

Clustering Face Carvings: Exploring the Devatas of Angkor Wat

Brendan Klare, Pavan Mallapragada, Anil K Jain¹
Michigan State University
East Lansing, MI, U.S.A.
{klarebre,pavanm,jain}@cse.msu.edu

Kent Davis
DatASIA, Inc
Holmes Beach, FL, U.S.A.
kdavis@datasia.us

Abstract—We propose a framework for clustering and visualization of images of face carvings at archaeological sites. The pairwise similarities among face carvings are computed by performing Procrustes analysis on local facial features (eyes, nose, mouth, etc.). The distance between corresponding face features is computed using point distribution models; the final pairwise similarity is the weighted sum of feature similarities. A web-based interface is provided to allow domain experts to interactively assign different weights to each face feature, and display hierarchical clustering results in 2D or 3D projections obtained by multidimensional scaling. The proposed framework has been successfully applied to the devata goddesses depicted in the ancient Angkor Wat temple. The resulting clusterings and visualization will enable a systematic anthropological, ethnological and artistic analysis of nearly 1,800 stone portraits of devatas of Angkor Wat.

Keywords—Facial carvings; Angkor Wat; Devatas; clustering; facial components; archaeology

I. INTRODUCTION

Ancient archaeological sites contain thousands of highly detailed sculptures carved over several generations. A significant number of sculptures contain intricate human face carvings which in essence, describe the ethnological, artistic and social environment of the time. Archaeologists and other experts manually compare and group similar facial carvings to study the history and progress of a society. When dealing with a large collection of sculptures, experts must resort to computational techniques, since human analysis is often subjective and infeasible [1]. Clustering and visualization are exploratory data analysis [9] techniques that assist in forming objective hypotheses that later can be confirmed by experts.

The notion of similarity between face carvings is critical to generate meaningful clusters. In order to measure this similarity, face carvings must be represented using features that capture the salient aspects of a face. Earlier approaches [4], [8] for clustering face carvings were limited due to the unavailability of computational resources and modeling methodology, so they relied on rather primitive features to represent faces. 3D models have also been used

to measure the similarity of carvings from the Bayon temple using laser range sensors [6]. Automatic face recognition has made substantial progress over the last three decades [11], and state-of-the art approaches use more complex models to represent detailed aspects of faces.

In this paper, we propose an exploratory framework for clustering and visualization of face carvings. Face carvings are represented using shape information in the form of local point distribution models built from different facial components (eyes, nose, mouth, etc.). The similarity between two face carvings is computed as the weighted combination of the similarities between corresponding facial components. This measure of similarity is flexible and allows experts to generate different clusterings by changing the weights of different facial components. An interface is provided where the expert can visualize the variations in clustering by interactively changing the importance of different facial components, and choose specific clusterings to form hypothesis of interest. Multi-dimensional scaling is used to map the faces into a two- or three-dimensional space for visualization.

The proposed framework is demonstrated on a collection of 252 digital photographs of face carvings from the 12th century Hindu temple of Angkor Wat in modern day Cambodia [5]. The Angkor Wat temple contains carvings of 1,796 contemporary 12th century women (referred to hereafter as *devatas*) depicted in detailed portraits (see Figure 1). According to archaeologists and ethnologists, the women appear to represent a range of ethnicities residing in this region. They may, therefore, represent social and political relationships among different corners of the Khmer Empire. Identifying their origins could indicate the political or cultural dominance of certain regions and provide a genetic fingerprint of the empire’s influence. Probable ethnicities of these devatas include (i) Khmer (Central Cambodia), (ii) Lao (NE Thailand & Laos), (iii) Thai-Lanna (North Central Thai), (iv) Chinese (East Asian), (v) Indian (South Asian), (vi) Cham (Coastal Vietnam), (vii) Primitive (Mountain Tribes) and (viii) Archipelago (Malaysian-Indonesian).

II. FACE CARVING REPRESENTATION

Traditional appearance based face recognition approaches, namely Eigenfaces (PCA) and Fisherfaces (LDA), extract

¹Anil K. Jain’s research was partially supported by World Class University (WCU) program through the National Research Foundation of Korea funded by the Ministry of Education, Science and Technology (R31-2008-000-10008-0).

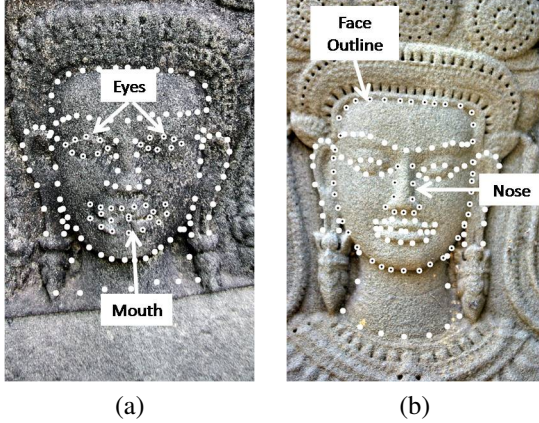


Figure 1. Two different face carvings of devatas with 140 landmarks used to represent each face. Subsets of these manually labeled points are used to describe different face features such as (a) eyes and mouth and (b) nose and face outline.

features from the pixel intensities on the face. However, the texture information used in appearance based methods cannot be leveraged in representing the images of face carvings because of: (i) the nature of stone carvings (different stone materials result in different textures), and (ii) degradation of carvings due to exposure to harsh weather conditions for over 800 hundred years. This restricts the face carvings of Angkor Wat to be represented using the shape information alone.

A. Face Landmark Points

The proposed approach uses a set of *landmark points* to represent the shape of a face. For a given face image I_i , let $P_i = (p_1^x, p_1^y, \dots, p_n^x, p_n^y)^T$ be the set of n landmark points, where p_j^x, p_j^y are the x and y -coordinates of the j -th landmark point. Further, let $p_j = (p_j^x, p_j^y)$. For the relatively low image quality of devata faces of Angkor Wat, a set of $n = 140$ landmark is manually marked. Example of two such landmark annotations can be seen in Figure 1. While typically 68 points have been used to model a face, we use a larger number of points to include the ears and neck. Typically, the process of extracting the landmark set P_i from subject i can be automated using the Active Shape Model (ASM) [2]. However, since an ASM relies on the texture information to build a robust model, it is not applicable to face carvings due to the lack of reliable texture information. To minimize the impact of errors in manual landmark placement, we first trace the boundary of each facial component. A new set of landmarks is then resampled along the contour at equal spacings. In our applications three times as many points along the contour were resampled which resulted in reliable landmarks for matching two faces. In the case of damaged carvings, we approximated the location of the corresponding landmark locations.

B. Procrustes Analysis

The similarity between two sets of landmarks can vary due to their translation, rotation, and the scale variations with respect to a canonical image. Therefore, it is necessary to align the face images such that the similarity computed quantifies the difference due to shape alone. Procrustes analysis [3] has been successfully used to normalize the face images by rigidly aligning multiple sets of points to minimize the distance between the corresponding landmarks. The landmarks points in each image are compensated for the differences in their translation, rotation, and scale using Procrustes analysis as follows:

- 1) The translational component is removed by subtracting the mean of the landmarks from each of the points, i.e. $p_i \leftarrow p_i - \frac{1}{n} \sum_{i=1}^n p_i$.
- 2) The scale is normalized by fixing concatenated vector of points in P_i to unit norm, i.e. $P_i \leftarrow P_i / \|P_i\|_2$.
- 3) The rotational differences are normalized by declaring the first set of points P_1 to have angle $\theta_1 = 0$, and using the least squares minimization to find each θ_i such that the distance between P_1 and P_i is minimized.

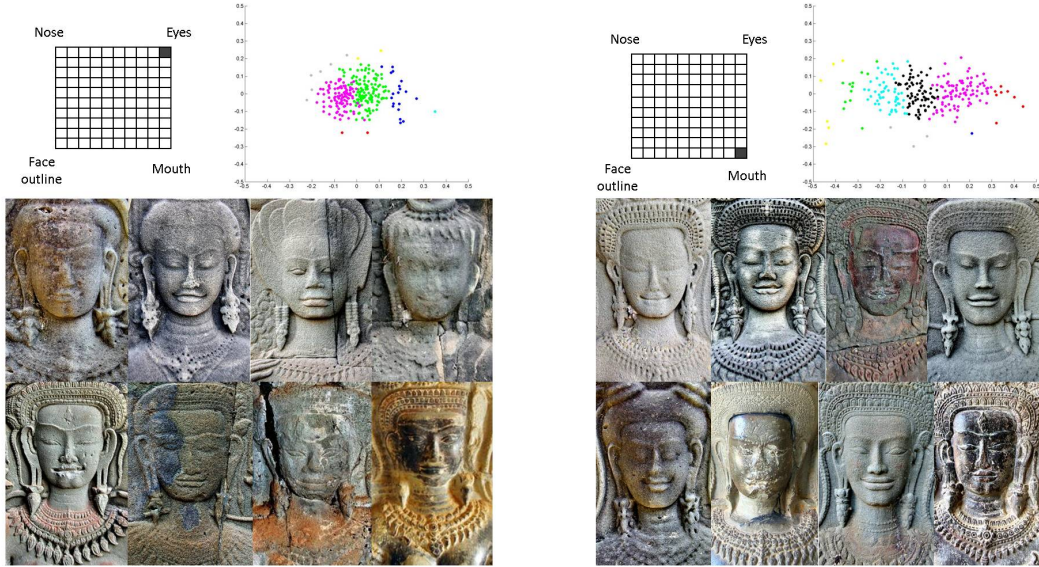
C. Point Distribution Model

After normalizing each point set, face carvings are represented using a Point Distribution Model (PDM). PDMs have been shown to be successful for representing faces in ASMs [2] and face recognition between sketches and photographs [10]. Since only the shape information is persistent when matching sketches to photographs, it resembles the current scenario where texture information is not usable. We follow an approach similar to [10] where the distances were computed by comparing local features extracted from facial components (eyes, nose, mouth, etc.), as well as the global relationship between the locations of components.

A PDM represents a set of landmarks using a statistical representation obtained by PCA. In particular, given n facial landmark points and point sets from m faces, data is first centered at the origin by subtracting the global mean μ . The mean centered data points $\hat{P}_i = P_i - \mu$, for $i = 1, \dots, m$ are collected into the columns of the pattern matrix $\Phi \in \mathbb{R}^{2n, m}$, where the i th column of Φ is the centered vector P_i . The covariance matrix $C = \frac{1}{m-1} \Phi \Phi^T$ is computed, and the top d eigenvectors $V \in \mathbb{R}^{2n, d}$ of C are selected such that 95% of the data variance is retained. For any subject i , the shape P_i is represented as $\mathbf{b}_i = V^T \hat{P}_i$.

D. Global and Local Features

At a coarse level, the similarity between a pair of face images is determined by (i) the local similarities between the components, and (ii) the global arrangement of the components relative to each other.



(a) Example clusters based on shape of the eyes (b) Example clusters based on shape of the mouth

Figure 2. Clustering results using different weight vectors w . The square grids in the top row denotes the set of possible weights an archaeologist may browse through to see clusterings obtained using those weights. The closer the user is to a corner in the grid, the higher the weight of the associated face component. The two examples show the prototypes of 8 clusters obtained with weight vectors that only use the (a) eyes, and (b) mouth.

Local Point Models: Given a set of K facial components (e.g. nose, mouth, eyes, etc.), each component is independently aligned using Procrustes analysis and a PDM is constructed for the component. We denote the coefficient vector for each component by \mathbf{b}_i^k , and all coefficient vectors together as $\mathbf{b}_i = [\mathbf{b}_i^1, \mathbf{b}_i^2, \dots, \mathbf{b}_i^K]$. The similarity between two face carvings is computed using the weighted Euclidean distance, where different facial components are assigned different weights.

Global Point Model: Local models capture only the similarity between corresponding components of the face. However, the relative placement of these components is also important in characterizing a face. The global relation between the components is captured as follows: Given K components, the $K(K-1)/2$ pairwise distances between the center of mass of all possible pairs are computed.

III. CLUSTERING FACE CARVINGS

Given K facial components (Section II), the similarity matrix for component k is

$$S_k(i, j) = 1 - \frac{\|b_i^k - b_j^k\|_2 - \min(S_k)}{\max(S_k) - \min(S_k)} \quad (1)$$

Data clustering is performed by combining the K local similarity matrices $S_k \in \mathbb{R}^{m, m}$ ($k = 1 \dots K$), and the global similarity matrix S_g into a single similarity matrix $S = w_g S_g + (1 - w_g) \sum_{k=1}^K w_k S_k$, where $0 \leq w_g \leq 1$, and $\sum_{k=1}^K w_k = 1$

The feature weights are introduced to emphasize certain facial features for data exploration. Different weights result

in different clusterings, providing different insights into the facial portrait collection.

A. Visualization

Visualization forms a key aspect of data analysis. In our application it assists a domain expert to browse through different clusterings obtained by using different weightings of facial components. A web-based interface was developed to allow an expert to alter the facial component weight vector w and visualize clusterings as the weights are altered. A discrete set of weights is used to display a K -sided convex polygon, in which each of the K vertices corresponds to one of the K facial features. When the user selects a point within the polygon, the weight vector w is determined based on the inverse distance of the selected point from each of the K vertices. As the user interactively modifies the weights, an updated clustering is generated and displayed allowing the user to investigate the groupings as function of weights. For each different weightings, the similarity matrix S is used to project the data points in either a 2 or 3 dimensional space using multidimensional scaling [7].

IV. EXPERIMENTAL RESULTS

The proposed clustering and visualization framework was applied to a subset of 252 facial images of *devatas*, collected from the West Gopura (or entrance pavilion) of the Angkor Wat temple. While the landmarks for many different facial components were marked, in this study we used only four of the major facial components (eyes, nose, mouth, and face outline) for clustering the devatas into 8 groups (possibly

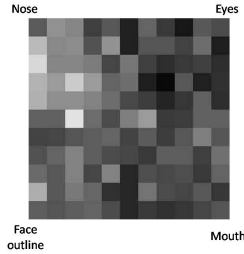


Figure 3. (Scaled) Heat map indicating the percentage of pairwise labels satisfied by each weighted combination of facial components. The brighter the location, the larger the number of constraints satisfied. These weightings can be explored further using the interactive web page¹.

corresponding to the 8 ethnic groups listed in Sec. I). We performed clustering using complete-link clustering, which allows for dendrogram analysis by domain experts.

Figure 2 shows two separate clustering results with different selection of weights. The figure also shows the associated 2-d projections of the 252 faces. The two 8-cluster groupings are based on just using the eyes or the mouth in computing the face similarities. An interactive web page of our proposed clustering framework is also available¹, where users can specify the weights interactively. Once a clustering of interest is found, it can be further explored by clicking on any displayed cluster prototype to view images of all carvings in the cluster.

A. Evaluation

Since the ground truth cluster labels are not available for the devata images, external cluster evaluation indices cannot be applied. While it is difficult to manually provide cluster labels for all the images, it is relatively easier to determine if a pair of images should be put in the same cluster or not. A group of faculty and students at the Khmer Arts Academy in Phnom Penh, Cambodia, have labeled a set of 243 randomly chosen pairs of images. Figure 3 shows a heat map of the fraction of pairwise constraints satisfied by the clusterings using different weighted combinations of the facial components. The brighter the block, the larger the number of pairwise constraints satisfied. The brightest block (6th row, 3rd column) corresponds to around 50% of the human pairwise labels satisfied by the clustering (note that for 8 clusters of equal size, only 12.5% of the constraints can be satisfied by random clustering). Utilizing the pairwise labels as side-information during clustering can further enhance the performance of clustering in terms of this evaluation metric. Further, Figure 3 reveals that face outline and nose are considered important facial components for manually labeling the pairs. Further, the objective is not to strictly adhere to the human categorization, but to suggest possible non-obvious groupings that can help the experts draw interesting conclusions.

¹<http://www.cse.msu.edu/~klarebre/angkor/cluster/index.html>

The different clusters generated by our framework are currently being analyzed by archaeologists to discover possible answers to long standing questions regarding the devatas of Angkor Wat. Using the interactive web page, the current hypotheses being examined include: (i) Do the Angkor Wat devatas represent the different ethnic groups listed in Section I? (ii) Does the location of certain devatas within the temple have any meaning (in order to further answer this question, we are currently labeling the remaining devata carvings with landmarks), and (iii) Is there any correlation between face carvings that may indicate how many different sculptors were used in crafting the devata carvings.

V. CONCLUSIONS

A method for grouping face carvings in archaeological sites is presented that: (i) computes the local distance along the predominant modes of variation for different facial components, (ii) allows domain experts to change the importance of each component by assigning different weights in computing the similarity, and (iii) provides a visualization tool to illustrate how the clusters change with different weights. The proposed framework results in many meaningful clusters of devata face carvings of the magnificent Angkor Wat temple. These clusterings suggest hypotheses to domain experts for their validation. Future work will attempt to find richer set of clusters using semi-supervised clustering techniques, by incorporating instance level pairwise constraints provided by experts into the grouping process.

REFERENCES

- [1] J. A. Barcelo, editor. *Computational Intelligence in Archaeology*. Information Science Reference, 2008.
- [2] T. F. Cootes et al. Active shape models—their training and application. *Computer Vision Image Understanding*, 61(1):38–59, 1995.
- [3] J. C. Gower. Generalized procrustes analysis. *Psychometrika*, 40(1):33–51, 1975.
- [4] E. Guralnick. The proportions of some archaic greek sculptured figures: a computer analysis. *Computers and the Humanities*, 10:153–169, 1976.
- [5] H. Jessup and T. Zephir. Sculpture of angkor and ancient cambodia: Millennium of glory. In *National Gallery of Art, Washington DC*, 1997.
- [6] M. Kamakura et al. Classification of bayon faces using 3d models. In *Virtual Systems and Multimedia*, 2005.
- [7] J. B. Kruskal. Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, 29(1):1–27, 1964.
- [8] G. Siromoney, S. Govindaraju, and M. Bagavandas. Temple carvings of southern india. *Perspectives in Computing*, 5(1):34–43, 1985.
- [9] J. W. Tukey. *Exploratory Data Analysis*. Addison Wesley, 1 edition, 1977.
- [10] P. Yuen and C. Man. Human face image searching system using sketches. *IEEE Trans. Systems, Man and Cybernetics*, 37(4):493–504, July 2007.
- [11] W. Zhao et al. Face recognition: A literature survey. *ACM Comput. Surv.*, 35(4):399–458, 2003.