

# Personal Authentication Based on 3D Configuration of Micro-feature Points on Facial Surface

Takao Yoshinuma<sup>(✉)</sup>, Hideitsu Hino, and Kazuhiro Fukui

Graduate School of Systems and Information Engineering, University of Tsukuba,  
Tsukuba, Japan

yoshinuma@cvlab.cs.tsukuba.ac.jp, {hinohide,kfukui}@cs.tsukuba.ac.jp

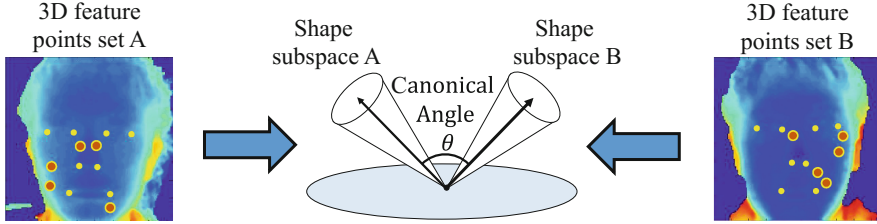
**Abstract.** This paper proposes a personal authentication method based on the 3D configuration of micro-feature points such as moles and freckles, plus common feature points such as corners of eyes, edges of mouth and nostrils. The basic idea behind the proposed method is the assumption that such 3D configuration can be uniquely determined by individuals. To compare two configurations of feature points effectively, the concept of 3D shape subspace in a high-dimensional vector space is introduced. With this idea, the task of comparing the sets of feature points is converted to that of measuring the structural similarity between the corresponding shape subspaces. The validity of the proposed method is demonstrated through experiments with feature points from actual face images. In addition, the performance limit of the method is explored using sets of artificially generated feature points.

**Keywords:** Personal authentication · Facial feature · Shape subspace · Grassmann discriminant analysis

## 1 Introduction

This paper presents a personal authentication method based on the 3D configuration of micro-feature points, such as moles and freckles, on a facial surface, and facial feature points such as corners of eyes, edges of mouth and nostrils. Common appearance-based face recognition methods [1–7] often fail to distinguish a pair of faces with similar appearance by its nature. In contrast, the most remarkable advantage of the proposed method is that it can distinguish even a pair of twins with the completely same appearance, since the 3D configuration of micro-features includes both inherent and acquired features.

The proposed method is motivated by a 3D object recognition method based on the structure similarity between shape subspaces [8]. In this method, each set of feature points is compactly represented by a linear subspace, which is called shape subspace as shown in Fig. 1. A shape subspace of an image is known to be generated by applying factorization method [9] to the trajectories of a set of feature points, which are extracted and tracked from sequential images. This converts the task of comparing two given sets of feature points to that of measuring



**Fig. 1.** A conceptual diagram of the proposed framework. Sets of  $T$  3D feature points are extracted from RGB-D images to constitute shape subspaces in  $T$ -dimensional vector space, and their similarity is evaluated by the notion of canonical angles.

the structural similarity between two corresponding shape subspaces, by using the canonical angles between them [8]. Suppose there are several different classes (e.g., different persons), and shape subspaces are already enrolled in the system for each class. Then, an input set of feature points is classified into the class with the highest similarity.

The above framework based on the shape subspace can work well under stable lighting condition and no occlusion. However, the framework has the following two issues: (1) stable extraction and tracking of a set of feature points are required, while they are difficult in practical situations, (2) classification ability is insufficient for discriminating persons with very similar 3D structure of feature points.

To address the first issue, we propose an effective method to generate a shape subspace from a single depth image captured by an RGB-D sensor. This method does not require feature points tracking. For the second issue, we enhance the classification ability by introducing a powerful subspace-based classification method called Grassmann discriminant analysis (GDA [10]). An additional issue is that facial feature detection is often error-prone, which may reduce the performance of our method. To deal with this issue, we generate multiple perturbed shape subspaces from an input depth image by repeating the random sampling of reliable feature points, instead of generating one subspace from the depth image.

The validity of the proposed method is demonstrated through simple experiments with feature points from actual face images. In addition, the performance limit of the method is explored using sets of artificially generated feature points, simulating the situation of distinguishing similar faces.

## 2 Preliminary

The core idea of the proposed method is the use of structural similarity between shape subspaces, which is measured by using the canonical angles between these subspaces. In the following, we firstly explain how to calculate canonical angles from two given sets of facial feature points. Then, we outline the algorithm of Grassmann discriminant analysis for realizing a highly accurate classifier using shape subspaces.

## 2.1 Definition of Structural Similarity with Canonical Angles

Consider an  $n_A$ -dimensional subspace  $\mathcal{S}_A$  and an  $n_B$ -dimensional subspace  $\mathcal{S}_B$ , where  $n_A \leq n_B$ . The principal canonical angle  $\theta_1$  is uniquely defined by [11]

$$\cos^2 \theta_1 = \sup_{\mathbf{u} \in \mathcal{S}_A, \mathbf{v} \in \mathcal{S}_B} \frac{\mathbf{u}^\top \mathbf{v}}{\|\mathbf{u}\|^2 \|\mathbf{v}\|^2}, \quad (1)$$

where  $\|\cdot\|$  denotes the norm of a vector. Let  $\mathbf{Q}_A$  and  $\mathbf{Q}_B$  denote the orthogonal projection matrices of the subspaces  $\mathcal{S}_A$  and  $\mathcal{S}_B$ , respectively. Then,  $\cos^2 \theta$  for the canonical angle  $\theta$  between  $\mathcal{S}_A$  and  $\mathcal{S}_B$  is equal to the eigenvalue of  $\mathbf{Q}_A \mathbf{Q}_B$  or  $\mathbf{Q}_B \mathbf{Q}_A$  [11]. The largest eigenvalue corresponds to the smallest angle  $\theta_1$ , whereas the second largest eigenvalue corresponds to the smallest angle  $\theta_2$  in a direction perpendicular to that of the largest canonical angle. The values  $\cos^2 \theta_l (l = 3, \dots, n_A)$  are calculated in the same manner.

The simplest classification method using the canonical angles is known as the mutual subspace method (MSM [3]). In this method, the structural similarity  $\varphi$  between  $\mathcal{S}_A$  and  $\mathcal{S}_B$  is defined by

$$\varphi = \frac{1}{n_A} \sum_{l=1}^{n_A} \cos^2 \theta_l. \quad (2)$$

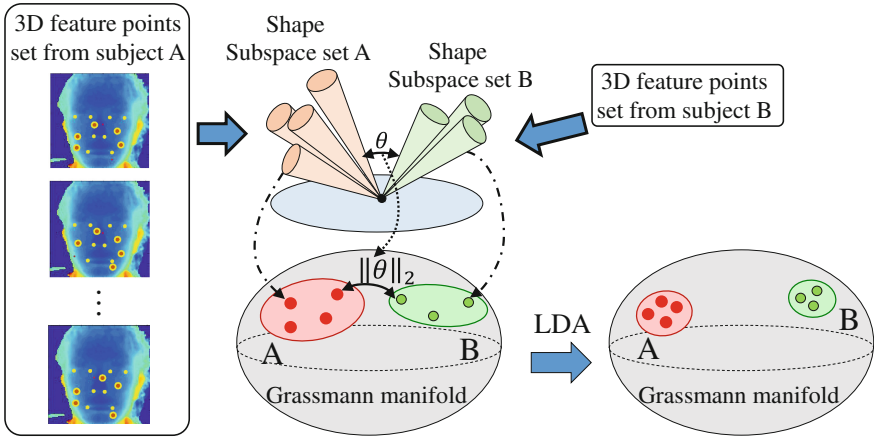
If two shape subspaces coincide completely with each other,  $\varphi$  is 1, since all canonical angles are zero. The similarity  $\varphi$  gets smaller as the two spaces deviate. The similarity  $\varphi$  is zero when the two subspaces are orthogonal to each other. In the MSM, we calculate the similarities between an input subspace  $\mathcal{S}$  and reference subspaces  $\mathcal{S}_i (i = 1, 2, \dots)$ . The input subspace is classified into the class with the highest similarity. As explained in detail later, the shape subspaces are 3-dimensional subspaces, hence we can compute three canonical angles. By using them, the similarity between shape subspaces  $\mathcal{S}_A$  and  $\mathcal{S}_B$  is defined as

$$\text{sim}(\mathcal{S}_A, \mathcal{S}_B) = \frac{1}{3} \sum_{l=1}^3 \cos^2 \theta_l. \quad (3)$$

## 2.2 Grassmann Discriminant Analysis

The structural similarity calculated by the MSM is effective for discriminating two sets of facial feature points. However, its classification ability for the structural similarity is insufficient in discriminating persons with very similar feature points configurations.

A linear subspace considered in the MSM is regarded as a point on a Grassmann manifold [12], and the concept underlying the MSM is the computation of distance between points on the Grassmann manifold. To improve the classification ability of the MSM, we introduce the notion of Grassmann discriminant analysis (GDA [10]), which is a generalization of linear discriminant analysis



**Fig. 2.** Conceptual diagram of the use of Grassmann discriminant analysis. Each shape subspace is considered as a point on a Grassmann manifold. Distance between points on the manifold corresponds to the canonical angle between the subspaces.

(LDA [13]) for dealing with a set of data points as a single datum as shown in Fig. 2. Grassmann discriminant analysis can find the classification axis that maximizes the class separability taking the relative similarity between subspaces into account.

Grassmann discriminant analysis is an extension of linear discriminant analysis by replacing vectors for samples with subspaces generated from the set of samples. For this extension, GDA utilizes the kernel trick [14] with the kernel function on the space of subspaces defined as

$$k(\mathcal{S}_i, \mathcal{S}_j) = \text{sim}(\mathcal{S}_i, \mathcal{S}_j), \tag{4}$$

where  $\text{sim}$  is defined by Eq.(3). We then define the empirical kernel feature map [15] of a shape subspace  $\mathcal{S}$  as

$$\mathbf{k}(\mathcal{S}) = (k(\mathcal{S}, \mathcal{D}_{1,1}), k(\mathcal{S}, \mathcal{D}_{1,2}), \dots, k(\mathcal{S}, \mathcal{D}_{2,1}), \dots, k(\mathcal{S}, \mathcal{D}_{C,N_C}))^\top, \tag{5}$$

where  $\mathcal{D}_{c,l} (c = 1, \dots, C, l = 1, \dots, N_c)$  are subspaces consist of training samples,  $C$  is the number of the classes (i.e., each person), and  $N_c$  is the number of the samples of class  $c$  (i.e., sets of feature points). For the details of GDA, refer to [10].

### 3 Proposed Framework

In this section, we explain the detail of the proposed framework for person authentication. First, we extract facial feature points by applying the circular separability filter [16] for a gray scale image of a face. Then, we obtain the 3D coordinates of the feature points from a depth image captured with the gray scale image at the same time.

A shape subspace is spanned by the three column vectors of a  $T \times 3$  matrix  $\mathbf{S}$ , which is defined by

$$\mathbf{S} = (\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_T)^\top = \begin{pmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ \vdots & \vdots & \vdots \\ x_T & y_T & z_T \end{pmatrix}, \quad (6)$$

where  $\mathbf{d}_t = (x_t \ y_t \ z_t)^\top$ , ( $1 \leq t \leq T$ ) denotes the positional vector of the  $t$ -th feature point.

In the next subsection, we explain how we extract the feature points on a face. Then, we explain the way we correspond two sets of feature points extracted from two face images.

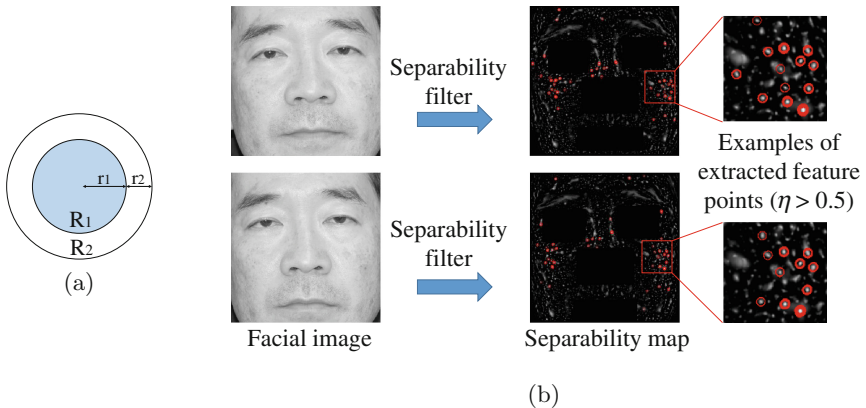
### 3.1 Feature Extraction with Separability Filter

The separability filter computes the separability  $\eta$  of two regions of an image as shown in Fig. 3(a). The separability  $\eta$  ( $0.0 \leq \eta \leq 1.0$ ) of two regions  $R_1$  and  $R_2$  in an image is calculated as follows.

$$\eta = \frac{\sigma_b^2}{\sigma_T^2}, \quad (7)$$

$$\sigma_b^2 = \frac{q_1}{q} (\bar{P}_1 - \bar{P})^2 + \frac{q_2}{q} (\bar{P}_2 - \bar{P})^2, \quad (8)$$

$$\sigma_T^2 = \frac{1}{q} \sum_{P_i \in (R_1 \cup R_2)} (P_i - \bar{P})^2 = \overline{P^2} - (\bar{P})^2, \quad (9)$$



**Fig. 3.** Circular separability filter (a) and examples of separability map and extracted feature points ( $\eta > 0.5$ ) (b). In (b), salient features such as eyes, nose, and mouth region are excluded. From another image of the same subject, many feature points are extracted in the almost same position.

where  $\sigma_b^2$  is the between-class variance,  $\sigma_T^2$  is the total variance,  $q_1, q_2$  are the numbers of pixels in  $R_1$  and  $R_2$ , respectively, and  $q = q_1 + q_2$ .  $P_i$  is the image feature at pixel  $i$ ,  $\bar{P}_1, \bar{P}_2$  are the mean values of the image features in  $R_1$  and  $R_2$ .  $\bar{P}, \bar{P}^2$  are the mean value and the mean of square of the image features from both regions.

By applying the separability filter on the whole facial image, we can obtain a separability map, where each local maximum point corresponds to the center point of a circular object, such as eyeballs, nostril or moles. The separability map and the example of the extracted feature points are shown in Fig. 3(b). Most of feature points are obtained accurately. The separability can be regarded as a measure of reliability of detection. In addition to them, the separability filter can also extract the corners of eyes and the edges of mouth stably and precisely [18]. See [16–18] for the details of the separability filter.

### 3.2 Corresponding Feature Points using Autocorrelation Matrix

Although a shape subspace can be easily generated as the column space of the matrix  $\mathbf{S}$  as mentioned previously, the generated shape subspace changes if the order of the feature points changes. Therefore, we need to correspond points between two images before calculating the similarity of the shape subspaces.

To deal with this issue, we introduce an effective method for corresponding points from two images based on autocorrelation matrices, which is an extension of the method proposed in [8]. The orthogonal projection matrix was used for corresponding feature points in [8]. In contrast, the autocorrelation matrix is used in this paper. As the autocorrelation remains more information of the feature points, we can conduct the corresponding points more stably. The method is based on the fact that if two shape subspaces are close with respect to canonical angles, the two corresponding autocorrelation matrices can also be close. In the particular case that the two shape subspaces coincide completely, the corresponding autocorrelation matrices also coincide completely after changing the order of the feature points appropriately.

Let  $\mathbf{S}_i$  and  $\mathbf{S}_j$  be shape matrices defined by Eq. (6) for two shape subspaces  $\mathcal{S}_i$  and  $\mathcal{S}_j$ , respectively. We define autocorrelation matrices of these subspaces as

$$\mathbf{A}_i = \mathbf{S}_i \mathbf{S}_i^\top, \quad \mathbf{A}_j = \mathbf{S}_j \mathbf{S}_j^\top. \quad (10)$$

By iteratively comparing rows of the autocorrelation matrices, we correspond the points in the shape subspaces as follows. Elements of autocorrelation matrices  $\mathbf{A}_i$  and  $\mathbf{A}_j$  are sorted in column-wise manner, and  $\mathbf{A}_i$  and  $\mathbf{A}_j$  are re-defined by the resultant sorted matrices. Then, for each row of  $\mathbf{A}_i$ , find the row index of  $\mathbf{A}_j$  such that the Manhattan distance from the row of  $\mathbf{A}_i$  is minimum. The pair of rows from  $\mathbf{A}_i$  and  $\mathbf{A}_j$  found by minimizing the Manhattan distance is associated. This matching procedure is repeated until the residual error defined between the two autocorrelation matrices is lower than a threshold.

### 3.3 Overall Procedure of the Proposed Framework

We summarize the procedure of the proposed framework, considering the case of classifying an input shape subspace of an unknown person into  $C$  person classes. Given a set of  $N'_c$  RGB-D images (gray image and depth image for each person).

#### Learning Phase

1. We detect a set of feature points from each gray scale image. To reduce the influence of the error in detecting feature points as described in Sect. 1, we generate multiple perturbed shape subspaces by randomly selecting feature points from the gray images repeatedly. Consequently, the number of the feature sets of each person class,  $N'_c$ , increases to  $N_c$ .
2. We set reference shape matrices  $\mathbf{S}_{c,l}$  ( $c = 1, \dots, C, l = 1, \dots, N_c$ ) in Eq. (6), and then generate autocorrelation matrices  $\mathbf{A}_{c,l}$  in Eq. (10) from them. These autocorrelation matrices are used as the references in the testing phase.

#### Testing Phase

1. We set input shape matrix  $\mathbf{S}_{input}$  from a pair of gray scale image and its corresponding depth image of an unknown person. Then, we generate the autocorrelation matrix  $\mathbf{A}_{input}$  to correspond the feature points.
2. We conduct the corresponding process between the input shape matrix  $\mathbf{S}_{input}$  and reference shape matrices  $\mathbf{S}_{c,l}$ . In this process, the input shape matrix  $\mathbf{S}_{input}$  is used as the reference. Namely, the row elements of  $\mathbf{S}_{c,l}$  are sorted based on the that of input matrix  $\mathbf{S}_{input}$ .
3. After completing the corresponding process, we calculate the similarity among each reference subspaces using Eq. (3). With the similarities, we calculate a discriminative space (its basis) on the Grassman manifold by applying the GDA to the Gram matrix, which is calculated from the similarities among all the reference subspaces.
4. We project each subspace onto the discriminative space, and calculate the distance between the projected input subspace  $\mathcal{S}_{input}$  and reference subspace  $\mathcal{S}_{c,l}$ . Finally, the input subspace is classified to the class of the nearest reference subspace.

## 4 Experimental Results and Consideration

We conducted four experiments. (1) To verify the effectiveness of the proposed framework, we evaluated its basic performance for real face data, (2) to show the validity of using the information of 3D configuration, we compared the performances of the methods with 3D configuration and 2D configuration of feature points, (3) to address the issue that the detected facial feature points include errors, we evaluate the effectiveness of using a large number of subspaces obtained by repeated random sampling from the reliable candidates of the feature points, (4) to evaluate the performance limit of the proposed method, we emulate the tough situation of distinguishing between twins.

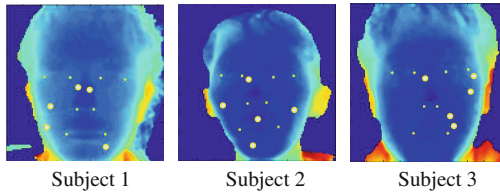
## 4.1 Experiment 1

**Experimental Condition.** We captured a pairs of a depth image and its corresponding gray scale image by using an RGB-D sensor (Microsoft Kinect v2) from 16 subjects. The sensor was about 0.6 m away from the subjects sitting on a chair.

It is desirable to extract 3D coordinates of feature points directly from the pairs. However, it was difficult to extract very unclear and weak feature points, due to the comparatively low resolution of  $512 \times 424$  pixels. To avoid the problem, we also captured gray scale images with higher resolution of  $4,000 \times 6,000$  pixels by using a single-lens reflex camera. The camera was about 1.2 m away from the subjects. We captured 5 high resolution images from subjects in one session. The subjects were asked to stand up and sit down at the beginning of each session to produce natural fluctuations in face direction and position. We repeated the session four times to totally capture 20 images.

We cropped a facial region of  $120 \times 120$  pixels from the low resolution image and then extracted eight common feature points: the four corners eyes, two nostrils, and two edges of mouth. For the high resolution image, we cropped a facial region of  $1,200 \times 1,200$  pixels from the whole image and detected micro-features by applying the separability filter to the region except that of eyes, nostrils, and a mouth as shown in Fig. 3(b). The detected micro-feature is selected by the descending order of the separability. Besides, we manually extracted the eight facial feature points. Figure 4 shows an example of the extracted feature points. Finally, based on the correspondence relation of the positions of the eight facial feature points of the high- and low- resolution images, we obtained 3D coordinates of the feature points.

In many cases, the features on facial surface are affected by factors such as genetic or age of subjects. Therefore, the number of obtainable feature points depends on the subjects. Later in this section, we will see the dependencies of the classification error on the number of extracted feature points.



**Fig. 4.** Examples of the extracted feature points. Small points are edge of eyes, nostril, and corners of the mouth, which are considered as commonly extracted feature points.

The above evaluation procedure is summarized as follows:

1. We obtain the feature points from the images taken from 16 persons. From each person, 20 images are taken, hence the total number of the collected images is 320.



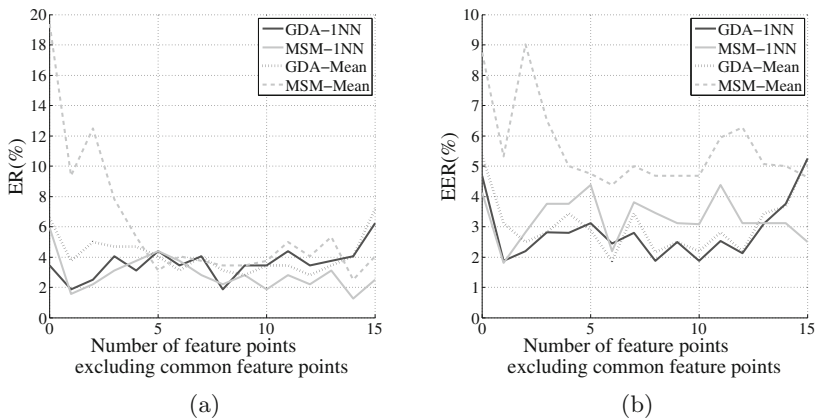
2. We make  $N = 16$  persons' shape subspaces by using the obtained feature point sets.
3. Since 20 images were taken from each person, we evaluate the classification accuracy by 20-fold cross validation. That is, we used one image from each of  $N$  different persons as test dataset, and the rest  $19 \times N$  subspaces from  $N$  persons as training datasets for constructing shape subspaces for  $N$  classes.

The MSM and GDA were used as classifiers and their accuracies were compared. We considered two simple methods in the classification process:

**1-NN:** a test sample (a set of feature points) is assigned the class for that of the closest subspace in the registered subspaces. The closeness is measured based on the Euclidean distance in the empirical kernel feature space induced by the kernel function defined in Eq. (4).

**Distance to mean:** mean points of subspaces for each class is calculated, and a test sample is assigned the class with the closest mean point. For calculating mean points of subspaces, distance is calculated based on the Euclidean distance in the feature space induced by the kernel function defined in Eq. (4).

**Results and Discussion.** Figure 5(a) and (b) show the error rates (ER) and equal error rates (EER), respectively, obtained by the MSM followed by 1NN method (MSM-1NN), the MSM followed by classification based on the distance to mean vector method (MSM-Mean), GDA followed by 1NN method (GDA-1NN), and GDA followed by the classification based on the distance to mean vector method (GDA-Mean).



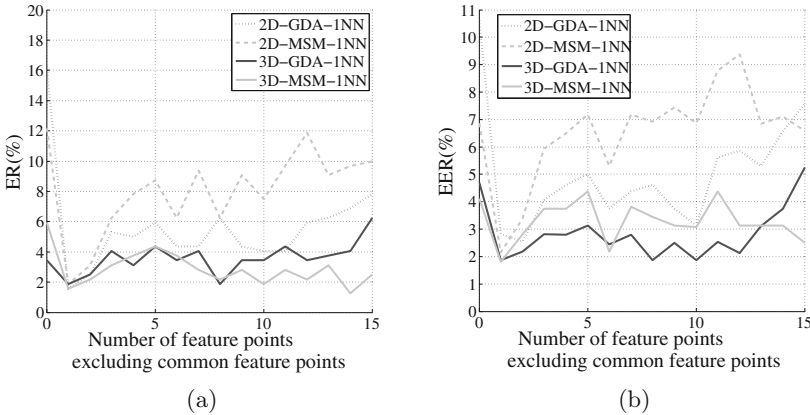
**Fig. 5.** Error rates (a) and equal error rates (b) of classification. GDA- and MSM-1NN in the legend indicate 1-nearest neighbor method in GDA and MSM, respectively. GDA- and MSM- Mean indicate classification method based on the mean of the distance from all the sample data in each class.

From Fig. 5, MSM-Mean does not show good performance compared to the other three methods. In the sense of ER, the other three methods are comparatively the same. From Fig. 5, we can see that GDA based two methods outperform the other two MSM based methods. The better performance of GDA based methods can be contributed to the fact that GDA can find the most discriminative classification axes in the same manner as LDA, which cannot be obtained by using a simple subspace based method such as the MSM. Finally, it is our surprise that the number of extracted feature points does not have a significant effect on the classification error rates based on GDA. This suggests that, the proposed method is applicable even for person with only few feature points such as moles or freckles on a facial surface.

### 4.2 Experiment 2

In our proposed framework, the shape subspace is constructed by using the 3D shape matrix defined in Eq. (6). To see the effectiveness of the use of 3D structural information for classification, we performed the following simple experiment. Here, we compared the performances of two methods with the 3D shape matrix in Eq. (6), and with 2D shape matrix which is defined by removing the  $z$  axis of the 3D shape matrix. Classification methods are the same as those used in the previous section.

The plots of ER and EER of classification results are shown in Fig. 6(a) and (b), respectively. From these results, we can see that the use of 3D structural information contributes to the improvement of the classification accuracy.



**Fig. 6.** Error rates (a) and equal error rates (b) of classification. One nearest neighbor based method with GDA and MSM using either 2D or 3D shape matrices are compared.

### 4.3 Experiment 3

In this experiment, we use the same dataset and settings as Experiment 1 to evaluate the effectiveness of using enhance references with perturbed subspaces, which is described in Sect. 3.3. For this purpose, we compared two types of the methods: “maximum separability” without the enhanced reference references, and “random sampling method” with the enhancement.

In maximum separability, 10 feature points with the 10 highest separabilities were simply selected. Thus, the same as Experiment 1, only 19 shape subspaces were used as the references for each person. In contrast, in “random sampling method”, 10 feature points were randomly selected from a set of 15 candidates of feature points with the 15 highest separabilities. By repeating this process 15 times, we obtained 15 sets feature points. In this case,  $19 \times 15$  shape subspaces were used as the reference for each person.

For the random sampling, the average of 5 trials was used as the final result. Table 1 shows the evaluation results of the both methods. In the random sampling, the ER of MSM is 1.19%, and that of GDA is 1.63%. In contrast, in the maximum separability, the ER of MSM is 1.88%, and that of GDA is 3.44%. From these results, we can see clearly the advantage of the random sampling method. Thus, we conclude that the reference enhancement by the random sampling is valid to solve the problem of the unstable detection of facial feature points.

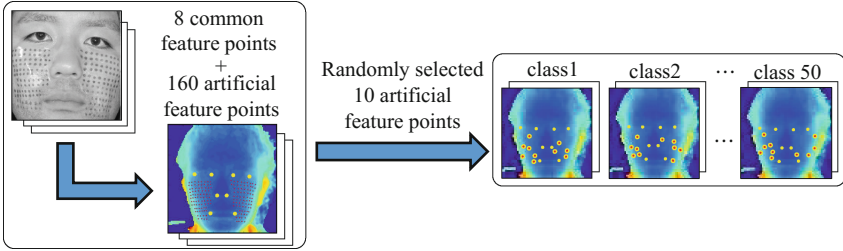
**Table 1.** Error rates and equal error rates of Experiment 3. Maximum separability in the table indicates the feature points selection method that chooses 10 feature points from the descending order of the higher separability. Random sampling indicates the method that chooses 10 feature points from 15 candidate feature points which have higher separability 15 times to increase the reference data.

Maximum separability				Random sampling			
ER (%)		EER (%)		ER (%)		EER (%)	
MSM	GDA	MSM	GDA	MSM	GDA	MSM	GDA
1.88	3.44	3.08	1.88	<b>1.19</b>	1.63	2.32	<b>1.75</b>

### 4.4 Experiment 4

In this experiment, we consider a more difficult recognition problem of classification between twins. To simulate such tough situation, we generated artificially sets of feature points on a face and explored the performance limit of the proposed method by using them. We should note that the conventional view-based methods in principle cannot distinguish them completely.

**Experimental Condition.** We took 15 almost frontal face images, which has 160 synthetic micro-feature points (Fig. 7, left), from on a subject. These synthetic feature points were generated by pasting a thin film with dots on the face of a subject. In addition to these micro-feature points, we extracted 8 common feature points, four edges of two eyes, two edges of a mouth, and two nostrils (Fig. 7). We assume that the eight facial feature points are stably extracted, and the artificially generated feature points has the ground truth locations.



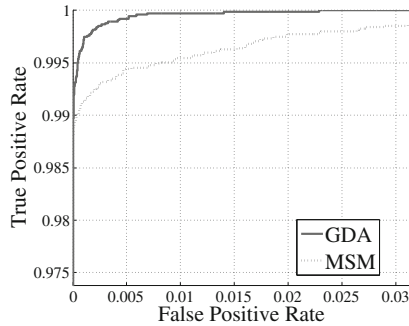
**Fig. 7.** A schematic diagram of the procedure of Experiment 4. A thin film with markers is pasted on a face to generate synthetic feature points. Then we obtain eight common feature points (the edge of eyes, nostril, and corner of the mouth) and 160 artificial feature points on the film. We select 10 artificial feature points randomly 50 times to generate artificial twins datasets.

Finally, we randomly selected 10 points from the 160 micro-features and obtained shape subspaces from them. The random selection was repeated 50 times to obtain  $N(N = 50)$  shape subspaces (Fig. 7, right). These 50 classes correspond to very similar 50 persons with exactly the same eight facial feature points and with different micro-feature points. By conducting the same procedure for each of the 15 face images, each of the 50 person can have 15 sample images.

We performed a classification experiments in the same manner as the Experiment 1. One out of 15 sample feature set was drawn from each class and constituted a test set. The rest  $14 \times 50$  sets of feature points were used as the training data, and in the same manner as in Experiment 1, the ER and EER of both MSM and GDA based classifiers were evaluated by 15-fold cross validation. The evaluation was repeated 10 times by varying the seed of random number generator for generating the 160 random feature points. The experimental results are summarized in Table 2 and the ROC curve is depicted in Fig. 8. From Table 2, we can see that the proposed method has favorable classification ability. The ER

**Table 2.** Error rates and equal error rates of simulation for Experiment 4.

ER (%)		EER (%)	
MSM	GDA	MSM	GDA
0.08	<b>0.00</b>	0.55	<b>0.19</b>



**Fig. 8.** ROC curve of recognition simulation of the situation of distinguishing between very similar appearance persons.

of the MSM is 0.08 %, while that of GDA is 0.00 %. In addition, from Fig. 8, the results become better when we use GDA as a classifier than in the case when we use the MSM. From the above results, we can conclude that our method with GDA has high classification accuracy even if the objects to discriminate have similar salient feature points such as those observed for twins, when we utilize the micro-feature points and apply GDA using 3D structural shape subspace similarity.

## 5 Concluding Remarks

In this paper, we proposed a novel framework for personal authentication based on the 3D configuration of micro-feature points, such as moles and freckles, on a facial surface. The proposed framework is instantiated by using the separability filter for feature extraction, feature point association based on the autocorrelation matrix, the notion of shape subspace, and Grassmann discriminant analysis. The usefulness and validity of the proposed framework are examined through a set of simple experiments. From the experimental results, we can confirm that RGB-D images contain features that can be used to form the shape subspace, and GDA offers highly accurate classification results.

The aim of this paper is to introduce a novel framework for person authentication, hence there remains a number of problems to be addressed. Making the feature extraction procedure fully automatic is the most important problem to be addressed. Also, our future works include large scale experiments to evaluate the proposed method, and comparison to state-of-the-art methods in the literature. The reproducibility of feature extraction is not guaranteed, even for the same person. Sensitivity of the proposed method to the feature points extraction should be also examined in more detail, and the way to stably extract feature points from the same person should be also investigated.

**Acknowledgement.** Part of this work is supported by JSPS KAKENHI Grant Number 25870811.

## References

1. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 586–591 (1991)
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997)
3. Fukui, K., Yamaguchi, O.: Face recognition using multi-viewpoint patterns for robot vision. In: 11th International Symposium of Robotics Research, pp. 192–201 (2005)
4. Gross, R., Matthews, I., Baker, S.: Appearance-based face recognition and light-fields. *IEEE Trans. Pattern Anal. Mach. Intell.* **26**(4), 449–465 (2004)
5. Heseltine, T., Pears, N., Austin, J., Chen, Z.: Face recognition: a comparison of appearance-based approaches. In: VIIth Digital Image Computing: Techniques and Applications, pp. 59–68 (2003)
6. Xianwei, L., Haiyang, Z.: A survey of face recognition methods. In: The 2nd International Conference on Computer Science and Electronics Engineering, pp. 1405–1407 (2013)
7. Li, S.Z., Jain, A.K.: *Handbook of Face Recognition*. Springer New York, Inc., Secaucus (2005)
8. Igarashi, Y., Fukui, K.: 3D object recognition based on canonical angles between shape subspaces. *ACCV 2010. LNCS*, vol. 6495, pp. 580–591. Springer, Heidelberg (2010)
9. Tomasi, C., Kanade, T.: Shape and motion from image streams under orthography: a factorization method. *Int. J. Comput. Vis.* **9**(2), 137–154 (1992)
10. Hamm, J., Lee, D.D.: Grassmann discriminant analysis: a unifying view on subspace-based learning. In: The 25th International Conference on Machine Learning, pp. 376–383 (2008)
11. Chatelin, F.: *Eigenvalues of Matrices*. Wiley, New York (1993)
12. Mattila, P.: *Geometry of Sets and Measures in Euclidean Spaces*. Cambridge University Press, New York (1995)
13. Fisher, R.A.: The use of multiple measurements in taxonomic problems. *Ann. Eugenics* **7**, 179–188 (1936)
14. Shawe-Taylor, J., Cristianini, N.: *Kernel Methods for Pattern Analysis*. Cambridge University Press, New York (2004)
15. Schölkopf, B., Mika, S., Burges, C.J.C., Knirsch, P., Müller, K.R., Rätsch, G., Smola, A.J.: Input space versus feature space in kernel-based methods. *IEEE Trans. Neural Netw.* **10**(5), 1000–1017 (1999)
16. Ohkawa, Y., Suryanto, C.H., Fukui, K.: Fast combined separability filter for detecting circular objects. In: 12th IAPR Conference on Machine Vision Applications, pp. 99–103 (2011)
17. Fukui, K., Yamaguchi, O.: Facial feature point extraction method based on combination of shape extraction and pattern matching. *Syst. Comput. Jpn.* **29**(6), 49–58 (1998)
18. Yuasa, M., Yamaguchi, O.: Real-time face blending by automatic facial feature point detection. In: 8th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2008), pp. 1–6 (2008)