Multimedia Analysis and Access of Ancient Maya Epigraphy

Rui Hu, Gulcan Can, Carlos Pallán Gayol, Guido Krempel, Jakub Spotak
Gabrielle Vail, Stephane Marchand-Maillet, Jean-Marc Odobez, Daniel Gatica-Perez

Abstract—We present an integrated framework for multimedia access and analysis of ancient Maya epigraphic resources, which is developed as an interdisciplinary effort involving epigraphers and computer scientists. Our work includes several contributions: definition of consistent conventions to generate high-quality representations of Maya hieroglyphs from the three most valuable ancient codices, currently residing in European museums and institutions; a digital repository system for glyph annotation and management; as well as automatic glyph retrieval and classification methods. We study the combination of statistical Maya language models and shape representation within a hieroglyph retrieval system, the impact of applying language models extracted from different hieroglyphic resources on various data types, and the effect of shape representation choices for glyph classification. A novel Maya hieroglyph dataset is contributed, which can be used for shape analysis benchmarks, and also to study the ancient Maya writing system.

I. INTRODUCTION

The Maya culture is one of the major Precolumbian civilizations that developed in ancient Mesoamerica. It began to flourish during the Pre-classic period (2000 BC to AD 250), reached its peak during the Classic period (AD 250-900), and continued throughout the Post-classic (AD 1000-1519) period. Ancient Mayan languages were recorded by means of a highly sophisticated system of hieroglyphic writing, comprising several thousand hieroglyphic signs, which has left us with an exceptionally rich artistic legacy.

The vast majority of ancient Maya texts were produced during the Classic Period, throughout which hieroglyphic texts were carved or painted on various media types, including stone monuments, architectural elements such as columns, lintels, capstones, and mural paintings, as well as personal items such as precious stones, ceramic vessels, bones, etc., in order to record diverse dedicatory, historical, astronomical, and mythological events. A rare type of Maya textual source is the so-called codex, which are screenfold books made of bark paper, coated with a film of lime plaster upon which textual and icon information was painted using a brush. Only three of these books are known to have survived the Spanish Conquest. Although their exact provenience and dating is not entirely known, they were produced in all likelihood within the greater Peninsula of Yucatan at some point during the Post-classic period. See Fig.1 for an example.

Maya texts are typically composed of glyph blocks arranged in double columns. The most common reading order of glyph blocks is from left to right and from top to bottom within these double columns (see green arrows in Fig.1). One glyph block (also referred to as 'block' in the rest of the paper) could
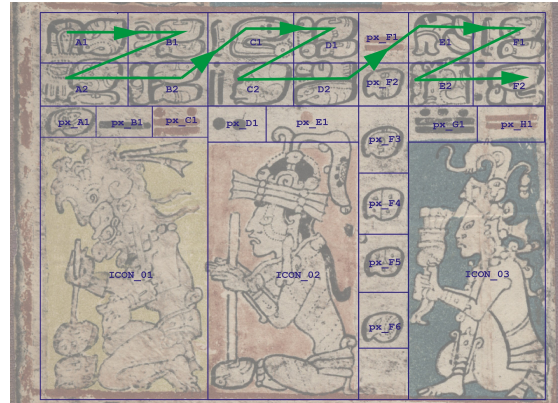

Fig. 1. Detailed template of Dresden Codex page 6b, showing individual constituting elements (glyph blocks, captions, calendric signs, and icons) framed by blue rectangles. Green arrows indicate reading order of the blocks. (by Carlos Pallán based on Förstemann 1880 and 1892 facsimiles.) High resolution images of the whole codex are available at SLUB website [4].

contain a single or multiple glyphs, see Fig. 5 for blocks and their segmented individual glyphs. Due to the large variety of resources at their disposal, Maya scribes could choose among several writing conventions to render a particular Maya term. As a result, graphic conventions within a single block can vary greatly, it could correlate with a phoneme (syllabic sign), an individual word (logograph), or even an entire sentence. Individual glyph recognition is a key and challenging step of Maya script decipherment.

Maya decipherment has undergone nearly 200 years of scholarly research [20]. While Maya archaeologists have discovered and documented a vast number of hieroglyphic monuments and their related context, epigraphers have achieved significant progress in deciphering the hieroglyphic script, and historical linguists have determined the languages recorded. Over 1000 signs have been classified thus far by scholars in several catalogs. It has been estimated that approximately 80% of signs occurring in the known hieroglyphic corpus can be read phonetically [18].

Maya hieroglyphic analysis requires epigraphers to spend a significant amount of time browsing existing catalogs to identify individual glyphs from each block, as a necessary step for generating transcriptions, transliterations, and translations of Maya texts. Technological advances in automatic analysis of digital images and information management are allowing the possibility of analyzing, organizing, and visualizing hieroglyphic data in ways that could facilitate research aimed at advancing hieroglyphic analysis. However, there are several challenges for automatic Maya hieroglyphic data analysis.

First, the available digitized Maya glyph data is limited. Second, the surviving Maya scripts have often lost their visual quality over time. Third, glyphs segmented from blocks are often partially missing due to occlusion. Finally, due to the artistic flexibility, glyphs of a same sign category may vary with time, location and styles; at the same time, glyphs of different categories may share similar visual features. In this paper, we address automatic glyph recognition as image retrieval and classification problems.

Our work is a multidisciplinary effort, where computer scientists work closely with Maya epigraphers to design, develop, and evaluate computational tools that can robustly and effectively support the work of Maya hieroglyphic researchers. Our contributions include: (1) novel Maya glyph datasets with unique historical and artistic value, which could potentially be used as shape benchmarks; (2) shape-based glyph retrieval and classification methods; (3) an in-depth study of the statistical Maya language model for automatic glyph retrieval; and (4) a multimedia repository for data parsing and annotation.

II. RELATED WORK

Computer vision algorithms have shown potential to provide new insights into the realm of digital humanities. Various systems have been proposed to aid the analysis of cultural, historical, and artistic materials, which can significantly facilitate the daily work of scholars in the field.

The automatic analysis of historical manuscripts is the domain most related to our work. A large body of literature in this field examines the digitization and automatic analysis of cultural heritage data, produced from medieval times to the early 20th century [13], [6]. The methodologies include applying machine vision algorithms for page layout analysis, text line extraction, character recognition, and information retrieval. However, the application of computer vision technologies for ancient manuscript analysis, such as Maya writing, is still a novel field. Previous work by our team contributed one of the first studies of visual Maya hieroglyph analysis, and addressed glyph retrieval as a shape matching problem [22], [23].

Shape-based visual information retrieval has been used for searching natural image datasets [10], [16], trademark images [19], technique drawings [14], 3-D objects [11], hand drawn images or clip-arts [26]. Traditional shape-based retrieval systems include curve fitting [9], point-to-point matching [7], and grid based matching [10]. These methods either do not scale well over large datasets, or only offer limited flexibility over shape variations. Recently, local shape descriptors [5], [16], [22] have been proposed and used in a Bag-of-Visual-Words (BoW) framework for shape-based retrieval. Such methods can scale sub-linearly with appropriate search structures.

Automatic Maya glyph retrieval has been addressed in [22], where the Histogram-of-Orientation Shape Context (HOOSC) descriptor was developed. HOOSC combines the underlying formulation of the Shape Context (SC) [7] with the benefits that the Histogram of Oriented Gradients (HOG) [8] provides. It was also applied in [5] for Egyptian hieroglyph analysis.

Previous studies have framed the Maya hieroglyph retrieval problem without considering any specific structure of Maya

Fig. 2. Left: Cropped glyph block. Middle: clean raster image. Right: reconstructed vectorial representation.

writing as a language. In contrast, language modelling has been widely used in machine translation and speech recognition. Glyph context information has recently been applied in [5] for Egyptian hieroglyph analysis with limited performance improvement. To the best of our knowledge, our previous work [17] was the first to incorporate glyph context information in Maya hieroglyph retrieval with significant accuracy improvement. However, the language model applied in [17] was at an incipient stage. It contained incomplete binary co-occurrence information of glyph pairs, extracted from the classic Thompson Maya hieroglyph catalog [25] dating from the 1960s. In this work, we extract a refined glyph co-occurrence model, and test it on datasets of two different resources.

III. DATA SOURCES AND DIGITAL REPOSITORY

Two main data sources are considered in our work: the ancient Maya hieroglyphic books (codices) and monumental inscriptions. In this section, we first introduce the two data sources, and then explain the novel data processing approach that we proposed to produce high-quality representation and annotation of ancient codical data.

A. Data sources

Given the inherent difficulties in the direct study and examination of the original hieroglyphic codex materials, the codical data sources comprise the existing primary documentation of the three extant ancient codices, known as the Dresden, Madrid and Paris codex, respectively. This documentation consists of reproductions, facsimiles, photographs, digital scans, and online resources as described in [15], [2]. The Dresden Codex is held in the state library in Dresden, Germany [4]. The Madrid codex is stored at the Museo de América in Madrid, Spain. The Paris codex resides at the Biblioteque Nationale de France [3]. While the exact provenience and dating of the Maya codices remains uncertain, most contemporary scholars consider that they were made within the northern Yucatan peninsula during the late Post-classic period.

The monumental inscription data sources comprise a variety of carved stone monuments and architectural elements. Besides differences in media and format, it has distinctive historical characteristics as compared to the codical sources. Their dating falls several centuries earlier than the codices, and they stem from several parts of the Maya region, whereas the codices are restricted to the northern Yucatan peninsula. Furthermore, monumental inscriptions number in the thousands as opposed to only three extant codices. Thus the monumental sign repertoire is far better represented than the codical one.

B. Digital Multimedia Repository

In this section, we introduce our data processing approach to generate high-quality digitization and annotation of the three Maya codices, which we refer to as the Digital Multimedia Repository (DMR) of Maya hieroglyphic texts and icons. Ancient Maya scribes usually divided Codex pages into smaller sections by red bands/lines; these sections are referred to by modern scholars as *t'ols*; each *t'ol* being further divided into frames relevant to the specific dates, texts, and imagery depicted. Frames contain glyph blocks (organized in a grid-like pattern), calendric glyphs, captions, and icons. Fig. 1 shows an example *t'ol* (register) from Dresden Codex "segmented" into main elements. The DMR approaches the codices at different levels of detail : 1) entire codex overview; 2) thematic sections; 3) almanacs; 4) *t'ol*s; 5) frames; 6) individual elements (main text glyph-blocks; calendric glyph-blocks, captions, icons); and 7) individual signs or individual iconic elements.

With several software applications, we generate high-quality digitization from the raw image data. Specifically, we first conduct image enhancement, noise-reduction, and up-sizing of images to 400% of their original size; the enhanced text area is then cropped into glyph blocks (Fig. 2, left); we generate clean-raster images from the cropped blocks, by separating the cultural information (brushstrokes) from background noise and preservation accidents (Fig. 2, middle); we then generate high-quality vectorial images in various formats, and by reconstructing the broken lines and missing strokes through thorough comparative analysis, we also generate reconstructed forms (Fig. 2, right). Epigraphers require 15-30 mins to generate a clean raster for a block, depending on complexity and preservation factors of the original data, and one to two hours more to further produce the high-quality reconstructed vectorial glyphs.

For the annotation of these visualizations, we developed an online server-like capture tool (relying on Filemaker Pro), allowing all partners real-time annotation and feedback capabilities. Among the several interconnected tables and templates of this tool, it is worth mentioning the Glyph-Concordance tool that we developed and that radically reduces the amount of time required to annotate glyph-strings within the codices under several different taxonomical systems. It provides automatic field translation and auto-completion functionalities for any individual sign (grapheme) or sign-string, maintaining maximum consistency between the annotations in four scholarly glyph catalogs [25], [12], [20], [27] that we have incorporated and cross-correlated. By enabling advanced multiple queries involving cross-referencing among the above mentioned catalogs, the system allows to overcome the inherent limitations of having a language model based solely on the Thompson catalog. It also increases compatibility with one of our partner's website [2], which contains annotations for all glyph-strings of up to four symbols occurring within the codices and that was used to build a co-occurrence model (see Section V-B). Lastly, several tables and layouts of the DMR are currently being translated into computer science specifications to develop an advanced interface for data parsing and visualization.
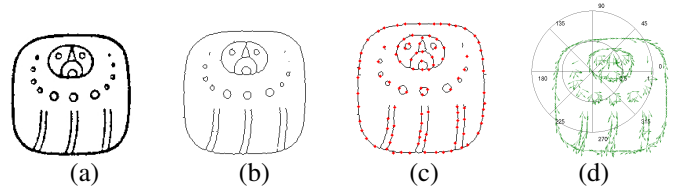

Fig. 3. (a) Original glyph (taken from [25]); (b) thinning; (c) the pivot points (in red) ; (d) HOOSC spatial quantization of a given pivot.

IV. GLYPH REPRESENTATION

Traditional shape descriptors [7], [21] may not be sufficient to capture the visual richness of Maya glyphs. We rely on the HOOSC descriptor, which performed well for Maya hieroglyph analysis [22], [17]. We now describe the pre-processing and feature extraction steps that we follow.

Maya glyphs are often composed of strokes with different degrees of thickness. Thus, contour extractors sometimes generate "double" contours from the internal and external stroke lines, which can result in noisy descriptors. Therefore, we apply a thinning algorithm [23] to pre-process the binary shape (Fig.3 (b)). To reduce computation cost, we compute the local descriptor only at a set of points (called pivots) obtained through uniform sampling of the points along the thinned contour (see Fig. 3(c)).

HOOSC was proposed in [22] as a robust shape representation for Maya hieroglyphs. It combines the strength of Histogram of Orientation Gradient (HOG) [8] with circular split binning from the Shape Context descriptor (SC) [7].

For each pivot point, the HOOSC is computed on a local circular space centered at the pivot's location, partitioned into 8 orientations and 2 rings, as illustrated in Fig.3(d). The radius of the outer ring is the average pairwise distance between each pair of points along the contour; the inner ring covers half of this distance. An 8-bin histogram-of-orientation gradient is calculated within each spatial region. This results in a richer representation than [7], [21], where a simple counting of points or a sum of the unit gradient vectors of all points falling within a region is computed. The HOOSC descriptor for a given pivot point is then the concatenation of histograms of the sixteen regions, which forms a 128-dimensional feature vector. As suggested in [22], we apply a per ring normalization.

V. AUTOMATIC MAYA HIEROGLYPH RETRIEVAL

In this section, we present an automatic Maya glyph retrieval system, combining shape and glyph context information. Experimental results show the performance of our method.

A. Shape-based glyph retrieval

We adapt the bag-of-words (BoW) model for glyph retrieval, which has been established as a framework for scalable image retrieval [24]. Specifically, we apply k-means clustering on the set of HOOSC descriptors extracted from all glyphs in the database. The resulting k clusters are referred to as 'visual words' and define the vocabulary of the system. A histogram representing the count of each visual word is then computed as a global descriptor for each glyph.

T0501	T0502	T0668	T0757	T0102	T0103
/b'a/	/ma/	/cha/	/b'a/	/ki/	/ta/

Fig. 4. Thompson numbers, visual examples, and syllabic values of glyph pairs. Each pair contains two different signs of similar visual features. All images are taken from [25].

Given a query G and a database glyph D, represented by histograms H^G and H^D generated from the BoW system, we compute the Cityblock distance to measure the dissimilarity between G and D:

$$d(G, D) = \sum_{1 \leq i \leq k} |H^G(i) - H^D(i)|, \qquad (1)$$

where each histogram is normalized so that $\sum_{1 \leq i \leq k} H(i) = 1$.

B. Glyph co-occurrence model

Using shape alone to distinguish different glyphs is often problematic for many reasons. First, different signs often share similar visual features, see Fig. 4. Furthermore, glyphs of the same sign category vary with time, location, and individual styles. For example, observe the last two blocks in the top row of Fig. 5. Although visually different, the two blocks are actually composed of two same glyphs (T0668 and T0102, see Fig. 4 for images of the two glyphs separately), pronounced as *cha-ki* and representing the name of the rain god. Fig. 6 shows six glyph pairs, each represents two glyphs of a same category, but has different visual features. Finally, the surviving Maya scripts often lose their visual quality over time. We propose to use glyph co-occurrence to complement visual information and help improve retrieval accuracy.

Maya glyph blocks were frequently composed of combinations of individual signs. Glyph co-occurrence within single blocks could therefore encode valuable information. While the reading order within a block usually follows the basic rule of left-to-right and top-to-bottom, several exceptions occur, particularly in the Madrid and Paris codices. Our methodology converts each glyph block into a linear string of individual signs, according to the reading order determined by our team's epigraphers, as shown in Fig. 5. We consider the first-order co-occurrence of neighboring glyphs as the glyph context information to build a statistical Maya language model.

Two glyph co-occurrence models (i.e. the Thompson co-occurrence model and the Vail co-occurrence model) are extracted from different sources. Their general form is:

$$C(S_i, S_j) = \begin{cases} f_n & \text{if sign } S_i \text{ appears before sign } S_j, \\ \alpha & \text{otherwise}, \end{cases}$$
$$(2)$$

where f_n represents the normalized frequency that sign S_i appears before sign S_j, and $\alpha \in [0, 1]$ is a smoothing factor that accounts for missing co-occurring glyph pairs in the two models, which we explore in the experiments. Note that $C(S_i, S_j) \neq C(S_j, S_i)$.

From a computational point of view, the difference between the Thompson and the Vail models are just variations of the co-occurrence table. However, they are really different from an archeological point of view, as one features the Classic period monumental glyphs while the other features the Post-classic codices. In section V-D, we test both of them on two different query sets to investigate the impact that this difference of writing conventions between the codex and the monument data has on the retrieval results.

1) Thompson co-occurrence model: It is extracted from the classic Thompson hieroglyphic catalog [25], which covers 892 signs extracted from codices and monuments. Despite its outmoded taxonomy, it remains one of the most comprehensive and widely used lists of Maya glyphs. Thompson categorizes signs into affixes, main signs, and portraits (this categorization is no longer used in recent research on Maya writing). Affixes often co-occur with main signs, portraits, or affixes to form blocks. In his book, Thompson provides two glyph co-occurrence tables for affix signs, distinguishing whether they are used as prefix or postfix. However, no frequency information is given in these tables (we thus set $f_n = 1$ for valid co-occurrence), and co-occurrence between main signs and portraits is not listed. There are 4574 glyph pairs included in this model, which correspond to a sparsity of 99.4%.

2) Vail co-occurrence model: This co-occurrence model is extracted from our partner's on-line Maya hieroglyphic database [2] containing state-of-the art description and interpretation of the three surviving Maya codices.

Using the database, we extract the co-occurrence information of neighboring glyphs within single blocks. The resulting model contains a complete set of co-occurring pairs that are known today from the three surviving codices, as well as their occurrence frequency. In total, the database contains 336 different glyph signs. There are 1818 glyph pairs, which co-occur with frequencies f from 1 to 188, corresponding to a sparsity of 99.8% (since we consider 892 signs from the Thompson catalog). More than half of these pairs only co-occur once. Around 93% of the pairs appear less than 10 times. We normalize f with the following function:

$$f_n = 1 + log10(f). \qquad (3)$$

3) Thompson & Vail co-occurrence model: Additionally, we build a third model by considering all co-occurrence pairs from the two former models. We disregard the frequency information of the Vail model and generate a binary model, which we refer to as 'Thompson & Vail' co-occurrence. It contains 5600 co-occurred glyph pairs, which correspond to a sparsity of 99.3%. We expect this model to perform better across different Maya datasets (e.g. codices and monument).

C. Incorporating Context Information

We now explain how to incorporate this information in the shape-based retrieval system.

1) Sequence model: Denote by $G_{1:n} = [G_1, \ldots, G_i, \ldots, G_n]$ the observed glyph string, and by $S_{1:n}$ the sequence of recognized states, where S_i indicates the sign category annotated for glyph G_i. Considering the glyph string $G_{1:n}$ as a first-order Markov chain, the probability of

Fig. 5. 1st row: six glyph block examples from the codex dataset; 2nd row: four glyph strings segmented from the raster (black) and vectorial (blue) representation of the first two blocks shown individually in the 1st row; 3rd row: six glyph blocks from the monument dataset; 4th row: three glyph strings cropped from the first three blocks shown in the 3rd row separately.

labelling it to a sequence of states $S_{1:n}$ is:

$$P(S_{1:n}|G_{1:n}) \propto P(G_1|S_1) \prod_{2 \leq i \leq n} (P(G_i|S_i)P(S_i|S_{i-1})),$$

(4)

where $P(S_i|S_{i-1})$ denotes the transition probability. Here we directly use $C(S_{i-1}, S_i)$ to approximate this probability. $P(G_i|S_i)$ refers to the likelihood of glyph G_i being labelled as sign S_i. To encode this term we use the visual similarity between G_i and the glyph example of S_i in the database, and define $P(G_i|S_i) \propto e^{-d(G_i, S_i)/\lambda}$, where $d(G_i, S_i)$ is computed using Eq.(1), and λ is a scale factor empirically set to the average distance of the top 50 ranked results for all queries.

2) Retrieval system: When only shape information is considered, the score of a query glyph G_i being labelled by sign D, is computed by their shape likelihood:

$$Score^{sh}(S_i = D) \propto P(G_i|S_i = D).$$

(5)

In our model, we propose to rank the glyphs according to

$$Score^{sh+context}(S_i = D) = \max_{S_{1:i-1;i+1:n}} P(S_{1:n}|G_{1:n}),$$

(6)

which means the following: given $S_i = D$, find the sequence of labels $S_{1:n}$ that provides the maximum probability to label $G_{1:n}$, under the model in Eq.(4), and use this probability as score to rank the database glyphs. This can be efficiently computed using the Viterbi algorithm. Thus, the score of the glyph G_i being recognized as $S_i = D$ now takes into account all observed glyphs in the string, with the effect that a glyph D that normally co-occurs with glyphs that are visually likely at neighboring positions will receive a higher weight.

D. Experimental results

Below we present the datasets, experimental setting and retrieval results.

1) Glyph Datasets: Two datasets, namely the Codex and Monument datasets, were used as query sets to retrieve from a common database.

Codex dataset. This dataset was produced by epigraphers in our team (see Section III) and is available to download from [1]. It contains glyph blocks from the three surviving Maya codices, along with their segmented individual glyphs and corresponding annotations. See the first two rows of Fig. 5
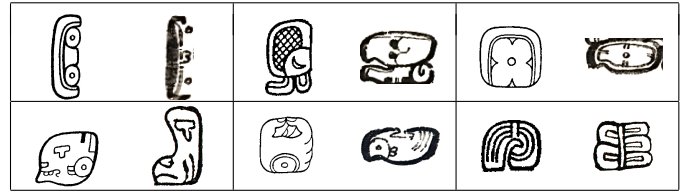

Fig. 6. Six pairs of glyph sign: T0001, T0158, T0544, T0668, T0671 and T0145 (from left to right and top to bottom). The left one of each pair is from the Thompson catalog, the right one is an example from the codex dataset.

for example. To study the impact of visual data quality on the retrieval performance, we considered two sub-datasets. The first one is a 'codex small' dataset, composed of 151 glyphs segmented from 60 blocks, for which we have both the vectorial and raster representations. Remember that producing high-quality vectorial representations (including reconstruction of missing parts) is time consuming, compared to raster images, which reflect the actual visual content but are affected by degradations. The second subset is intended to assess the validity of the reported performance by using a larger corpus (termed 'codex large') comprising only the raster representation of 587 glyphs from 224 blocks.

Monument dataset. It contains 127 glyphs of 40 blocks extracted from stone monument data, and is derived from a quite different data source than the codex data in terms of Maya historical period, media type, and glyph generation process. Samples are shown in the last two rows of Fig. 5. The data consisted of line drawings of glyphs manually traced on top of multiple layers of enhanced photographs taken at sites at night under raking-light illumination to bring out different levels of detail. Given each block, we manually drew a tight bounding box around individual glyphs to extract query glyphs. The queries may be affected by adverse effects, like background noise, additional strokes from neighboring glyphs, or partially missing strokes due to glyphs overlapping within blocks.

Retrieval database. We scanned and segmented all the glyphs from the Thompson catalog [25] to form the retrieval database. In total, it contains 1487 glyph images belonging to 892 different categories. Thus, a category is usually represented by a single glyph image, and sometimes by multiple ones, each representing a different visual instance of the glyph category. Fig. 4 shows glyph images of six different categories.

2) Experimental setting: For each glyph query, we extract the rank of the true glyph in the retrieval results, and use the average of these ranks over all queries as performance measure (the lower the average ranking, the higher the accuracy).

We studied the impact of several factors on the performance, including the vocabulary size of the BoW representation, the smoothing factor α used to build the co-occurrence models (Eq. 2), and the co-occurrence models.

3) Results and discussion: Results are presented in Fig. 7 and Fig. 8, in which 'T', 'V' and 'T&V' refer to the 'Thompson', 'Vail' and 'Thompson & Vail' co-occurrence models respectively.

Shape-based glyph retrieval. Looking first at the impact of data origin and quality on glyph retrieval performance when only shape information is considered, the following observations can be made. First, as expected, higher quality vectorial repre-
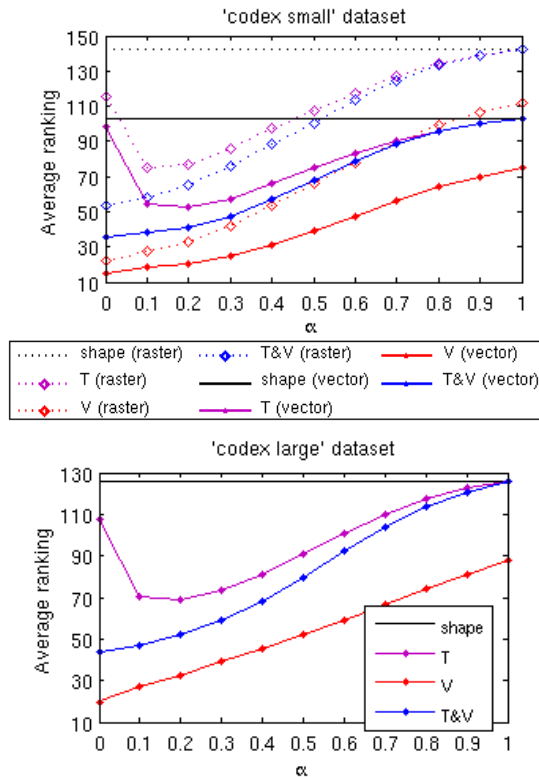
Fig. 7. Average ranking on the codex datasets with varying smoothing factor α (in Eq.2), and fixed vocabulary size 5000, (top) 'codex small' and (bottom) 'codex large' datasets.

Fig. 8. Average ranking on the monument dataset, (top) with varying smoothing factor α (in Eq.2), and fixed vocabulary size 5000; (bottom) with varying vocabulary size and fixed $\alpha = 0.2$ for 'T' and 'T&V' models; $\alpha = 1$ for 'V' model.

sentations result in higher retrieval accuracy (103 for vectorial vs 142 for raster images, see plain and dotted horizontal lines in Fig. 7 (top)). Second, by comparing the shape-based retrieval results of the monument data (see horizontal line in Fig.8 (top)) and the codex data (see plain and dotted horizontal lines in Fig.7), we can see that, despite the presence of distracting elements (background noise, line strokes from neighboring glyphs, etc.) the shape retrieval accuracy on the Monument data (86) is higher than on the Codex data. This reflects the higher visual similarity between the glyphs in the monument dataset and those from the retrieval database. As glyphs in the retrieval database are extracted from the Thompson catalog, which largely relies on monument data, as compared to signs from the Codex data, which often exhibit more variability as illustrated in Fig. 6. For the same reason, using finer HOOSC descriptor quantization (i.e., using a larger BoW vocabulary) consistently improved the results on the monument data (Fig. 8(bottom)), whereas it had no impact on the Codex data (curves not shown).

Incorporating context information. As can be seen from Fig. 7 and Fig. 8, the average retrieval rankings obtained using different co-occurrence models and smoothing factors are usually significantly lower than when using only shape. For instance, on the small codex dataset, the Vail model (with $\alpha = 0$) can reduce by as much as 130 and 90 the average ranking for the raster and vectorial representation, respectively, whereas on the monument data (Fig. 8 (top)), the gain is smaller (around 20 for the Thompson model with $\alpha = 0.2$).

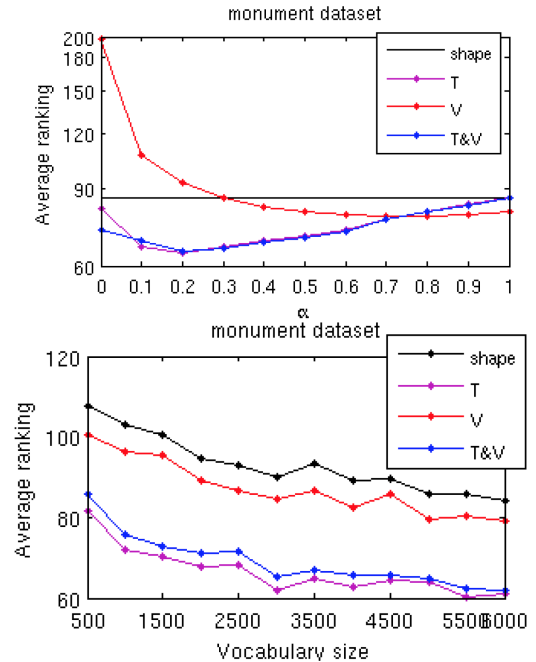Differences across models as well as the optimal smoothing factors mainly reflect the agreement between the source of the block queries and the data used to build the co-occurrence models. Thus, on one hand, the Vail model achieves the best accuracy on codex datasets (Fig. 7), but under-performs on the monument data (Fig. 8), even degrading the shape-only results for low smoothing factors. Since this model is purely built from the codices, this may imply that the Maya writing on codices and monuments follows different glyph co-occurrence conventions. On the other hand, the Thompson model, built from a mixed source of monument and codex data, offers a much smaller gain when applied to the Codex data, but still performs well on monument data. Altogether, these two models are advantageously combined in the more versatile 'T & V' model.

No smoothing factor ($\alpha = 0$) is needed when applying the Vail model on Codex data, since it covers all known co-occurrence instances of the codices; whereas the Thompson model that relies only on incomplete data sources misses some co-occurrence pairs and thus requires a smoothing factor (typically $\alpha = 0.2$). In general, all the above remarks remain valid when considering the large Codex data (Fig. 7 (bottom)).

As a final remark, one can notice on Fig. 7 (top) that the retrieval performance differences between the vectorial and raster representation becomes less important when using a co-occurrence model. In this context, the raster representation can be used as an compromise between data production efficiency and retrieval accuracy.

VI. SHAPE-BASED GLYPH CLASSIFICATION

There are use cases in which inferring the correct category of a glyph is important. In this section we thus study the single glyph classification task, first presenting the classification methods, and then discussing the obtained results.
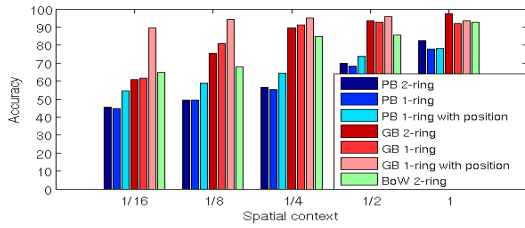
Fig. 9. Classification accuracy of the BoW method (green bar), and the proposed method at the pivot level (pivot-based 'PB' results, blue bars) and the glyph level (glyph-based results 'GB', red bars), with various settings to compute the HOOSC descriptor.

A. Glyph Classification methods

The objective is to build a classifier that categorizes a test shape into one of the N_G categories. As baseline, we use the method of [23], where glyphs are represented using the global BoW representation. A test glyph gets the class label of its nearest neighbor (using the BoW cityblock distance in Eq. 1) in the training set.

As an alternative, we propose a method that categorize an unknown glyph by first identifying the category of its individual local pivot points. Specifically, for a given glyph, we first compute the HOOSC descriptor at each pivot point and classify it using a K nearest neighbor method. Then, in a second step we classify the glyph as the category that receives the largest number of votes from the individual pivots.

B. Experimental Results

1) Dataset: We used a subset of glyphs from the monumental inscriptions that were used in [23]. We only consider glyph categories which contain more than 30 glyphs. The resulting dataset is composed of 10 glyph categories with 25 training images per class and 125 test images in total. The groundtruth of the glyph category is provided by our team scholars.

2) Experimental Setting: We used 300 equidistant pivots where we compute the HOOSC descriptor. Note that here, we extracted the orientation from the raw images preprocessed by a continuous Gaussian orientation filter, as this produced more stable results than applying the thinning pre-processing.

We considered three settings to compute the HOOSC descriptor: (1) 2 rings and 8 radial bins, see Fig. 3(d); (2) 1 ring and 8 radial bins, see Fig. 10; (3) case (2) with added position information, i.e., the HOOSC descriptor is augmented with the relative position (defined within $[0, 1] \times [0, 1]$) of the pivot point within the glyph bounding box.

Furthermore, for each of the three settings, we considered five spatial context (defined by the radius of the outer ring in HOOSC computation): 1/16, 1/8, 1/4, 1/2, and 1, all defined as a proportion to the mean of the pairwise distance between pivot points (see gray circles in Fig. 10), as we are interested in studying the impact of the spatial scope used to compute the HOOSC descriptor on the classification performance. Indeed, while large scopes used in previous works (and the retrieval Section) led to good results when dealing with clean glyph inputs, there are situations where smaller scopes would be useful, e.g. when dealing with damaged glyph shapes (the damage will affect most of the descriptors when using a large scope), or if we wanted to identify which local part of
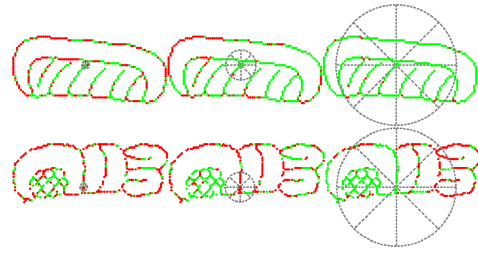

Fig. 10. Classified pivots using HOOSC 1-ring with position and spatial context 1/16 (left), 1/4 (middle) and 1 (right) respectively. Green (resp. red) points indicate correctly (resp. incorrectly) classified pivots.

the glyph is a 'diagnostic' feature, i.e., a discriminant visual element that scholars rely on to distinguish a glyph.

3) Results and Discussion: Fig. 9 shows the classification results obtained using the BoW method and the proposed method ('glyph-based' results, denoted GB) for different spatial context sizes and partition settings. In order to better understand the proposed method, we also show the 'pivot-based' (denoted PB) classification accuracy, i.e., the percentage of pivot points whose descriptor is correctly classified as the category of its associated glyph.

First, from the results of the 'pivot-based' method (blue bars), we can notice that the performance degrades almost linearly as the spatial context decreases, but remains well above chance level (10%) even for small spatial extent (1/16). Interestingly, as this context gets smaller, the incorporation of the spatial position (PB 1-ring with position) allows to boost performance by 10% as compared to the case without position (PB 1-ring). Furthermore, while two rings are useful as the spatial context is large, it is not superior than one ring in terms of PB performance and actually degrades the GB performance when smaller spatial context is considered (e.g. 1/4 to 1/16).

Secondly, the performance w.r.t. spatial context at the glyph level (red bars) does not decrease as dramatically as at the pivot level, indicating that misclassified points, even if they dominate, usually get distributed over all other classes rather than a single one. Hence the pivots predicted with true labels may win in the voting phase. For GB 1-ring with position, the classification remains as high as 94% with a spatial context of 1/8. Note that this is not the case of the BoW approach (green bars), whose performance degrades as the spatial context decreases, performing worse than the proposed approach with spatial radius larger than 1/4, and can not keep up with the 1-ring with position results at smaller spatial scopes.

Fig.10 illustrates the pivot classification results for two glyphs over three spatial context levels. We can see that the number of pivots classified correctly increases with the spatial context. It also shows that while some local structures are recognized at most scales (diagonal lines for the top glyph, hatches for the bottom one), there are structures that still remain confusing among glyph classes, even at the larger contexts (e.g. the pivots near the 'ears' in the bottom glyph).

We can conclude that a two-step approach where class-information is used to categorize the descriptor (rather than simple quantization in BoW) brings more robustness as the spatial context decreases (and may bring even more robustness when dealing with partially damaged glyphs), and that incorporating the relative position of pivots is important, as

the same local shape structure might be observed at different positions for different glyph categories.

VII. CONCLUSION AND FUTURE WORK

This paper presented an approach to capture and produce high-quality multimedia data from valuable historical codex data sources, upon which we propose algorithms for automatic Maya hieroglyph analysis.

We defined consistent conventions to generate high-quality representations of the ancient Maya hieroglyphs, as well as a data model which not only provides a convenient platform for epigraphers to annotate and analyze data, but also serves as a bridge between epigraphers and computer scientists for data parsing and analysis. A novel codex dataset is contributed under the proposed system.

We then addressed two automatic glyph analysis tasks with value to support epigraphers' daily work, namely glyph retrieval and glyph classification. Regarding retrieval, two Maya language models were extracted from different data sources and incorporated into a shape-based automatic glyph retrieval framework. Our study showed that glyph co-occurrence encode valuable information of the Maya writing system, which can be used to complement the visual automatic analysis. The retrieval results also showed that the Maya writing on codices and monuments follows different glyph co-occurrence conventions. Finally, we studied the effect of shape representation choices in the classification task.

Our future work includes automatic Maya text area detection, as well as detection and segmentation of blocks and glyphs, which will facilitate the daily work of scholars when more data becomes available. In another direction, we are working on designing a visualization interface to allow manipulation of Maya data in a systematic and flexible way. Data will be displayed as clusters in various feature spaces (from low-level visual features to high-level semantic spaces); analyzed with different levels of context information (within block co-occurrence, surrounding text, icons); and visualized in various resolutions and positions by zooming and panning. We expect the traditional Maya hieroglyph decipherment to benefit from such functionalities.

ACKNOWLEDGMENT

REFERENCES

[1] codex dataset. www.idiap.ch/dataset/maya-codex.
[2] The Maya Hieroglyphic Codices. http://www.mayacodices.org/.
[3] Paris Codex. http://gallica.bnf.fr/ark:/12148/btv1b8446947j/f1.zoom.r=Codex%20Peresianus.langDE.
[4] Saxon State and University Library Dresden (SLUB) library. http://digital.slub-dresden.de/werkansicht/dlf/2967/1/.
[5] M. Franken and J. v. Gemert. Automatic egyptian hieroglyph recognition by retrieving images as texts. In *ACM MM*, pages 765–768, 2013.
[6] N. Arvanitopoulos and S. Süsstrunk. Seam carving for text line extraction on color and grayscale historical manuscripts. In *InterNational conference on frontiers in handwriting recognition*, 2014.
[7] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *PAMI*, pages 509–522, 2002.
[8] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *CVPR*, pages 886–893, 2005.
[9] A. Del Bimbo and P. Pala. Visual image retrieval by elastic matching of user sketches. *PAMI*, pages 121–132, 1997.
[10] M. Eitz, K. Hildebrand, T. Boubekeur, and M. Alexa. A descriptor for large scale image retrieval based on sketched feature lines. In *SBM*, pages 29–36, 2009.
[11] M. Eitz, R. Richter, T. Boubekeur, K. Hildebrand, and M. Alexa. Sketch-based shape retrieval. *ACM Transactions on Graphics*, 31(4):31:1–31:10, 2012.
[12] E.B. Evrenov, Y. Kosarev, and B.A. Ustinov. *The Application of Electronic Computers in Research of the Ancient Maya Writing*. USSR, Novosibirsk, 1961.
[13] A. Fischer, H. Bunke, N. Naji, J. Savoy, M. Baechler, and R. Ingold. The HisDoc project. automatic analysis, recognition, and retrieval of handwritten historical documents for digital libraries. In *InterNational and InterDisciplinary Aspects of Scholarly Editing*, 2012.
[14] M. J. Fonseca and J. A. Jorge. Towards content-based retrieval of technical drawings through high-dimensional indexing. In *1st Ibero-American Symposium in Computer Graphics*, pages 263–270, 2002.
[15] D. Gatica-Perez, C. Pallan-Gayol, S. Marchand-Maillet, J.M. Odobez, E. Roman-Rangel, G. Krempel, and N. Grube. The MAAYA project: Multimedia analysis and access for documentation and decipherment of Maya epigraphy. In *Digital Humanities Conference*, 2014.
[16] R. Hu and J. P. Collomosse. A performance evaluation of gradient field hog descriptor for sketch based image retrieval. *CVIU*, pages 790–806, 2013.
[17] R. Hu, C. Pallan-Gayol, G. Krempel, J.M. Odobez, and D. Gatica-Perez. Automatic maya hieroglyph retrieval using shape and context information. In *ACM MM*, pages 1037–1040, 2014.
[18] H. Kettunen and C. Helmke. *Introduction to Maya Hieroglyphs - XVII European Maya Conference*. 2012.
[19] W. H. Leung and T. Chen. Trademark retrieval using contour-skeleton stroke classification. In *ICME*, pages 517–520, 2002.
[20] M. J. Macri and G. Vail. *The New Catalog of Maya Hieroglyphs, vol II, the codical texts*. University of Oklahoma Press, 1901.
[21] G. Mori, S. J. Belongie, and J. Malik. Efficient shape matching using shape contexts. *PAMI*, 27(11):1832–1837, 2005.
[22] E. Roman-Rangel, C. Pallan-Gayol, J.M. Odobez, and D. Gatica-Perez. Analyzing ancient maya glyph collections with contextual shape descriptors. *IJCV*, pages 101–117, 2011.
[23] E. Roman-Rangel, C. Pallan-Gayol, J.M. Odobez, and D. Gatica-Perez. Searching the past: an improved shape descriptor to retrieve Maya hieroglyphs. In *ACM MM*, pages 163–172, 2011.
[24] J. Sivic and A. Zisserman. Video Google: A text retrieval approach to object matching in videos. In *ICCV*, pages 1470–1477, 2003.
[25] J. Eric S. Thompson. *A catalog of Maya Hieroglyphs*. University of Oklahoma Press, 1962.
[26] C. Wang, J. Zhang, B. Yang, and L. Zhang. Sketch2Cartoon: composing cartoon images by sketching. In *ACM MM*, pages 789–790, 2011.
[27] G. Zimmermann. *Die Hieroglyphen der Maya Handschriften. Abhandlungen aus dem Gebiet der Auslandskunde*. Band 62- Reihe B, Universität Hamburg. Cram, De Gruyter & Co., 1956.

Rui Hu is a Postdoctoral Researcher at Idiap Research Institute in Switzerland. Email: rhu@idiap.ch

Gulcan Can is a PhD. Candidate at Idiap Research Institute in Switzerland. Email: gcan@idiap.ch

Carlos Pallán Gayol is a Research Associate in the Department of Anthropology of the Americas, at the University of Bonn. Email: s5capall@uni-bonn.de

Guido Krempel is a Research Associate in the Department of Anthropology of the Americas, at the University of Bonn. Email: guidokrempel@gmail.com

Jakub Spotak is a PhD. Candidate in the Faculty of Arts, at the Comenius University in Bratislava, Slovakia. Email: spotak.jakub@gmail.com

Gabrielle Vail is a Research Scholar in the Division of Social Sciences, New College of Florida in Sarasota, Florida. Email: gvail@ncf.edu

Stephane Marchand-Maillet is Associate Professor at the University of Geneva, Switzerland. Email: stephane.marchand-maillet@unige.ch

Jean-Marc Odobez is the Head of the Perception and Activity Understanding group at Idiap, and Maitre d'Enseignement et de Recherche at EPFL, Switzerland. Email: odobez@idiap.ch

Daniel Gatica-Perez is the Head of the Social Computing Group at Idiap and Professeur Titulaire at EPFL, Switzerland. Email: gatica@idiap.ch