend on the nature of the shapes one deals with.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Retrieval models*; I.4 [**Image Processing and Computer Vision**]: Feature Measurement—*Feature representation, size and shape*

1. INTRODUCTION

One of the many ways in which computer vision and multimedia retrieval can help improve the work of archaeologists and curators is by developing a machine that can rank automatically digital versions of historical materials according to visual similarity. Such a tool will have many positive implications within the realm of Cultural Heritage, starting with a significant decrease in the usual time spent looking manually at traditional printed catalogs (e.g., [16, 28]). This is in particular the case for the understanding of pre-Columbian cultures like the Maya, who developed highly sophisticated writing and counting systems based on hieroglyphs. Our collaborative efforts between archaeologists and multimedia researchers seek to develop tools capable of efficiently retrieving two specific types of highly encoded visual information: namely, Maya hieroglyphs (syllabic and word glyphs) and Maya iconographic elements (Maya art). Specifically on this paper, we deal with a subclass of the Maya hieroglyphic system: that of syllables or syllabographs.

1.1 Our contributions

The first contribution of this paper is the introduction of a new dataset of Maya syllabic hieroglyphs for automatic visual analysis. This dataset has been collected by expert manual identification, annotation, and segmentation of hieroglyphs that appear in Maya inscriptions of the Mexican territory, it has been generated by the National Anthropology and History Institute of Mexico (INAH) through the AJIMAYA project, and it has been complemented with a few glyphs taken from other sources [16, 20, 28]. In total, it comprises 24 positive syllabic classes containing 1200+ glyphs, and other 2100+ examples in a negative class. To the best of our knowledge, this is the largest Maya syllabic dataset ever analyzed with computer vision techniques.

The second and main contribution of this work is the improvement of the Histogram of Orientation Shape Context (HOOSC) [23], a shape descriptor recently proposed to retrieve complex shapes such as Maya hieroglyphs. Similar to traditional shape descriptors like the Shape Context [3] (SC), the HOOSC takes as input a set of points. This robust descriptor combines the log-polar regional segmentation of the Shape Context with a distribution of orientations (HOG-like) description [7], and it has proven to work well in retrieval tasks with small databases in the order of a few hundred glyphs. Our improvements include: 1) a refor-

mulation of the input to be described, going from a set of points sampled from the raw and typically thick contours of the shape, to a set of points taken from thinned versions of them. This addresses the problem of inaccurate description that arises as a consequence of the rough lines often found in Maya hieroglyphs that generate duplicate contours; 2) traditionally, only a subset of points from the input is used in the description to avoid expensive computations, in contrast we propose an efficient way to integrate the complete set of points, leading to more robust descriptors while still avoiding redundancy and high computing cost; 3) we increased the discriminative power of the descriptor by using only the most informative portion of the spatial context; 4) as complement to the implicit position that the HOOSC already features, we included the explicit relative self-position within the global shape of each described point.

We conducted several experiments to validate our improvements on the Maya syllabic dataset. Our results show that their adequate combination results in better retrieval performance, with an increment of approximately 20% of average precision, in comparison with the original HOOSC [23]. Complementary experiments with the MPEG-7 Core Experiment CE-Shape-1 test set [12] allowed us to validate which of the proposed improvements are suitable to describe other shape data sources in function of the nature of the shapes at hand.

1.2 The Maya source

The pictorial collection we analyze is not a modern construct, it was devised several hundred years ago by a now extinct civilization: the ancient Maya culture. The Maya is generally regarded as the epitome of ancient (pre-industrial) civilizations in the Americas, with many of its achievements comparable to those of the Old-World cultures that developed in Egypt, Greece, Rome, Sumer, and Babylon, to name but a few. Of concern here are the Maya hieroglyphic writing and Maya visual narrative or iconography (a substrate of Maya Art), which are often more sophisticated than its Old-World counterparts.

Roughly outlined, the ancient Maya was one of the several civilizations belonging to a cultural super-area called Mesoamerica, which encompassed the major parts of what are now the countries of Mexico, Guatemala, Honduras, Belize, and El Salvador (see Fig. 1). The Maya culture began to flourish during a chronological period called the Preclassic (c.a., BC 2000 - 250 AD), in a region labeled as the Maya lowlands, which encompassed an area roughly the same size of modern Germany. Although their development was differential according to region and speed, generally speaking, their heyday is regarded to have occurred during the subsequent Classic period (c.a., AD 250 - 900), and it was then when their hieroglyphic writing and the highly encoded iconographic imagery attained the levels of sophistication and consistency that we can rightly regard as a coherent, self-contained visual system, capable of conveying speech and ideas with admirable precision, even when compared with our "new-era" devices for information exchange: e.g., alphabets, syllabaries, graphic conventions, and so forth.

In a nutshell, any ancient script could be defined as a system for the visual recording of information through signs (graphemes) related in some way to the meanings (lexemes) and sounds (phonemes) that conform any given speech [5]. Only briefly, we note here that linguists generally ascribe
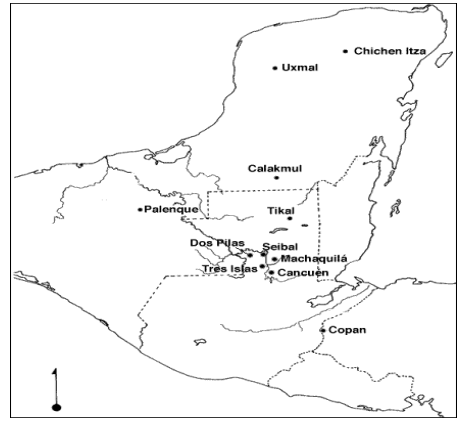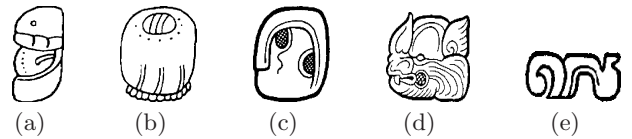

Figure 1: Maya region.

Figure 2: Examples of syllabographs: (a) 'a, and (b) b'a; and logographs: (c) KAB', (d) SUUTZ', and (e) K'AHK'. Images from [16, 28].

the Maya writing system to the class of the so-called logo-syllabic writing systems, to which a large number of other ancient-world scripts belong, such as the Anatolian from Syria, or the Hiragana from Japan. These writing systems are primarily composed of two distinct categories of signs: syllabographs and logographs. The former are visual signs which encode only specific phonetic value (i.e., phonemes) and almost always comprise a consonant-vowel or single aspirated vowel structure. On the other hand, logographs encode both sound and meaning (roughly equivalent to the notion of "word-signs"), and the vast majority of them have a consonant-vowel-consonant structure, noting that the embedded vowel could be either simple or complex, thus making possible forms like KAB' (earth); SUUTZ' (bat) and K'AHK' (fire). Fig. 2 shows as visual examples 2 syllables and 3 logographs.

Currently, almost 1000 different glyphs (semantic classes) have been identified, including both syllabographs and logographs. However, only 80% have been deciphered and are readable today. Note that these semantic classes might be represented with instances containing high level visual variability, which increments as the temporal and spatial gaps increase. For the purpose of creating an image dataset with the highest applicability for retrieval, and under the logic of a progressing scheme, we decided to focus exclusively on syllabic signs reserving logographs for later stages, going from relative simplicity towards increased complexity, which in future work would ultimately lead to the retrieval of queries performed over Maya iconographic elements. Thus, for the set of glyphs herein presented, we reach 24 classes, relying on their higher frequency of occurrence over other syllabic signs, thus facilitating the labor of manual localization, segmentation, extraction, and annotation by archaeologists; and allowing to have enough material for experimentation.

1.3 Outline

The remaining of this paper is organized as follows. Sec-

tion 2 highlights some of the most recent related work. Section 3 introduces the novel dataset used in this work. Section 4 describes the HOOSC and presents the improvements we propose. Section 5 describes the protocol we followed to test our method. In section 6 we present and discuss the results. Finally, in section 7 we present our conclusions.

2. RELATED WORK

Retrieval from image and shape representations have been approached in several ways. In [17] local viewpoint-invariant features are computed for specific areas automatically detected, providing robustness to image clutter, partial visibility, occlusion, and changes in viewpoint or lighting conditions. A review of shape representation and retrieval is found in [31], where the authors compare descriptors which mainly differ according to whether they describe contours or regions, and according to the locality scope of the description.

Some works rely on the robust Shape Context descriptor [3] and tackle the resulting set-to-set matching problem using linear programming optimization methods [32]. With similar approach, shapes can be represented by sets of local contour segments organized as trees, and used to perform search of shapes based on a particle filtering framework [15]. Descriptive object shape models can be learned combining long salient bottom-up contours [26], where a latent SVM learning formulation tunes the scores of a many-to-one contour matching approach used to deal with the random fragmentation that might occur in the contours. However, these techniques might not be computationally efficient enough when the set size increases as a consequence of dealing with complex shapes. The generalized version of the Shape Context explored in [18] uses quantized descriptors and local orientation information, which results in faster retrieval implementations and obtains better results than the original SC. In general, modeling objects by bags-of-visterms has proven very efficient [21, 25, 29], even though some issues related to the loss of spatial information arise with this representation. A bag-of-features approach able to retains spatial information has been proposed recently in [6].

Skeletal shape representations have been also studied. In [27], both the geometric and topological information of 3-D objects is combined in the form of a skeletal graph, using graph matching techniques to match and compare skeletons. The shape recognition problem is also approached with a graph matching algorithm in [1], based on object silhouettes where geodesic paths between skeleton endpoints are compared without considering the topological graph structure. Appearance-based object representations of local descriptors are explored in [8, 19, 24] to describe shape images as visual vocabularies of boundary fragments. For instance, in [8] a local segment network representation is used. More recently, shape retrieval has been boosted with graph techniques like local diffusion process [30] and graph transduction [2], achieving very good retrieval results. However, rather than dealing with shapes, these methods focus on retrieval of silhouettes with no internal details, and very often with closed and well defined boundaries.

In the specific field of automatic visual analysis of historical and cultural datasets, the work in [4] investigates how to formulate region-of-interest queries, and perform retrieval with relevance feedback. In [13], a system to retrieve paintings and photos of art objects using content and metadata is presented. Description and retrieval of Chinese characters has been broadly studied. For instance, the work in [14], detects visual patterns and trends in image collections of ancient Chinese paintings, lending to a correct identification of the artist style. In [33], Chinese calligraphy characters are retrieved using contour shapes and interactive partial-distance-map-based high-dimensional indexing that speeds up the performance. Another interesting work in cultural heritage is presented in [10], where artist identification is achieved using wavelets that characterize brushstrokes of several van Gogh paintings.

Finally, with an archaeological approach, a set of rules such as single symmetry axis and morphology is used in [9] to recognize a single polymorphic Mesoamerican symbol by describing its variations as sets of discrete curves. Previous works on description and retrieval of Mayan hieroglyphs achieved competitive results in small datasets up to a few hundreds of glyphs [22, 23]. In particular, in [23] the HOOSC descriptor is proposed and shown to be promising.

3. DATA

The complex manual process commonly followed by archaeologists to obtain digital versions of the hieroglyphs, and the laborious manual search needed to rank them according to visual similarity, are two of the main motivations to conduct this research. We found it relevant to present a brief outline of how the data has been compiled and organized. Namely, the overall process is as follows:

1. At archaeological sites, digital photographs of inscriptions are taken during the night. Instances with variations in the illumination are gathered to achieve high levels of detail (e.g., to study eroded inscriptions).

2. Line drawing are obtained by tracing the inner features of the inscriptions on top of multi-layers photographs.

3. Manual segmentation, search, and identification of hieroglyphs is done consulting existing glyphic catalogs.

4. Experts transcript manually the identified glyphs. i.e., map the phonetic value of each Maya sign into alphabetical conventions.

5. When needed, transliteration is performed to represent ancient Maya speech into modern alphabetic forms.

6. Morphemes and lexemes are obtained through morphological segmentation.

7. Grammatical analysis is done to indicate the function of each of these elements.

8. Finally, the translation of ancient Maya text into a modern target language is achieved, e.g., English.

Figure 3 shows the first and second steps of this process.

The goal of this work is to improve the third step of this process. To that purpose, we collected a dataset which comprises 1270 Maya glyphs belonging to 24 syllabic classes, plus 2128 extra glyphs that do not correspond to any of the 24 classes and that are grouped in a negative class. The glyphs have been gathered from different archaeological sources, including the AJIMAYA project of INAH, the Macri and Looper syllabic catalog [16]; the Thompson catalog [28]; and the website of the Foundation for the Advancement of
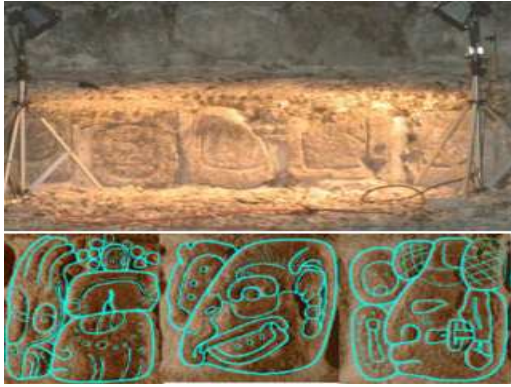
Figure 3: First two steps in the collection process.

Table 1: Thompson numbers, visual examples, and sounds for the classes of the Maya syllabic dataset.

T1	T17	T23	T24	T25	T59
/u/	/yi/	/na/	/li/	/ka/	/ti/
T61	T82	T92	T102	T103	T106
/yu/	/li/	/tu/	/ki/	/ta/	/nu/
T110	T116	T117	T126	T136	T173
/ko/	/ni/	/wi/	/ya/	/ji/	/mi/
T178	T181	T229	T501	T534	T671
/la/	/ja/	/'a/	/b'a/	/la/	/chi/

Mesoamerican Studies, FAMSI [20]), generating what to our knowledge is the largest dataset of Maya glyphs that has been analyzed with automatic techniques.

As simple as it might sound, this task required non-trivial work of archaeologists expert in Mayan iconography, who spent several months looking manually for the images in complex inscriptions. A dataset like this cannot be produced by non-trained annotators. Table 1 shows one visual example of each positive class, along with their Thompson number which traditionally is used as identifier, and their syllabic value, i.e., their sound. Note that this dataset posits many challenges in terms of visual complexity due to the richness in internal details of its elements, their variability, the fact that some of the classes might be visually similarity, and that conversely some glyphs inside each class might not be as similar as expected in visual terms.

The instances in this dataset correspond to glyphs that often appear in inscriptions from 4 main subregions of the Maya area (Petén, Usumacinta, Motagua, and Yucatán), and the participation of archaeologists in our team helped validating the localization and segmentation of each instance within the inscriptions. Finally, each glyph was manually aligned to the orientation most commonly seen for its class.

We divided the dataset into two subsets. Approximately 80% of the glyphs from each class are in the first subset (candidates, denoted by G_C), comprising 1004 instances and leaving the remaining 266 glyphs ($\approx$20%) in the seconds
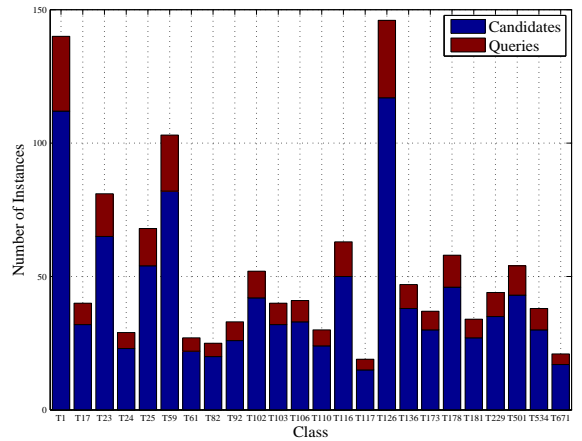

Figure 4: 1270 glyphs distributed over 24 classes. Number of candidates in blue and number of queries in red. (Best viewed as pdf).

subset (queries, denoted by G_Q). Fig. 4 shows the number of glyphs in each class.

The dataset features many of the complex phenomena presented in Maya glyphs. For instances, the class T534 shown in Table 1 is a "pars pro toto" version of T178 (i.e., a fraction of the original sign containing diagnostic features that account for the whole), both referring to syllable *la*. Since we address the retrieval problem from a visual perspective rather than semantic, and taking advantage of the availability of data, we decided to treat them as two different classes.

The negative class was gathered from the same sources, taking at random as many glyphs as possible. Note that some of the glyphs in this class might be logographs.

4. OUR APPROACH

This section explains the HOOSC descriptor introduced in [23], which is motivated to cope with several drawbacks from the Shape Context (SC) [3] and Generalized Shape Context (GSC) [18] descriptors. This section also explains the approach used to perform retrieval using the HOOSC as shape descriptor. Finally, we present the improvements we propose to the HOOSC.

4.1 HOOSC descriptor

The HOOSC is a robust shape descriptor that combines the log-polar regional formulation of the Shape Context (SC) [3], with a distribution of orientations (HOG-like) description [7]. According to [23], for a given shape whose contours are represented by a set P of N points, a HOOSC descriptor $hoosc_i$ is a vector that describes the point p_i as a function of the distribution of the local orientations of the $N-1$ remaining points.

This descriptor is computed on a log-polar space whose origin ($\theta = 0, \rho = 0$) corresponds to the position of p_i. The remaining points are distributed over 12 angular intervals accounting for a complete perimeter, and 5 distance intervals (that we refer to as *rings*) covering in total as twice as the average pairwise distance between every pair of points in the shape P, thus resulting in a space of 60 log-polar regions as shown in Fig. 5.

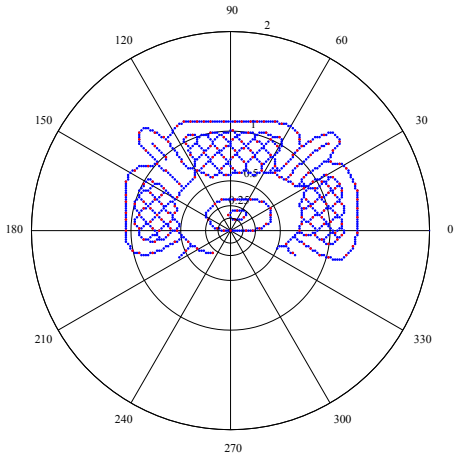More precisely, let us denote by P_i^r the subset of points

Figure 5: Pivots (red) and points (blue) on the 60 log-polar regions (12 orientations and 5 rings) used for shape description.

falling within the r-th region with respect to the point p_i:

$$P_i^r = \{p_j \in P : p_j \neq p_i, (p_j - p_i) \in R_r\}, \qquad (1)$$

where $p_j - p_i$ means vector difference, and R_r denotes one of the 60 regions in the log-polar grid indexed by r.

The region R_r is further characterized by a histogram of the local orientations in P_i^r. Since the histogram is encoded with 8 bins, the final $hoosc_i$ descriptor has 480 dimensions [1]. To take into account uncertainty in orientation estimation and to avoid hard binning effects, the distribution of local orientation is calculated through a kernel-based approach for orientation density estimation. More precisely, the density for angle θ in the log-polar region R_r of point p_i, is denoted by $h_i^r(\theta)$,

$$h_i^r(\theta) = \sum_{p_j \in P_i^r} \mathcal{N}\left(\theta; \theta_j, \sigma^2\right), \qquad (2)$$

where $\mathcal{N}\left(\theta; \mu, \sigma^2\right)$ is the value at angle θ of a Gaussian of mean μ and variance σ^2 ($\sigma = 10°$ works well in practice). The actual value of the 8-bins orientation histogram in bin $[a, b]$ is obtained by integrating the density $h_i^r(\theta)$ within this interval. Independent normalization of each of the 5 rings is suggested in [23]. The final HOOSC descriptor $hoosc_i$ for point p_i is the concatenation of all the 60 histograms h_i^r after the normalization:

$$hoosc_i = \left[h_i^1, h_i^2, \ldots, h_i^{60}\right]. \qquad (3)$$

Note that SC and GSC [3, 18] compute either the number of points in each region or its dominant local orientation, resulting in 60 and 120 dimensional descriptors respectively. However, previous attempts demonstrated that those methods are not as suitable to describe Maya hieroglyphs effectively [22, 23].

4.2   Shape retrieval with the HOOSC

Computing the HOOSC descriptor for every point of a given shape can be thought of as describing the shape from

---

[1]Though both log-polar bins and orientation bins are components of histograms, we use the terms region and bins to refer to them respectively.

different perspectives, which allows for a robust representation of the shape. However, direct comparisons of shapes is difficult because the number of points might differ from one shape to other, and solving the point-to-point correspondence problem is computationally expensive in some cases.

Often, the k-means algorithm is used to quantize the descriptors and build a bag of visual words or visterms (bov), such that shapes are more efficiently described. Two shapes can be further compared by simply computing the distance between their respective bov [18, 23]. To perform shape retrieval, we rank the bov of candidate-shapes according to their L1 similarity with respect to the bov of a given query-shape, as proposed in [23].

4.3   Improving the HOOSC descriptor

Several limitations arise as consequences of the HOOSC construction (large dimensionality, redundancy in the description, the need to find a trade off between the number of points for the description and computational efficiency), and the nature of the considered shapes (thick and thin lines, noisy shapes, hatched drawings, instance variability, complex internal details). In the following, we explain four improvements we propose to the HOOSC descriptor that address these limitations and that are key to achieve good retrieval results when dealing with Maya syllabic instances, as shown in section 6.

Thinned contours as input. Computing HOOSC descriptors for a set of points sampled along the raw contours of a shape works well with silhouettes and shapes whose internal details are not of crucial importance. This also performs well when accurate contours can be easily extracted. However, the Maya glyphs often present lines of different degrees of thickness, both along the contours as well as in their internal details. Thus, the use of contour extractors sometimes generates "double" contours as shown in Fig. 6(a), which can result in noisy descriptors and in an increase of intra-class variability.

We propose the use of thinning algorithms [11] to preprocess the binary shapes and estimate thinned versions of their contours, as the one shown in Fig. 6(c). This often provides more input to the HOOSC descriptor.

Pivot points. Instead of computing descriptors for each of the points in the whole input set P, usually only a uniformly sampled subset of points P' is considered in order to make the description computationally efficient [3, 23]. This means that each point p_i' is described using all the other points in P' to compute the HOOSC histograms explained in section 4.1. Fig. 6(b) illustrates in red a subset of points sampled from original set in 6(a). Attention is needed to ensure that P' is large enough to describe the shape in a reliable way.

We propose to compute HOOSC descriptors at each point in the subset P' as a function of *all* the points in the original set P, rather than only on those belonging to P'. The resulting descriptors will be more accurate, yet will remain computationally efficient. We call the points in P' "pivots" to differentiate them from the points in P, which are simply referred to as "points". Fig. 6(d) shows in red the selected pivots from the set of points in 6(c).

Spatial span of the descriptor. As shown in Fig. 5, the most internal regions of the log-polar space usually include very few points (sometimes only the point to be described is in those regions). Also very often, many of the external regions are empty or only contain points that are close to the

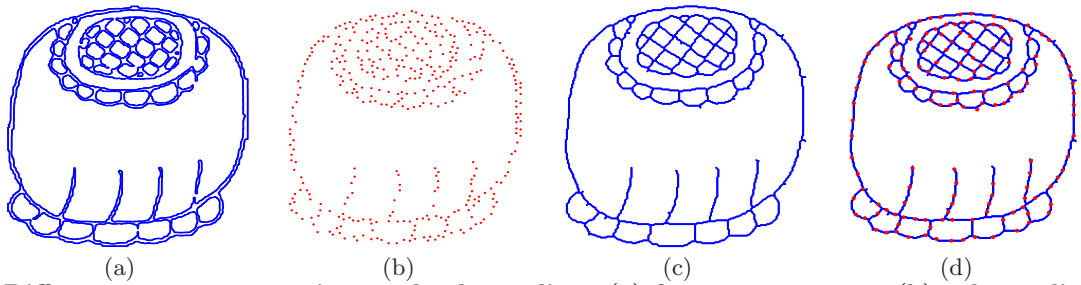(a)   (b)   (c)   (d)

Figure 6: Different contour extractions and subsampling. (a) from raw contour. (b) subsampling of points from the contours in (a). (c) from thinned contours. (d) the pivots of (c) in red. (Best viewed as pdf).

inner boundary of the ring. These two facts might result in several empty or noisy sections of the 480-D HOOSC vector that can affect the representation accuracy.

Since the most discriminative information is found in the intermediate spatial scope of the log-polar space, we use only the 288 dimensions of the descriptor that correspond to rings 2, 3, and 4. This reduces the dimensionality of the vector while improving its discriminative power.

Explicit spatial position. The HOOSC implicitly encodes continuous information about the position of each pivot within the shape, e.g., few observations in the lower regions of the log-polar grid means that the pivot point is located towards the bottom of the image.

The descriptive ability of the HOOSC can be improved further if the position is explicitly incorporated within the descriptor. We concatenate the coordinates (x_i, y_i) of each pivot p'_i, as two additional dimensions in the descriptor, thus representing the relative position within the bounding box encapsulating the glyph, i.e., within the interval [0,1]. With this normalization, the distance between the information explained by the position of two descriptors weights approximately twice the distance between the information contained in each of their rings.

5. EXPERIMENTS

In this section we describe the experimental protocol followed to evaluate the improvements proposed in section 4.

5.1 Evaluated methods

On the syllabic Maya dataset, we evaluated the Generalized Shape Context (GSC) [18] and five different variants of the HOOSC descriptor: HOOSC0 is the original method explained in section 4.1. The HOOSC1 takes as input thinned versions of the shape contours. The HOOSC2 describes the subsampled pivot points with respect to the whole set of points. HOOSC3 trims the log-polar space and only uses intermediate spatial scope (i.e., only rings 2, 3, and 4). Finally in HOOSC4, the relative self-position is explicitly incorporated within the description. As a consequence of the modifications proposed in section 4.3, the vectors named as HOOSC3 and HOOSC4 only have 288 and 290 dimensions respectively, instead of 480 as is the case of the HOOSC0, HOOSC1, and HOOSC2. Table 2 summarizes the differences between the 5 evaluated methods, note that these improvements are key to achieve good retrieval performance.

5.2 Evaluation protocol

We followed [23] to decide the number of pivot points to be described. With GSC and HOOSC0 we use one tenth

Table 2: The 5 HOOSC variants evaluated in this paper in the Maya syllabic dataset. Improvements are highlighted in blue.

HOOSC	0	1	2	3	4
Contours	Raw	Thin	Thin	Thin	Thin
Pivots from	Pivots	Pivots	Points	Points	Points
Rings	1:5	1:5	1:5	2:4	2:4
self-position	NO	NO	NO	NO	YES

of the number of points in the original raw contours, constraining it when possible to be at least 100 points, i.e., $max(10\%, 100)$. Sampling from thinned contours usually results in less points than sampling from the double lines generated by the raw contour. To avoid very sparse pivot-sets when sampling from thinned contours, we increased the sampling rate to $max(20\%, 150)$, obtaining on average the same number of pivots per glyph than in the raw contour cases: 161.7 and 169.5 respectively.

From the subset of candidate-glyphs (denoted G_C, see end of section 3), we randomly selected 1500 descriptors (i.e., GSC or HOOSC#) in each of the 24 positive classes, and using k-means clustered them into 2500 visual words. Then, we estimated the bov of each glyph in G_C and G_Q, and performed retrieval experiments to evaluate the retrieval precision. Our results are reported as Mean Average Precision (mAP). More precisely, for each method mentioned in Table 2 we implemented the following protocol:

1. Compute the descriptors, (i.e., GSC or HOOSC#).

2. Learn a visual vocabulary, using only descriptors in G_C of the 24 classes. This ensures that the vocabulary does not contain information about the queries in G_Q.

3. Describe every glyph in G_C and G_Q as a bov distribution over the resulting vocabulary.

4. Query from G_C using each glyph of G_Q, rank the retrieved vector, and compute the retrieval precision.

Additionally, we repeated the retrieval experiment adding to G_C all the glyphs of the negative class, and representing them by their bov computed over the visual vocabulary. This experiment is referred to as "24+1".

5.3 Applicability to other datasets

As already stated, our research is mainly motivated by the current needs of the archaeological community and the complexity of the glyphs they need to deal with. That said, to further asses which of our improvements are suitable and

Table 3: HOOSC variants evaluated in the A-MPEG-7 dataset. Improvements are highlighted in blue.

HOOSC	0	5	6	7	4
Contours	Raw	Raw	Raw	Raw	Thin
Pivots from	Pivots	Points	Points	Points	Points
Rings	1:5	1:5	2:4	1:5	2:4
self-position	NO	NO	NO	YES	YES

Table 4: mAP for the 6 methods evaluated in the Maya dataset: GSC, and HOOSC0 to HOOSC4.

Classes	GSC	HOOSC				
		0	1	2	3	4
24	0.236	0.350	0.422	0.492	0.502	**0.538**
24 + 1	0.195	0.201	0.281	0.341	0.341	**0.374**

beneficial to describe and retrieve shape data of different nature, we conducted experiments on the MPEG-7 Core Experiment CE-Shape-1 test set [12]. This dataset comprises 70 classes of silhouettes with 20 instances each. We divided it randomly at the same rate as the Maya dataset: 80% G_C and 20% G_Q, and repeated the experimental protocol described in section 5.2.

The original HOOSC descriptor was not designed to handle neither rotation nor reflexion since hieroglyphs under this transformation may have different meanings. Thus, we started by manually aliging 283 instances ($\approx$20%) that were either rotated or reflected, and tested the HOOSC variants summarized in Table 3. In our experiments we refer to this version of the data as A-MPEG-7. Later, we tested the best HOOSC variant on the original unaligned MPEG-7 dataset. To handle the rotation issue, we followed the same approach used in [3], where the tangent vector at each pivot is treated as the positive x-axis of the reference log-polar, thus resulting in a theoretical rotation invariant descriptor.

6. RESULTS

We present the results obtained with our improvements on retrieval experiments with the Maya syllabic dataset. We also show results of the generalization of these improvements to retrieve common shapes, i.e., the MPEG-7 dataset.

6.1 Maya syllabic glyphs.

The first row of results in Table 4 shows the mean Average Precision (mAP) of each method evaluated using the 24 positive classes. The original method (HOOSC0) obtains a precision 12% higher than the GSC in absolute terms. Changing the input to be thinned contours makes the description more robust and leads to better retrieval results (HOOSC1). The mAP of HOOSC2 shows that taking into account all the points for the description improves the results. We also notice that computing the descriptors at the complete set of points P, does not provide substantial improvements (results not shown). When we trim the HOOSC to only its intermediate distance scope (rings 2, 3, and 4) as in the case of the HOOSC3, the resulting descriptors are shorten while leading to a slightly higher precision. Finally, the explicit addition of the self-position (HOOSC4) allows for a mAP result of 0.532, for a total improvement of almost 20% in absolute terms with respect to the original HOOSC.

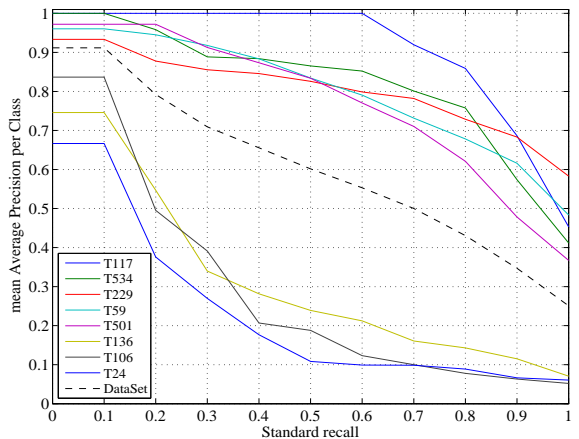The classes might vary in terms of size and visual com-

Figure 7: mAP precision vs standard recall for the whole collection (dashed line), plus the corresponding results for the 5 classes with highest average precision and the three with lowest average precision.

plexity. We present in Fig. 7 the per-class average precision AP versus the standard recall. These curves correspond to the 5 classes with the highest AP and the 3 with the lowest AP. Although class T117 has very few instances (see Fig. 4), it is the one with highest average precision. This is because it contains unique features that are not shared with any other class, such as its vertical orientation and the circles in the right hand side. Similar trends occur with classes T534 (inverted face), T229 (one circle in a superior section, and some circles in a vertical arrangement on the left hand side), T59 (concentric circles and quasi parallel lines), and T501 (circles and lines in specific internal regions).

The curve of class T136 degrades relatively fast because its instances are often confused with class T126 (Table 5). We observed a similar behavior with class T24 which is confused with classes T1, T17, and T23. In the case of class T106, the high variability among their instances, which could be split into two visual subclasses, results in a relative low precision. Despite the relative low precision for few classes, note that on average the precision is acceptable as shown in the dashed line in Fig. 7, and in the examples of Table 5.

Finally, the second row of results in Table 4 presents the mAP when the 2128 elements in the negative class are incorporated within the pool G_C. Note that the degradation roughly follows the same trend as before, keeping the method with our improvements as the best one in both cases.

6.2 Visual retrieval machine.

One of the long term objectives of our research is the implementation of an accurate visual retrieval system for Maya hieroglyphs. This tool will allow archaeologists to quickly search in large corpus for instances of visual queries. Fig. 8 shows an example of the preliminary version of this system. When such a tool is further improved, it will ameliorate the amount of time invested by archaeologists in manual searches, it also will help in the training of scholars learning about the Maya writing system. A video demo of this preliminary tool is available in the suplementary material.

6.3 Results on the MPEG-7 dataset.

Previous methods tested on this dataset for retrieval tasks

Table 5: Retrieval results. The first and second columns show the name of each class and one random query, followed by its Top 15 retrieved candidate-glyphs in ascending order from left to right. Relevant glyphs are enclosed in a gray square.

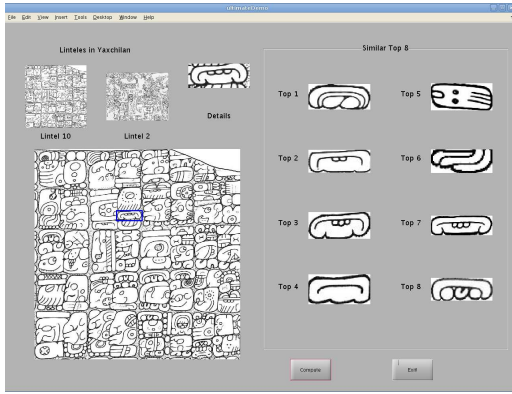| Class	Query	Top 15 retrieved vector

Figure 8: The segmented instances that are most similar to a selected glyph are retrieved from a database (Best viewed as pdf).

Table 6: mAP and bes for the methods evaluated in the A-MPEG-7 dataset.

Classes	GSC	HOOSC				
		0	4	5	6	7
mAP	0.813	0.848	0.790	0.852	0.849	**0.867**
bes	0.882	0.905	0.848	0.908	0.906	**0.918**

are compared via the *Bull's eye* score (bes) [2], we present our results in terms of both mAP and bes. Table 6 shows the results for the GSC and the evaluated variants of the HOOSC on the A-MPEG-7 dataset.

The HOOSC4 which uses our 4 improvements and that works the best with the Maya hieroglyphs does not perform as well with the MPEG-7 dataset. The reason is due to two factors: a) as these shapes are dominantly filled and clean convex silhouettes with very well defined boundaries, the morphological thinning transformation results in a loss of information and in descriptors with lower discriminative power. For these shapes, directly sampling the descriptors from the raw contours produces better results as shown with the HOOSC5 to HOOSC7 in table 6; b) unlike to the Maya hieroglyphs, where using the 5 *rings* (the whole spatial scope) adds noise to the description, using the 5 *rings* with the MPEG-7 silhouettes does not harm the description and the retrieval performance remains competitive as shown by HOOSC5 and HOOSC6. We can see that computing descriptors at pivots with respect to the whole set of points, and incorporating the relative self-position in the descriptor provide good results (HOOSC7).

Finally as explained in section 5.3, we incorporated robustness to rotation and experimented with the original unaligned shape instances achieving results of 0.733 of mAP and 0.811 of bes with the HOOSC7. Examples of this experiment are shown in Table 7, we can see that while rotated instances are well retrieved, the HOOSC is not robust to reflected shapes yet. It is important to mention that the quality of the hieroglyphs varies drastically due to the nature of the documents from which they are extracted. Thus, some improvements of the HOOSC descriptor (thinning and using only rings 2, 3, 4) are specifically designed to deal with noise (e.g., due compression). In contrast, the MPEG-7 shape dataset is cleaner.

7. CONCLUSIONS

Table 7: 15 queries randomly chosen from the MPEG-7 dataset and their corresponding Top 7 retrieved candidates via the HOOSC7 method.

We have compiled a new digital set of Maya hieroglyphs that comprises 3400+ instances distributed over 24 positive and 1 negative classes. This dataset presents several challenges for automatic visual description, and to the best of our knowledge is the largest one that has been analyzed with automatic tools.

We analyzed the Histogram Of Orientation Shape Context (HOOSC) descriptor with a set of retrieval experiments of Maya hieroglyphs. We proposed four improvements to the descriptor that allow for better retrieval results, roughly achieving 20% absolute improvement in terms of retrieval precision compared to the original HOOSC descriptor. Overall, our results demonstrate that relevant elements are retrieved first for most of the cases, and that only a few of them fail, either because of their intra-class variability or because of the high visual similarity across some classes. To validate the generalization of our improvements, we evaluated them on a general shape dataset, the MPEG-7 dataset. We found that two out of these four improvements proposed to describe complex shapes are also suitable to describe convex and clean silhouettes. Although the HOOSC was not designed to handle rotations, we found that this feature is easy to incorporate. Future work is still required to provide the HOOSC descriptor with robustness against reflection.

Overall, we believe that the proposed descriptor is suitable for general shapes and that it will be able to handle other shape datasets, to the condition that the right combinations of the options presented here is used, depending on the target shapes. We think that our results will motivate its implementation into several systems to support queries of scholars in archaeology and, in the long term, from general audiences like visitors to museums.

8. ACKNOWLEDGMENTS

9. REFERENCES

[1] X. Bai and L. J. Latecki. Path Similarity Skeleton Graph Matching. *IEEE Trans. PAMI*, 30(7):1282–1292. 2008.

[2] X. Bai, X. Yang, L. J. Latecki, W. Liu, and Z. Tu. Learning Context-Sensitive Shape Similarity by Graph Transduction. *IEEE Trans. PAMI*, 32(5):861–874. 2010.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape Matching and Object Recognition Using Shape Contexts. *IEEE Trans. PAMI*, 24(4):509–522. 2002.

[4] N. Boujemaa, V. Gouet, and M. Ferecatua. Approximate Search vs. Precise Search by Visual Content in Cultural Heritage Image Databases. *Proc. MIR Workshop ACM-MM*. 2002.

[5] J. K. Browder. Place of the High Painted Walls: The Tepantitla murals and the Teotihuacan writing system. PhD thesis. University of California. 2005.

[6] Y. Cao, C. Wang, Z. Li, L. Zhang, and L. Z. Spatial-bag-of-features. *Proc. IEEE CVPR*. 2010.

[7] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. *Proc. IEEE CVPR*. 2005.

[8] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of Adjacent Contours for Object Detection. *IEEE Trans. PAMI*, 30(1):36–51. 2008.

[9] Y. Frauel, O. Quesada, and E. Bribiesca. Detection of a Polymorphic Mesoamerican Symbol Using a Rule-based Approach. *Pattern Recognition*, 39(7):1380–1390. 2006.

[10] C. R. Johnson, E. Hendriks, I. Berezhnoy, E. Brevdo, S. Hughes, I. Daubechies, J. Li, E. Postma, and J. Z. Wang. Image Processing for Artist Identification - Computerized Analysis of Vincent van Gogh's Painting Brushstrokes. *IEEE Signal Processing Magazine, Special Issue on Visual Cultural Heritage*, 25(4):37–48. 2008.

[11] L. Lam, S.-W. Lee, and C. Y. Suen. Thinning Methodologies a Comprehensive Survey. *IEEE Trans. PAMI*, 14(9):869–885. 1992.

[12] L. J. Latecki, R. Lakamper, and T. Eckhardt. Shape Descriptors for Non-rigid Shapes with a Single Closed Contour. *Proc. IEEE CVPR*. 2000.

[13] P. H. Lewis, K. Martinez, F. S. Abas, M. F. A. Fauzi, S. C. Y. Chan, M. Addis, M. J. Boniface, P. Grimwood, A. Stevenson, C. Lahanier, and J. Stevenson. An Integrated Content and Metadata Based Retrieval System for Art. *IEEE Trans. Image Processing*, 13(3):302–313. 2004.

[14] J. Li and J. Z. Wang. Studying Digital Imagery of Ancient Paintings by Mixtures of Stochastic Models. *IEEE Trans. Image Processing*, 13(3):340–353. 2003.

[15] C. Lu, L. J. Latecki, N. Adluru, X. Yang, and H. Ling. Shape Guided Contour Grouping with Particle Filters. *Proc. IEEE ICCV*. 2009.

[16] M. Macri and M. Looper. The New Catalog of Maya Hieroglyphs. Vol. 1 The Classic Period Inscriptions. University of Oklahoma Press : Norman. 2003.

[17] K. Mikolajczyk and C. Schmid. Scale and Affine Interest Point Detectors. *IJCV*, 60(1):63–86. 2004.

[18] G. Mori, S. Belongie, and J. Malik. Efficient Shape Matching Using Shape Contexts. *IEEE Trans. PAMI*, 27(11):1832–1837. 2005.

[19] A. Opelt, A. Pinz, and A. Zisserman. A Boundary-fragment Model for Object Detection. *Proc. ECCV*. 2006.

[20] M. Pitts and L. Matson. Writing in Maya Glyphs. Foundation for the Advancement of Mesoamerican Studies, Inc. 2008.

[21] P. Quelhas, F. Monay, J. M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. V. Gool. Modeling Scenes with Local Descriptors and Latent Aspects. *Proc. IEEE ICCV*. 2005.

[22] E. Roman-Rangel, C. Pallan, J.-M. Odobez, and D. Gatica-Perez. Retrieving ancient maya glyphs with shape context. *Proc. IEEE ICCV, Workshop on eHeritage and Digital Art Preservation*. 2009.

[23] E. Roman-Rangel, C. Pallan, J.-M. Odobez, and D. Gatica-Perez. Analyzing Ancient Maya Glyph Collections with Contextual Shape Descriptors. *IJCV, Special Issue in Cultural Heritage and Art Preservation*, 94(1):101âĂŞ117. 2011.

[24] J. Shotton, A. Blake, and R. Cippola. Contour-based Learning for Object Detection. *Proc. IEEE ICCV*. 2005.

[25] J. Sivic and A. Zisserman. Video Google: A Text Retrieval Approach to Object Matching in Videos. *Proc. IEEE ICCV*. 2003.

[26] P. Srinivasan, Q. Zhu, and J. Shi. Many-to-one Contour Matching for Describing and Discriminating Object Shape. *Proc. IEEE CVPR*. 2010.

[27] H. Sundar, D. Silver, N. Gagvani, and S. Dickinson. Skeleton Based Shape Matching and Retrieval. *Proc. IEEE International Conference on Shape Modeling and Applications*. 2003.

[28] J. E. S. Thompson. A Catalog of Maya Hieroglyphs. University of Oklahoma Press : Norman. 1962.

[29] J. Willamowski, D. Arregui, G. Csurka, C. R. Dance, and L. Fan. Categorizing Nine Visual Classes Using Local Appearance Descriptors. *Proc. ICPR, Workshop on Learning for Adaptable Visual Systems*. 2004.

[30] X. Yang, S. Koknar-Tezel, and L. J. Latecki. Locally Constrained Diffusion Process on Locally Densified Distance Spaces with Applications to Shape Retrieval. *Proc. IEEE CVPR*. 2009.

[31] D. Zhang and G. Lu. Review of Shape Representation and Description Techniques. *Pattern Recognition*, 37(1):1–19. 2004.

[32] Q. Zhu, L. Wang, Y. Wu, and J. Shi. Contour Context Selection for Object Detection: A set-to-set Contour Matching Approach. *Proc. ECCV*. 2008.

[33] Y. Zhuang, Y. Zhuang, Q. Li, and L. Chen. Interactive High-dimensional Index for Large Chinese Calligraphic Character Databases. *ACM TALIP*, 6(2). 2007.