Analyzing Ancient Maya Glyph Collections with Contextual Shape Descriptors

**Edgar Roman-Rangel · Carlos Pallan ·
Jean-Marc Odobez · Daniel Gatica-Perez**

Abstract This paper presents an original approach for shape-based analysis of ancient Maya hieroglyphs based on an interdisciplinary collaboration between computer vision and archeology. Our work is guided by realistic needs of archaeologists and scholars who critically need support for search and retrieval tasks in large Maya imagery collections. Our paper has three main contributions. First, we introduce an overview of our interdisciplinary approach towards the improvement of the documentation, analysis, and preservation of Maya pictographic data. Second, we present an objective evaluation of the performance of two state-of-the-art shape-based contextual descriptors (Shape Context and Generalized Shape Context) in retrieval tasks, using two datasets of syllabic Maya glyphs. Based on the identification of their limitations, we propose a new shape descriptor named Histogram of Orientation Shape Context (HOOSC), which is more robust and suitable for description of Maya hieroglyphs. Third, we present what to our knowledge constitutes the first automatic analysis of visual variability of syllabic glyphs along historical periods and across geographic regions of the ancient Maya world via the HOOSC descriptor. Overall, our approach is promising, as it improves performance on the retrieval task, has been successfully validated under an epigraphic viewpoint, and has the potential of offering both novel insights in archeology and practical solutions for real daily scholar needs.

Keywords Cultural heritage · Maya civilization · Archeology · Epigraphy · Image retrieval · Visual similarity · Shape descriptor · Histogram of orientation

E. Roman-Rangel (✉) · J.-M. Odobez · D. Gatica-Perez
Idiap Research Institute, 1920 Martigny, Switzerland
e-mail: eroman@idiap.ch

J.-M. Odobez
e-mail: odobez@idiap.ch

D. Gatica-Perez
e-mail: gatica@idiap.ch

E. Roman-Rangel · J.-M. Odobez · D. Gatica-Perez
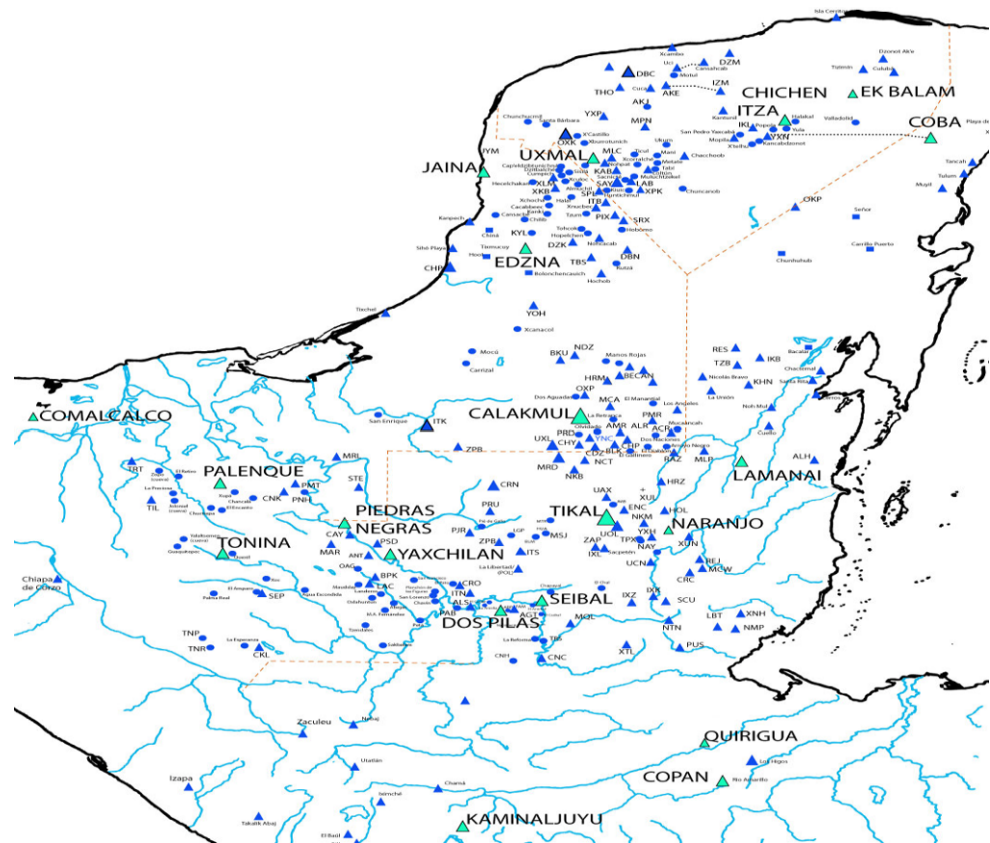École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

C. Pallan
National Anthropology and History Institute of Mexico (INAH), Mexico, Mexico
e-mail: pallan.carlos@gmail.com

1 Introduction

The study of invaluable historical materials with digital technologies is a field of multidisciplinary nature and significant to the society at large. Automatic and semi-automatic vision algorithms embedded in search and browsing tools can significantly facilitate the daily work of scholars in the humanities and the arts, including historians, archaeologists, anthropologists, linguists, curators, and photographers, who often need to consult, annotate, and catalog pictographic documents or photographed pieces. Furthermore, computer vision algorithms have the potential of providing new insights about existing theories in the humanities through the recognition and discovery of patterns and connections within and across pieces in a digital collection. Finally, visual search and browsing could boost the creation of educational systems useful for both professionals and the general public.

The ancient Maya civilization is regarded as one of the major cultural developments that took place in the

Fig. 1 The Maya region with main sites in larger font

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, which ranks among the most visually sophisticated ever created in world history. Maya hieroglyphs have been studied by western scholars for over two centuries, and today the field of Maya iconographic and hieroglyphic (epigraphic) analysis remains very active worldwide, given the complexity of the Maya writing system and the high frequency rate with which new archaeological sites and monuments continue to be discovered, which in turn increases the availability of source material for the research community devoted to its study, thus providing additional depth and richness to our knowledge of this ancient culture.

Our paper contains three contributions. First, we describe our interdisciplinary approach, which aims at developing automated techniques to work on a series of problems related to the analysis, documentation, and preservation of Maya pictographic data, tasks that to our knowledge have not been previously attempted and that integrate work in computer vision and archeology. Our source of data is the Hieroglyphic and Iconographic Maya Heritage project (AJIMAYA), an ongoing project under the auspices of Mexico's leading institution on Cultural Heritage management and research, the National Institute of Anthropology and History (INAH). As such, AJIMAYA is focused on collecting images and manual line drawings of all known hieroglyphic Maya monuments within the Mexican territory, thus representing one of the largest Maya pictorial collections and image data bases ever gathered. This project is currently working in archaeological sites located on three Mexican states (Tabasco, Chiapas, and Campeche, shown in Fig. 1).

Second, we analyze two versions of the Shape Context (SC) (Belongie et al. 2002; Mori et al. 2005), a shape descriptor successfully used for several other sources of shape data; and introduce the Histogram of Orientation Shape Context (HOOSC), a new shape-based descriptor that is better suited for the description of Maya glyphs. The Shape Context (Belongie et al. 2000) creates a robust, point-based description of global shape by encoding proximity and orientation relations of both adjacent and distant points, thus building a compact description that can be used to estimate visual similarity and perform matching between shapes. The Generalized Shape Context (GSC) (Mori et al. 2005) combines the discriminative power of the SC with the bag-of-words approach which results in a computationally efficient technique to compute glyph similarity. In turn, the HOOSC extends the GSC using a histogram of orientations and distance-based normalization to better describe the kind of rough thick contours often found in Maya glyphs, which also allows for reliably description of low resolution images. Our proposed method is evaluated in the context of automatic analysis of Maya syllabic phonetic signs (syllabo-

grams), which constitute one of the main building blocks within ancient Maya scripts. We demonstrate that the investigated shape-based descriptors are overall a promising methodology to handle Maya glyphs, which are very challenging by nature due to the complexity and wide variety of their pictorial content. We also show that our method leads to a more robust and computationally efficient representation of glyph shapes.

Third, we evaluate the proposed glyph descriptors based on real and current cultural heritage needs. We first do so with a retrieval task which uses two glyph datasets generated by different definitions of visual and semantic relevance. In archeology, glyph retrieval is useful to rank glyphs in a digital catalog for search purposes, a common task for epigraphy students and researchers whose time for execution would thus be ameliorated. Later, we analyze the visual variability of glyphs over three historical periods and four geographical regions of the Maya territory. A current research question in Maya epigraphy concerns the evolution of writing styles over temporal and geographical boundaries. Our analysis shows that glyphs present different degrees of variation, and that certain periods and regions appear to have wider variability than others, thus providing valuable information compared to what is currently known about this subject in archeology.

The paper is organized as follows. Section 2 describes our interdisciplinary approach. Section 3 describes related work. Section 4 presents the Shape Context algorithm. Section 5 introduces the HOOSC descriptor proposed in this paper. Section 6 describes the datasets used for experiments. Section 7 presents the experimental protocol. Section 8 discusses our results. Finally, Sect. 9 offers some concluding remarks and discusses open problems.

2 Our Interdisciplinary Project

The invention of writing was a rare event in world's history, only witnessed five different times: in Egypt, the Indus valley, Mesopotamia, China, and the Americas. In one way or another, all other writing systems derive from these regions.

Over 200 years of scholarly tradition focused on the Maya system have revealed many of its secrets, through a progression started with its mathematical cycles and astronomical tables, reaching its underlying phonetic keys and intimate functioning (Justeson et al. 1985), which in turn opened the gate to the study of Maya history (Berlin 1958; Proskouriakoff 1960), and derived aspects such as ancient geopolitics (Martin and Grube 2000) and religion (Taube 1992).

2.1 The Maya Writing System

The Maya culture originated some time during the transition period from the middle to the late Preclassic period (c.a.,

1,500–400 BC) in parts of present-day Mexico, Guatemala, Belize, and Honduras, arguably reaching its cultural peak during the late Classic period (c.a., 600–900 AD).

Maya writing was not an isolated phenomenon, but stems from a larger phonetic tradition that developed in southern Mesoamerica (Stuart et al. 2005). Some of the earliest Maya inscriptions date from the late Preclassic (c.a., 400 BC–250 AD), and originated in the Guatemalan Petén district, the Salama Valley, and the Pacific coastal region, and at sites such as San Bartolo, Takalik Abaj and El Baul. The script remained operational during at least 17 or 18 centuries. It was until later that it spread into the western and northern lowlands (the Usumacinta, Campeche, Belize-Quintana Roo and Yucatan). By late-Classic times (c.a., 600–900 AD), the usage of the Maya script became commonplace throughout the entire Maya lowlands, and has been detected on a few but steadily growing number of transitional lowland-highland sites as well. During the Terminal Classic (c.a., 800–950 AD), writing and scribal traditions continued, albeit on a fairly diminished capacity, with the exception of new and revitalized northern centers like Chichen Itza, Uxmal, Mayapan, and Santa Rita Corozal, all of which continued to operate after the so-called "Maya Collapse" (Sharer 1996).

The Maya script belongs to the family of the so-called logosyllabic writing systems. This term describes systems composed of two functionally distinct types of phonetic signs: logographs (word signs with meaning) and syllabic signs or syllabograms (used only to convey sound without meaning). In practice, several signs of both types could be arranged inside a single glyph-block, where usually logographs are phonetically complemented by syllables, either on initial position (i.e., as a prefix or superfix) or in final position (i.e., as a postfix or suffix). In general terms, syllables have a consonant-vowel structure (denoted as CV, for instance *b'a*, which is a sound with no meaning) while logographs alternate between consonant-vowel-consonant (CVC) and different compounds (i.e., CVC–VC, CVC–CVC, etc., for instance B'AHLAM meaning "jaguar"). Even though approximately 1,000 distinct signs have been identified thus far, only a maximum of 800 were used at any given time, of which approximately 80–85% have been deciphered and are readable today.

For the research described in this paper we focus on syllabograms, thus excluding logographs. The reason is that logographs comprise over 80% of all known signs, thus presenting a far greater degree of variability, both through time (diachronic variation) and geography (synchronic variation). Conversely, working with syllabograms allows for very challenging, yet more manageable sample that presents increased consistency. To exemplify the degree of complexity that makes such a strategy desirable, as our knowledge of the governing principles of the Maya script has advanced, so
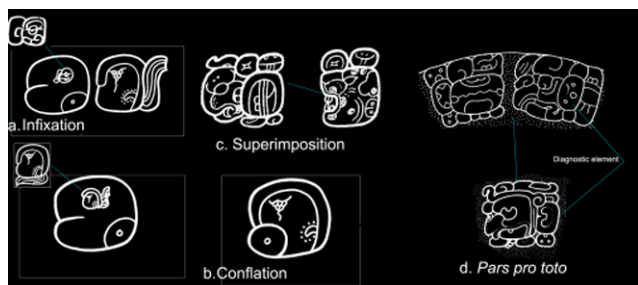
Fig. 2 Examples of complexity in the Maya writing system

Fig. 3 Paired columnar format for reading order of Maya inscriptions, where letters designate the columns and numbers the rows

Fig. 4 Example of the first and second steps in the digitalization process for Maya photographs

has our awareness of a number of its underlying non-linear possibilities, which include resources like conflation (when two signs are visually fused, retaining each its same relative size); infixation (when a sign is reduced in size and inserted within another); superimposition (when one sign partially covers another which main elements remain visible as background); and pars pro toto (when only a fraction containing diagnostic features of any given sign replaces the whole). Examples of these phenomena are shown in Fig. 2.

Typically, Maya glyphs are arranged on a paired columnar format which is referred to as the system of coordinates, where letters designate the columns and numbers the rows. In such a grid, a standard reading order for a text of 4 columns and 2 rows would be: A1, B1, A2, B2, C1, D1, C2, D2, as exemplified in Fig. 3. Longer texts usually conform to the same general pattern (with occasional exceptions).

2.2 Work in Data Collection

For some sites, the AJIMAYA project has already compiled a full photographic record of monuments, each of which has to undergo an eight-fold methodological treatment that generated some of the data used in this paper. Figure 4 shows the first and second steps of such a process consisting in:

1. Digital photographs, taken at night under raking-light illumination to bring out the level of detail that facilitates the study of eroded monuments.
2. Line drawings, traced on top of multiple layers of enhanced photographs under different light conditions, to capture the inner features that are diagnostic towards their subsequent identification.
3. Manual identification of glyphic signs with the aid of existing glyphic catalogs.
4. Manual transcription, i.e., rendering the phonetic value of each Maya sign into alphabetical conventions.
5. Transliteration, which involves representing ancient Maya speech into alphabetic form.
6. Morphological segmentation, which breaks down recorded Maya words into their minimal grammatical constituents (morphemes and lexemes).
7. Grammatical analysis, which uses common conventions to the fields of historical and structural linguistics, to indicate the function of each segmented element.
8. Translation, which involves rendering ancient Maya text on a modern target language, e.g., English.

2.3 Goals of the Collaboration

AJIMAYA's substantive goals encompass the safekeeping, preservation, study, and dissemination of monuments and written records from Mexico's Maya archaeological and cultural heritage. This project has been partially inspired by the pioneering work of Ian Graham (Graham 1975), who along with his collaborators at Harvard University's Peabody Museum of Archeology and Ethnology, provided the scholarly community with a comprehensive repository of photographs and drawings of such cultural items.

Our collaboration inserts itself as part of the general goals mentioned above. We target the application of computer vision technology in order to support Maya hieroglyphic decipherment. As a result of this fusion between technological and cultural heritage studies, we hope to obtain an array of research tools tailored-made to meet specific research needs for archeology. One of such useful tools would address one of modern decipherment's strongest needs: developing a more refined and updated hieroglyphic catalog,

which could be periodically updated, with the ability to incorporate input from multiple colleagues working in different parts of the world. Despite pioneering efforts on this regard by Thompson (1962) and Macri and Looper (2003), one of the challenges that limits the potential of existing catalogs operates at the taxonomic level: epigraphy needs an upgraded system for classifying the 800+ known signs, which separates the consensually 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 sign are readily accessible, in order to better assess its function within the script as a whole.

In order to succeed in creating these and other research tools, one of the basic abilities that needs to be developed is that of performing queries of a particular glyph, which could retrieve the contexts where equivalent instances occur. Taking a step further, we would like to develop the ability of automatically detecting not only specific instances of particular signs, but also their variants, such as their allographs (different signs indicating the same phonetic value), their homophones (signs with similar sound, which meaning could differ), and their polivalencies (signs that could take more than one phonetic value, depending on the context).

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

3 Related Work

The work in this paper relates to both the use of computer vision techniques for cultural heritage preservation or analysis, and to the development of computer vision techniques for shape analysis and retrieval. In the following, we discuss related works in those areas.

3.1 Automatic Visual Analysis of Cultural Heritage

The use of computer vision techniques for automatic indexing and retrieval of 2-D imagery in cultural heritage has been the topic of important research in the past recent years. One of the most representative works in this domain is Lewis et al. (2004), in which a system for retrieval of paintings and photos of art objects, using content and metadata, was developed by a multidisciplinary team involving a group of European museums. The project identified two relevant user tasks for specialists: 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 region-of-interest queries appears in Boujemaa et al. (2002). Techniques that address other relevant tasks, for instance the detection of visual patterns and trends in image collections, like the characterization of ancient Chinese paintings by artist style (Li and Wang 2003), have also been proposed. It is well known, however, that classic content-based image retrieval (CBIR) techniques are often limited to handle variations of illumination, appearance, and viewpoint (Smeulders et al. 2000). In this work, we address the retrieval of complex shapes using more recent techniques for visual entity representation (bag-of-words constructed from contextual shape features), which are novel within this field.

3.2 Image and Shape Representations for Retrieval Applications

The representation and matching of visual entities has experienced much progress since the invention of robust local viewpoint-invariant features, computed over automatically detected local areas, to represent objects and scenes (Mikolajczyk and Schmid 2004). The local character of these features provides robustness to image clutter, partial visibility, and occlusion, while their invariant nature addresses several issues related to changes in viewpoint and lighting conditions. In this context, one research direction has focused on modeling objects by histograms of quantized local descriptors (bags-of-visterms) (Sivic and Zisserman 2003; Willamowski et al. 2004; Quelhas et al. 2005), which allows for fast retrieval applications in large collections. However, the main limitation of these approaches is that the spatial information is lost given the assumptions for bag construction. In addition, the above works represent the visual content with appearance descriptors which might not be well adapted to the shape/binary images used in our work.

Representation of shape has a long history. The work in Zhang and Lu (2004) 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 local structure. Global representations like Fourier or moment descriptors are usually sensitive to variations in some regions of the shape. Structural approaches, which represent shapes by sets of local contour segments potentially organized as trees do not suffer from this drawback (Zhu et al. 2008; Lu et al. 2009). As an example, such a representation is used by Zhu et al. (2008), where the search for a shape in an image is formulated as a set-to-set matching problem, relying on Shape Context descriptors (Belongie et al. 2000)

built from control-points. The resulting hard combinatorial problem of matching control points and foreground contour selection, is solved using linear programming optimization methods. A similar matching formulation is used by Lu et al. (2009) but is resolved in a particle filtering framework. Despite the use of these recent optimization schemes, the set representation quickly becomes large when dealing with complex shapes like our glyphs, and make these techniques computationally very expensive for shape comparison.

Recently, the use of shape to represent objects regained attention. However, the focus is on tasks other than shape retrieval, namely model learning from few training images (Jiang et al. 2009), shape discovery (Lee and Grauman 2009), model matching in images (Zhu et al. 2008; Lu et al. 2009), or exemplar-dependent shape distance learning (Frome et al. 2007). In most cases, these works rely on finding feature correspondences and/or shape alignment using techniques that are too slow for retrieval applications in large databases (e.g., the thin plate spline robust point matching algorithm; Jiang et al. 2009). In addition, due to their different goals, they mainly address the representation of images of 3-D real objects, currently handling object categories with limited complexity, like apple logos, mugs, or bottle shapes, which are substantially different than our Maya glyphs.

Taking inspiration from the success of the appearance-based local descriptors for object representation, recent work has investigated similar approaches to represent shape information in images, by exploiting vocabularies of boundary fragments in different ways (Shotton et al. 2005; Opelt et al. 2006; Ferrari et al. 2008). As an example, Ferrari et al. (2008) used a quantized version of a local segment network (mainly segment pairs) as a structural shape representation. 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 do not provide appropriate explicit matching scores for retrieval applications. In contrast, the shape-context representation (Belongie et al. 2000, 2002) provides a holistic shape representation through descriptors that integrate shape and geometric discrepancy measures, providing a good framework for shape comparison. In addition, its generalized version (Mori et al. 2005) uses contour information and quantized descriptors, and lends itself to fast retrieval implementations. We have followed this approach, proposing several improvements (a better contour orientation representation, using an histogram constructed with a kernel-based approach rather than a dominant orientation descriptor, and a more appropriate normalization scheme), which are shown to significantly improve retrieval results.

4 Shape Context Algorithm

Here we summarize the Shape Context algorithm (Belongie et al. 2000, 2002) which is the starting point for our work.

4.1 Descriptors

Shape Context (SC) takes as input a finite set of n 2-D 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 of the shape P, its histogram sc_i^P, called shape context descriptor, is computed as the distribution of the relative position of the $n - 1$ points with respect to p_i. In other words, the kth entry $sc_i^P(k)$ of sc_i^P is defined as:

$$sc_i^P(k) = \frac{|P_i^k|}{|P| - 1}, \quad k = 1, \ldots, K, \tag{1}$$

where $|\cdot|$ denotes the set cardinality operator, K is the number of bins in sc_i^P, and P_i^k denotes the set of points of P falling in the spatial bin k with respect to the point p_i:

$$P_i^k = \{p_j \in P : p_j \neq p_i, (p_j - p_i) \in \text{bin}(k)\}, \tag{2}$$

where $p_j - p_i$ denotes vector difference. Such a histogram is computed on 5 normalized distance intervals which in total span twice the mean pairwise distance of all the points in the image, and 12 angular intervals covering a complete perimeter around the reference point. This results in a histogram of $K = 60$ bins, which are uniform in log-polar space, making the descriptor more sensitive to nearby points than to the farthest ones. The SC descriptor construction process is illustrated in Figs. 5(a)–5(h).

According to Belongie et al. (2002), measuring similarity between two shapes P and Q can be done using the χ^2 statistics for each pair of points p_i of P and q_j of Q as:

$$C_{ij}(p_i, q_j) = \chi^2\left(sc_i^P, sc_j^Q\right)$$
$$= \frac{1}{2} \sum_{k=1}^{K} \frac{[sc_i^P(k) - sc_j^Q(k)]^2}{sc_i^P(k) + sc_j^Q(k)}. \tag{3}$$

In Belongie et al. (2002), a distance between shape tangents at points p_i and q_j was also exploited. In our experiments however, it produced no improvement.

4.2 Point-to-Point Matching and Shape Similarity Index

A scalar measure of similarity (similarity index) between two shapes P and Q can be obtained through point-to-point correspondences. One way to do so, is to find the permutation Π of the vector of points P that minimizes the sum of the matching costs given by:
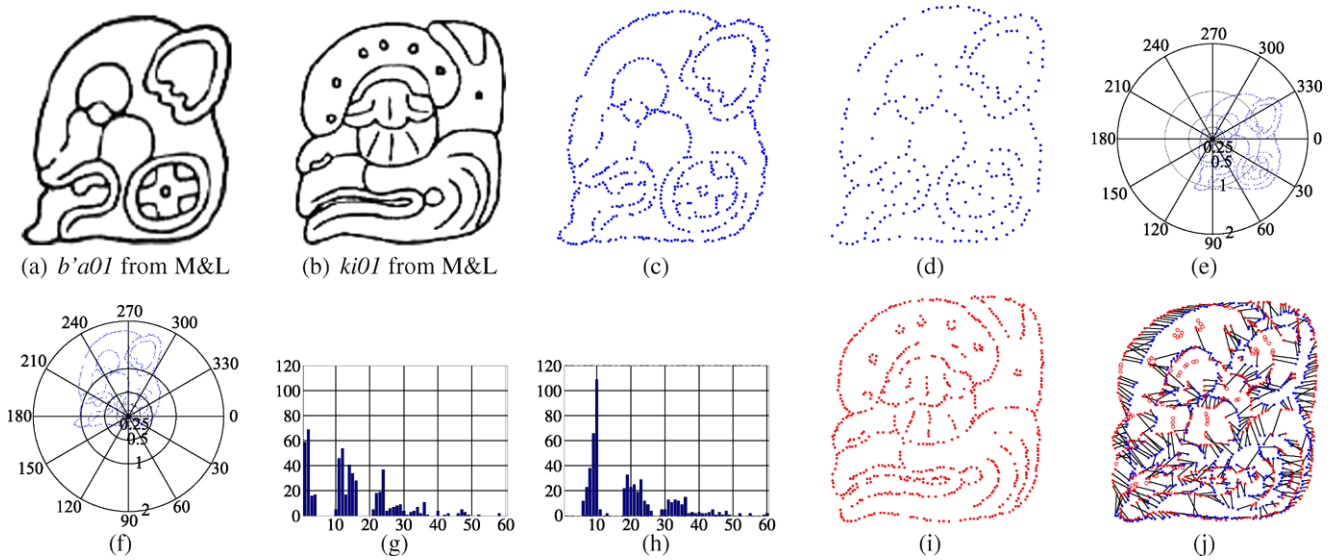
$$M(\Pi) = \sum_i C(\pi_i, q_i), \tag{4}$$

(a) *b'a01* from M&L   (b) *ki01* from M&L   (c)   (d)   (e)

(f)   (g)   (h)   (i)   (j)

Fig. 5 Shape Context process for description and matching: (**a**) glyph *b'a01*, (**b**) glyph *ki01*, both taken from Macri and Looper (2003). (**c**) and (**d**) point representation of the contours of (**a**) at 2 sampling rates, 10% and 5% respectively. (**e**) and (**f**) log-polar binning spaces for two sample points. (**g**) and (**h**) the SC histogram descriptors for the two sample points. (**i**) glyph *ki01* sampled at rate 10%. (**j**) matching between *b'a01* and *ki01*, correspondences are connected by a *line* while dummy matches are presented by *disconnected circles*

where π_i is the ith entry of the vector Π.

A constraint to this minimization problem is that the number of points should be the same for the two glyphs, i.e., $|P| = |Q| \rightarrow |\Pi| = |Q|$. If the restriction is not satisfied, one can make use of "dummy handlers". Assuming that $|P| < |Q|$, we can add dummy points to P to have the same length in both vectors, and fill the corresponding entries in matrix C with a dummy value ε_d. In this way, points in Q with expensive matching cost are forced to match dummy points in P. Nevertheless, all points in P are still forced to have a real match in Q, which might not be desirable in some cases. To allow outliers in P, we introduce dummy handlers in both sets P and Q by increasing the dimension of the cost matrix C up to $\lceil m(1 + \psi) \rceil$, where $m = \max(|P|, |Q|)$ and ψ is the dummy rate, i.e., the fraction of points in the biggest set allowed to have no real match. Any new entry in C is then set to the value ε_d. Figure 5(j) illustrates the best point-to-point matching between two similar glyphs.

By visual inspection on the several glyph retrieval experiments, we noticed that the similarity index proposed in Belongie et al. (2002) is not discriminative enough, as they compute the average of the minimum pairwise matching scores allowing one-to-multiple point matches. To alleviate this issue we rather use the average of the matching costs suggested in Roman-Rangel et al. (2009),

$$d_{sc} = \frac{1}{\lceil m(1+\psi) \rceil} \sum_{i=1}^{\lceil m(1+\psi) \rceil} C(\pi_i, q_i) = \frac{M(\Pi)}{\lceil m(1+\psi) \rceil}. \quad (5)$$

This function discriminates better between visually similar and non-similar glyphs. It also takes into account all the dummy assignments which might contribute significantly to the score depending on the number of points in the two shapes to be compared.

While the SC is a powerful descriptor, it is computationally expensive and therefore other faster options are desirable. This motivated our proposed approach which is presented in the next section.

5 Our Approach

The need to optimize the permutation Π of points to compute the similarity between any two shapes does not lend the original SC method to fast retrieval. In Mori et al. (2005) this problem is alleviated working with a bag of quantized descriptors (called *Shapemes*), which also improves the discriminative ability of the descriptor by adding shape orientation information. We describe this approach below discussing its limitations, and presenting the novelties we introduce in our proposed method.

5.1 Generalized Shape Context (GSC)

Incorporating contour orientation in the cost function only at the reference point as proposed by Belongie et al. (2002) appears to be limited (Mori et al. 2005; Roman-Rangel et al. 2009). In contrast, replacing the simple counting of points in each spatial bin k by the sum of the unit gradient vectors
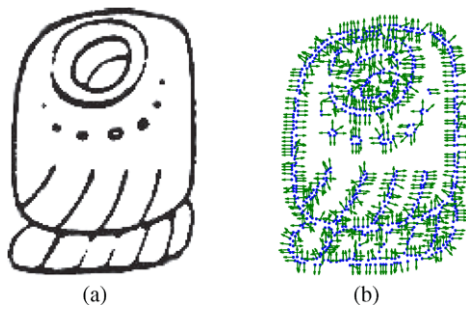
Fig. 6 (**a**) One instance of glyph *b'a*. (**b**) Opposite gradient directions resulting from the double sides of wide contours

t_j of all points p_j falling within that bin, provides with a richer description of shapes. This leads to the definition of the *Generalized Shape Context* (GSC) descriptor gsc_i^P of a point p_i, in which each spatial bin k is described by its 2-D dominant orientation vector $gsc_i^P(k)$:

$$gsc_i^P(k) = \sum_{p_j \in P_i^k} t_j, \qquad (6)$$

where P_i^k, the set of points falling into bin k relative to the point p_i, is defined in (2), and t_j denotes the gradient vector. As a result, using the same 60 spatial bins than in the SC case, the resulting GSC has a dimension of 120.

5.2 Histogram of Orientation Shape-Context (HOOSC)

The GSC descriptor suffers from several limitations, some due to the specific data we use, while others are more general. We describe below the improvements we propose.

5.2.1 Edge Orientation Representation

Unlike other image sources where contours emerge from the contrast between two object regions, Maya glyphs are frequently drawn by traces in which internal and external contours must be considered as the same part of a shape element. Unfortunately, glyph shapes usually have very rough contours as they come from photographs taken from stones where the original inscription has very wide traces or is eroded. Therefore, the extraction of the contours might result in double lines as illustrated in Fig. 6, one internal line and one external line whose gradients will often have opposite direction and whose contribution to (6) cancel each other, thus resulting in unreliable estimates. Very poor quality images may also produce the same effect.

To handle this issue, we only exploit information about contour orientation, not contour direction. We thus normalized the contour gradient using only angles between the interval $[0, \pi)$ to avoid term cancellations when summing the gradients of the same "stroke" represented by double lines.

5.2.2 Histogram of Orientation

As a consequence of the above, the algebraic sum of gradient vectors in (6) does not make sense anymore, as modulo π orientation vectors are difficult to add. In addition, in many cases the notion of dominant orientation is not fine enough, as spatial bins (especially outer ones, see for instance Fig. 5) often contain sets of edges with different orientations. Thus, following a successful trend in computer vision (e.g., SIFT or HOG features), the contours in each bin can be more precisely represented using a histogram of local orientations. To do so and to take into account uncertainty in orientation estimation while avoiding hard binning effects, we propose to consider the orientation measure as a probability distribution instead of a point estimate, and to resort to an (unnormalized) kernel-based approach for orientation density estimation. That is, the density for angle θ in the spatial bin k of point p_i is estimated as:

$$f_i^k(\theta) = \sum_{p_j \in P_i^k} \mathcal{N}(\theta; \theta_j, \sigma^2), \qquad (7)$$

where $\mathcal{N}(\theta; \mu, \sigma^2)$ denotes the value at angle θ of a Gaussian of mean μ and variance σ^2. Empirically $\sigma = 10$ degrees performs well. Then, the value of the orientation histogram in bin $[a, b]$ is simply obtained by integrating the density $f_i^k(\theta)$ within this interval (modulo π). A new parameter N_θ is introduced to control the number of orientation bins. Empirically, $N_\theta = 4$ or $N_\theta = 8$ bins allows for a suitable description, which given the 60 regions of the log-polar spatial binning, leads to the definition of the HOOSC, a histogram of orientation shape-context descriptor of 240 or 480 dimensions for each point p_i of the shape P, and denoted as $hoosc_i^P$.

5.2.3 Normalization of the Histogram of Orientation

Finally, the radial distances of the spatial binning, which are quantized in 5 intervals following a logarithmic law, produce an important discrepancy between the unnormalized orientation histograms computed in each spatial bin: as the area of the central bins are much smaller than that of the bins in the outer circles, the amount of observations (and thus the magnitude of the corresponding histogram of orientation entries) are much larger on the outer bins. Thus, when comparing two descriptors, too much emphasis is given to observations farther away from the reference point. The GSC method suffers such an issue. To avoid it, we perform a local and independent normalization for each of the distance intervals of the descriptor referred from now on as *rings* (see Fig. 5), meaning that the contribution of each point to the full descriptor effectively depends on how far it is from the reference point.

5.3 Shapeme-Based Similarity Measure

We rely on a quantization approach in which all the considered shape descriptors are extracted and clustered, and then represented by an index that points to the cluster they belong to. Such an approach has been shown to allow the design of fast retrieval applications (Sivic and Zisserman 2003). In the context of shape retrieval, Mori et al. (2005) called the obtained vocabulary of cluster centers as *Shapemes*.

In practice, for a given descriptor type (GSC or HOOSC), we performed a K-means clustering into N_{sh} shapemes, and then represented each glyph P by the normalized distribution of quantized descriptor indices over N_{sh}, called bag-of-shapemes and denoted by $bosh^P$. To compare two glyphs P and Q, we compute the distance between their bag-of-shapemes according to the L1 norm, which empirically performed better than other alternatives:

$$d_{sh}(P, Q) = \|bosh^P - bosh^Q\|_1$$

$$= \sum_{l=1}^{N_{sh}} |bosh^P(l) - bosh^Q(l)|. \tag{8}$$

6 Data

Two datasets of syllabic Maya glyphs are used in our work. The first one is taken from a relatively well known catalog of Maya hieroglyphs (Macri and Looper 2003), and the second is provided by AJIMAYA. Several Maya syllables might have more than one associated visual representation (variant or allograph), meaning that the same syllable could be represented by two or more glyphs where visual similarity is neither obvious nor necessarily found. Conversely, other glyphs associated to distinct syllables could present visual similarities among them.

6.1 Macri and Looper (M&L) Dataset

Our first dataset taken from Macri and Looper (2003), consists in 297 syllabic Maya glyphs distributed over 80 syllables. Despite the fact that some syllables might be rendered by different glyphs, each glyph itself is represented by a single holistic instance, thus excluding multiple images for the same instance. The glyphs are resized to fit 256 pixels in their largest dimension while keeping proportionality in the smallest axis. Since glyphs have tremendously varying degrees of detail and complexity, we decided to represent all the contours not with a fixed number of points as in Belongie et al. (2002), but with a percentage of the total number of points in each of the raw images.

From the collection, we randomly selected the 22 glyphs shown in Fig. 7 to use them as queries. For each of them, we manually labeled as relevant all the similar glyphs found
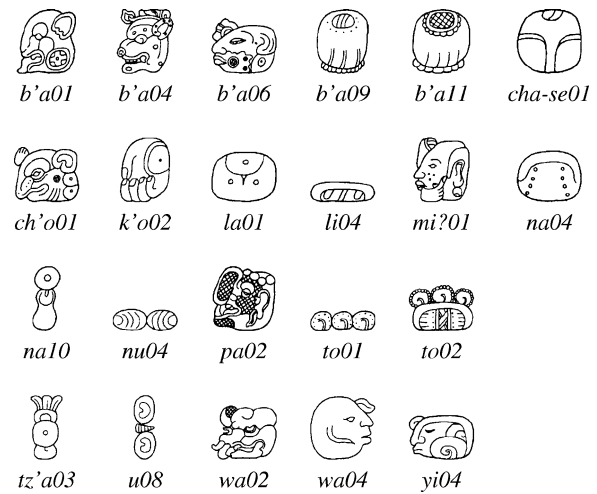

Fig. 7 Set of 22 Maya syllabic query glyphs from M&L dataset with names constructed using their corresponding sound, plus an identifying number added at the end of each glyph

in the collection, purely based on visual similarity. Note that the set of queries is diverse: glyphs are often considerably different from each other.

6.2 INAH Dataset

The second dataset was manually extracted from files provided by INAH. Each file contains a table representing different instances of the same glyph, with variations along temporal and regional axes. The synchronic or time dimension is divided in 3 periods: period 1 corresponding to the Early Classic (200–500 AD), period 2 to the Late Classic (600–800 AD), and period 3 to the Terminal Classic (800–950 AD); while the diachronic or regional dimension is organized in 4 main regions: the central region (Petén), the southern region (Motagua), the western region (Usumacinta), and the northern region (Yucatán peninsula) (refer to Fig. 1), thus allowing for up to 12 possible instances per glyph. The dataset consists in 84 images distributed over 8 glyphs. Figure 8 shows the 8 tables that form this dataset, note that not all the entries in the tables are filled as some of them either could not be found or would require much more time to be located within the known corpus.

Although a majority of syllables could be arranged in such a tabular form, it is advisable at this stage to focus the analysis on signs with a very high rate of occurrence within the hieroglyphic corpus, thus yielding glyph tables with as few gaps as possible. As simple as it might sound, the selection process of finding the most suitable candidates for analysis can often prove overly difficult and time-consuming, given that it relies almost entirely on specialized epigraphic knowledge and abilities that take many years to develop, such as visual memory. It is precisely because of
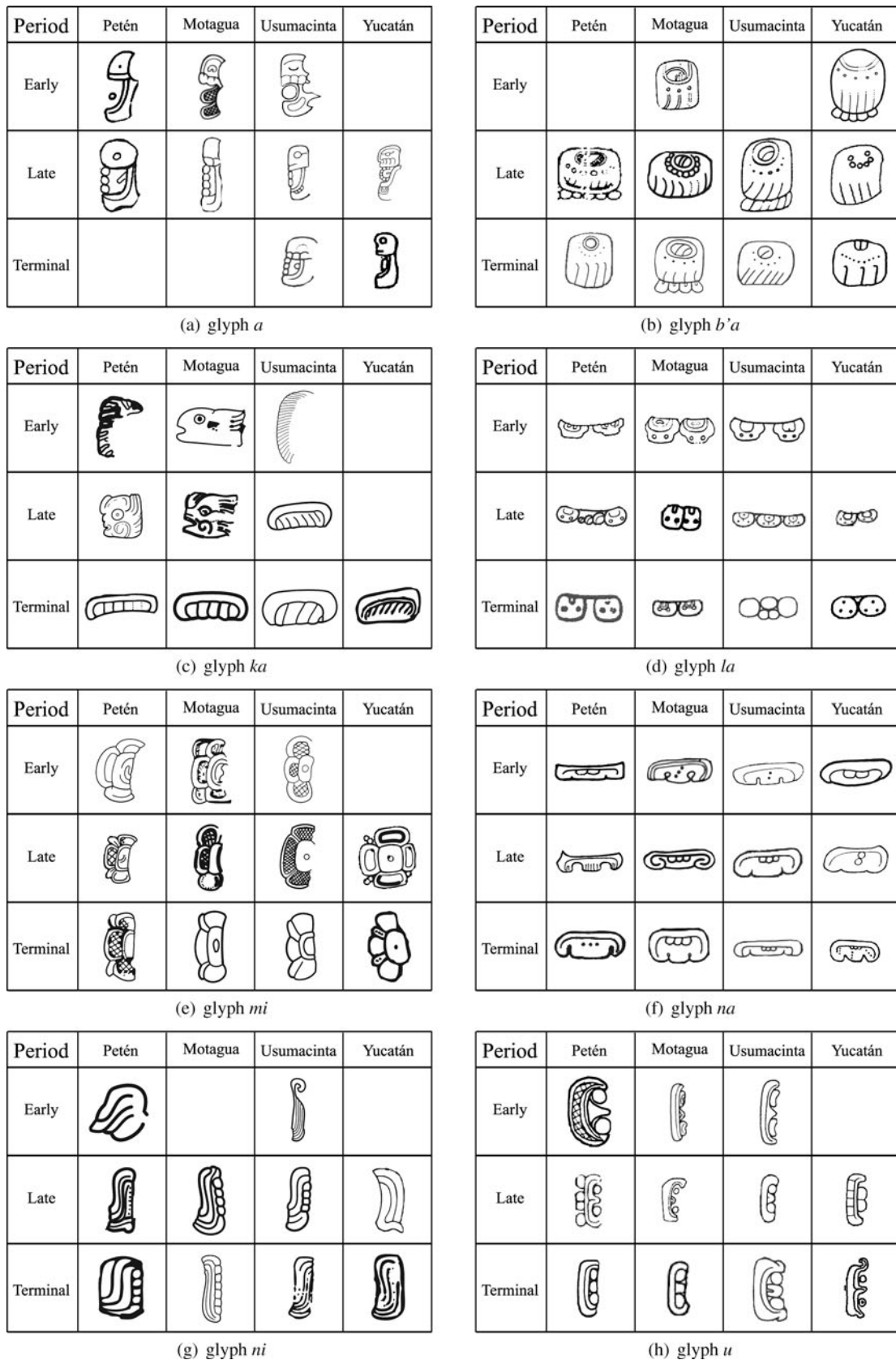
(a) glyph *a*

(b) glyph *b'a*

(c) glyph *ka*

(d) glyph *la*

(e) glyph *mi*

(f) glyph *na*

(g) glyph *ni*

(h) glyph *u*

Fig. 8 Glyph tables for the 8 syllables in the INAH dataset. The instances correspond to the Early, Late and Terminal Classic Periods, and to four main regions of the ancient Maya civilization. Missing instances either could not be found or would require more time to be located within the known corpus. Note that several instances are quite noisy

these difficulties that the field of Maya studies needs a re-search tool which could automatically retrieve such candidates.

The images in this set are roughly double the resolution than those of M&L dataset. However, at this point it was not possible to get a consistent high-quality from all of the samples, due to the varying qualities of the source materials within the AJIMAYA database and the different degrees of erosion that each text presents, ranging from the almost illegible to the fairly well preserved instances.

7 Experimental Protocol

In this section we describe three tasks we performed with the two datasets described in Sect. 6 as well as with one extra dataset constructed as a mixture of them.

1. *Glyph retrieval in the M&L dataset.* Using SC descriptors and the index similarity in (5), we performed a series of experiments to analyze the effects of the contour sampling rate n and the dummy rate ψ. Later with (8), we explored the impact of three parameters under the bag-of-shapemes approach: the descriptors used to build the vocabulary (GSC or HOOSC), the number of consecutive rings used to build the descriptor, and the number N_{sh} of words in the shapeme vocabulary. To evaluate retrieval quality we made use of the mean average precision measure (*mAP*) (Baeza-Yates and Ribeiro-Neto 1999) in all cases.

2. *Glyph retrieval in a mixed dataset.* To further evaluate our HOOSC method, we extracted from the M&L dataset the 8 glyphs that correspond to the syllables presented in the INAH dataset, and used them as queries to retrieve glyphs from the latter set. Our goal is to conduct a retrieval task where the relevance judgment of the queries is not based only on visual similarity but on semantics; i.e., the expert knowledge which establishes that all the instances relevant to a query correspond to the same syllable.

3. *Assessing glyph variability across periods and regions.* The last task consists in analyzing trends of visual variability of the instances for the 8 glyphs in the INAH dataset. Such an evaluation was done both at the intra and interclass levels with the goal of finding out whether there are glyphs showing more consistency across historical periods and regions, as well as being able to numerically characterize visual similarity among the glyphs.

   We consider this evaluation important as it addresses one of the main needs of scholars working on archaeological and epigraphic related studies; they look for automatic tools to help them classify known and new hieroglyphs, as well as a better understanding of the visual relationship among instances of the same sign and

Table 1 Tested combinations with Shape Context. Values for sampling rate n, dummy rate ψ, and mean Average Precision *mAP* computed over the 22 queries. The best result obtained is shown in bold

Case	n	ψ (%)	*mAP*
a	max(100, 5%)	0	0.301
b	max(100, 5%)	10	0.315
c	max(100, 5%)	20	0.319
d	max(100, 10%)	0	0.311
e	max(100, 10%)	10	0.319
f	max(100, 10%)	15	**0.322**

with respect to other signs. Examples of these needs can be found in two seminal taxonomical and paleographical contributions conducted by Grube (1989) and Lacadena (1995), which constitute ground-breaking efforts to account for the amount of synchronic and diachronic variability found within the Maya script.

8 Results

In this section we present experimental results for each of the tasks described in Sect. 7, along with a discussion about the performance of our framework.

8.1 Results for Glyph Retrieval

In the initial stage of our experiments we did an analysis to assess how well the glyph contours in the M&L dataset are represented at different sampling rates. We tried representing the contours with 2%, 3%, 5%, 10% and 15% of the number of points in the original shapes, and observed that with ratios less than 5% many glyphs are represented by less than 100 points, thus yielding rather poor representations. Conversely, while 10% and higher percentages produce robust representations, they also make the SC 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 minimum bound of 100 points, that is, $n = \max(100, 5\%)$ and $n = \max(100, 10\%)$.

8.1.1 Shape Context

Table 1 shows the retrieval results for different parameter values. In all cases, the dummy assignment cost was set to $\varepsilon_d = 0.25$.

It is clear that a more populated representation of the contours results in a more accurate description. Combinations *a* and *d* correspond to the original SC, with dummy assignments only in the smallest set of points but not in the largest one. From then on we can see that adding dummy assignments in both shapes increases the precision, and the more we add the higher it is, i.e., $mAP(a) < mAP(b) < mAP(c)$,

Table 2 Estimated *mAP* computed over 22 queries in the M&L dataset for different combinations of N_{sh} (vocabulary size) and the number of consecutive rings N_{rings} (spatial context) used to build the HOOSC descriptor. The best result is shown in bold

N_{rings}	ρ	N_{sh}				
		100	300	500	800	1000
2	0.25	0.192	0.239	0.278	0.279	0.264
3	0.50	0.232	0.291	0.296	0.308	0.292
4	1.00	0.260	0.296	0.337	0.321	0.330
5	2.00	0.315	0.350	0.379	**0.390**	0.387

and $mAP(d) < mAP(e) < mAP(f)$. Intuitively, SC now relies on a higher proportion of correct matches, while dummy matching cost are assigned to the unreliable, often wrong, matches. Note that the sampling rate is higher in case *d* than in case *c*, however the precision in *c* is higher as it makes use of dummy handlers. Case *f* provides the best *mAP* at 0.322, which is 10 times higher than random ranking which would give a *mAP* equals to 0.030. In absolute terms, the *mAP* of 0.32 is relatively low, and is due to the complex nature of the glyphs with many internal contours that create confusion.

On the other hand, we notice that increasing the value ε_d does not improve the *mAP* (such results are not shown here). The original similarity index of Belongie et al. (2000) and the first one proposed in Roman-Rangel et al. (2009) (not shown), lead to lower retrieval performance. An important restriction for the SC is that the higher the dummy rate, the slower is to find the permutation Π, reason for which we did not increase further the dummy rate ψ.

8.1.2 Shapemes

Table 2 shows the *mAP* results when varying two parameters of the HOOSC descriptor: the shapeme vocabulary size N_{sh}, and the number of consecutive rings N_{rings} of the spatial context considered to build the descriptor. We observe a trend to improve performance when increasing the vocabulary size up to 800 shapemes, after which the precision starts to degrade. Also worth of attention is that a global description performs better than a local one, no matter the vocabulary size (e.g., 300 shapemes with 5 rings is better than 800 shapemes with only 4 rings or less), which indicates that adding context is critical. Using 800 shapemes and the whole spatial context, our approach increases *mAP* up to 0.390, thus providing a relative improvement of 21.1% over the original SC.

Figure 9 shows examples of the retrieved glyphs with the Shapemes-HOOSC approach. Note that several retrieved glyphs belong to the same semantic class as their queries (i.e., they correspond to the same sound). However, glyphs that share an overall silhouette are sometimes confused by the HOOSC approach, even if they are rich in internal details (e.g., *yi04* and *pa02*). For the queries shown in Fig. 9,
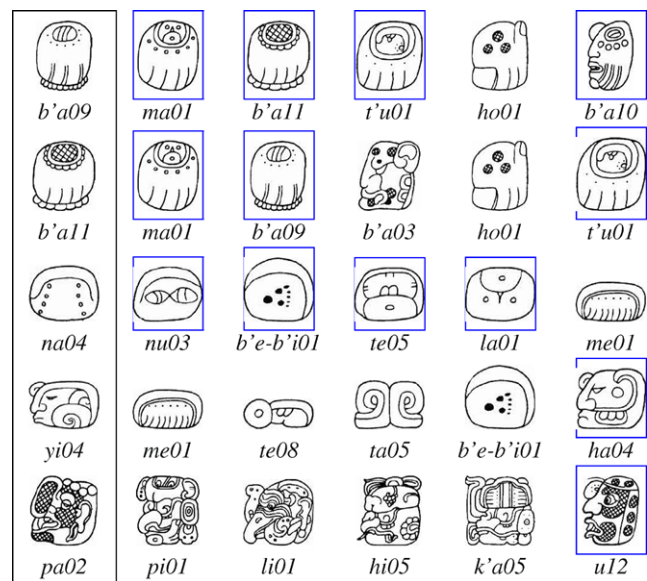

Fig. 9 (Color online) Examples of retrieved glyphs with *Shapemes*-HOOSC on the M&L dataset. The first column shows the three queries with highest *AP* and the two queries with lowest *AP* (in descending order). Then, each row shows the top 5 retrieved glyphs with the relevant ones enclosed by a *rectangle*

the HOOSC retrieves at least one relevant element in the top 5 retrieved vector, while SC does not. Note that the query *b'a11* (second row in Fig. 9) retrieves signs which previous glyphic catalogs find logical to place adjacent to one another, thus regarding them as pertaining to the same class from an iconic point of view (i.e., the purely visual representation of any given sign, independently of its associated phonetic or semantic values). For instance, Thompson (1962) places the signs *ma01* right next to the *b'a* syllable (Thompson names T502 and T501 respectively), while Macri and Looper (2003) intuitively arranged them exactly in the same order as it was automatically generated (Macri & Looper names XE2 and XE3).

8.1.3 Impact of Descriptor Type on Retrieval

We also evaluated the specific contribution of the modifications we proposed. More specifically, we compared five cases:

– SC: This corresponds to the best combination of the SC algorithm (case *f* earlier in this section).
– *GSC*: This is the original approach proposed by Mori et al. (2005) which sums all the local tangent angles within a bin and represents it by the dominant orientation.
– *GSC-mod(π)*: This is the same as *GSC*, but with orientation angles computed on the contours over the $[0, \pi)$ range only, to avoid mutual cancellation when using data with rough contours or double lines.
– HOOSC-*mod(π)*: This represents the contour spatial region by a histogram of orientation, with $N_\theta = 4$ or $N_\theta = 8$.

Table 3 Different approaches explored for retrieval

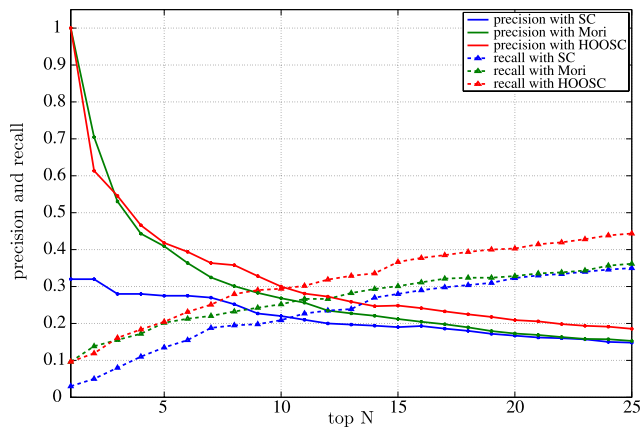| Descriptor type		Approach	mAP
SC		case f	0.322
GSC		*GSC*	0.149
		GSC-mod(π)	0.279
HOOSC	$N_\theta = 4$	*mod(π)*	0.315
	$N_\theta = 8$	*mod(π)*	0.320
	$N_\theta = 8$	*mod(π), normalized*	**0.390**

Fig. 10 (Color online) Average precision (continuous lines) and average recall (dotted lines) curves for the retrieval task in the M&L dataset with three methods; in *blue* the approach with best *mAP* in SC-based retrieval (case *f*), in *green* the GSC method, and in *red* the HOOSC method

– HOOSC-*mod(π)-normalized*: The same as above, but with a descriptor normalization conducted independently for each spatial ring, and $N_\theta = 8$.

All the shapeme approaches used $N_{sh} = 800$ and 5 rings, i.e., full spatial context. Table 3 shows the *mAP* results. GSC performs surprisingly poorly (0.15), both with respect to SC (0.32) and our method (0.39). We noted that, as discussed in Sect. 5.2, opposite gradients happened often, thus resulting in mutual cancellations of the directions and imprecise descriptions. When considering only orientation rather than direction, *GSC-mod(π)* increases the performance but still remains below that of SC. The use of the histogram of orientation provides better results, with a slight advantage when using more bins. Table 3 shows the need for a proper normalization procedure to avoid giving too much weight to the contribution of outer rings in the descriptor.

When looking at the top N retrieval results in each experiment (Fig. 10), we observe that the GSC approach (referred to as Mori) performs similarly to the HOOSC method at lower N, but then degrades more rapidly as more glyphs are retrieved, achieving the same performance as SC.

8.1.4 Computational Time

Besides improving the *mAP*, our approach is much faster than SC. Given a query image, the shapeme approach needs to compute a set of HOOSC descriptors and a bag-of-shapemes, and then compare it against each of the bag-of-shapemes for the elements in the dataset. On the other hand, the SC approach requires to calculate the set of point descriptors, and find a set of point to point correspondences (i.e., solving the assignment problem) for each of the set descriptors within the collection. In practice, the bag-of-shapemes approach takes only 0.05 seconds for the whole process, while SC does it in 227.4 seconds (3′48″), which results in a speedup factor of ≈4500 using a standard desktop machine. All the experiments were conducted with non-optimized Matlab code for our approach, and the publicly available Matlab code of SC.

Given that our approach relies on a bag-of-shapemes representation, it can be easily scaled up to thousands of glyphs using approximate NN-search or compressed features, as demonstrated by several works in the image retrieval domain, e.g., Jégou et al. (2008). We thus believe it can be used in real time applications for scholars.

8.2 Results for Mixed Dataset Glyph Retrieval

The results in this section and Sect. 8.3 were obtained using our HOOSC method with values $N_{sh} = 800$ and $N_{rings} = 5$. Figure 11 presents the top 5 retrieved glyphs for each query. We can notice that despite the complexity of the glyphs, the retrieved images are often relevant. It is clear however, that results are worse for glyphs presenting high variability among their respective instances. This is the case for query *a01* where 4 out of the 5 retrieved glyphs belong to syllable *mi*, as syllables *a* and *mi* have a similar visual structure with vertical alignment divided in three main sections, and *a* presents high visual variability such as small circles whose position might vary from one image to another. A similar trend, although less frequent, occurs with syllables *b'a* and *u*. On the other hand, for most of the queries the results are very satisfactory (e.g., syllables *ka*, *la*, *mi*, *ni*, *na*, *u*) as the HOOSC descriptor seems suitable to represent complex Maya glyphs. Overall, we obtained a *mAP* of 0.61, which is higher than in the previous experiments due to the use of a smaller dataset with more consistent content.

8.3 Glyph Variability over Historical Periods and Geographic Regions

We conducted an analysis of the visual variability of each glyph class in the INAH dataset. It was done by computing the pairwise distance d_{sh} between all the 84 glyphs, and then by computing means and variances according to intra or interclass criteria.
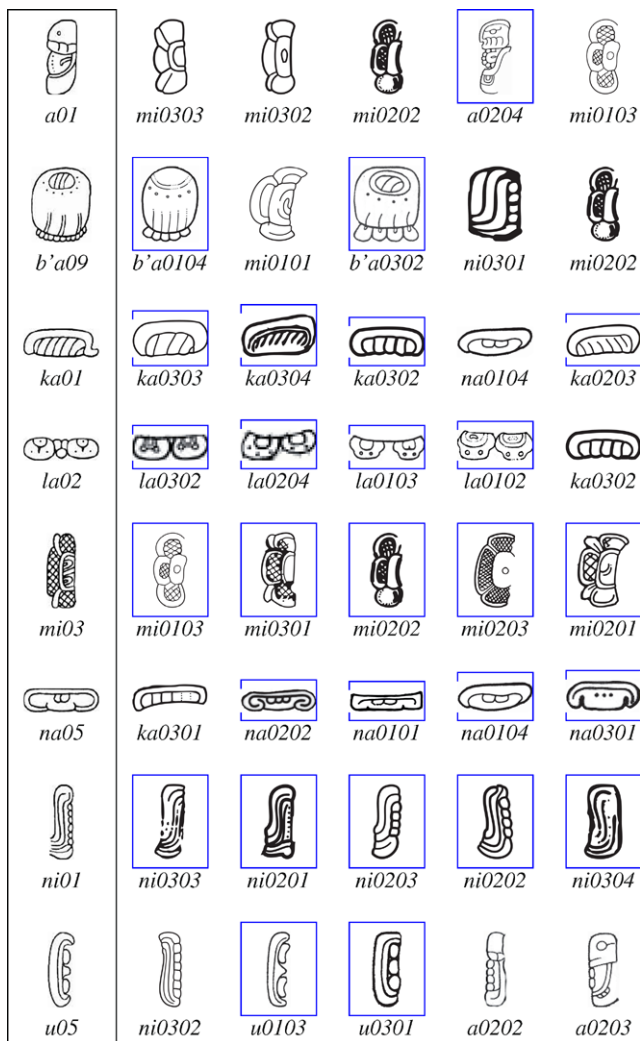
Fig. 11 (Color online) Retrieved glyphs from INAH dataset using 8 images from M&L dataset as queries. The first column shows the query, and the next five columns show the top 5 retrieved glyphs. The relevant glyphs to each query are enclosed by a *rectangle*

8.3.1 Intra-class Analysis

The results show that glyph *mi* has overall the lowest visual variability (average of 0.192), which is consistent with its good retrieval results (see Fig. 11). Conversely, glyph *ka* presents the maximum visual variability with a mean distance of 0.300. This can mainly be explained by the presence of two visual subclasses for this glyph. The rest of the glyphs have very similar intra-class distance values ($\approx$0.23), meaning that they have similar degree of variability within the INAH dataset.

8.3.2 Analysis over Historic Periods

Analyzing in more detail the intra-class variability over time, we observed that the first historic period (Early Classic, 200–500 AD) presents the *highest* variability for 5 of

Table 4 Periods in which syllables exhibit their highest and lowest intraclass average variabilities. The table provides also, the average variability for each period along with its standard deviation

Period	Syllables		Variability	
	Highest	Lowest	Average	Std
Early	*a, ka, mi, ni, u*	*la*	0.277	0.063
Late	*la, na*	*b'a, ni*	0.238	0.036
Terminal	*b'a*	*a, ka, mi, na, u*	0.228	0.028

Table 5 Regions for which syllables exhibit their highest and lowest intraclass average variabilities. The table provides also, the average variability for each region along with its standard deviation

Region	Syllables in region with		Variability	
	Highest	Lowest	Average	Std
Petén	*la, na*	*a, ka*	0.251	0.039
Motagua	*a, b'a, ni, u*	*mi*	0.258	0.057
Usumacinta	*ka*	*ni*	0.349	0.028
Yucatán	*mi*	*b'a, la, na, u*	0.214	0.033

the 8 glyphs, whereas the third period (Terminal Classic, 800–950 AD) has the highest variability for only 1 of them. These results are shown in Table 4. In the second column of Table 4 (labeled *highest*), we show the list of syllables having their highest variability on each period, along with the average variation and corresponding standard deviation for every period. Reversely, we found that only 1 glyph has its lowest variability in the first period, while they are 5 in the last one.

Overall, this analysis might suggest that for the small set of glyphs in our dataset the visual representation went through a stabilization process across time (i.e., on average the variability goes from 0.277 to 0.238 and then to 0.228, also with decreasing standard deviation). This initial result would have to be validated with more data in the future.
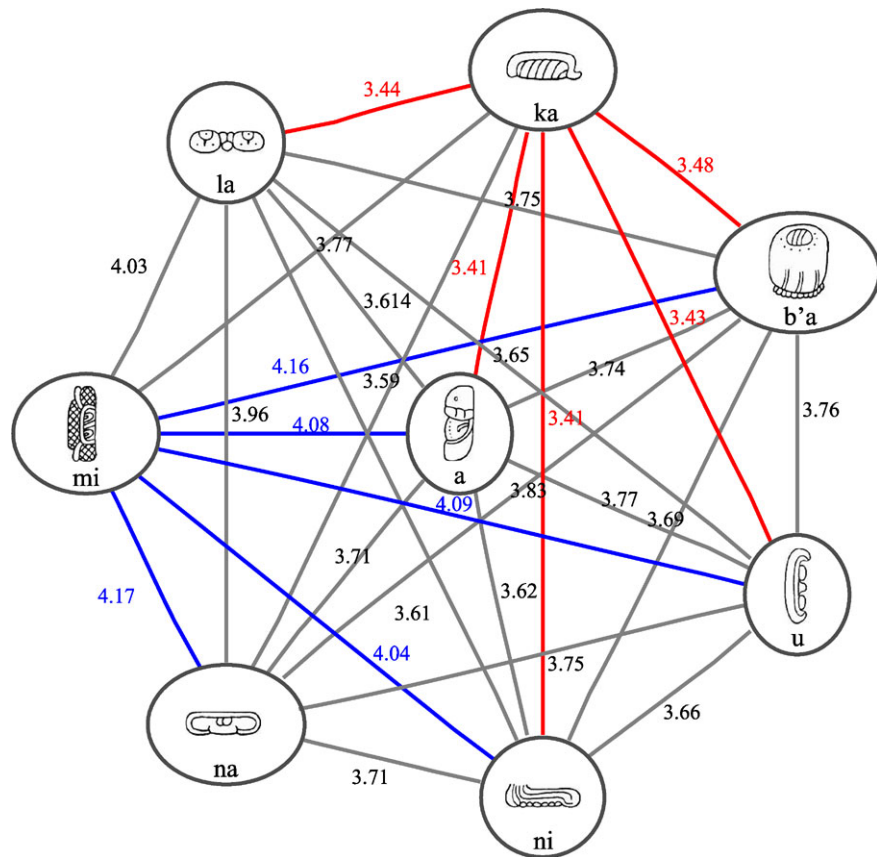
8.3.3 Analysis over Geographic Regions

Table 5 shows the average variability across each of the regions. The Motagua region seems to be the most varying since half of the glyphs show the highest variability for this region, consistent with its average variability which is the second highest with a value of 0.258. On the other hand, the Yucatán region appears to be the less diverse region as half of the glyphs have their lowest variability in it. This is also reflected in its average variability (0.214).

8.3.4 Inter-class Analysis

We define the distance between two classes A and B to be the average of all the distances from each instance of class

Fig. 12 (Color online)
Interclass similarity graph.
Every node contains an example
of the glyphs corresponding to a
class, every edge is weighted by
the similarity between the two
nodes it connects. The edges are
colored to indicate similarity
level; *red*: very different
(smaller values), *blue*: very
similar (higher values), *gray*:
middle term

A to each instance of class B. We use the inverse of this distance as link strength. A graph representation is shown in (Fig. 12).

The set of link strengths varies from 3.40 to 4.17, which implies that the relative difference between the most and the less similar pair of syllables is only 18.3%. Interestingly, the five edges with smallest edge weights (red in Fig. 12) connect syllable *ka*, whereas the five edges with highest edge weights (in blue) all connect syllable *mi*.

8.3.5 Archaeological Analysis

It appears to us significant that despite radically different methodologies, our initial results of analysis over historical periods appears compatible with previous conclusions posited by Lacadena (1995), concerning what he calls the paleographical distance (the amount of variability in signs), which he found to consistently diminish when moving towards the Terminal Classic and beyond (equivalent to our period 3), while augmenting towards the Early Classic (equivalent to our period 1). This said, we hesitate at the moment to derive further conclusions of historical significance based on the analysis over geographic regions that we have performed, since the available literature dealing with this particular topic using more orthodox methodologies than ours is rather scant. Furthermore, the number of

sites hitherto used as source of glyphic signs within the INAH dataset cannot yet be regarded as fully representative for each general region, compelling us to reserve any further interpretations for future works, when a substantially increased dataset, now under compilation, is available.

9 Conclusions and Future Work

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

Two complex datasets of Maya syllabic glyphs, unique in nature and highly valuable for the archeology scholarly community, are used in the experiments conducted in this work, obtaining encouraging results. It is important to note that the experiments were conducted with only few instances for each syllable, making our results to be constrained to a small portion of the whole potential variability that might exist. Nevertheless, our methodology would in principle be

applicable to larger corpora of syllables. Based on the results obtained so far, we hope that the potential of automatic vision technologies could be refined to a degree that would make them useful to retrieve glyphs modified by one of the phenomena mentioned in Sect. 2.

We studied two main variants of a robust shape descriptor: the Shape Context (SC) and the Generalized Shape Context (GSC). Both of them proved to generate reliable descriptions of Maya hieroglyphs, the former being more effective in retrieval experiments. During these analyses we noted a number of their drawbacks for the description and comparison of shapes, and proposed improvements to overcome them. In the GSC case, our improvements lent us to propose the HOOSC descriptor which proved to be more robust. Both GSC and HOOSC are cleverly complemented by the bag of shapemes approach; we studied the quality of the retrieval results for different vocabulary size and degrees of locality. Using the HOOSC descriptor for retrieval tasks, we obtain a relative improvement of 21.1% over the SC approach and 161.7% over the GSC. The bag of shapemes approach also showed to be thousands of times faster than the point-to-point comparison of the original SC.

We also presented what to our knowledge is the first study of glyphic visual similarity over historical periods and geographical regions, finding out that Maya glyphs tend to have less visual variability in subsequent periods, suggesting a gradual convergence in their representation within a given region, an observation that coincides with the previous conclusions of Lacadena (1995). However, as glyphs started to disseminate across different regions, they were enriched with new visual features that increased their variability as the analysis across regions showed. Whether these overall trends will hold true as more syllable sets are gradually incorporated into the analysis remains to be seen. So far we prefer a cautious approach before deriving broader interpretations from such a restricted set of data, as recommended by Lacadena (1995), especially when taking into account the unevenness in the chronological and geographic distribution of a substantial portion of signs within the script.

Our methodology allowed to construct a similarity graph to compare classes of glyphs, i.e., syllables, which might help archaeologists and epigraphers categorize new hieroglyphs discovered in the future. The analysis of the visual variability of the INAH dataset, points to discover common patterns that eventually will allow to build a software that archaeologists could use in query and classification jobs. However, further work is clearly needed towards improving the description of the glyphs, making it robust enough to handle slight differences among instances of the same syllable and at the same time sufficiently discriminant to avoid confusions.

Among the many research threads to study in the future, the first one will be to increase the dataset, populating the temporal-regional tables to have stronger statistical support in future analyses, and enriching the set of syllables itself. We would like to conduct analysis with more uniform samples of glyphs, which may involve producing vectorial tracings of each glyphs, and possibly even restoring damaged portion of signs if there is enough intrinsic and comparative information to allow for it. Second, we plan to explore new methods to describe better the Maya corpus, gradually incorporating not only more syllables, but also logographs (word signs). New work for shape description has to 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. Overall, we expect to be able to refine the bag of shapemes approach for describing and comparing hieroglyphs, through the introduction of more structured information that improves the oversimplifying bag assumption, and that might discover diagnostic features that better discriminate between glyphs belonging to different syllables.

It almost goes without saying that the challenges involved in building such tools, or more advanced ones, appear almost as formidable as the potential benefits it would provide to the Cultural Heritage community.

References

Baeza-Yates, R., & Ribeiro-Neto, B. (1999). *Modern information retrieval*. Reading: Addison Wesley.

Belongie, S., Malik, J., & Puzicha, J. (2000). Shape context: a new descriptor for shape matching and object recognition. In Proc. neural information processing systems, Denver, pp. 831–837.

Belongie, S., Malik, J., & Puzicha, J. (2002). Shape matching and object recognition using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *24*(4), 509–522.

Berlin, H. (1958). El glifo emblema en las inscripciones Mayas. *Journal de la Societé des Américanistes*, *47*, 111–119.

Boujemaa, N., Gouet, V., & Ferecatua, M. (2002). Approximate search vs. precise search by visual content in cultural heritage image databases. In *Proc. ACM MM-MIR*, Paris.

Ferrari, V., Fevrier, L., Jurie, F., & Schmid, C. (2008). Groups of adjacent contours for object detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *30*(1), 36–51.

Frome, A., Singer, Y., Sha, F., & Malik, J. (2007). Learning globally-consistent local distance functions for shape-based image retrieval and classification. In *Proc. int. conference on computer vision (ICCV)*, Rio de Janeiro.

Graham, I. (1975). *Introduction to the corpus, corpus of Maya hieroglyphic inscriptions*. Cambridge: Peabody Museum of Archeology and Ethnology.

Grube, N. K. (1989). Die Entwicklung der Mayaschrift: Grundlagen zur Erforschung des Wandels der Mayaschrift von der protoklassik Bus zur spanischen Eroberung. Ph.D. thesis, Universität Hamburg.

Jégou, H., Douze, M., & Schmid, C. (2008). Hamming embedding and weak geometric consistency for large scale image search. In *European conference on computer vision, LNCS*, Vol. I, pp. 304–317.

Jiang, T., Jurie, F., & Schmid, C. (2009). Learning shape prior models for object matching. In *Proc. computer vision and pattern recognition (CVPR) conference*, Miami.

Justeson, J., Norman, W., Campbell, L., & Kaufman, T. (1985). *The foreign impact on lowland Maya language and script, vol.* 53. Tulane: Tulane University Press.

Lacadena, A. (1995). *Evolución formal de las grafías escriturarias Mayas: implicaciones históricas y culturales*. Ph.D. thesis, Universidad Complutense de Madrid.

Lee, Y., & Grauman, K. (2009). Shape discovery from unlabeled image collections. In *Proc. computer vision and pattern recognition (CVPR) conference*, Miami.

Lewis, P. H., Martinez, K., Abas, F. S., Fauzi, M. F. A., Chan, S. C. Y., Addis, M., Boniface, M. J., Grimwood, P., Stevenson, A., Lahanier, C., & Stevenson, J. (2004). An integrated content and metadata based retrieval system for art. *IEEE Transactions on Image Processing*, *13*(3), 302–313.

Li, J., & Wang, J. Z. (2003). Studying digital imagery of ancient paintings by mixtures of stochastic models. *IEEE Transactions on Image Processing*, *13*, 340–353.

Lu, C., Latecki, L. J., Adluru, N., Yang, X., & Ling, H. (2009). Shape guided contour grouping with particle filters. In *Proc. int. conference on computer vision (ICCV)*, Kyoto.

Macri, M., & Looper, M. (2003). *The new catalog of Maya hieroglyphs, vol. 1: the classic period inscriptions*. Norman: University of Oklahoma Press.

Martin, S., & Grube, N. (2000). *Chronicle of the Maya kings and queens: deciphering the dynasties of the ancient Maya*. London: Thames and Hudson.

Mikolajczyk, K., & Schmid, C. (2004). Scale and affine interest point detectors. *International Journal of Computer Vision*, *60*(1), 63–86.

Mori, G., Belongie, S., & Malik, J. (2005). Efficient shape matching using shape contexts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *27*(11), 1832–1837.

Opelt, A., Pinz, A., & Zisserman, A. (2006). A boundary-fragment model for object detection. In *Europ. conf. computer vision*, Graz.

Proskouriakoff, T. A. (1960). Historical implications of a pattern of dates at Piedras Negras, Guatemala. *American Antiquity*, *25*, 454–475.

Quelhas, P., Monay, F., Odobez, J. M., Gatica-Perez, D., Tuytelaars, T., & Gool, L. V. (2005). Modeling scenes with local descriptors and latent aspects. In *Proc. international conference on computer vision*, Beijin.

Roman-Rangel, E., Pallan, C., Odobez, J. M., & Gatica-Perez, D. (2009). Retrieving ancient Maya glyphs with shape context. In *Proc. IEEE workshop on eheritage and digital art preservation at 12th international conference on computer vision workshops*, Kyoto.

Sharer, R. (1996). *Daily life in Maya civilization*. Greenwood: Westport.

Shotton, J., Blake, A., & Cippola, R. (2005). Contour-based learning for object detection. In *Proc. international conference on computer vision*, Beijin.

Sivic, J., & Zisserman, A. (2003). Video google: a text retrieval approach to object matching in videos. In *Proc. international conference on computer vision*, Nice.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., & Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *22*, 1349–1380.

Stuart, D. S., MacLeod, B., Martin, S., & Polyukhovich, Y. (2005). Glyphs on pots: decoding classic Maya ceramics. In *Sourcebook for the 29th Maya hieroglyphic forum*. Austin: The University of Texas at Austin, Department of Art and Art History.

Taube, K. A. (1992). *The major gods of ancient Yucatan*. Washington: Dumbarton Oaks Research Library.

Thompson, J. E. S (1962). *A catalog of Maya hieroglyphs*. Norman: University of Oklahoma Press.

Willamowski, J., Arregui, D., Csurka, G., Dance, C. R., & Fan, L. (2004). Categorizing nine visual classes using local appearance descriptors. In *Proc. ICPR workshop on learning for adaptable visual systems*.

Zhang, D., & Lu, G. (2004). Review of shape representation and description techniques. *Pattern Recognition*, *37*, 1–19.

Zhu, Q., Wang, L., Wu, Y., & Shi, J. (2008). Contour context selection for object detection: a set-to-set contour matching approach. In *Proc. European conference on computer vision*, Marseille.