Retrieving Ancient Maya Glyphs with Shape Context

Edgar Roman-Rangel[1,2]   Carlos Pallan[3]   Jean-Marc Odobez[1,2]   Daniel Gatica-Perez[1,2]

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

{eroman, odobez, gatica}@idiap.ch, carlospallan@yahoo.com

Abstract

We introduce an interdisciplinary project for archaeological and computer vision research teams on the analysis of the ancient Maya writing system. Our first task is the automatic retrieval of Maya syllabic glyphs using the Shape Context descriptor. We investigated the effect of several parameters to adapt the shape descriptor given the high complexity of the shapes and their diversity in our data. We propose an improvement in the cost function used to compute similarity between shapes making it more restrictive and precise. Our results are promising, they are analyzed via standard image retrieval measurements.

1. Introduction

The preservation, analysis, and management of valuable and fragile historical and cultural materials with digital technologies is a field of multidisciplinary nature, and significant to the society at large. The benefits of using visual and multimedia analysis techniques in this domain are manifold. Automatic and semi-automatic content-based analysis methods can provide scholars in the humanities (historians, anthropologists, archaeologists, and linguists) and the arts (curators, art historians, and photographers) with tools to facilitate some of their daily work -consulting, organizing, annotating, and cataloging pieces- but also to obtain new insights about specific theories through the recognition and discovery of patterns and connections within and across pieces in a collection. Furthermore, automated analysis techniques can boost the creation of educational systems for public access and retrieval of digital versions of ancient media. The careful and efficient use of these digital collections, with potential impact in local development and tourism, also has a definite economic value.

This paper describes interdisciplinary work between archaeological and computer vision teams, on automatic visual analysis of ancient Maya media, more specifically of hieroglyphs. The ancient Maya civilization has been regarded as one of the major cultural developments that took place in the New World, as reflected by their impressive achievements, encompassing the artistic, architectural, astronomical, and agricultural realms. Paramount among these is their refinement of a fully-phonetic writing system that ranks among the most visually sophisticated ever created in world history.

We present a study of automatic shape-based retrieval of ancient Maya glyphs. Our paper contains three contributions. First, we describe our interdisciplinary approach, which has brought together computer scientists and archaeologists to work on a series of problems related to analysis, documentation, and preservation of a unique source of materials. Our source of data is the result of the AJIMAYA project[1], an ongoing effort at INAH (*Instituto Nacional de Antropología e Historia de México*) to collect high-resolution images and high-quality manual line drawings of all the known hieroglyphic Maya monuments within the Mexican territory, numbering in the thousands, and spread along several of the southern states of Mexico (see Fig 1), including several UNESCO's World Heritage Sites.

Second, we analyze in detail the Shape Context [2, 3], a state-of-the-art shape descriptor, as a method of visual representation of syllabic glyphs, which constitute one of the fundamental structures of the ancient Maya writing system. To our knowledge, the automatic visual analysis of Maya writing materials has not been conducted before. Third, we evaluate the shape descriptor in a glyph retrieval task, which shows that, while the descriptor is indeed effective in several cases, it faces significant challenges due to the nature of the glyphs, in terms of the complex and wide variety of content. In this sense, our work can be seen as establishing a novel testbed for robust computer vision analysis methods.

The paper is organized as follows. Section 2 describes the project in more detail. Section 3 describes related work in the specific context of automatic retrieval. Section 4 summarizes the Shape Context algorithm. Section 5 presents the

---

[1]The acronym stands for *Acervo Jeroglífico e Iconográfico Maya* (Hieroglyphic and Iconographic Maya Heritage), a project started in 2006.
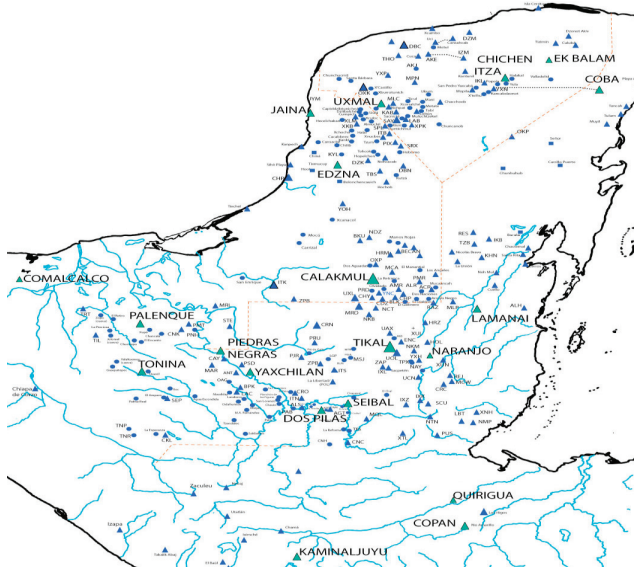
Figure 1. The Maya region with main sites in larger font.

experimental protocol. Section 6 presents and discusses the results. Finally, section 7 offers some concluding remarks and open issues.

2. Project

The invention of writing was a rare event in world's history, only witnessed in five different times and regions, in Egypt, the Indus valley, Mesopotamia, China, and the Americas. In one way or another, all other writing systems derive from the aforementioned, including the Indo-European runic and alphabetic traditions.

Stories about how ancient writing systems such as the Egyptian or the Maya came to be deciphered are regarded as one of the most fascinating aspects of human cultural achievements, where ancient and modern intellects reach one another. Over 200 years of scholarly tradition focused on the Maya system have revealed its secrets through a progression that started with its mathematical cycles and astronomical tables, until reaching its underlying phonetic keys and intimate functioning [8], which in turn opened the gate to the study of Maya history [4] and derived aspects such as ancient geopolitics [11] and religion [18].

2.1. The Maya writing system

The Maya culture originated from the middle to late Preclassic periods (*c.a.* 1,500 BC) in parts of what is now Mexico, Guatemala, Belize, and Honduras, and it reached its climax during the late Classic period (*c.a.* 600-900 AD). The Maya writing system remained operational during 17 or 18 centuries. Even though at least 1,000 distinct signs have been identified so far, only a maximum of 800 were used at any given time, of which approximately 80-85% can be read today.
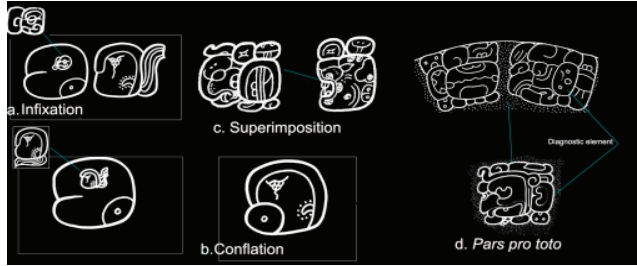

Figure 2. Examples of complexity in Maya writing system.

The Maya writing is part of a larger and deeper phonetic tradition that developed between Mexico's Gulf coast and Isthmus, and southern portions of Mesoamerica (Fig. 1). The earliest datable Maya glyphic records appear at lowland sites such as Tikal in the Peten, along with highland or Pacific coastal sites such as Takalik Abaj. From there it diffused across a number of sites through the Guatemalan Peten to Belize, Honduras, and the Mexican states of Campeche and Quintana Roo. By late Classic times, the Maya writing became widespread throughout the Maya lowlands, although its usage was almost completely discontinued in the highlands. Its practice continued, albeit on a fairly diminished capacity, until the Terminal Classic period (*c.a.* 800-1000 AD), except for a number of new and/or revitalized centers (like Chichen Itza) which would continue to operate under new political, religious and military order, where not only Maya, but also other Mesoamerican ethnic groups seemed to have participated.

The Maya scripts belong to the family of the so-called logo-syllabic writing systems. This term describes a system composed of two, functionally distinct types of phonetic signs: logographs and syllables. The former implies "word-signs", *i.e.* those which convey both meaning and sound (*e.g.* B'AHLAM, "jaguar"). The latter only represents sound without any semantic connotation. Several different signs can be arranged inside a single "glyph-block", where usually logographs are phonetically complemented by syllables, either on initial (prefix or superfix) or in final position (postfix or suffix). Phenomena that illustrate the complexity of the writing system are: *conflation*, which occurs when two signs are visually fused, each retaining its same relative size; *infixation*, which involves one sign being reduced in size and inserted within another; *superimposition*, that takes place when one sign partially covers another; and *pars pro toto*, which makes use of one or more diagnostic features of any given sign in order to mean the whole. Examples of these phenomena are presented in Fig 2. Typically, Maya glyphs are arranged on a paired columnar format, which is referred to as a system of co-ordinates, where letters designate the columns and numbers the rows. In such a grid, a standard reading order for a text comprising 4 columns and 2 rows would be: A1, B1, A2, B2, C1, D1, C2, D2.

2.2. Goals of the collaboration

AJIMAYA's substantive goals encompass the safeguard, preservation, management, study, and dissemination of the Mexican Maya archaeological and cultural heritage. Although Maya hieroglyphic monuments are regarded as one of the most important segments of Mexico's cultural heritage, epigraphic studies in Mexico are fairly recent compared with scholarly traditions in Europe and North America. AJIMAYA has been partially inspired by the pioneering work of Ian Graham [7].

Our collaboration inserts itself as part of the above-mentioned general goals, with the challenging target of the application of computer vision technology to support Maya decipherment. Within the epigraphic community, there is a shared view about the need for a more refined and updated hieroglyphic catalog, one that can be periodically updated, with the ability to incorporate input from multiple colleagues working on different parts of the world. One of the challenges involved operates at the taxonomical level: epigraphy needs an upgraded system for classifying the 800+ known signs, which separates the deciphered from the undeciphered ones, where the authorship of each specific decipherment is unambiguously indicated, and where all known contexts of occurrence of each particular grapheme are readily accessibles in order to better assess its function. A basic ability that needs to be developed is that of performing queries of a particular sign, which could retrieve the contexts where equivalent instances occur. Taken a step further, we would like to develop the ability of automatically detecting not only specific instances of particular signs, but also their variants, their allographs, their homophones, and their polivalencies, among other related signs.

One of the most exciting long-term goals envisioned is to develop a sensitive enough tool to compare at a glance thousands of inscriptions, in order to detect meaningful patterns that could then be correlated with phonetic substitutions. If this could be achieved, it could become one of the primary tools used by researchers for the decipherment of the remaining 15-20% of the signs in the hieroglyphic corpus whose phonetic values remain unknown.

2.3. Work in data collection

The AJIMAYA project is currently working in three states of Mexico (Tabasco, Chiapas and Campeche). For some of these sites, the project has already compiled a full photographic record, which serves as the foundation for elaborating vectorial line drawings of the inscriptions, all of which have to undergo an eight-fold methodological treatment: (1) digital photographs are taken at night under raking-light illumination, to bring out the level of detail that facilitates the study of eroded or worn out monuments; (2) line drawings are traced on top of multiple lay-

Figure 3. Example of digitalization process for Maya images.

ers showing the same signs photographed under different light conditions, to capture the highest possible amount of inner details and features that are diagnostic towards their subsequent identification; (3) glyphic signs are then manually identified with the aid of existing glyphic catalogs; (4) transcription involves the annotation of Maya phonetic signs in latin alphabet form; (5) transliteration renders ancient Maya speech into alphabetic form; (6) morphological segmentation breaks down recorded Maya words into their minimal grammatical constituents; (7) grammatical analysis makes use of conventions that are common to the fields of historical and structural linguistics; and finally (8) ancient Maya text can be rendered into a target language *e.g.* English. Fig. 3 shows the first and second steps of this process.

3. Related work

The use of computer vision techniques for automatic indexing and retrieval of 2D imagery in cultural heritage has been the topic of important research in the past recent years. In the following, we discuss some related work in this area.

One of the most representative works in this domain is [9], in which a system for retrieval of paintings and photos of art objects, using content and metadata, was developed by a multidisciplinary team jointly with a group of European museums. The project identified two relevant user tasks from specialists' feedback: *query-by-example*, and *cross-collection search*, and proposed algorithms based on adaptation and extension of techniques developed in the mid-90s (*e.g.* color coherence vectors). A discussion about variations of visual query formulations in cultural heritage collections, and ideas on how to approach them with similar emphasis on queries by region-of-interest appears in [5]. It is well known, however, that classic content-based image retrieval (CBIR) techniques are often limited to handle variations of illumination, appearance, and viewpoint [17].

More recently, the representation and matching of visual entities has experienced much progress given the recent in-

vention of robust local viewpoint-invariant features, computed over automatically detected local areas, to represent objects and scenes [12]. The local character of these features provides robustness to image clutter, partial visibility, and occlusion, while their invariant nature addresses naturally the issues related to changes in viewpoint and lighting conditions. In this context, one emerging research direction has focused on modeling objects by histograms of quantized local descriptors (bags-of-visterms) [14, 16], which allows for fast retrieval applications in large collections. However, the main limitation of these approaches is that the spatial information is lost. In addition, the above works represent the visual content with appearance descriptors which might not be well adapted to our shape/binary images.

Representation of shapes has a long history. The work in [19] provides a recent review of this domain. Roughly speaking, shape descriptors differ according to whether they are applied to contours or regions, and whether the shapes are represented globally or by their structure. Global representations like Fourier or moment descriptors are usually sensitive to alterations or variations in some regions of the shape. Structural approaches, which represent shapes by trees of local segments, do not suffer from this drawback. However, the tree representation becomes quickly large when dealing with complicated shapes like our glyphs, and make these techniques computationally very expensive for shape comparison.

Taking inspiration from the success of the appearance local descriptors for object representation, recent work has investigated similar approaches to represent shape information in images, by exploiting in different ways vocabularies of boundary fragments [6, 13, 15]. However, these techniques were specifically designed for object detection or classification, and are thus associated with discriminant learning approaches (like adaboost) and sliding-window or voting methods for localization, and therefore did not provide appropriate explicit or implicit matching scores for retrieval applications. In contrast, the shape-context algorithm [2, 3] allows to represent the shape through local descriptors[2] that integrate shape and geometric discrepancy measures, providing a good framework for shape comparison.

4. Shape Context Algorithm

In this section we describe how the Shape Context works [2, 3]. We explain how to compute the descriptors, to measure dissimilarity between two shapes and how to find the point-to-point correspondences between them.

---

[2]Strictly speaking, the Shape Context descriptor h_i computed at point p_i includes information of the global shape; however, the emphasis of the descriptor is on segments of the contours near p_i.
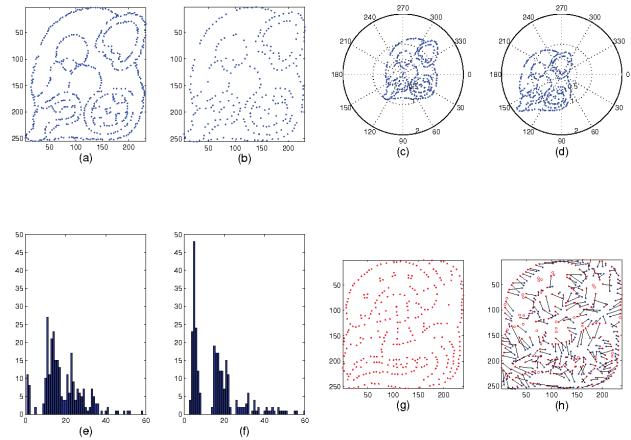

Figure 4. Shape Context process for description and matching. (a): glyph *b_a01* (see Fig. 5) with 10%(553) sampled points. (b): glyph *b_a01* with 5%(277) sampled points. (c) and (d): log-polar spaces for two different points chosen as examples. (e) and (f): the histograms describing the shape context for the two example points, respectively. (g): 5%(309) sampled points for the glyph *ki01* retrieved as the most similar one for *b_a01*. (h): matching between the query and the retrieved glyph, correspondences are connected by a black line while dummy matches are presented by disconnected circles.

4.1. Descriptors

Shape Context is a method proposed by Belongie *et al*. [2] that takes as input a finite set of n points, which represent the contours of a given shape, and describes it as a set of n histograms. More precisely, for each point p_i in the image P, its histogram h_i called shape context descriptor, is computed as the distribution of the relative position of the $n-1$ points with respect to the point in question. Such a distribution lies on 5 normalized distances and 12 angles, resulting in 60 bins which are uniform in log-polar space making the descriptor more sensitive to nearby points than to the farthest ones.

It is intuitive that the larger n the better the representation of the contours and the more accurate the description of the shape. In Figure 4 (a) and (b) we show the representation at two different values of n for the first glyph in Figure 5. Also, Figure 4 (c) and (d) present the log-polar space for two points randomly taken from the set in (b), the picked point in each case appears in the center of the space. Figure 4 (e) and (f) show the shape context descriptors for the selected points in (c) and (d) respectively.

4.2. Shape dissimilarity

Once the Shape Context is computed, it is possible to measure the dissimilarity or cost between two shapes P and Q. According to [3], by using the χ^2 test statistic like in Eq. 1, one can calculate the difference between each point p_i (*i.e.* histogram g_i) in the first set P against each point q_j

(*i.e.* histogram h_j) in the second set Q:

$$C_{ij}^{sc} \equiv C^{sc}(p_i, q_j) = \chi^2(g_i, h_j)$$

$$= \frac{1}{2} \sum_{k=1}^{K} \frac{[g_i(k) - h_j(k)]^2}{g_i(k) + h_j(k)}. \qquad (1)$$

It is possible to make the cost stricter by incorporating the difference of local tangent angles. This is done in [3] by updating C with Eq. 2;

$$C_{ij} = (1 - \beta) C_{ij}^{sc} + \beta C_{ij}^{tan}, \qquad (2)$$

where C_{ij}^{sc} is the cost matrix computed in Eq. 1, $C_{ij}^{tan} = 0.5 \left(1 - \cos\left(\theta_i^P - \theta_j^Q\right)\right)$ is the matrix of local tangent differences with θ_t^R being the local angle corresponding to the $t - th$ point in shape R. We set $\beta = 0.1$ such as in [3].

4.3. Matching point-to-point

By imposing a point-to-point matching, for each point p_i it is possible to know which is its best correspondence q_j,

$$H(\pi) = \sum_i C(\pi_i, q_i), \qquad (3)$$

where π_i is the $i - th$ entry in the vector Π, which is nothing but a permutation of vector P such that Eq. 3 is minimized. A constraint to this minimization problem is that the number of points should be the same for the two glyphs *i.e.* $|P| = |Q| \rightarrow |\Pi| = |Q|$, where $|P|$ indicates set cardinality. If the restriction is not satisfied, one can make use of dummy handlers, as done in [3] to deal with outliers. Assume that $|P| < |Q|$ then, add dummy points to P to have the same length in both vectors, and fill the corresponding entries in matrix C with a dummy threshold ϵ_d. In this way, all the points in Q with the worst matching cost will be forced to match with dummy nodes in P.

Nevertheless, all point in P are forced to have a match in Q. To allow as well outliers in P, we introduce dummy handlers in both sets P and Q by increasing the dimension of the cost matrix C up to $\lceil m(1 + f) \rceil$, where $m = \max(|P|, |Q|)$ and, f (dummy rate) is the fraction of points in the biggest set that we allow to have no match. Any new entry in C is then set to ϵ_d. Section 6 gives more details about the tuning of f. Figure 4 (g) shows the set of points representing the contours of the most similar glyph to the one in (b), and (h) shows the point-to-point match between these two glyphs.

4.4. Dissimilarity index

It is possible to rank a collection of glyphs w.r.t. their dissimilarity to a given query glyph. Finding a scalar value for the dissimilarity between two shapes (*dissimilarity index*) is possible by several ways [2, 3],

$$d_B = \max\left(\frac{1}{|Q|} \sum_{q \in Q} \min_{p \in P} C, \frac{1}{|P|} \sum_{p \in P} \min_{q \in Q} C\right), \qquad (4)$$

where $C = C(p, q)$, is the cost matrix histogram-to-histogram defined in Eq. 2.

By visual inspection in several ranked retrieved glyphs, we noticed that Eq. 4 is not a powerful enough discriminative function, as two similar glyphs might have indices whose difference is bigger than that for two dissimilar ones. This is due to the fact that it takes into account just the minimum values of point-to-point matching costs, and not the real costs of matching correspondences $C(\pi, q)$. We proposed two functions with small, yet important, differences from Eq. 4. The first proposed variation is,

$$d_{rm} = \frac{1}{|Sreal|} \mathrm{tr}(C_{Sreal}) \qquad (5)$$

where $Sreal$ is the subset of point-to-point costs belonging to only real matching points, C_{Sreal} is a square matrix $|Sreal| \times |Sreal|$, extracted from C (Eq. 2) which is permuted according to Eq. 3, such that, the total cost equals the mean of those costs corresponding to only real matches.

This function makes a clear distinction between similar and non-similar glyphs by giving different enough dissimilarity values. However, since it does not take in account dummy assignments, which might be a considerable amount depending on the number of points in the two shapes to compare, it could generate a small value when two shapes have similar overall shapes but with important different internal details. To deal with this issue we proposed the following function,

$$d_{sc} = \frac{1}{|Stotal|} \mathrm{tr}(C_{Stotal}) \qquad (6)$$

where $Stotal$ is the whole set of point-to-point costs including dummy assignments, *i.e.* $|Stotal| = \lceil m(1 + f) \rceil$ (see section 4.3). We decided to make use of Eq. 6 to perform our experiments, since it gives the high discriminative power needed by our data.

5. Experimental protocol

In the next subsections we describe how we obtained and organized the data used for experiments. We also describe the way in which we performed the experiments.

5.1. Data

For the experiments we used a prescaned collection of 297 Syllabic Maya Glyphs from [10]. Maya syllables might
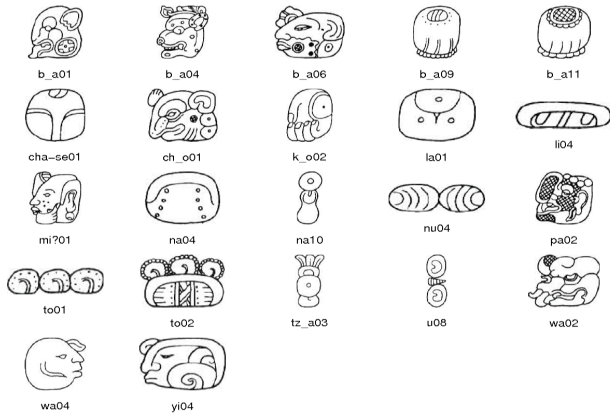
Figure 5. Set of 22 Maya syllabic query glyphs with their corresponding sounds. Since one syllable might have more than one glyphic representation, we added a consecutive-per-syllable number at the end of the name to have an unique identifier for each glyph.

Figure 6. Example of retrieved glyphs. First column shows the three queries with highest *AP* and the two with lowest one (in descending order). Then, each row has the top 5 retrieved glyphs with the relevant ones enclosed in a blue rectangle. This ranking is obtained with the best combination (case *l*) explained in Table 1.

have more than one associated glyph. It is important to mention that glyphs associated to the same syllable might be visually different, while others associated to distinct syllables could present visual similarity among them.

The glyphs were scaned at their current size in [10], and stored as binary images (white background and black contours). Given that some glyphs have square shapes and others are elongated either on vertical or horizontal orientation, we resized them to fit 256 pixels in their largest dimension keeping proportionality in the smallest axes. Since glyphs have a different degree of detail and complexity, we decided not to represent all the contours with the same number of points, but with a percentage of the total number of points presented in each of the raw images. For each glyph we extracted three random sets of points along its contours whose size $n = 100$, $n = \max(100, 5\%)$, and $n = \max(100, 10\%)$. In Section 5.2, we argue why this adaptive approach gives better results than the fixed one.

From the collection, we randomly selected 22 glyphs to use them as queries. For each of them, we labeled as relevant all the similar glyphs found in the collection, purely based on visual similarity. The set of queries is diverse: queries are considerably different among each other. Figure 5 shows the 22 selected queries.

5.2. Performance measure

We performed a series of experiments to analyze the effects of the sampling rate n, the dummy rate f, and the dummy threshold ϵ_d. To evaluate retrieval quality we made use of the mean average precision measure (*mAP*) [1]. We also did visual inspection of the top retrieved glyphs ranked by the dissimilarity index d_{sc}. Figure 6 presents an example of the top 5 retrieved glyphs for the three queries with the
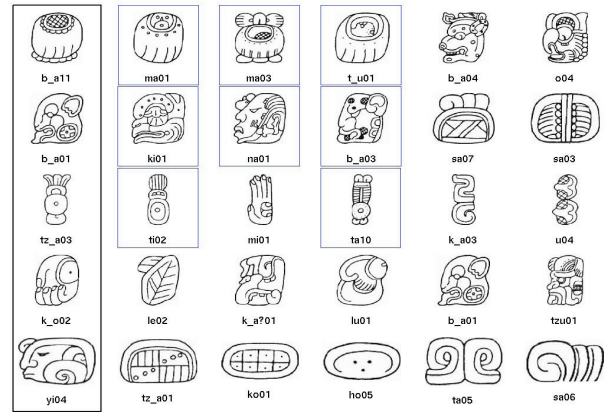
highest average precision (*AP*) [1] and for the two with the lowest one. Section 6 gives more details about the results of our experimentation.

Before starting our experiments, we did an analysis to assess how well the contours are represented at different sampling rates. We tried representing the contours with 2%, 3%, 5%, 10% and 15% of the total number of points in the original shapes. We observed that with 2% and 3% many glyphs are represented by less than 100 points, such a representation is very poor. 10% and higher percentages produce a robust representation but make the computation slower. Empirically, 5% is a good trade-off between accurate representation and efficiency. For the experiments, we decided to use both 5% and 10% sampling rates with a bound of 100 points, that is, $n = \max(100, 5\%)$ and $n = \max(100, 10\%)$.

6. Results

Using the cost function in Eq. 6, sampling rates of $n = \max(100, 5\%)$ and $n = \max(100, 10\%)$, and in few cases incorporating local appearance by Eq. 2, we tested the 12 combinations shown in Table 1, where n represents the sampling rate, f whether or not we added dummy handlers, ϵ_d is the value for the dummy assignments, β is the percentage used for the combination of shape dissimilarity and local tangent dissimilarity, and *mAP* is the mean average precision computed for each combination.

Comparing the *mAP* obtained with combinations *a* and *b*, we see that adding handlers for dummy matches increases the precision. This is because the algorithm becomes able to characterize the portions that should match between compared shapes and those that should not. The same occurs for

Case	n	$f(\%)$	ϵ_d	β	mAP
a	5%	0	—	0.00	0.306
b	5%	10	0.25	0.00	0.315
c	$\max(100, 5\%)$	0	—	0.00	0.301
d	$\max(100, 5\%)$	10	0.25	0.00	0.315
e	$\max(100, 5\%)$	10	0.25	0.01	0.316
f	$\max(100, 5\%)$	15	0.25	0.01	0.317
g	$\max(100, 5\%)$	15	0.35	0.01	0.310
h	$\max(100, 5\%)$	20	0.25	0.01	0.319
i	$\max(100, 5\%)$	20	0.25	0.00	0.319
j	$\max(100, 10\%)$	10	0.35	0.00	0.315
k	$\max(100, 10\%)$	10	0.25	0.00	0.319
l	$\max(100, 10\%)$	15	0.25	0.00	0.322

Table 1. Tested combinations. Values for sampling rate n, dummy rate f, dummy threshold ϵ_d, percentage β for combination of shape appearance, and *mAP* computed over the 22 queries. $f = 0$ implies that no dummy threshold is needed ($\epsilon_d =$—).

cases c and d. Increasing the dummy rate f gives slightly better results: compare cases i with d, e with f with h, and k with l. Related to dummy handlers, cases f and g show that increasing the threshold of ϵ_d does not improves the *mAP*. Similar behavior is found by taking a look at cases j and k.

Even though no further improvement is achieved in the *mAP* after setting the lower bound for n, we noted by visual inspection that using less than 100 points results in ambiguous descriptors. In the other hand, at sampling rate of 10% we get slightly better results than at 5%, which is intuitive since the description of the shapes becomes richer. Comparing cases d and e, and cases h and i, we see that incorporating local appearance makes no change in the *mAP* results.

Case l provides the best *mAP* at 0.322, more than 10 times higher than random ranking which would give a *mAP* equals to 0.030. The relative low *mAP* is due to the complex nature of the glyphs which usually have a lot of internal contours that create confusion.

We also can compare the various parameter combination cases by looking at Figure 7, where we present the average precision and average recall curves for the three cases with highest *mAP* and the two with lowest one in Table 1. Even when precision curves have similar behavior among them, it is possible to see that those corresponding to the best cases stay, in general, above the ones representing the worst cases. Nevertheless, they asymptotically converge since the sizes of the subsets are small compared with the size of the whole collection. In a similar way, recall curves start near each other and around top 25 diverge, showing better performance those corresponding to cases with higher *mAP* in table 1. However, the differences are negligible.

Figure 8 shows the standard precision-recall curves for the four queries with the best values of *AP* and the two
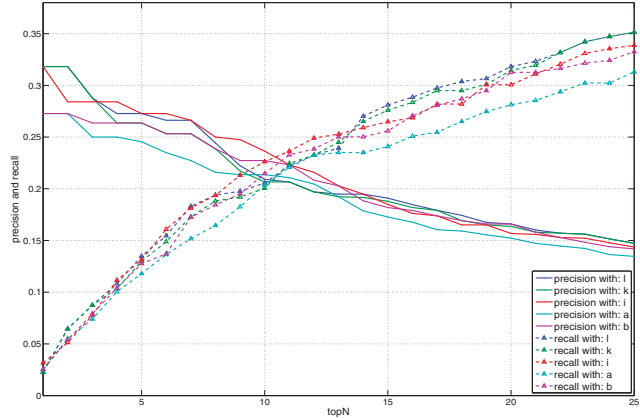

Figure 7. Examples of average precision and average recall curves for the three cases with highest *mAP* in Table 1 and for the two cases with lowest one. Each curve shows, from top 1 to top 25, the average over the 22 queries for the retrieved set of glyphs in each case.
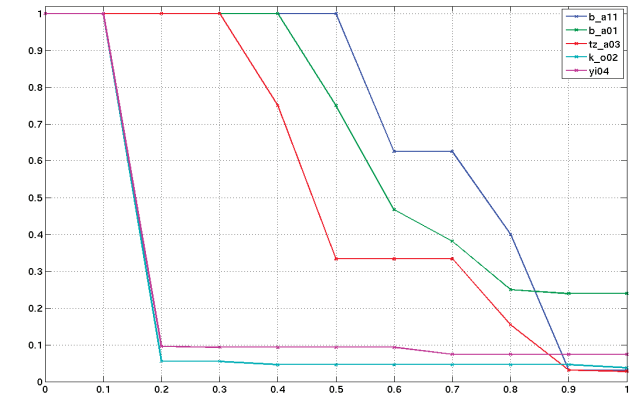

Figure 8. Standard precision-recall, in the best combination l, for the four queries with best *AP* and the two with the worst one. The legend of each line shows its *AP*.

with the worst one for the best parameter combination l. As we can see, the first two lines corresponding respectively to syllables b_a11 and b_a01, stay with 100% of precision until 40% and 50% of recall because these queries have relevant glyphs that are very similar and therefore retrieved with high ranking. For these queries, figure 6 presents the top 5 retrieved glyphs, 3 are relevant in both cases.

For the third curve, query tz_a03, we have a special case since it is a vertical shape and in Figure 6 we see that the retrieved glyphs are vertical as well. We found the same kind of behavior for queries $na10$ and $u08$, and also for horizontal shapes like those in queries $li04$ and $nu04$ (see Figure 5 for reference about the queries). This means that the Shape Context strongly takes into account the overall shape; the method should rely on dummy handlers to give more importance to inner details, which might be restrictive depending in the final number of points used to compute the permutation matrix $C(\pi, q)$.

The one to last line in Figure 8, query *k_o02* which is a hand, performs poorly due to the fact that we chose its relevant set under semantic criteria rather than visual similarity. More precisely, we included in the set all the hands found in the collection even those with different shapes like extended palms, fist, fist with thumb up or down, *etc*. The last curve, glyph *yi04*, gives poor results since it is the unique human face in the collection with its overall shape and pointing to the left, and in this case we also included faces with several shapes as relevant glyphs. However, retrieved glyphs for these two examples have similar overall shapes, as seen in Figure 6.

7. Conclusion and future work

We have presented novel work towards the integration of computer vision techniques in the analysis of ancient Maya cultural heritage materials. Our interdisciplinary approach is challenging and rich, as it genuinely addresses needs and open problems in archaeology, and poses a number of questions for computer vision research, resulting in a unique opportunity to integrate knowledge in archaeology, linguistics, and computing.

We studied a robust shape descriptor in a very complex dataset of Maya syllabic glyphs, obtaining encouraging results. The initial strategy of restricting queries only to segmented syllabic glyphs has been useful, however, this kind of signs represents only 10% of the total, and the analysis of other kind of Maya glyphs still remains as an open issue. As part of our study, we also proposed an improvement of the algorithm to compute dissimilarity between shapes with considerable amount of inner details. Based on the results obtained so far, we hope that the potential of automatic vision technologies could be refined to a degree that would be useful to retrieve glyphs modified by one of the phenomena mentioned in section 2, and to detect distinctive patterns of phonetic substitutions within Maya scripts.

Among the many research threads to study in the future, the first one will be to increase the number of syllabic queries making use of multiple annotators, along with the evaluation of other shape descriptors. Also, new analysis might be performed, such as handling transformations like rotation and reflection. Unlike traditional problems in vision, where invariance with respect to the above characteristics is desirable, they might be inadequate as rotated and mirrored Maya glyphs often have different meanings. Another area of research will be the analysis of the evolution of shape in glyphs over geographical regions and historical periods, using the large data set that is available from the AJIMAYA project.

References

[1] R. Baeza-Yates and B. Ribeiro-Neto. *Modern Information Retrieval*. Addison Wesley, May 1999.

[2] S. Belongie, J. Malik, and J. Puzicha. Shape context: A new descriptor for shape matching and object recognition. In *in Proc. (NIPS)*, pages 831–837, December 2000.

[3] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE T-PAMI*, 24(4):509–522, April 2002.

[4] H. Berlin. El glifo emblema en las inscripciones mayas. *in Journal de la Societé des Américanistes*, (47):111–119, 1958.

[5] N. Boujemaa, V. Gouet, and M. Ferecatua. Approximate search vs. precise search by visual content in cultural heritage image databases. *Procc. ACM MM-MIR*, 2002.

[6] V. Ferrari, L. Fevrier, F. Jurie, and C. Schmid. Groups of adjacent contours for object detection. *IEEE T-PAMI*, 2008.

[7] I. Graham. *Introduction to the Corpus, Corpus of Maya Hieroglyphic Inscriptions*. Peabody Museum of Archaeology and Ethnology, Cambridge, 1975.

[8] J. Justensona, W. Norman, L. Campbell, and T. Kaufman. *The Foreign impact on Lowland Maya Language and Script*. Number 53. Tulane University Press, 1985.

[9] 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 T-IP*, 13(3):302–313, 2004.

[10] M. Macri and M. Looper. *The New Catalog of Maya Hieroglyphs*, volume 1 The Classic Period Inscriptions. University of Oklahoma Press : Norman, 2003.

[11] S. Martin and N. Grube. *Chronicle of the Maya Kings and Queens: Deciphering the Dynasties of the Ancient Maya*. Thames and Hudson, 2000.

[12] K. Mikolajczyk and C. Schmid. Scale and affine interest point detectors. *Int. Journal of Computer Vision*, 60(1):63–86, 2004.

[13] A. Opelt, A. Pinz, and A. Zisserman. A boundary-fragment model for object detection. In *Europ. Conf. Computer Vision*, 2006.

[14] P. Quelhas, F. Monay, J. M. Odobez, D. Gatica-Perez, T. Tuytelaars, and L. V. Gool. Modeling scenes with local descriptors and latent aspects. In *In Proc. ICCV*, pages 883–890, 2005.

[15] J. Shotton, A. Blake, and R. Cippola. Contour-based learning for object detection. In *ICCV*, 2005.

[16] J. Sivic and A. Zisserman. Video google: A text retrieval approach to object matching in videos. *In Proc. ICCV*, 2:1470, 2003.

[17] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *IEEE T-PAMI*, 22:1349–1380, 2000.

[18] K. A. Taube. *The Major Gods of Ancient Yucatan*. Dumbarton Oaks Research Library, 1992.

[19] D. Zhang and G. Lu. Review of shape representation and techniques. *Pattern recognition*, 37:1–19, 2004.