Classification of Bayon Faces Using 3D models

Mawo KAMAKURA, Takeshi OISHI, Jun TAKAMATSU and Katsushi IKEUCHI

Institute of Industrial Science, The University of Tokyo

4-6-1 Komaba, Meguro, Tokyo, 153-8505, JAPAN
Email:{mawo, oishi, j-taka, ki}@cvl.iis.u-tokyo.ac.jp

Abstract. Digital 3D models of historic buildings or cultural heritage objects are useful for preservation. Not only can we store them permanently, but the models can supply a clear guideline for the restoration process. 3D models also provide sufficient information about geometrical characteristics that may help archaeologists to inspect and classify the objects. Currently, we are working on a 3D digital-archiving project of the Bayon Temple. It is a building of stonework that was built in the 12th century in Cambodia. It is famous for its towers with four faces at the four cardinal points. According to research by JSA (Japanese government team for Safeguarding Angkor), the faces can be classified into three groups based on subjective criteria. In this paper, we explore a more objective way to classify the faces by using measured 3D geometrical models. After alignment of 3D faces in the same coordinate system, orientation, and normalization, we captured in-depth images of each face and then classified them by several statistics methods.

1. Introduction

Over the past years, much research has been done on automatically obtaining 3D shapes of art objects and cultural heritage objects using laser range sensors. The performance of computers has improved rapidly, so research in the fields of image processing and computer vision have also advanced. Measuring real world objects and converting them to 3D digital models are well-known applications of computer vision. We are currently working on digital preservation of large-scale cultural heritage objects by using computer vision techniques and laser range sensors. We have worked on the Great Buddha Project[1] and have preserved cultural heritage objects such as the Kamakura, Nara, and Atchana Great Buddha and at the same time have researched and developed advanced modeling techniques.

Currently we are working on the Bayon Digital Archiving Project[2]. The Bayon is a temple constructed in the 12th century and is located in the center of the Angkor Tom. Fig. 1 shows pictures of Bayon temple and Bayon Face. An enormous site with a size of 100m x 100m and towers reaching about 50m at the highest, it is famous for its towers, each with four faces carved on each side. The Bayon Digital Archiving Project started in 2003 and until now four missions have been executed. In addition to modeling of the whole site, we have been modeling and constructing libraries of the faces.

There are 52 towers in Bayon with faces on them, and 173 faces have endured damage or collapse. According to research by the JSA (Japanese government team for safeguarding Angkor)[3], the 173 faces can be classified into three types: Deva, Asura, and Devata. It is known that the Deva type face is dominant. However, this classification is based on subjective evaluation.

Our work is on measuring and preserving the 3D shape of the Bayon faces using laser range sensors. Fig. 2 shows one of the laser range sensors. Fig. 3 shows the measurement result of the entire Bayon Temple. We have completed measurement of all 173 faces in the previous four missions. Fig. 4 shows the picture of a face and the measurement result of the 3D face model. From the 3D models acquired from the measurement, we made a more objective classification of the faces. By doing so, we expected to scientifically confirm the classification results of JSA, and if not, to present a new classification which would be impossible to achieve by human eyes.

The outline of the rest of this paper is as follows. Section 2 gives a description of the Bayon faces. Section 3 explains the pre-classification normalization process and the classification methods. In Section 4 we present experimental results. Finally, in Section 5 we summarize and conclude this paper.

Fig.1 left: pictures of the Bayon temple, right: picture of Bayon face

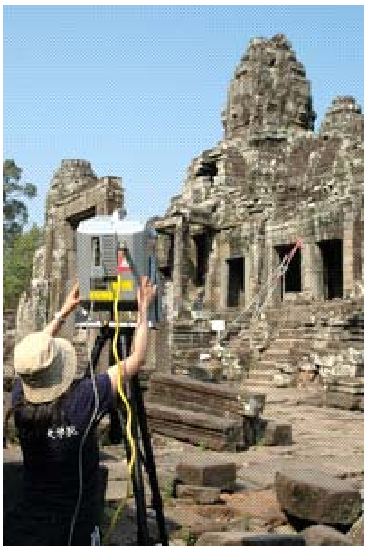


Fig.2 picture of laser range sensor

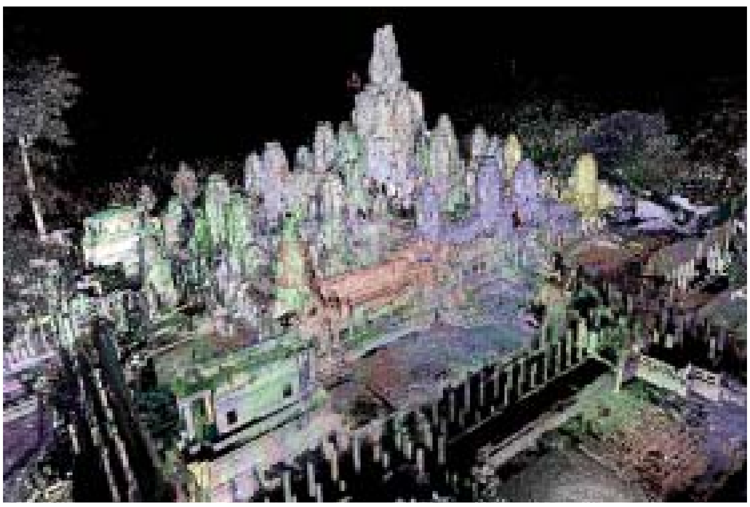


Fig.3 measurement result of entire Bayon temple

Fig.4 left: picture of face, right: measurement result of 3D face model

2. Bayon face

The Art History team of JSA investigated the Bayon faces in the towers of the Bayon Temple to see if they might provide insight into the purpose of the construction of the temple. According to the JSA, as a result, the faces can be classified into roughly three types: Deva, Asura, and Devata[3]. These faces are classified based on the outlines of the faces. Some degree of regularity could be identified in the locations where each of the types was found. JSA found that the faces looking out from temple are all the Asura type; those on the inner facing center tower are the Deva type; and those looking toward the central sanctuary are the Devata type. The four-faced tower is thus a composite of guardian deities, giving protection of the deities Deva and Asura, with the goddess Devata attendant on the main deities. Fig.5 shows pictures, 3D models and shape lines of three types Bayon faces. The each typical face is 35N (Asura), 51S (Deva), 50E (Devata).

The Deva face is calm and noble, representing God. It is plump and rounded in shape. The Devata face is a comparatively narrow face with a harsh expression. The Asura face is angular in shape with a square jaw and a rather grim, heavy expression (Devil). It is known that the Deva type face is the dominant one of these three types.

However, on some of the four-faced towers of the Bayon temple are many other faces that are difficult to classify with any accuracy. This is thought to be related to the division of labor between different craftsmen, and the differences in their techniques. Also, differences may have been created by other influences like weathering and destruction. If JSA examined the degree of completion of the faces on the towers, JSA found that although some examples are very nearly finished, overall there is high number of unfinished examples. Furthermore, the Bayon temple was built in a short period of time, so the work must undoubtedly have been divided between many different craftsmen. As mentioned above, it can be gathered that the same group completed the four faces in a single tower. It is highly likely that the same group completed two or more towers at the same time. Furthermore, consistent variations can be identified in the shapes and expressions of the faces that enable us to classify them roughly into the three types listed above: Deva, Devata, and Asura. However, other places were found where the craftsmen could not have worked with any awareness of creating one of these three types. There are many instances of clear differences between the left and right sides of the face, suggesting the possibility that separate groups of craftsmen worked on left and right sides. Given this situation, 173 faces of four-faced towers of Bayon temple are classified, but the classification is not exact.

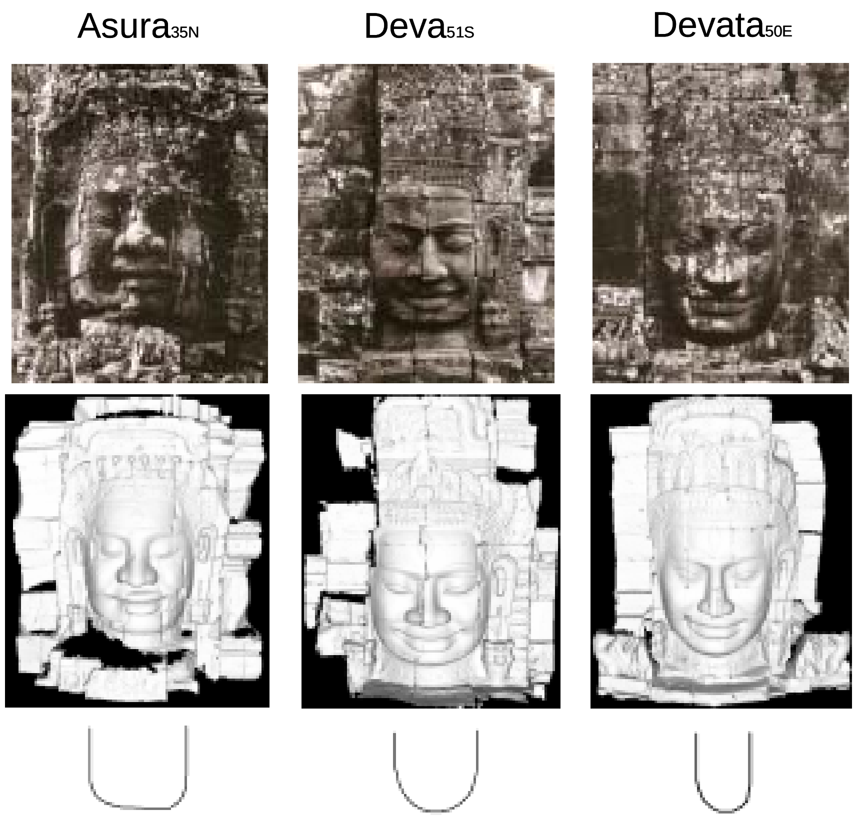


Fig.5 pictures, 3D models and shapes line of 3 types of Bayon faces.

3. Classification technique

3D models with Bayon faces acquired by measurement are used and classified. First of all, we convert all the faces into homogeneous in-depth images, that is, normalization. Afterward, we classify them by several statistic techniques using the converted images.

3.1. Normalization

First of all, we normalize an arbitrary 3D Bayon face as a standard face. This normalization removes differences between the sizes and also between the directions of the faces in the same size images. It also suppresses the influence of data excluding the face to a minimum. The 3D Bayon face model as a standard face is in a consistent position and orientation, and its in-depth image from an appropriate viewpoint is displayed in the entire specified area. According to the research by JSA, faces are classified based on the outline of the face. So we fit the outline of the face with image size manually. At this time, the moving matrix is M_{ref}.

Next, we obtain a transformation matrix M_{tar},i (i=1,...,N) for displaying all face images similar to a standard face. N is number of faces being classified. First, we obtain three coordinate points of a characteristic face; two points from the inner corners of the eyes and one from a point between the mouth and the nose. Let the points of a characteristic standard face model be (x_1, x_2, x_3) and the points of characteristic normalized face models be (y_{1i}, y_{2i}, y_{3i}). We obtain R: rotation matrix, t: translation matrix, and c: variable of expansion and reduction to minimize the squared distance between these two points. This theorem is an absolute orientation problem[4]. Shown below is the technique for solving the transformation matrix[5]. Fig.6 shows the outline of normalization method.

Let X={x_1, x_2, …, x_n} and Y={y_1, y_2, …, y_n} be corresponding point patterns in m-dimensional space. The average of the squared distances is,

$$e^2(R,t,c) = \frac{1}{n}\sum_{i=1}^{n}\|y_i-(cRx_i+t)\|^2 \tag{1}$$

Transformation parameters (R, t, c) to minimize the equation are given as follows:

$$R = USV^T \tag{2}$$

$$t = \mu_y - cR\mu_x \tag{3}$$

$$c = \frac{1}{\sigma_x^2} tr(DS) \tag{4}$$

$$\sigma_x^2 = \frac{1}{n}\sum\|x_i - \mu_x\|^2 \tag{5}$$

$$\sigma_y^2 = \frac{1}{n}\sum\|y_i - \mu_y\|^2 \tag{6}$$

where μ_x and μ_y are mean vectors of x_i and y_i, UDV^T is the singular value decomposition of a covariance matrix between x_i and y_i, S in (4) must be chosen as

$$S=\begin{cases} I & \text{if } \det(U)\det(V)=1 \\ \mathrm{diag}(1,1,\cdots,1,-1) & \text{if } \det(U)\det(V)=-1 \end{cases} \tag{7}$$

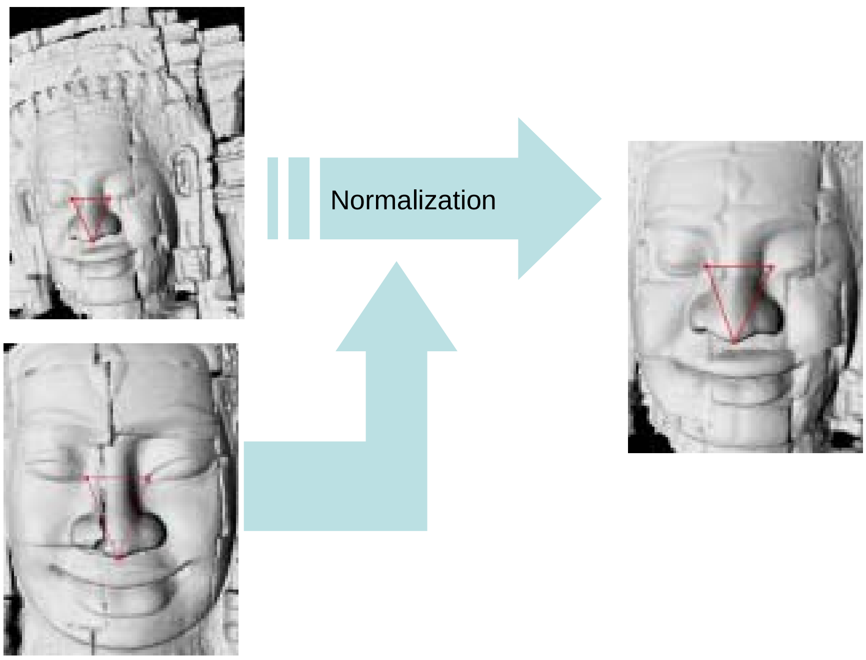


Fig.6 outline of normalization method

3.2. 3D Shape Analysis

In this paper, we examined two types of analyses: supervised and unsupervised analyses. The purpose of the supervised analysis is to clarify the differences among the given classes. JSA has already classified all faces into three types based on its subjective evaluations. Through such an analysis, we can verify correctness of the process of JSA's classification, and then objectively evaluate the differences using statistical analysis methods.

In contrast, the purpose of the unsupervised analysis is to discover new knowledge through classification of the faces without any a priori standards. This analysis may be able to produce a novel and detailed classification. As a result, it may reveal undiscovered historical secrets. In this section, we first describe these analysis methods.

At the end of this section, we describe a method to visualize classification criteria. Although conventional methods for object recognition only pay attention to improvement in recognition ability, it is more important in this paper to clarify the criteria.

3.2.1. Supervised Analysis: Linear Discriminant Analysis

Consider a sample space R^m. Given two classes G_1, G_2, discriminant analysis provides us with a scalar function $f(\mathbf{x})$ to decide which class any points $\mathbf{p}$ belong to ; if $f(\mathbf{x}) > 0$, $\mathbf{p}$ belongs to Class G_1 and if $f(\mathbf{x}) < 0$, $\mathbf{p}$ belongs to Class G_2.

In this paper, we use a linear function $f(\mathbf{x}) = \mathbf{n} \cdot \mathbf{x} + d$ as the classification function. The reason is that the dimension of the sample space, that is, image size ($= 64 \times 64$, in this case), is much greater than the number of samples; there are only 173 faces in the Bayon temple. It is preferred aht the dimension and parameters of the function are small in order to prevent a so-called "over-fitting" problem.

Roughly speaking, the parameters of the function $\mathbf{n}, d$ can be determined by maximizing S_B/S_T, where S_B and S_T are intraclass and interclass variances, respectively. Concretely, it is necessary only to solve simultaneous linear equation in $S\mathbf{n} = \overline{\mathbf{x}}^{(1)} - \overline{\mathbf{x}}^{(2)}$, where $S = \frac{1}{n_1 + n_2 - 2}\left(\sum_{i=1}^{n_1} \left(\mathbf{x}_i^{(1)} - \overline{\mathbf{x}}^{(1)}\right)\left(\mathbf{x}_i^{(1)} - \overline{\mathbf{x}}^{(1)}\right)^T + \sum_{i=1}^{n_2} \left(\mathbf{x}_i^{(2)} - \overline{\mathbf{x}}^{(2)}\right)\left(\mathbf{x}_i^{(2)} - \overline{\mathbf{x}}^{(2)}\right)^T \right)$, n_i is the number of samples including Group *i*, $\mathbf{x}_j^{(i)}$ is the *j*th sample of Group *i*, and $\overline{\mathbf{x}}^{(i)}$ is the average of Group *i*.

Also $d = -\frac{1}{2}\left(\mathbf{n} \cdot (\overline{\mathbf{x}}^{(1)} + \overline{\mathbf{x}}^{(2)})\right)$. Because the matrix S is not full-rank matrix, we solve the equation using the singular value decomposition (SVD) method while minimizing $|\mathbf{n}|$.

3.2.2. Unsupervised Analysis: Hieratical Cluster Analysis

Cluster analysis provides us with some classification of samples according to distances among them. In this paper, we employ agglomerative hierarchical cluster analysis. This analysis begins with each sample being considered as each cluster and then proceeds to combine the nearest two clusters until all samples belong to one cluster. As the result of this analysis, we obtain a dendrogram as shown in Fig.9. Unfortunately, we cannot determine the correct number of clusters from this analysis only. However, distances between the combined two clusters are useful to determine the number.

Before using the analysis, it is necessary to define the distance between any two samples. We naturally define it as the Euclidean distance in the distance image space. It is also necessary to define the distance between the combined cluster and the other cluster. In this paper, we calculate the distance based on the Ward method, which is superior in practical use.

3.2.3. Visualization of Differences between two Classes

As mentioned above, it is quite important to clarify the classification criterion, especially for this paper. Suppose we have two classes and a linear discriminant function to classify them. The value of a linear discriminant function expresses the distance between a sample point and the hyper-plane expressed by the function. The greater the distance becomes, the more similar the sample looks to faces in the class G_1, vice versa. That is, variance of the distance helps visualize the differences between the two classes. Because the variance is equal to displacement of the point along the direction of the normal of the plane, the direction just expresses the classification criterion. The visualization is quite easy.

4. Experiments

Original 3D models of faces used in this experiment are obtained by the following steps: We first measured 3D models of Bayon faces by the laser range sensors, Cyrax2500[6] and Vivid910[7]. Next, we applied the alignment[8] and refinement[9] methods to these models. After obtaining the models, we made their in-depth images using the normalization method as mentioned above. The sizes of these images were 64 x 64. In this experiment we assume that the image data is only a 4096 dimensional vector and that usual vector operations are applicable for the data. In this experiment we used 88 images. We preformed the following two analyses: the supervised and unsupervised analyses.

4.1. Supervised analysis

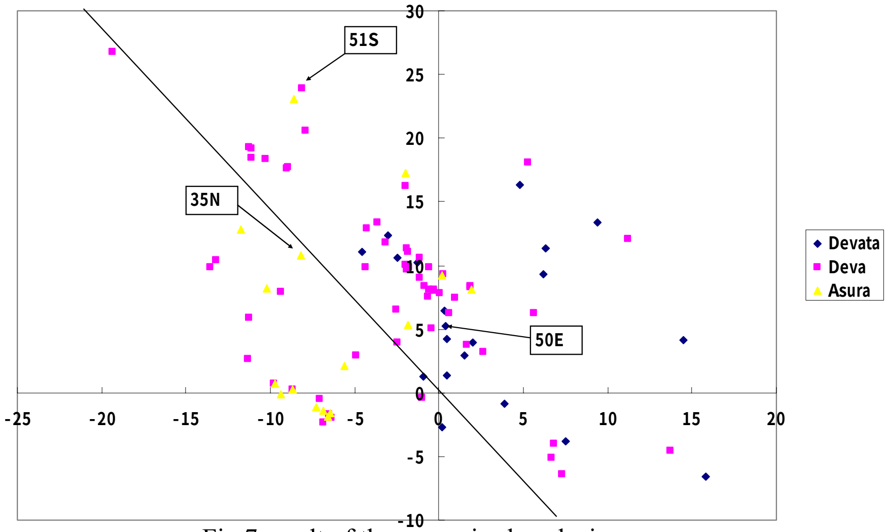
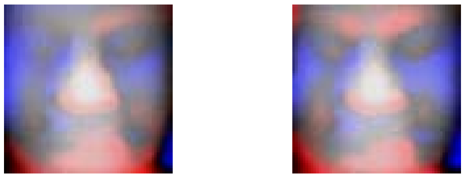


Fig.7 result of the supervised analysis

In this experiment, we preformed the supervised analysis based on the JSA's classification. Fig.7 illustrates the result of the analysis. The graph is obtained by projecting all vector points of faces on the 2D flat surface that are determined by two linear discriminant functions in order to most clearly express the classification result.

The former discriminant function classifies Devata (goddess) and Deva (god) and the separate plane corresponds to the y-axis. That means the right side area of the Y axis is a female area and the left side area a male area. Although Asura images are not used for determining the function at all, almost all of Asura (male) data are in the male area.

The latter function classifies Deva (god) and Asura (devil) and the separate plane corresponds to the diagonal line. That means the upper side area of the diagonal line is a god area and the under side area is a Devil area. Because Devata is a goddess, not a devil, almost all of Devata data are in the god area.



left: Devata → Deva   right: Deva → Asura
Fig.8 visualization of differences between two classes

Additionally, Fig.8 shows visualization of differences among three classes. Two images reveal that when some face of the former group is morphed to become similar to a face of the latter group, the blue area is dented and the red area bulges. About the left side image, we find that the bite of Deva is dented and the chin bulges more than Devata's. In the same manner, we also find about the right side image, the bite of Asura is dented and the chin bulges more than Deva's. We find this result corresponds to characteristics of the three types.

4.2. Unsupervised analysis

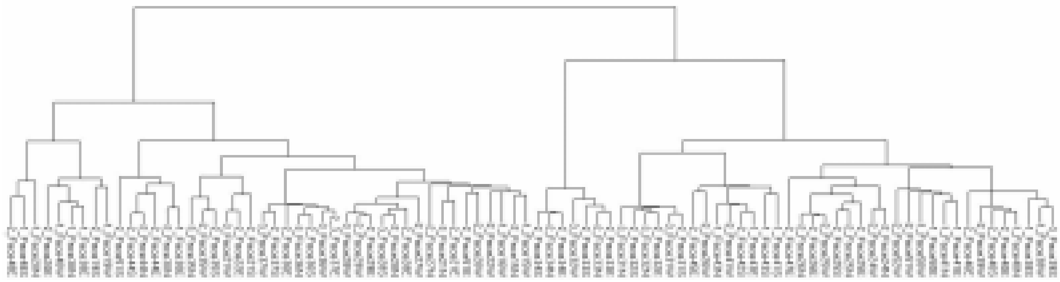


Fig.9 dendrogram obtained by unsupervised analysis

The dendrogram as shown in Fig.9 is the result of unsupervised analysis. This dendrogram shows that, as expected, the face data of the same tower is in the same cluster.

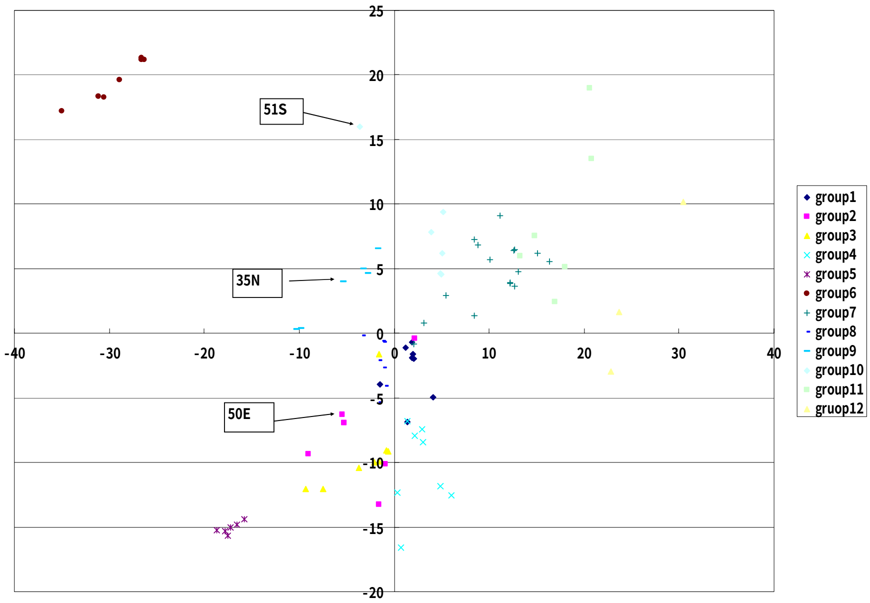


Fig.10 scatter diagram of unsupervised analysis result

In order to examine the result of this analysis, we used the scatter diagram shown in Fig.10. Because each vector is 4096 (not two) dimensional, a map is needed to draw the diagram. We determine the map using the PCA (principal component analysis). PCA can determine the map to minimize the reduction of Euclidean distances caused by the map. Each color expresses one cluster.

In this analysis, we don't have any a priori knowledge. This diagram shows distributions of the clusters. Although this diagram includes some outliers (group: 5, 6, 11 and 12), almost all clusters concentrate in one area. Especially three clusters including the typical faces are near the boundary and respectively far away from each other. That is, we find that JSA correctly selected the more distinguished faces as the typical faces.

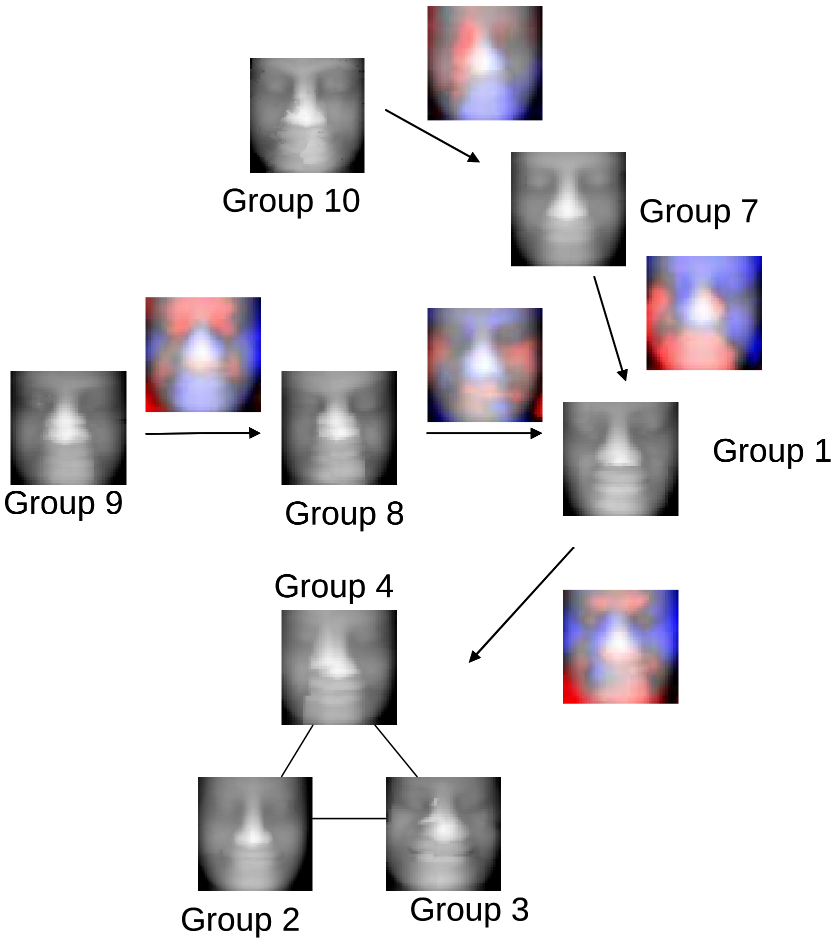


Fig.11 visualization of the clusters and their differences

Fig.11 shows visualization of the clusters and their differences. Each gray image in the figure expresses an average face of the corresponding cluster. The images on transitions express differences between the two clusters linked by arrows. Faces included in group 1 look unlike the three typical faces. That proves it may be impossible to classify all the faces into the three types in first place.

5. Conclusion

In this paper, we classified the Bayon faces using their precise 3D models. As mentioned above, the Art History team of JSA classified all faces into three types, but this classification was based on subjective criteria. Therefore, we tried classification based on objective criteria using 3D models.

For our purpose, we first measured the 173 faces by laser range sensors. As the result of measurement of the faces, we made libraries of the Bayon faces. In this experiment, we used these face model libraries.

We tried to classify using two analysis methods: One was supervised analysis based on linear discriminant analysis and the other was unsupervised analysis based on hierarchical cluster analysis.

We actually classified the three types using supervised analysis. Furthermore, we visualized differences among the three types of faces by our objective criteria. The result justified JSA's classification criteria. The classification of our result and that of JSA's were almost same, however, these two classifications did not coincide perfectly. In our future work, we should determine whether the reason is caused by failure of our analysis or by JSA's misclassification.

Also we showed classification of all the faces using unsupervised analysis. This analysis assumes no a priori knowledge. In this analysis, we illustrated the justification of the selection of the typical faces and found that these pieces of data are not on an overlapping area of clusters, but on an area where the boundary is relatively clear. This analysis also showed difficulties in classification. We determined that it may be impossible to classify all the faces into three categories; there are some faces unlike Deva, Devata, or Asura. In the future, we should further investigate results obtained by our analysis methods.

References

[1] Katsushi Ikeuchi, Atsushi Nakazawa, Kazuhide Hasegawa, Takeshi Oishi, "The Great Buddha Project: Modeling Cultural Heritage for VR Systems through Observation," The second IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR2003), Oct 2003.

[2] K. Ikeuchi, K. Hasegawa, A. Nakazawa, J. Takamatsu, T. Oishi and T. Masuda, “Bayon Digital Archival Project,” In Proceedings of the Tenth International Conference on Virtual System and Multimedia, pages 334-343, November 2004.

[3] http://www.angkor-jsa.org/

[4] Berthold.K.P.Horn, “Closed-form solution of absolute orientation using unit quaternions”, Journal of the Optical Society of America A, vol.5, no.7, pp.1127-1135, 1987.

[5] Shinji Umeyama, “Least-Squares Estimation of Transformation Parameters Between Two Point Patterns”, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol.13, no.4, April 1991.

[6] http://www.cyra.com

[7] http://konicaminolta.com/products/instruments/vivid/vivid910.html

[8] T. Oishi, A. Nakazawa, R. Kurazume and K. Ikeuchi, “Fast Simultaneous Alignment of Multiple Range Images using Index Images,” In Proc. 3D Digital Imaging and Modeling, 2005(to appear).

[9] Ryusuke Sagawa, Takeshi Oishi, Atsushi Nakazawa, Ryo Kurazume, Katsushi Ikeuchi, "Iterative Refinement of Range Images with Anisotropic Error Distribution," Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.79-85, 2002.