Discriminant Analysis about Sex and Age Based on Facial Features Using Soft Computing

Tetsuya Miyoshi and Nobuyuki Kumazawa
Toyohashi Sozo College
20-1, Matsushita Ushikawa Toyohashi Aichi,440-8511,JAPAN
miyoshi@sozo.ac.jp

***Abstract* - This paper describes a comparison study of gender and age-group classifications using three classifiers, such as Fisher's Linear Discriminant (FLD), Neural Networks (NN), and the Classification and Regression Tree (CART), in which geometrical facial features are used as the input sources. The performance of the three classifications is evaluated and compared. Important key factors for the classification are extracted using three classifiers, and the similarities and differences in the classification results obtained using the three classifiers are discussed.**

keywords: gender classification, aging effect, face processing, face recognition.

I. Introduction

Visual communication plays an important role in human communication and interaction. As the need for automation in various Human-Computer Interaction (HCI) systems has been recognized, a number of face and gesture identification systems have been proposed and developed [1][2][3][4][5]. Furthermore, computer vision systems for human monitoring are likely to play an increasingly important role in our lives. Humans can quickly identify the gender, age, state of health, and emotions from faces and gestures and rapidly categorize these faces into different types on the basis of their appearance. If a recognition system of faces and gestures similar to that used by humans could be developed and built into computer vision systems, the application of HCI could expand to a wide area.

Generally speaking, pattern recognition techniques can be divided into two main categories: (1) those employing geometrical features and (2) those using gray-level information. There are two types of face image-processing methods with respect to the input sources. One is an appearance-based method using the gray-level picture as the input source, and the other is a geometric feature-based method using the coordinate of facial feature points, which are extracted by a system that detects the location of facial features. Although it is rather difficult to automatically extract facial components accurately using a computer system, several algorithms to extract the facial components have been proposed [6][7]. On the other hand, since the input source of the appearance-based method is the vector's very large dimensional number, which is equal to the number of pixels making up the image, it exhausts the computer resources, such as the calculating time and memory. In this study, a gender and age-group classification based on geometric facial features will be reported.

Early research in the face recognition field focused on the extraction of geometrical features for the description of the shape of facial components, including the mouth, eyes, and brow. Previous studies on gender classification have used high-resolution images with hair information and relatively small data sets for their experiments. Although, recently, a significant amount of research on face image processing and facial recognition has been presented, e.g., [1][2][3][4][5], few gender and age-group classifications have been presented [8][9][10]. Especially, little research on the effects of aging on the face has been reported. It is difficult to reveal the effects of aging on a face because there is significant variation in the effects of aging.

It is important to reveal the effects of aging on facial images. The age-related factor considerably affects the performance of a face recognition system. Face recognition systems are trained using a number of subject faces. The aging process causes a facial image to significantly change; as a result, the recognition performance using the classifier trained with previous images decreases. If the effects of aging could be included when recognizing a face image, the robustness of the recognition of the face would increase.

Appearance variation due to aging displays some unique characteristics. One of the uses of coordinate transformation to impose age-related changes on a human face is a cardioidal transformation, which characterizes the craniofacial shape of children [13]. Burt and Perrett investigated the process of aging using face composites from different age groups and caricature algorithms [11]. McArthur and Apatow reported that large eyes, low vertical placement of features, and short features, either singly or in combination, served to decrease perceived impressions of a stimulus person's physical strength, social dominance, and intellectual astuteness [12].

This paper describes a comparison study of gender and age-group classifications using three classifiers, such as Fisher's Linear Discriminant (FLD), Neural Networks (NN), and Classification and Regression Tree (CART), in which geometrical facial features are used as the input sources. The performance of the three classifications is evaluated and compared. Important key factors for the classification are extracted using three classifiers, and the similarities and differences in the classification results obtained using the three classifiers are discussed.

II. Classification Methods

In this study, a gender and age-group classification analysis was carried out using artificial NN, CART, and FLD, and the classification performance of the three classifiers and extracting rules in each model for the gender and age-group classifications were compared. In this section, the calculation algorithms and extracting methods of the classification rules of the three classifiers are summarized.

A. Neural netowark classifier

Neural network models have been used in the wide area of pattern recognition, including human face recognition. A neural network basically represents the relationship between inputs $\boldsymbol{x}$ and outputs $\boldsymbol{y}$, that is, the function which maps the inputs $\boldsymbol{x}$ into the outputs $\boldsymbol{y}$. In gender classification based on facial geometrical features, geometrical features, such as eye height, brow height, mouth width, and nose width, are the input $\boldsymbol{x}$, and the gender, i.e., male/female, is the desired output $\boldsymbol{y}$. The neural network is mathematically represented as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x},\boldsymbol{w}) \tag{1}$$

where $\boldsymbol{w}$ is the weight parameters included in the neural network model. Through training, the weight parameters w are determined so that the network can represent the input-output relationship, in other words, so that it can approximate the output by the neural network model. In the training procedure, the parameters $\boldsymbol{w}$ are initialized randomly and updated using the Back propagation learning rule to make the output $f(x_i, w)$ as close to the y_i as possible.

Although the neural network is accepted as a reliable method for data analysis, these models have their shortcomings. The major difficulty lies in the fact that the relationship between the inputs and outputs is not explicitly explained. In other words, it is difficult to interpret the significance of the input variables and understand the role played by the elements in the hidden layer from the trained network. To overcome this difficulty, a sensitivity measure that assesses the relative importance of input factors of the network to arrive at its target was proposed [18].

For a three-layer neural network with N input units, K hidden units, and M output units, the relative importance (RI_i) of the i-th component of the input vector can be estimated as

$$RI_i = \frac{\sum_{j=1}^{K} \left| \frac{w_{ij}^1 w_{jk}^2}{\sum_{i=1}^{N} \left| w_{ij}^1 \right|} \right|}{\sum_{k=1}^{M} \left| w_{jk}^2 \right|} \tag{2}$$

where w_{ij} is the weight from the i-th input node to the j-th hidden node and w_{jk} is the weight from the j-th hidden node to the k-th output node. The RI measure incorporates certain rates of change of the strengths of signals as they flow through the network.

B. Classification and regression tree(CART) [19]

The classification and regression tree is a tree-based classification and prediction method that uses recursive partitioning to split the training records into segments with similar output field values. CART starts by examining the input fields to find the best split, measured by the reduction in an impurity index that results from the split. The split defines two subgroups, each of which is subsequently split into two more subgroups until one of the stopping criteria is triggered.

Impurity refers to the extent to which subgroups defined by the tree have a wide range of output field (attribute) values within each group. Since the goal of the tree is to create subgroups in such a way that each subgroup has the same or similar output values, the input field to minimize the impurity of the node is chosen.

Gini is a general impurity measure based on probabilities of category membership for the branch. By letting im(D) be the impurity of database D and the probability p of a category, the impurity of the tree is defined by

$$im(D)=1-p^2-(1-p)^2 = 2p(1-p). \tag{3}$$

When the database is split by an input field(attribute) A, the impurity of a subgroup whose attribute value is a_i is defined by

$$im(A,a_i)=1-q_i^2-(1-q_i)^2 \tag{4}$$

where q_i is the probability of a category in the subgroup. The average of the impurity for groups with each attribute value is calculated by

$$im(A)=\sum_{i=1}^{n} \Pr(A=a_i) \quad im(A,a_i) \tag{5}$$

The attribute whose average is the smallest in all input attributes is chosen as the test attribute of the node. The main advantage of the tree classifier is the possibility to interpret the decision rule in terms of individual features. This makes decision trees attractive for interactive use by experts.

The decision tree represents the classification rule. Each data is categorized by testing with each node attribute. The upper attributes near the root node are more significant for the classification than the lower attribute.

C. Fisher's linear discriminant (FLD)

Fisher's linear discriminant (FLD) is an example of a class - specific subspace method that is able to find the optimal linear projection for a classification. FLD determines the linear projection that maximizes the ratio of between-class scatter to that of within-class scatter. FLD is one of the classifiers to construct decision boundaries directly by optimizing a given error criterion. The driving force of the training procedure is the minimization of a criterion such as the apparent classification error or the mean squared error (MSE) between the classifier output and some preset target value.

In FLD, the class is determined by the location of data with respect to the decision hyperplane, that is, the class is determined by the value of projection to the normal vector of the hyperplane. The component of the projection vector corresponding to an input value means the importance of the input value.

The larger the component is, the more significantly the input value affects the classification result.

D. Cross validation technique[20]

The goal of designing a recognition system is to classify future test samples which are likely to be different from the training samples. The classifier should be evaluated by its generalization ability which refers to its performance in classifying test patterns not to be used during the training stage.

The cross validation(CV) technique divides an observed data set into two parts, a training set to identify a model and a checking set to validate the trained model. In the case of *k*-fold CV, the observed data set is divided into *k* parts; one of them is a checking data set, and the other is a training data set. In a training procedure, a model is trained using a training data set, and, in an evaluating procedure, a model is evaluated with the square sum of errors for the checking set. This procedure is repeated *k* times using the different data sets. The average value of the square sum of errors over all *k* procedures is a cross-validation index. The model with the least averaged index is selected as an optimal model. In the leaving-one-out CV, for a given sample size *J*, a model is trained using *J-1* data and validated in the remaining case. This procedure is repeated *J* times, and the average of error for validity data is obtained.

In general, a division of the observed data decreases the reliability of parameters when the number of sample data is comparatively smaller than the number of parameters within a model. On the other hand, leaving-one-out has the advantage that almost all the available data are used for the identification of a model, but it is computationally expensive. In this study, a two-fold CV procedure was conducted for the gender classification because a relatively large number of training data were collected, and the ten-hold CV method was used for the age-group classification because the number of collected data was relatively small.

III GENDER CLASSIFICATION

Literatures on psychology investigate gender classification and physical differences [13][14][15]. Fewer studies have focused on gender classification, in comparison with the number of studies on facial recognition. We conducted three experiments using NN, CART, and FLD in order to classify gender based on geometric-based features, such as brow height and eye width.

A. Experimental facial images and faical features

Facial images from 100 males and 100 females were taken from the commemorative photo albums of a high school for the experiments of gender classification. Examples of the facial photographs used are shown in upper row of Fig. 6. The facial images wearing eye-glasses were exclued. Twenty-six points out of all facial points were selected to characterize the physical location of the facial components in order to calculate the facial features, as shown in Fig. 1. Although several algorithms extracting the facial feature points have been proposed [6][7], in this study, the extraction of the characteristic points in facial images was performed manually.

Since the size of facial photographs used were various, the faces had to be normalized so as to make the length of eye-nose distance unit. After the normalization process, we calculated 8 facial features, including the "nosewidth, h "eye height, h and "eye-nose area," from the coordinate of the feature points. The 8 facial features, "brow curve," "brow-eye area," "brow separation," "eye-nose area," "eye height," "nose width," "eye height/ brow height," and "brow-chin distance," were selected as the inputs of the gender classifier. The calculating method of these facial features is shown in Table 1.

B. Learning of gender classifiers

The gender classification was conducted using three classifi-

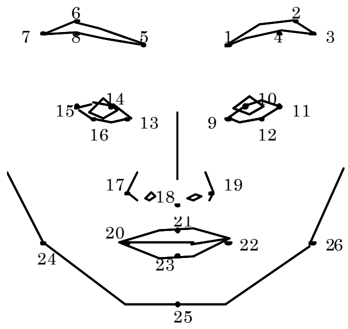

Fig. 1 Feature points of a face

Table 1 Facial features

facial geometrical features	definition
brow height	vertical distance of between 6 and 8
brow curve	curvature of points 5,6,7
brow-eye area	area of box 3,7,11,15
brow separation	horizontal distance between 1 and 5
eye-nose area	area of box 11,15,20,22
eye height	vertical distance of between14 and 16
nose width	horizontal distance of between 17 and 19
eye height/brow height	
brow-chin distance	vertical distance of between 5 and 25

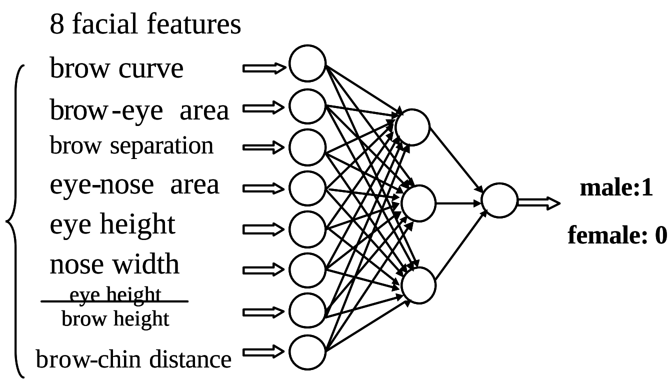

Fig. 2 Input variable and structure of neural network

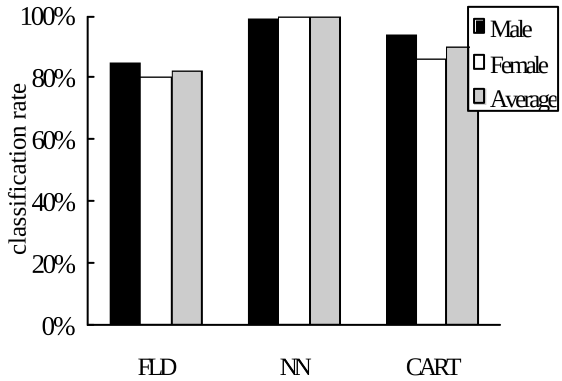

Fig. 3 Classification performance for the training data of three classifiers FLD, NN, and CART in gender classification.

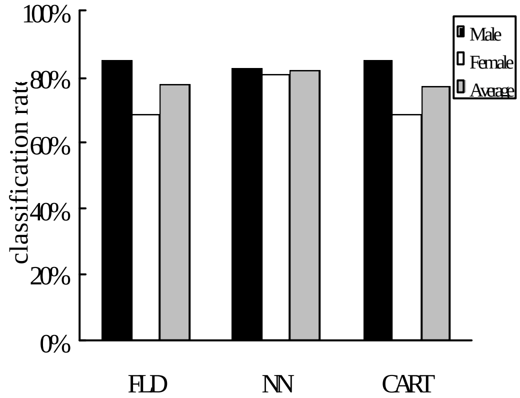

Fig.4 Classification performance for the checking data of three classifiers FLD, NN, and CART in gender classification.

ers, NN, CART, and FLD. The neural network architecture is shown in Fig.2, which consists of 8 input units, 3 hidden units, and 1 output unit. The number of hidden units was determined so as to minimize the error for the checking data. The classification performance was evaluated with the generalization performance using the cross-validation method. All faicial data(200) were divided into two groups, which included 100 data. Three classifiers were trained using the first group data, and the classification rate for the second group data was calculated. Training procedures using the three classifiers were repeated for the first group data and the calculation of the classification rate for the second group data was performed. The three classifiers, NN, CART, and FLD, were evaluated using the average of the classification rate for checking data.

C. Comparison among three classifiers

The classification rates for training data and checking data are plotted in Figs. 3 and 4, respectively. As shown in Fig. 3, the classification rates for the training data were relatively higher than the ones for checking data. Especially, the classification rate of NN was higher than that for the other methods, at almost 100%. The classification of NN indicated the best performance in the average of the classification rate among the three classifiers.

The tree generated by the CART algorithm is interpreted as the classification rules for the gender classification. Then, the significant key factors of facial features and classification rules for the gender classification were extracted, as shown in Fig. 5. The resultant tree demonstrates that the input picture is classified as female when the ratio of eye height to brow height is larger than 1.73 or the brow-eye area and nose width are relatively small.

The key factors for the gender classification extracted by each classifier are given in Table 2. As shown in Table 2, it is confirmed that common key factors were extracted such as geye height/ brow height h and gbrow-eye area. hThese common key factors imply that a face in which eye height was long was classified as female and a face in which the brow height was short was also classified as female.

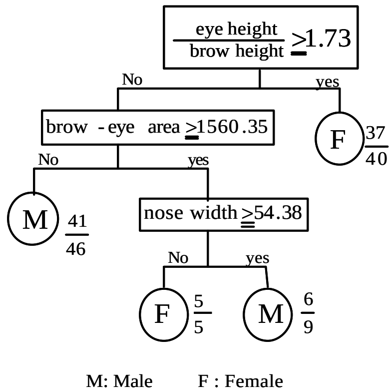

Fig. 5 Extracted decision tree for gender classification by CART

Table 2 Key factors of gender classification extracted usign three classifiers.

Classifier	key factor
FLD	eye height/brow height, brows－eyes area
NN	eye height/brow height, brows－eyes area
	brow separation, eyes－nose area
CART	eye height/brow height, brows－eyes area, nose width
	eye height/brow height, brow curve ,eyes －nose area

IV Age-Group Classification

A. Experimental facial images and faical features

For the investigation of age-related facial features, geometrical facial features are extracted from the facial images of three

Fig. 6 Examples of facial data for the age-group classification. Pictures on the upper , middle and lower row show the faces of teenagers, individuals over 20 and 30 years old, respectively.

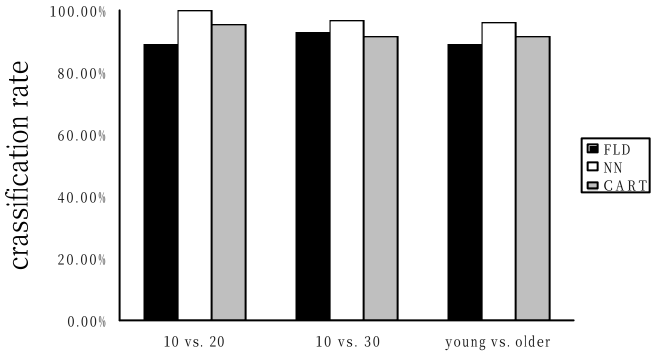

Fig. 7 Classification performance for the training data of three classifiers FLD, NN, and CART in age-grouped classification.

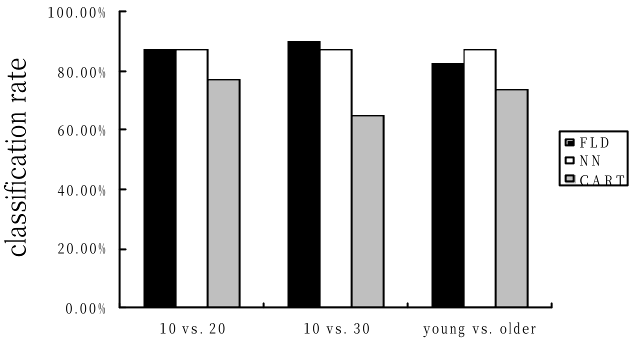

Fig. 8 Classification performance for the training data of three classifiers FLD, NN, and CART in age-grouped classification.

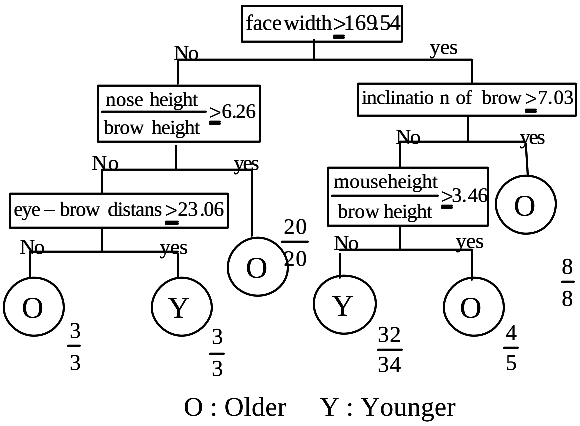

Fig. 9 Extracted decision tree for the age-group classification by CART.

Table 3 Key factors of the age-group classification extracted usign three classifiers.

	10 vs. 20 and 10 vs. 30	Teens vs. over Teen
FDL	brow height, inclination of eye, inclination of brow, outline-mouth distance	face width, inclination of brow, inclination of eye
NN	brow height, inclination of eye, inclination of brow, face width	face width, inclination of brow , nose height/brow height
CART	brow height, inclination of eye, inclination of brow, eye-nose distance, eye-mouth distance, face width	face width, inclination of brow, nose height/brow height

age-groups, (1) group 1, between 10 and 20 years old, (2) group 2, between 20 and 30 years old, and (3) group 3, over 30 years old. We took sets of facial images, which consisted of 10 males and 10 females in each age group, from photo album gJapanese face h [21] in which the Japanese faces and facial expressions are reported. Images of teenagers were selected from commemorative photo albums as described above. In total, there were 60 facial images. Samples of test facial images are shown in Fig. 6, in which the photos on the upper, middle, and lower rows are the faces of teenagers and individuals over 20 and 30 years old, respectively.

B. Learning age-group classifiers

The age-group classification was performed using three classifiers, as for the gender classification. The classification for three age groups was performed. Three kinds of two classes classification such as ones of group1 and 2, group1 and 3, group2 and 3, were conducted. It is well known that facial changes are significant in teenagers and that the aging process causes further significant facial changes. The fact suggests that the classification of teens and elders over twenty years old seems to result in good classification performance, and conversely, it is difficult to classify data over twenty into two groups, the group2 over twenty years old and group3 over thirty years old. Then, in addition to the three classification procedures, an experiment of age-group classification between the group of individuals under 20 years old (20 males and 20 females) and that of individuals over 20 years old (20 males and 20 females) was conducted. All classification performances were evaluated using 10-fold CV.

C. Classification results of the age-group classification

The classification results for the training data and checking data are shown in Figs. 7 and 8. In Figs. 7 and 8, the classification rates for the training and checking data were plotted in three kinds of classifications. A comparison of the results indicated that the classification performance for the checking data using FLD and NN was better than that by CART.

The example of the decision tree generated by the CART algorithm is shown in Fig. 9. The test attributes using upper nodes are significant factors for age-group classification. The key factors extracted by each classifier in the age-group classification are summarized in Table 3. Similar key factors were extracted in three classifiers. These extracted key factors show that a face with a relatively large width and a small inclination of the brow would likely be classified into the younger group, and conversely, a face with a relatively small width and a large ratio of the nose height to the brow would likely be classified into the older group.

V Conclusion

In this paper, a comparison of the classification performance and the key factors in gender and age classification using three classifiers, such as Neural Network, Classification & Regression Tree, and Fisher's Linear Discriminant, were discussed. In both classifications, the performance, i.e., the classification rate of the NN classifier, was the best among the three classifiers. The key factors for the gender and age classification extracted by each classifier are similar.

Reference

[1] R. Gross, I. Matthews, S. Baker: Appearance-Based Face Recognition and Light-Fields, IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 449-465, 2004

[2] J. Yang, D. Zhang, A.F. Frangi, J. Yang : Two-Dimensional PCA: A New Approach to Appearance-Based Face Representation and Recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence, pp. 131-137, 2004

[3] Z. Pan, G. Healey, M. Prasad, B.Tromberg : Face Recognition in Hyperspectral Images, IEEE Trans. on Pattern Analysis and Machine Intelligence, pp. 1552-1560, 2003

[4] K. Etemad and R. Chellappa : Discriminant Analysis for Recognition of Human Face Images, J. Optical Soc. Am. A, vol. 14, pp. 1724-1733, 1997.

[5] B. Moghaddam and A. Pentland : Probabalistic Visual Recognition for Object Recognition, IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 696-710, 1997.

[6] D. Chai and K.N. Ngan, Face Segmentation Using Skin Color Map in Videophone Applications, IEEE Trans. on Circuits and Systems on Video Technology, Vol. 9, No. 4, pp. 551-564, 1999.

[7] M.Yang, D.J. Kriegman, N. Ahuja : Detecting Faces in Images: A Survey, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, pp.34-58, 2002

[8] D.M. Burt and D.I. Perrett, "Perception of Age in Adult Caucasian Male Faces: Computer Graphic Manipulation of Shape and Color Information," Proc. Royal Soc. London, vol. 259, pp. 137-143, 1995.

[9] G.A. Khuwaja and M.S.Laghari: A Parameter-Based Combined Classifier for Invariant Facial Expression and Gender,Int.J.ofPattern Recognition and Artificial Intelligence, Vol.16, No.1, pp.27-51, 2002

[10] T.Kanno, M.Akiba, Y.Teramachi, H.Nagahashi, and Takeshi Agui: Classification of Age Group Based on Facial Images of Young Males by Using Neural Networks, IEICE Trans. on Information and Systems, Vol.E84-D, No.8, pp.187- 190, 2001

[11] A.Lanitis, C.J. Taylor, T.F. Cootes: Toward Automatic Simulation of Aging Effects on Face Images, IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol. 24, No. 4, pp.442-455, 2002

[12] L.Z.McArthur and K.Apatow: Impression of Baby-Faced Adults, Social Cognition, Vol.3, No.4, pp.315-342, 1983

[13] V. Bruce and A. Young : In the Eye of the Beholder : The Science of Face Perception, Oxford Univ. Pr. Published, 1998

[14] V.Bruce, A.M.Burton, N.Dench, E.Hanna, P.Healey, O.Mason, A.Coombes, R.Fright, and A.Linney : Sex Discrimination: How do we tell the difference between male and female faces?, Perception, vol.22, pp.131-52,1993

[15] A.M.Burton, V.Bruce, and N.Dench: What's the Difference between Men and Women?, Evidence from facial measurement, perception, vol.22, pp.153-176, 1993

[16] V.Bruce, A.M Borton, E.Hanna, P. Healey, O.Mason, A. Coombes, R. Fright, and A. Linney, Sex Discrimination: How Do We Tell the Difference between Male and Female faces? Perception, Vol. 22, pp.52~131, 1993

[17] V.Bruce, A.M Borton, E.Hanna, P. Healey, O.Mason, A. Coombes, R. Fright, and A. Linney, Sex Discrimination: How Do We Tell the Difference between Male and Female faces? Perception, Vol. 22, pp.52~131, 1993

[18] G.D. Garson, Interpreting neural-network connection weights, Al Expert, pp.47~51, 1991

[19]L.Breiman, J.Friedman, R.Olshen and C.Stone Classification and regression trees, Wadsworth, 1984

[20] M.Stone : Cross Validatory Choice and Assessment of Statistical Predictions, J. of the Royal Statistical Society, Vol. 36, No.2, pp.111-47, 1974

[21] N. Araki : Nihonjin no kao, Nihonjin-no-kao Project 2002