

## PAPER

# Class-Distance-Based Discriminant Analysis and Its Application to Supervised Automatic Age Estimation

Tetsuji OGAWA<sup>†a)</sup>, Member, Kazuya UEKI<sup>††</sup>, Nonmember, and Tetsunori KOBAYASHI<sup>†††</sup>, Member

**SUMMARY** We propose a novel method of supervised feature projection called class-distance-based discriminant analysis (CDDA), which is suitable for automatic age estimation (AAE) from facial images. Most methods of supervised feature projection, e.g., Fisher discriminant analysis (FDA) and local Fisher discriminant analysis (LFDA), focus on determining whether two samples belong to the same class (i.e., the same age in AAE) or not. Even if an estimated age is not consistent with the correct age in AAE systems, i.e., the AAE system induces error, smaller errors are better. To treat such characteristics in AAE, CDDA determines between-class separability according to the class distance (i.e., difference in ages); two samples with similar ages are imposed to be close and those with spaced ages are imposed to be far apart. Furthermore, we propose an extension of CDDA called local CDDA (LCDDA), which aims at handling multimodality in samples. Experimental results revealed that CDDA and LCDDA could extract more discriminative features than FDA and LFDA.

**key words:** FDA, LFDA, CDDA, LCDDA, dimensionality reduction, automatic age estimation

## 1. Introduction

Many researchers have attempted to develop techniques of estimating individual characteristics of humans, such as gender and age [1]–[5]; these techniques are useful in marketing surveys on consumer ages. We have developed automatic age estimation (AAE) systems that can estimate age from facial images [1], [2]. Appearance-based AAE requires dimensionality reduction that extracts information that is useful for estimating ages.

Fisher discriminant analysis (FDA) [6] has been frequently applied to supervised dimensionality reduction. The projection matrix in FDA is estimated such that the between-class scatter is maximized while the within-class scatter is minimized. Many researchers have studied manifold learning in recent years, which aims at preserving the neighborhood structure of input samples in feature projection, e.g., locally linear embedding (LLE) [7], [8], Isomap [9], and locality preserving projection (LPP) [10]. LPP in some tasks outperformed FDA [10], [11]. Many researchers have attempted to extend LPP [12]–[20]. Local Fisher discriminant analysis (LFDA) [16] has the advantages of both FDA and

LPP and outperforms FDA and LPP. In LFDA, data pairs of different classes are imposed to be far apart and nearby data pairs of the same class are imposed to be close. In addition, data pairs of the same class but which are far apart are not imposed to be close. As a result, LFDA has one very useful property that the multimodality of samples is not lost by dimensionality reduction.

FDA and LFDA, however, basically aim at projecting samples in an input space such that the samples of each class are separated in a projected feature space. Here, class distances, i.e., the difference between the class labels of two samples, are not explicitly considered in feature projection. Therefore, these methods are suitable for systems whose performance can only be measured in terms of accuracy of estimation, which indicates whether an estimate obtained from the system belongs to the correct class or not. In contrast, ages, which are considered as class labels in AAE systems, consist of integer values with order relations. This indicates that AAE systems should be evaluated in terms of errors, which are represented by the difference between the estimated and real ages. Smaller errors would be better in AAE systems, and these systems would be worse with an increase in errors. The above discussion indicates that for AAE systems, between-class separability should be determined according to the class distance (i.e., age difference) between two samples.

We propose a method of feature projection called class-distance-based discriminant analysis (CDDA) in this paper, which determines between-class separability according to the class distance. In CDDA, two samples with similar class labels (i.e., close ages) are projected such that they are close in a projected space, while two samples with distant class labels are projected such that they are far apart in a projected space. These characteristics make the performance of the system more intuitive for humans. Furthermore, we extended CDDA such that the multimodality of samples was not lost due to feature projection, as it was with LFDA. We called this method local class-distance-based discriminant analysis (LCDDA). These methods can be used for improving the performance of AAE systems because they can deal with class distances and class continuity at the feature-extraction stage.

The rest of this paper is organized as follows. Section 2 reviews existing methods of linear feature projection. Section 3 describes details on CDDA and LCDDA. Section 4 discusses the effectiveness of CDDA and LCDDA, which were experimentally investigated in terms of mean absolute

Manuscript received December 13, 2010.

Manuscript revised April 18, 2011.

<sup>†</sup>The author is with the Waseda Institute for Advanced Study, Waseda University, Tokyo, 169–8555 Japan.

<sup>††</sup>The author is with the NEC Soft, Ltd., Tokyo, 136–8606 Japan.

<sup>†††</sup>The author is with the Department of Computer Science, Waseda University, Tokyo, 169–8555 Japan.

a) E-mail: ogawa@pcl.cs.waseda.ac.jp

DOI: 10.1587/transinf.E94.D.1683

errors (MAEs) and cumulative scores. Finally, concluding remarks are made in Sect. 5.

## 2. Linear Feature Projection

A two-dimensional image with a height of  $h$  pixels and a width of  $w$  pixels is generally treated as an  $h \times w$  dimensional vector  $\mathbf{x}_i \in \mathbb{R}^d$  ( $d = h \times w$ ). This vector, however, has high dimensionality and redundant information. Therefore, the dimensionality of this vector has to be reduced and only informative features that are required for age estimation should be extracted. Existing methods of feature projection such as FDA [6], LPP [10], and LFDA [16] are reviewed in this section.

### 2.1 Formulation

Assume that a set of observed samples  $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^N$ ,  $\mathbf{x}_i \in \mathbb{R}^d$  and their class labels  $\mathbf{Y} = \{y_i\}_{i=1}^N$  are given. Using a  $d' \times d$  matrix  $\Phi$ , lower-dimensional feature samples  $\{\mathbf{z}_i\}_{i=1}^N$ ,  $\mathbf{z}_i \in \mathbb{R}^{d'}$ ,  $d' < d$  are computed with the following projection:

$$\mathbf{z}_i = \Phi^T \mathbf{x}_i \quad (1)$$

Our aim was to determine the low-dimensional projection matrix  $\Phi$ .

### 2.2 Fisher Discriminant Analysis (FDA)

Let  $\mathbf{S}^{(w)}$  be the within-class scatter matrix and  $\mathbf{S}^{(b)}$  be the between-class scatter matrix defined by

$$\mathbf{S}^{(w)} = \sum_{i=1}^l \sum_{j:y_j=i} (\mathbf{x}_j - \boldsymbol{\mu}_i)(\mathbf{x}_j - \boldsymbol{\mu}_i)^T \quad (2)$$

$$\mathbf{S}^{(b)} = \sum_{i=1}^l N_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \quad (3)$$

where  $N_i$  denotes the number of samples in class  $i$ ,  $\boldsymbol{\mu}_i$  denotes the mean of samples in class  $i$ , and  $\boldsymbol{\mu}$  denotes the mean of all samples given as:

$$\boldsymbol{\mu}_i = \frac{1}{N_i} \sum_{j:y_j=i} \mathbf{x}_j \quad (4)$$

$$\boldsymbol{\mu} = \frac{1}{N} \sum_{j=1}^N \mathbf{x}_j \quad (5)$$

FDA projection matrix  $\Phi_{\text{FDA}}$  is estimated so that the between-class scatter is maximized while the within-class scatter is minimized as:

$$\Phi_{\text{FDA}} = \arg \max_{\Phi} \left( \text{tr} \left[ \frac{\Phi^T \mathbf{S}^{(b)} \Phi}{\Phi^T \mathbf{S}^{(w)} \Phi} \right] \right) \quad (6)$$

Projection matrix  $\Phi$  is computed by solving a generalized eigenvalue problem as:

$$\mathbf{S}^{(b)} \Phi = \lambda \mathbf{S}^{(w)} \Phi \quad (7)$$

Here, the solution to FDA,  $\Phi_{\text{FDA}}$ , is given as:

$$\Phi_{\text{FDA}} = (\phi_1, \phi_2, \dots, \phi_l), \quad (8)$$

where  $\Phi_{\text{FDA}}$  denotes a  $d \times l$  matrix and  $\{\phi_i\}_{i=1}^l$  denote the eigenvectors associated with eigenvalues  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l$ .

Here,  $\mathbf{S}^{(w)}$  and  $\mathbf{S}^{(b)}$  are expressed as pairwise expressions, as [16]:

$$\mathbf{S}^{(w)} = \frac{1}{2} \sum_{i,j=1}^N A_{i,j}^{(w)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (9)$$

$$\mathbf{S}^{(b)} = \frac{1}{2} \sum_{i,j=1}^N A_{i,j}^{(b)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (10)$$

where

$$A_{i,j}^{(w)} = \begin{cases} 1/N_c & \text{if } y_i = y_j = c, \\ 0 & \text{if } y_i \neq y_j, \end{cases} \quad (11)$$

$$A_{i,j}^{(b)} = \begin{cases} 1/N - 1/N_c & \text{if } y_i = y_j = c, \\ 1/N & \text{if } y_i \neq y_j. \end{cases} \quad (12)$$

### 2.3 Locality Preserving Projection (LPP)

LPP preserves the geometrical structure of neighborhood samples before and after projection; nearby samples in the input space are embedded close in the projected (i.e., low-dimensional) feature space. LPP projection matrix  $\Phi_{\text{LPP}}$  is estimated as:

$$\Phi_{\text{LPP}} = \arg \min_{\Phi} \frac{1}{2} \sum_{i,j}^n A_{i,j} \|\Phi^T \mathbf{x}_i - \Phi^T \mathbf{x}_j\|^2 \quad (13)$$

$$\text{subject to } \Phi^T \mathbf{X} \mathbf{D} \mathbf{X}^T \Phi = \mathbf{I}, \quad (14)$$

where  $\mathbf{A}$  denotes an affinity matrix, the  $(i, j)$ -th component of which reflects the locality (i.e., geometrical distance) between  $\mathbf{x}_i$  and  $\mathbf{x}_j$ . By using the constraint described in Eq. (14), we can avoid trivial solution  $\Phi = \mathbf{0}$ .

LPP reduces to the following generalized eigenvalue problem.

$$\mathbf{X} \mathbf{L} \mathbf{X}^T \Phi_{\text{LPP}} = \lambda \mathbf{X} \mathbf{D} \mathbf{X}^T \Phi_{\text{LPP}} \quad (15)$$

Here, the solution to LPP,  $\Phi_{\text{LPP}}$ , is given as:

$$\Phi_{\text{LPP}} = (\phi_1, \phi_2, \dots, \phi_l), \quad (16)$$

where  $\Phi_{\text{LPP}}$  denotes a  $d \times l$  matrix and  $\{\phi_i\}_{i=1}^l$  denote the eigenvectors associated with the eigenvalues  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_l$ . Here,  $\mathbf{L} = \mathbf{D} - \mathbf{A}$  is a Laplacian matrix, and  $\mathbf{D}$  is a diagonal matrix given by

$$D_{i,i} = \begin{cases} \sum_j A_{i,j}, & i = j \\ 0, & i \neq j, \end{cases} \quad (17)$$

whose  $(i, i)$ -th component  $D_{i,i}$  represents the locality around  $\mathbf{x}_i$ ;  $\mathbf{x}_i$  becomes more informative with an increase in  $D_{i,i}$ .

$\mathbf{A}$  is arbitrarily defined such that  $\{\mathbf{z}_i\}_{i=1}^N$  would have the same structure as  $\{\mathbf{x}_i\}_{i=1}^N$  for neighborhood samples. In [10],

the affinity matrix is defined using a heat kernel as:

$$A_{i,j} = \begin{cases} \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma^2}\right), \\ \text{if } \mathbf{x}_i \in NN_j^K \text{ or } \mathbf{x}_j \in NN_i^K \\ 0, \text{ otherwise} \end{cases} \quad (18)$$

where  $\sigma$  denotes a tuning parameter that controls the decay of affinity and  $NN_i^K$  denotes a set of  $k$ -nearest neighbor samples of  $\mathbf{x}_i$ .

Scaling parameter  $\sigma$  in Eq. (18) cannot treat the different scales for each sample. A local scaling method [21] is proposed to equalize the density around each sample. The affinity weight is described as:

$$A_{i,j} = \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_i \sigma_j}\right), \quad (19)$$

where  $\sigma_i$  denotes the local scaling of input samples around  $\mathbf{x}_i$  computed as:

$$\sigma_i = \|\mathbf{x}_i - \mathbf{x}_i^{(k)}\| \quad (20)$$

Here,  $\mathbf{x}_i^{(k)}$  is the  $k$ -th nearest neighbor of  $\mathbf{x}_i$ , such that for all  $\mathbf{x}_j, y_i = y_j$ .

## 2.4 Local Fisher Discriminant Analysis (LFDA)

In LFDA [16], two adjacent samples of the same class are imposed to be close, and those of different classes are imposed to be far apart. In addition, two samples of the same class but distant in the input space are not imposed to be close. As a result, the application of LFDA ensures that the multimodality of samples is not lost by dimensionality reduction.

LFDA projection matrix  $\Phi_{\text{LFDA}}$  is computed as:

$$\Phi_{\text{LFDA}} = \arg \max_{\Phi} \left( \text{tr} \left[ \frac{\Phi^T \bar{S}^{(b)} \Phi}{\Phi^T \bar{S}^{(w)} \Phi} \right] \right), \quad (21)$$

where  $\bar{S}^{(w)}$  and  $\bar{S}^{(b)}$  correspond to the local within- and between-class scatter matrices. These matrices are expressed as pairwise forms as:

$$\bar{S}^{(w)} = \frac{1}{2} \sum_{i,j=1}^N \bar{A}_{i,j}^{(w)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (22)$$

$$\bar{S}^{(b)} = \frac{1}{2} \sum_{i,j=1}^N \bar{A}_{i,j}^{(b)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \quad (23)$$

where

$$\bar{A}_{i,j}^{(w)} = \begin{cases} A_{i,j} \cdot \frac{1}{N_c} & \text{if } y_i = y_j = c, \\ 0 & \text{if } y_i \neq y_j, \end{cases} \quad (24)$$

$$\bar{A}_{i,j}^{(b)} = \begin{cases} A_{i,j} \left( \frac{1}{N} - \frac{1}{N_c} \right) & \text{if } y_i = y_j = c, \\ \frac{1}{N} & \text{if } y_i \neq y_j. \end{cases} \quad (25)$$

In [16], affinity weight  $A_{i,j}$  was computed using the local scaling method described in Eq. (19).  $\Phi_{\text{LFDA}}$  can be optimized with the generalized eigenvalue problem as can FDA.

## 3. Class-Distance-Based Discriminant Analysis (CDDA)

Data pairs of the same and different classes are basically imposed to be close and far apart in FDA and LFDA. When evaluating AAE systems, even if the system induces errors, smaller errors are better, and the system would worsen with an increase in errors. To introduce this property to feature projection, we attempted to project two samples with similar class labels (i.e., close ages) such that they would be close in the projected feature space, while projecting two samples with class labels that were far apart (i.e., apart ages) such that they would be far apart in the projected feature space. We called this discriminant analysis taking into consideration the differences in class labels (i.e., differences in ages) of two samples as “class-distance-based discriminant analysis” (CDDA). From this discussion, CDDA has a human-intuitive property where class continuity, which is the order relation in ages, is introduced by dimensionality reduction; this method can increase between-class separability in achieving class continuity.

Furthermore, we attempt to extend CDDA such that class continuity of projected feature samples can be more accurately achieved and the multimodality of input samples can be preserved through dimensionality reduction. We call this extended method “local class-distance-based discriminant analysis” (LCDDA). The property of LCDDA can be realized by introducing the weight applied to LFDA, which is described in Eq. (19), to the affinity matrix for CDDA; this weight ensures that two samples that have close class labels but are far apart in the input space are not forced to be close [16].

### 3.1 Formulation of CDDA

In CDDA, an aggregated scatter matrix  $\bar{S}^{(a)}$  and a diffuse scatter matrix  $\bar{S}^{(d)}$  are defined as:

$$\bar{S}^{(a)} = \frac{1}{2} \sum_{i,j=1}^N A_{i,j}^{(a)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (26)$$

$$\bar{S}^{(d)} = \frac{1}{2} \sum_{i,j=1}^N A_{i,j}^{(d)} (\mathbf{x}_i - \mathbf{x}_j)(\mathbf{x}_i - \mathbf{x}_j)^T \quad (27)$$

Here,  $A_{i,j}^{(a)}$  denotes the affinity matrix that is taken to be equal to one if a data pair has the same age and zero if the age difference between the data pair is large.  $A_{i,j}^{(d)}$  denotes the affinity matrix that is taken to be equal to zero if a data pair has the same age and one if the age difference between the data pair is large. In the present study, we define  $A_{i,j}^{(a)}$  and  $A_{i,j}^{(d)}$  such that these matrices satisfy the aforementioned constraints as

$$A_{i,j}^{(a)} = \exp\left(-\frac{|y_i - y_j|^2}{t^2}\right) \quad (28)$$

$$A_{i,j}^{(d)} = 1 - \exp\left(-\frac{|y_i - y_j|^2}{t^2}\right) \quad (29)$$

Using  $\tilde{\mathbf{S}}^{(a)}$  and  $\tilde{\mathbf{S}}^{(d)}$ , CDDA projection matrix  $\Phi_{\text{CDDA}}$  is computed as:

$$\Phi_{\text{CDDA}} = \arg \max_{\Phi} \left( \text{tr} \left[ \frac{\Phi^T \tilde{\mathbf{S}}^{(d)} \Phi}{\Phi^T \tilde{\mathbf{S}}^{(a)} \Phi} \right] \right) \quad (30)$$

The optimization of  $\Phi_{\text{CDDA}}$  reduces to the same generalized eigenvalue problem as FDA and LFDA.

### 3.2 Formulation of LCDDA

LCDDA decreases the affinity weight for two samples with close class labels but is far apart in input space while CDDA gives a large affinity weight for two samples with close class labels, irrespective of the distance of input samples. The affinity weight for the aggregated scatter matrix (i.e.,  $A_{i,j}^{(a)}$ ) is defined for that purpose using not only the class-distance of the two samples but also the distance of two input samples. Here,  $A_{i,j}^{(a)}$  and  $A_{i,j}^{(d)}$  are defined as:

$$A_{i,j}^{(a)} = \exp\left(-\frac{|y_i - y_j|^2}{t^2}\right) \cdot \exp\left(-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{\sigma_i \cdot \sigma_j}\right) \quad (31)$$

$$A_{i,j}^{(d)} = 1 - \exp\left(-\frac{|y_i - y_j|^2}{t^2}\right), \quad (32)$$

where  $\sigma_i$  denotes the local scaling, which is the distance between  $\mathbf{x}_i$  and the  $k$ -th nearest neighbor sample of  $\mathbf{x}_i$ . Unlike LFDA, the  $k$ -th nearest neighbor sample is not limited to that of the same class as  $\mathbf{x}_i$ .

By computing the scatter matrices in Eqs. (26) and (27) by using the affinity matrices in Eqs. (31) and (32), LCDDA projection matrix  $\Phi_{\text{LCDDA}}$  is computed as:

$$\Phi_{\text{LCDDA}} = \arg \max_{\Phi} \left( \text{tr} \left[ \frac{\Phi^T \tilde{\mathbf{S}}^{(d)} \Phi}{\Phi^T \tilde{\mathbf{S}}^{(a)} \Phi} \right] \right) \quad (33)$$

## 4. Experiments on Age Estimation

We carried out experimental comparisons using AAE. We evaluated four methods of feature projection of FDA, LFDA, CDDA, and LCDDA. We set parameter  $k$  in Eq. (20) to five for LFDA and LCDDA, which yielded the best performance.

### 4.1 Facial Image Resources

We used a large-scale facial-image database to automatically estimate human characteristics, which was called the Waseda Human Computer Interaction Technology Database (WIT-DB), developed by Waseda University and NEC Soft, Ltd. It comprises frontal facial images of Japanese individuals taken in various environments (e.g., companies and universities). In addition, the database includes various lighting

**Table 1** Number of facial images we used.

age	female	male
-10	846	1,486
11-15	209	212
16-20	641	1,316
21-25	329	580
26-30	403	640
31-35	404	700
36-40	400	676
41-45	437	722
46-50	317	635
51-55	431	785
56-60	436	610
61-	613	1,130
total	5,466	9,492

conditions without occlusion. Most of the images have neutral facial expressions and some images show people smiling. The image size was  $64 \times 48$  pixels. The actual age of subjects ranged from 5 to 70 years, with a step size of one year (66 classes). Table 1 lists the number of facial images we used.

### 4.2 Classifiers

We used  $k$ -nearest neighbor ( $k$ NN) classifiers. We estimated the age of the subjects as the average of the class labels of  $k$ NN samples in this case. We set  $k$  to eight in this study.

We evaluated  $k$ NN-based AAE systems using two-fold cross validation tests for each gender. In the images of females, one and the other fold comprised 2,455 and 3,011 images, respectively. In the images of males, one and the other fold comprised 4,636 and 4,856 images, respectively.

### 4.3 Evaluation Measures

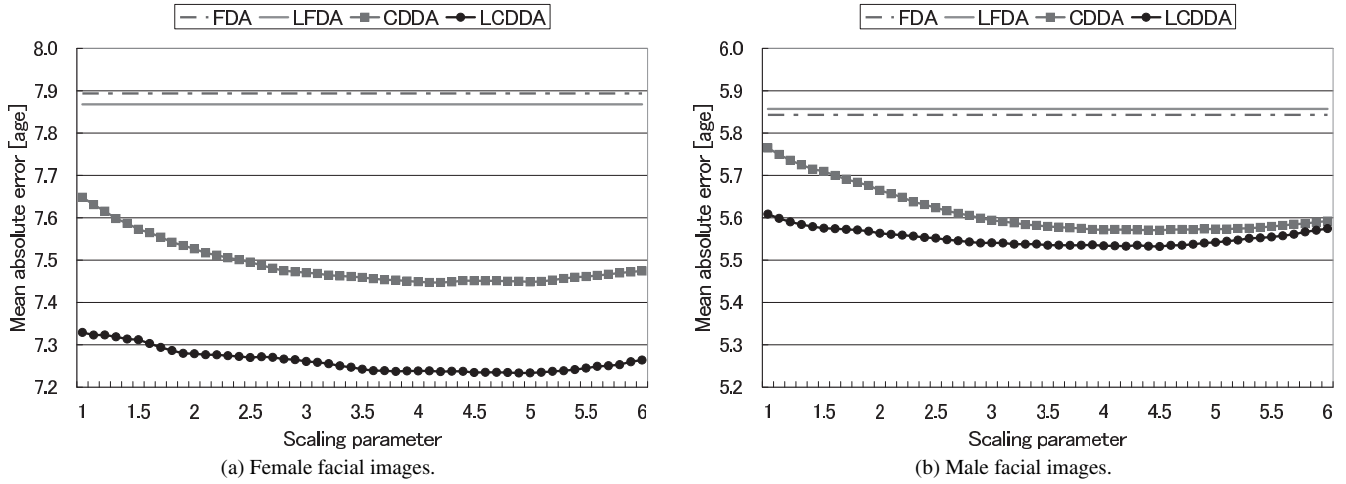
We used two objective evaluation measures, mean absolute errors (MAEs) [3], [4], [22] and cumulative scores [4], [5], which are frequently applied in studies on AAE.

#### 4.3.1 Mean Absolute Error

The MAE, which is the average of errors for all ages, was computed as:

$$\text{MAE} = \frac{1}{N_t} \sum_{i=1}^{N_t} |a_i - \hat{a}_i|, \quad (34)$$

where  $N_t$  denotes the number of test samples we used and  $a_i$  and  $\hat{a}_i$  correspond to the real and estimated ages of the  $i$ -th sample. Note that a small MAE indicates excellent performance in feature projection systems. Greatly overestimating the age of a 20-year-old person to be 45 years is more serious than slightly overestimating it to be 25 years in AAE systems. MAE can prove to be useful in such situations. MAE increases with an increase in the number of samples that induce large errors in the estimated age.



**Fig. 1** MAE as function of scaling parameter in affinity matrices of CDDA and LCDDA.

#### 4.3.2 Cumulative Score

The cumulative score was computed as:

$$\text{CumScore}(l) = \frac{N_{\leq l}}{N_t} \times 100 (\%), \quad (35)$$

where  $N_t$  denotes the number of test samples we used and  $N_{\leq l}$  denotes the number of test samples whose error levels (i.e., difference between real and estimated ages) are lower than  $l$ . The cumulative scores in this study were measured as a function of the error levels. The accuracy of AAE can be evaluated in detail by using the cumulative score as compared to the MAE. In addition, this criterion is useful in investigating the relation between the accuracy of systems and tolerance against errors.

It should be noted that a better cumulative score indicates that samples in the projected feature space are distributed with more accurate order relation in ages (i.e., class continuity) because the score is computed by using the estimated age given by the  $k$ -NN-based classifiers, which use the geometrical structure of neighborhood samples in the projected feature space.

#### 4.4 Experimental Results

Figure 1 plots MAE as a function of scaling parameter  $t$  used in Eqs. (28), (29), (31), and (32). This figure shows the average MAEs for dimensionality, ranging from 1 to 50, of the projected samples. The AAE systems based on CDDA and LCDDA yielded better MAEs than the conventional FDA- and LFDA-based AAE systems, irrespective of the scaling parameters. In addition, the LCDDA-based system yielded the best MAEs of the four feature projection methods, irrespective of the scaling parameters. Here, we determined optimal scaling parameter  $t$  of CDDA and LCDDA from this result; the CDDA-based system performed the best for  $t = 4.1$  for female facial images and 4.5 for those of males. The LCDDA-based system performed the best for  $t = 4.9$

for female facial images and 4.5 for those of males. In the rest of the experiments, these values were set to the scaling parameter to evaluate the CDDA- and LCDDA-based systems.

Figure 2 plots MAE as a function of the dimensionality of the projected samples for female and male facial images. This figure shows that the CDDA- and LCDDA-based systems yielded better MAEs than the conventional FDA- and LFDA-based systems, irrespective of the dimensionality of the projected samples. The MAEs from the LCDDA-based system exceeded that from the CDDA-based system especially for the female facial images, irrespective of dimensionality.

Table 2 lists the best MAEs and corresponding dimensionality of the projected samples for all methods of feature projection. The results indicate that the LCDDA-based system performed the best; it reduced the MAE of the FDA- and LFDA-based systems by 0.42 points for the former and 0.43 points for the latter, which were averaged for gender. In addition, the LCDDA-based system reduced the MAE of the CDDA-based system by 0.12 points.

Figure 3 plots cumulative scores as a function of the error levels for ages ranging from 0 to 30 years for the female and male facial images. The dimensionality of the projected feature samples used in this experiment were optimized in the MAE-based evaluation, which are listed in Table 2. The CDDA- and LCDDA-based systems yielded similar and slightly better scores than the FDA- and LFDA-based systems, irrespective of the gender or error level. The cumulative score was 60% in the CDDA- and LCDDA-based systems when the tolerance against error in age estimation was 7 years for the female facial images and 5 years for the male facial images, and 90% for the male and female facial images when the tolerance against error was 18 years for the former and 13 years for the latter.

As described in 4.3.2, a better cumulative score indicates that more accurate class continuity of samples is achieved in the projected feature space. Therefore, the results in Fig. 3 show that CDDA and LCDDA can more ac-

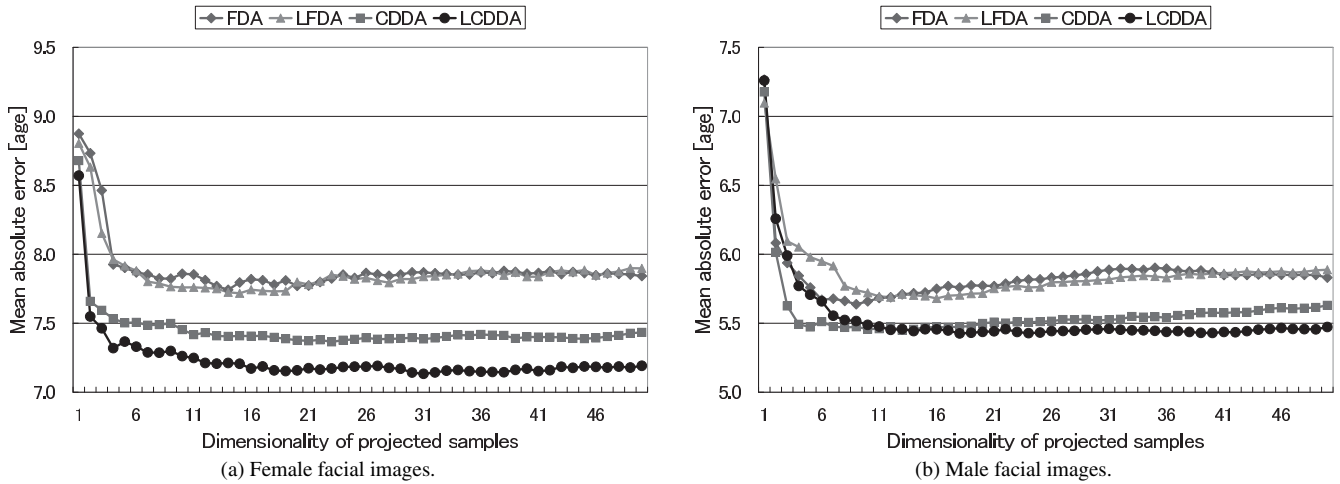


Fig. 2 MAE as function of dimensionality of projected samples.

**Table 2** Best MAE for all projection methods. “Dimension” and “Parameters” denote corresponding dimensionalities of feature vectors and tuning parameters (i.e.,  $k$  in Eq. (20) and  $t$  in Eqs. (28), (29), (31), and (32)).

Method	Female			Male		
	MAE	Dimension	Parameters	MAE	Dimension	Parameters
FDA	7.74	14		5.64	9	
LFDA	7.72	15	$k = 5$	5.68	16	$k = 5$
CDDA	7.35	23	$t = 4.1$	5.44	14	$t = 4.5$
LCDDA	7.13	36	$k = 5, t = 4.9$	5.42	18	$k = 5, t = 4.5$

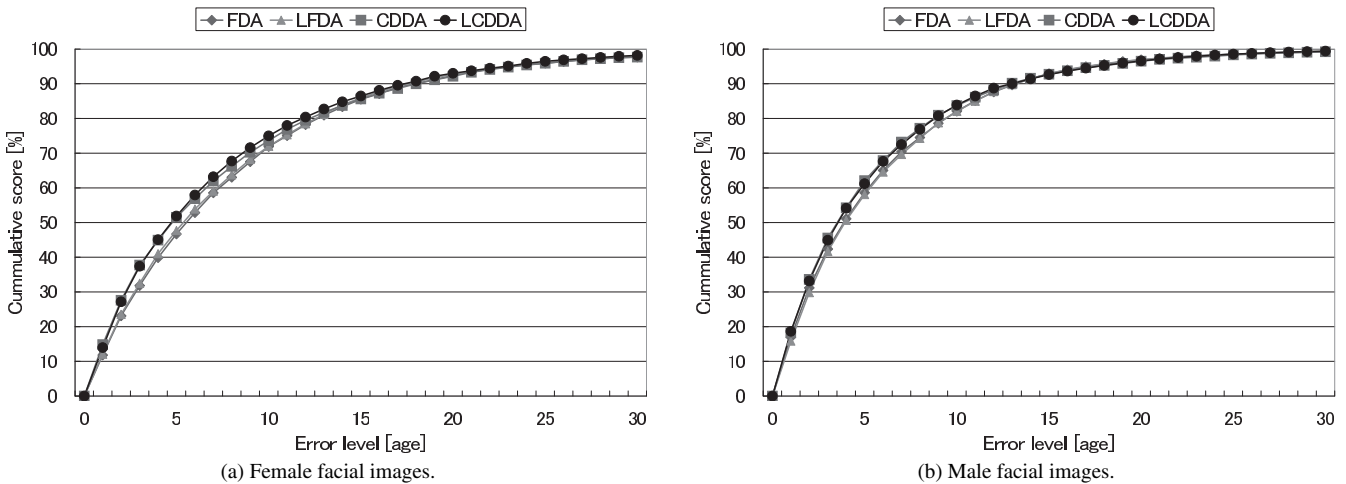


Fig. 3 Cumulative score as function of error level for ages ranging from 0 to 30 years.

curately achieve class continuity of the projected feature samples than FDA and LFDA. In addition, since LCDDA gave approximately the same cumulative score as CDDA, as shown in Fig. 3, CDDA and LCDDA may give the same performance in preserving class continuity. On the other hand, LCDDA improved the mean absolute errors compared to CDDA, as shown in Figs. 1 and 2. This improvement may have been derived from the difference between CDDA and LCDDA; the difference is represented by the effect of the weight in Eq. (19), which preserves multimodality through dimensionality reduction. Therefore, LCDDA can more precisely preserve the multimodality of input samples than

CDDA.

These experimental results indicate that CDDA and LCDDA had the best discrimination of the conventional methods, since they took into account not only the within- and between-class scatter but also the order relation of class labels.

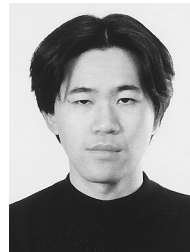
## 5. Conclusion

We proposed a new method of supervised feature projection called CDDA and applied it to AAE. We also proposed an extension of CDDA called LCDDA. These methods ex-

licitly utilized class distances of two samples and could extract discriminative features. The experimental results revealed that CDDA and LCDDA outperformed FDA and LFDA in terms of MAE and the cumulative score. The results showed that CDDA and LCDDA could achieve slightly more accurate class continuity for the projected feature samples than FDA and LFDA. In addition, LCDDA may more precisely preserve multimodality of input samples through feature projection than CDDA. Therefore, the proposed methods contribute to achieving human-intuitive and high-performance AAE systems.

## References

- [1] K. Ueki and T. Kobayashi, "Fusion-based age-group classification method using multiple two-dimensional feature extraction algorithms," *IEICE Trans. Inf. & Syst.*, vol.E90-D, no.6, pp.923–934, June 2007.
- [2] K. Ueki, M. Miya, T. Ogawa, and T. Kobayashi, "Class distance weighted locality preserving projection for automatic age estimation," *Proc. BTAS*, Oct. 2008.
- [3] A. Lanitic, C. Taylor, and T. Coates, "Toward automatic simulation of aging effects on face images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.24, no.4, pp.442–455, 2002.
- [4] X. Geng, Z. Zhou, Y. Zhou, G. Li, and H. Dai, "Learning from facial aging patterns for automatic age estimation," *Proc. ACM Multimedia*, pp.307–316, Oct. 2006.
- [5] Y. Fu, Y. Xu, and T.S. Huang, "Estimating human age by manifold analysis of face pictures and regression on aging features," *Proc. ICME2007*, pp.1383–1385, July 2007.
- [6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Academic Press, Boston, 1990.
- [7] S. Rowis and L. Soul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol.290, no.5500, pp.2323–2326, Dec. 2000.
- [8] L. Soul and S. Rowis, "Think globally fit locally; unsupervised learning of low dimensional manifolds," *JMLR*, vol.4, pp.119–155, 2003.
- [9] J.B. Tenenbaum, V. de Silva, and J.C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol.290, no.5500, pp.2319–2323, Dec. 2000.
- [10] X. He and P. Niyogi, "Locality preserving projections," *Advances in Neural Information Processing Systems*, vol.16, pp.153–160, 2004.
- [11] X. He, S. Yan, Y. Hu, P. Niyogi, and H. Zhang, "Face recognition using laplacianfaces," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.27, no.3, pp.328–340, March 2005.
- [12] C. Jian, L. Qingshan, L. Hanqing, and C. Yen-Wei, "Supervised kernel locality preserving projections for face recognition," *Neurocomputing*, vol.67, no.8, pp.443–449, 2005.
- [13] W. Yum, X. Teng, and C. Liu, "Discriminant locality preserving projections: A new method to face representation and recognition," *Proc. 2nd Joint IEEE International Workshop on VS-PETS*, pp.201–207, Oct. 2005.
- [14] D. Cai, X. Han, J. Han, and H.J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE Trans. Image Process.*, vol.15, no.11, pp.3609–3614, Nov. 2006.
- [15] G. Feng, D. Hu, D. Zhang, and Z. Zhou, "An alternative formulation of kernel LPP with application to image recognition," *Neurocomputing*, vol.69, no.13–15, pp.1733–1738, 2006.
- [16] M. Sugiyama, "Local Fisher discriminant analysis for supervised dimensionality reduction," *Proc. ICML*, pp.905–912, 2006.
- [17] H. Zhao, S. Sun, Z. Jing, and J. Yang, "Local structure based supervised feature extraction," *Pattern Recognit.*, vol.39, pp.1546–1550, 2006.
- [18] D. Hu, G. Feng, and Z. Zhou, "Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition," *Pattern Recognit.*, vol.40, pp.339–342, 2007.
- [19] Z. Lei and Z. Shanan, "Face recognition based on orthogonal discriminant locality preserving projections," *Neurocomputing*, vol.70, pp.1543–1546, 2007.
- [20] C. Sibao, Z. Haifeng, and K. Min, "2D-LPP: A two-dimensional extension of locality preserving projections," *Neurocomputing*, vol.70, pp.912–921, 2007.
- [21] L. Zelnik-Manor and P. Perona, "Self-tuning spectral clustering," *Advances in Neural Information Processing Systems*, vol.17, pp.1601–1608, 2005.
- [22] A. Lanitic, C. Draganova, and C. Christodoulou, "Comparing different classifiers for automatic age estimation," *IEEE Trans. Syst. Man Cybern. B, Cybern.*, vol.34, no.1, pp.621–628, 2004.



**Tetsuji Ogawa** received his B.S., M.S., and Ph.D. in electrical engineering from Waseda University in Tokyo, Japan, in 2000, 2002, and 2005. He was a Research Associate from 2004–2007 and a Visiting Lecturer in 2007 at Waseda University. He has been an Assistant Professor at Waseda Institute for Advanced Study since 2007. His research interests include stochastic modeling for pattern recognition, speech enhancement, and speech recognition. He is a member of the Information Processing Society of Japan and Acoustic Society of Japan.



**Kazuya Ueki** received his B.S. in information engineering in 1997 and M.S. in computer and mathematical sciences in 1999, from Tohoku University in Sendai, Japan. He received his Ph.D. in computer science from Waseda University in Tokyo, Japan, in 2007. He currently works at NEC Soft, Ltd. His research interests include image processing and face recognition.



**Tetsunori Kobayashi** received his B.S., M.S., and Ph.D. in electrical engineering from Waseda University in Tokyo, Japan, in 1980, 1982, and 1985. He was a Lecturer from 1985–87 and an Associate Professor from 1987–91 at Hosei University in Tokyo, Japan. He joined Waseda University as an Associate Professor in 1991. He has been a Professor since 1997. He was a Visiting Scientist at the Spoken Language System Group of the Laboratory for Computer Science at the Massachusetts Institute of Technology from 1994–1995. His research interests include perceptual computing and intelligent robotics. He is a member of the Information Processing Society of Japan, Institute of Electrical and Electronics Engineers, and Acoustic Society of Japan.