Pattern Recognition 48 (2015) 1161–1173

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Shape-based detection of Maya hieroglyphs using weighted bag representations ☆

Edgar Roman-Rangel *, Stephane Marchand-Maillet

Viper group - CVMLab, Department of Computer Science, University of Geneva, Switzerland

ARTICLE INFO

Article history:
Received 30 December 2013
Received in revised form
11 May 2014
Accepted 11 June 2014
Available online 19 June 2014

Keywords:
Shape-based object detection
Binary images
Bag-of-words
Maya hieroglyphs

ABSTRACT

This work addresses the problem of detecting individual visual patterns in binary images, and more precisely, individual syllabic signs in large inscriptions of Maya hieroglyphs with high levels of visual complexity. The data we use corresponds to a corpus that is of great interest for archaeologists, and it poses a difficult challenge in terms of visual complexity. We introduce a new weighting function, which helps constructing more robust bag-of-visual-words representations for detection purposes. This weighting function depends on the ratio of intersection of the local descriptors, and their respective distances to the center of the bounding box that is under evaluation. As shown by our results, the use of the proposed weighted bag representation improves the detection rate with respect to a traditional bag construction. Also, we validate the use of an ad hoc methodology to approach the detection scenario through a retrieval setup. Our results show that this approach achieves better detection performance than the traditional sliding-windows approach when only a few data is available for training, as is the case of the Maya hieroglyphs. To the best of our knowledge, our work is among the first contributions that addresses the problem of shape detection using binary images, since the previous attempts to detect shapes rely on the use of intensity images.

1. Introduction

The automatic detection of specific objects in images is one of the principal challenges in the field of computer vision [20]. Although this task corresponds to one of the main capabilities of human perception, its replication into automatic methods is arguably limited, as its complex process has not yet been totally understood [15].

Despite the many important milestones that have already been reached in the task of automatic detection [28,26,20], all these works have investigated the detection problem using intensity images, thus designing methods to exploit the local intensity details that these kinds of images contain. Different from those works, our research focuses on the detection of very specific elements, where the information available for local description is rather limited, and of binary nature. Therefore, different strategies need to be proposed. Moreover, most of the previous works focus on detecting visual objects of daily human use [20], which makes it relatively easy to collect enough data and develop statistical models that detect objects relying upon well trained classification techniques. In contrast, in our work, we address the problem of detecting instances of the ancient Maya writing system, which correspond to a specific type of data with rather limited availability [18], thus making it difficult to train those statistical models and to conduct detection experiments under the traditional sliding-windows approach [24,6,7].

The motivation behind this research stands upon the need of developing new tools to support the current manual process of identification of Maya hieroglyphs [17], in which even expert archaeologists often need to consult printed catalogs to identify the different individual signs [21,11,12].

Commonly, Maya inscriptions are very sparsely populated with text hieroglyphs. This is to say, the size of an individual hieroglyph is considerably small with respect to the size of the complete inscription that contains it, thus making laborious the manual detection process. Fig. 1 shows one example of a Maya inscription extracted from a stone monument. The complexity of this manual detection increases if we take into consideration that it is a common feature of the Maya writing system to arrange hieroglyphs at arbitrary orientations, sizes, and positions within the inscriptions. Therefore, the implementation of techniques for automatic detection of these complex hieroglyphs requires special attention.

☆This work is partially based on "Evaluating Shape Descriptors for Detection of Maya Hieroglyphs" by Edgar Roman-Rangel, Jean-Marc Odobez, and Daniel Gatica-Perez, which appeared in the Proceedings of the 5th Mexican Conference on Pattern Recognition (MCPR'13), Queretaro, Mexico, June 2013 [17].

* Corresponding author.

http://dx.doi.org/10.1016/j.patcog.2014.06.009
0031-3203/

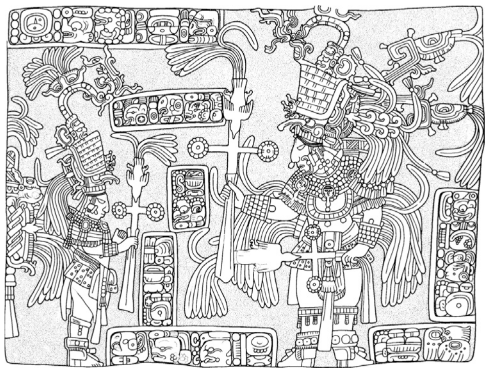

Fig. 1. Maya inscription from a monument in the archaeological site of Yaxchilan, Mexico. Note the level of sparsity of the text hieroglyphs with respect to the size of the whole inscription. ©Images provided by archaeologist Carlos Pallan (University of Bonn).

In this paper, we propose 3 improvements to an initial evaluation of methods for automatic detection of Maya hieroglyphs [17]. Namely, the contributions of this paper are the following:

1. *Weighted bag representations*: The main contribution of this work corresponds to a weighting function designed to construct bag representations of local shape descriptors. This weighting function relies on (a) an evaluation of the overlap of a local descriptor with the bounding box of the shape of interest, and (b) the proximity of the local descriptor to the center of the bounding box.
2. We investigate the effects of using different *levels of locality* for local descriptors, i.e., Histogram-of-Orientations Shape-Context (HOOSC) [18], which have proven to be effective for the task of shape retrieval, but lack of a procedure to estimate its locality span when the bounding box of the shape of interest is unknown.
3. We validate the effectiveness of the proposed weighted bag representation and the level of locality. To this end, we use an *ad hoc* methodology for the detection of shapes in weakly annotated data [17]. This method approximates the traditional sliding-windows detection approach by means of an image retrieval setup, where the sliding-window is replace by a set of pre-segmented subwindows. This method overcomes the issue of having sparsely annotated data, as is the case of the Maya inscriptions.

Our results show that (1) the use of the proposed weighted bag representation improves the detection rate, with respect to both the indiscriminate use of all the local descriptors, and an arbitrary thresholding approach previously proposed for bag construction [17]; (2) in the task of shape detection, the relatively new Histogram-of-Orientations Shape-Context (HOOSC) descriptor [18] outperforms the detection rate of the well known SIFT descriptor [10]; (3) a formal methodology to estimate the locality span of local descriptors is needed; (4) the use of a dense sampling procedure [23] is more appropriate for shape detection, as the use of interest point detectors [10,13] generates very sparse sets of points in binary images; and (5) the retrieval-based setup achieves better detection performance than the traditional sliding-windows approach when limited data is available for training.

The rest of this paper is organized as follows. Section 2 introduces the related work in shape description and object detection. Section 3 presents the weakly annotated dataset used in this work, and the formatting we devised to approximate a detection scenario. Section 4 discusses, what to our knowledge is, the unique previous work on object detection using only binary images [17]. Section 5 explains the proposed method to construct weighted bag representations. Section 6 explains the experimental protocol that we followed for shape detection. Section 7 presents statistical analysis of the previous and the proposed method. Section 8 discusses the detection results obtained with our proposed method. Finally, Section 9 provides conclusions and summarizes the findings of our work.

2. Related work

The representation of shapes is a research topic with long tradition [4,27,28,26,20]. In a nutshell, shape descriptors differ according to whether they are applied to contours or regions, and whether they describe global or local patterns of the shapes. For instance, (a) descriptors based on moments are relatively easy to compute, and they are robust against location, scale, and rotation variations [4]; and (b) Fourier descriptors work well for simple shapes of convex contours [27]. However, both of them perform poorly when dealing with affine transformations, and when the shapes correspond to complex structures with many local variations. Also, they need efficient approaches to normalize descriptors derived from different shape signatures [28].

Shape Context (SC) descriptors [1] incorporate robustness against affine variations, and are able to deal with shapes of high visual complexity [18]. These types of descriptors have proven to be effective in the task of Content-based Image Retrieval of complex shapes [1,14,18]. However, the size of the bounding box containing the shape of interest must be known to estimate the locality span of the local descriptors, which in principle is unknown on a detection setup. Therefore, despite their good descriptive potential, such descriptors are not straightforwardly suitable for detection purposes.

Several approaches have shown success in the task of detecting objects on gray-scale images [24]. The common framework for image detection implements a sliding-window approach, in which a classifier is used to evaluate sub-windows and decides whether or not they contain the element of interest [6,7,25]. However, such methods require having enough amount of data to train a classifier [2,5,19]. Another limitation for using traditional approaches based on intensity images [10] is that they rely on local regions of interest whose size is estimated using the information provided by local intensity changes [13], and this information is absent in binary images.

Common approaches to deal with the problem of detecting shapes address these issues by relying on shape information estimated from gray-scale images, i.e., by extracting contours and local orientations based on the local gradients of intensity images rather than using binary images [16], for instance, using a networks of local segment as descriptors, and performing detection of shapes belonging to classes that are relatively easy to differentiate in visual terms [3].

In contrast, in this work we address the problem of detecting complex shapes that exist as binary images. These shapes belong to visual classes that exhibit high levels of both inter-class similarity and intra-class variability, thus making the problem very challenging.

3. Dataset

In this section, we present the dataset of Maya hieroglyphs for which we want to detect instances. We explain the source of data and the format we gave it, and we also comment on the challenges this dataset poses in terms of visual complexity.