Combination of Face Direction Estimation and Face Recognition Using Four-Directional Features

Hitoshi Hongo*, Mamoru Yasumoto*, Yoshinori Niwa*, Kazuhiko Yamamoto+
* HOIP, Softopia Japan/JST, 4-1-7 Kagano, Ogaki, Gifu 503-8569 Japan
+ Gifu University, Faculty of Engineering 1-1, Yanagido, Gifu 501-1193 Japan
E-mail: hongo@softopia.pref.gifu.jp

Abstract

To identify people's faces from various directions, we propose a novel method that distinguishes people using a combination of face direction estimation and face recognition. Both the face direction estimation method and the face recognition method are appearance-based methods that use a linear discriminant analysis on the same features, the Four-Directional Features. Since these methods are constructed on the same framework, it is easy to combine them and extend the variable range of face directions. Our method consists of two hierarchical processes: the first is face direction estimation, and the second is face recognition. Estimating the face direction, we can use the face recognition discriminant space by limiting the range of face directions. Limiting the face direction can strengthen a face recognition discriminant space. Using the face recognition discriminant spaces extended the range of the face directions estimated, a total accuracy rate of face recognition was improved. Experiments showed that our method performed at an accuracy rate of 97.6 % for 105,000 images using 150 subjects and 35 different face directions in the range of ±45 degrees horizontally and ±30 degrees vertically.

1. Introduction

Recently, "intelligent environments" have been developed which can assist humans in accomplishing specific tasks by identifying a person, estimating gaze points, and determining individual attributes to examine their desires and current status [1, 2, 3]. These environments integrate many computer vision technologies to detect, track, and identify people [4, 5], recognize gestures [6], detect gaze points [7], and interpret human behavior [8]. In particular, face recognition methods that can distinguish many people in various poses are required in intelligent environments. This is because humans can move and face any direction in a room. In a typical office, for example, there exist several people including visitors. Although the type and number of people are unspecified, the system should recognize them as precisely as possible. This technology could be useful for security systems, surveillance systems, and appliance interfaces.

Some methods of face recognition based on PCA (Principal Component Analysis) and LDA [9, 10, 11, 12] have demonstrated the power of appearance-based methods in accuracy. However, these methods can only be performed reliably using the frontal view of the face.

To recognize non-frontal views of the face, several approaches are proposed. Georghiades et. al. [13] presented a generative appearance model and demonstrate its usefulness for image-based rendering and recognition. Murase and Nayar [14] have used PCA for solving the general problem of object recognition and pose estimation. Kurita [15] proposed a viewpoint-independent face recognition method using a mixture of viewpoint-dependent classifiers. This research has created new approaches and reported good results, however they only experimented on a fixed number of face directions using a small number of people (about 10 persons or less). It is expected that the accuracy rate decreases as the number of people and variations of face direction increase.

In this paper, we propose a hierarchical appearance-based method that identifies people using a combination of face direction estimation and face recognition. Our method uses linear discriminant analysis (LDA [16]) on "Four-Directional Features (FDF)" [17]. The proposed method estimates the face direction and then recognizes faces from the limited face directions. Since this technique is appearance-based, it can easily be extended to identifying categories of objects that are arranged in a well-defined configuration, such as faces, face directions, objects, and hand-signs [17]. In addition, multiple hierarchical processes have the capability to classify objects in detail.

To evaluate the performance of our method, we constructed a facial database that consists of 105,000 images using 150 persons and 35 different face directions in a range of ±45 degrees horizontally, and ±30 degrees vertically. The experimental results of our method are shown using this facial database.

2. Recognition method

We have proposed a face direction estimation that uses LDA on the FDF method [21]. We also used FDF that can achieve a high recognition rate for face recognition [17].

In this section, we describe our principal recognition methods with regard to face direction estimation and face recognition. First, we introduce the face detection method. In order to estimate facial direction and recognize faces, our method detects the facial region by using color information, and then extracts the FDF from that region. Next, we explain the feature of the FDF and how to extract these features. Finally, LDA applied to both the face direction estimation and the face recognition is presented.

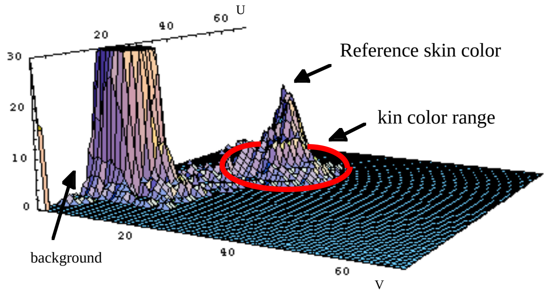
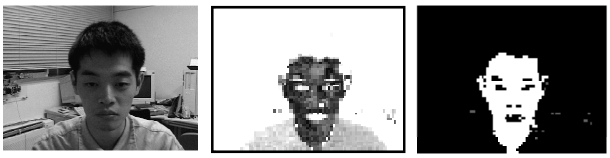


Figure 1. Color distribution



(a) Original image (b) color distance (c) result

Figure 2. An example of the detected skin region

2.1. Skin color detection

Color information is useful to detect a human face in various poses and from various viewpoints. Various methods of face detection using color information have been developed [19, 20]. To focus on face recognition from a facial image in this paper, we reduce some segmentation problems by enclosing the interior of the room with blue curtains. However, to be more reliable given individual differences and varying environmental conditions in a room, we use the proposed method of face detection using *u* and *v* values in the CIE-Luv color space [17]. The CIE-Luv color space consists of lightness (*L*) and chrominance values (*u* and *v*). Theoretically, the CIE-Luv color system is as capable as humans in representing the color distance between any two colors. Our method to extract skin color regions is described as follows. Only the skin colors of Asian subjects is considered in this paper.

First, a two-dimensional histogram of *u* and *v* values is made from the previous frame. From the two-dimensional histogram, our method determines the reference skin color. This is the color that covers the maximum number of pixels within the range of skin colors. Second, the *u* and *v* value of each pixel in the input image is converted to the color distance from the reference skin color. Finally, a histogram of the color distance is constructed from the above results and then the skin region is extracted by discriminant analysis.

Figure 1 shows the reference skin color and the color distribution of the input image. Figure 2 shows an example of the skin region detected with our method. Figure 2(b) shows the color distance from the reference skin color. The level of brightness denotes the distance from the reference skin color. Dark areas indicate a close match to skin colors. Figure 2(c) shows the results of the detected face region.

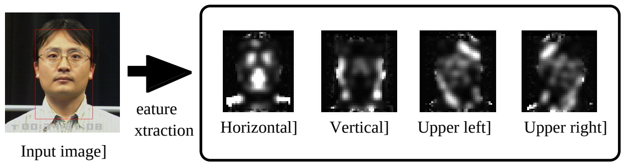


Figure 3. Four-Directional Features (FDF)

2.2. Four-Directional Features (FDF)

FDF is one of the robust features to distinguish patterns. Since FDF represents the edge direction information in low resolution, it is robust against deformation and noise and can be processed quickly. Extracting FDF is performed as follows:

(1) By applying Prewitt's operator in four directions, vertically, horizontally and along both diagonals, four edge images are constructed from the detected face region.

(2) Gaussian filters are applied to each edge image. The images are then normalized to 8×8 pixel low-resolution images.

(3) 256-dimensional feature vectors are made from this set of four images.

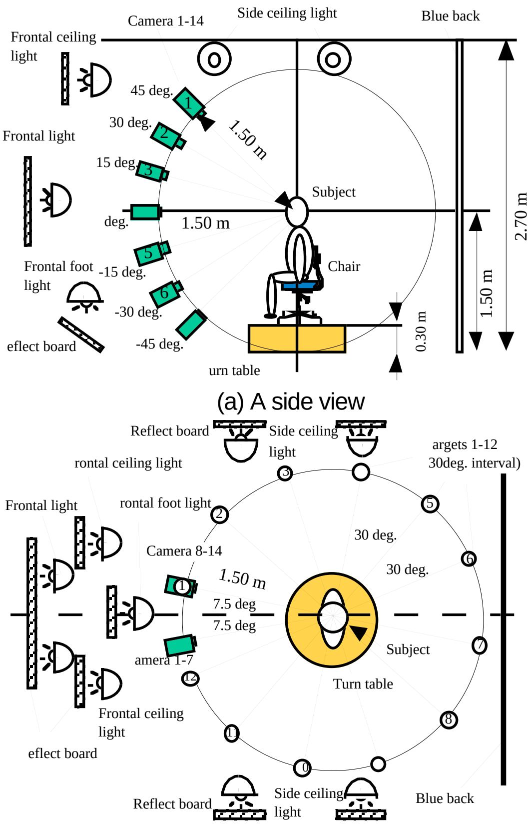
Figure 3 shows an example of FDF. Converting four edge images to low resolution has the advantage of keeping the shape information better than directly converting the original image.

2.3. Linear discriminant analysis (LDA)

LDA is used to construct a discriminant space in which the within-class variance is minimized and the between-class variance is maximized. Therefore, a pattern can be classified to the optimum class using LDA. In this linear discriminant analysis, the optimal coefficient matrix $\boldsymbol{A}$ is obtained by solving the eigenvalue problem. The optimal coefficient matrix $\boldsymbol{A}$ generates the best new features from the primitive features. Let the number of classes be denoted by K. The dimension of the discriminant space N is bounded by min(K-1, M), where M is the dimension of the FDF. In this work, the dimension M is 256 (4×8×8).

In the discrimination process, first, the feature vector $\hat{x}_i$ is transformed to y_i through the coefficient matrix $\boldsymbol{A}$. Second, the distance D_{ik} from y_i to the mean vector $\bar{y}_k$ of class C_k, where $D_{ik} = |y_i - \bar{y}_k|^2$, is calculated. Finally, the result is decided to the class C_k whose distance is minimum.

Within this framework, it can be easy to extend the identification to categories of objects that are arranged in a well-defined configuration, such as faces, face directions, objects, and hand-signs.



(a) A side view

(b) A ground plan

Figure 4. Multiple-camera system

3. Facial database

We have developed a multiple-camera system that captures facial images from various viewpoints simultaneously [18]. As shown in Figure 4, we arranged 14 cameras to gather facial images with various poses to ease the subjects' task. In this case, the subjects were only required to look in front of them. Each camera is gen-locked by a sync generator and a time code (VITC) is superimposed on its output. The image data is input as a 640 × 480 pixels, 24-bit color image at a rate of 30 frames/sec., and then recorded onto the hard disk of each of 14 PCs using motion-JPEG compression. Since the main PC controls all client PCs connected with each camera through Ethernet sockets, it is easy to operate the multiple-camera system to capture images from the different viewpoints of the cameras. The cameras can be arranged at any position for their purpose.

Using the multiple-camera system, we constructed the facial database to evaluate the performance of face recognition for variations of persons and face directions. The subjects sat on a chair, and then we made them face given targets. In this paper, the facial data set consists of 150 people and 35 directions of faces per person. The subjects' ages vary from 15 to 64 years old. Figure 5 shows examples of male and female facial images with intervals of five years. Figure 6 shows an example set of images the 35 face directions of one subject: 30, 15, 0, -15 and -30 degrees vertically, 45, 30, 15, 0, -15, -30, and -45 degrees horizontally. We gathered images twice in each direction in order to acquire the training data and the test data for direction, a total 2×10=20 images were recorded. The face database totals 105,000 (150×35×20) images.

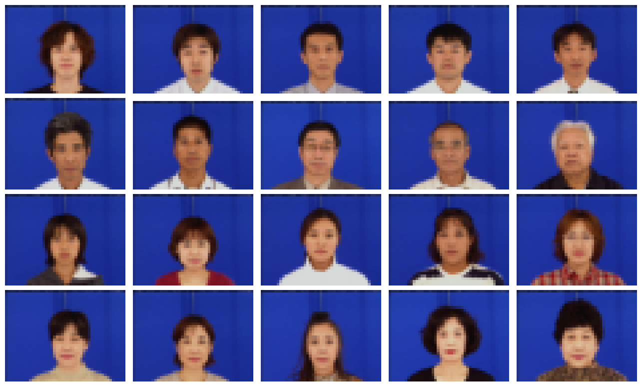


Figure 5. Examples of the facial database

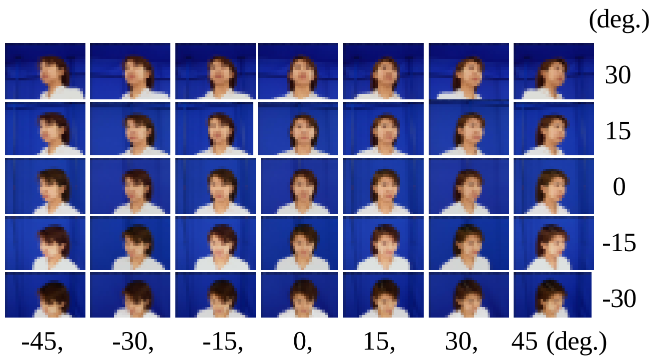


Figure 6. An example set of facial images with 35 face directions

4. Combination of face direction estimation and face recognition

It is expected that the accuracy rate of face recognition decreases as the number of persons and the variations of face direction increase. In this paper, we propose a method that identifies a person using the combination of the face direction estimation and of the face recognition.

To evaluate the validity of our proposed method, we conduct three kinds of experiments. In the first experiment, we test the performance for a variable number of persons and the variation of face directions. In the second experiment, we estimate the face direction for unspecified persons. Finally, in the third experiment we report the end result using our hierarchal method that combines face direction estimation and face recognition.

4.1. Exp-1: Face recognition with a variable number of people and face directions

Figure 7 shows the accuracy rates with a variable number of people and face directions. In this experiment, we analyzed the face recognition performance for range of face directions, a small range, a middle range and a wide range. D09 is a small range which consists of nine directions of faces within the range of ±15 degrees in vertical direction and ±15 degrees in the horizontal direction. Thus, these facial images show near frontal views of a face. D25 is a middle range which consists of 25 directions of faces within the range of ±30 degrees and ±30 degrees, and D35 is a wide range which includes 35 directions of faces. It was expected that the accuracy rate would decrease as the number of persons and the variations of face directions increased. In the case of D09, our method achieved an accuracy rate of 99.5 % even if the number of persons was 150. We confirmed that it was easy to recognize a near frontal view of a face within ±15 degrees. The results suggest that limiting the direction of a face can strengthen a face recognition discriminant space. They also suggest that the accuracy rate of our method decreases as the number of persons increases in the range of face directions over ±30 degrees.

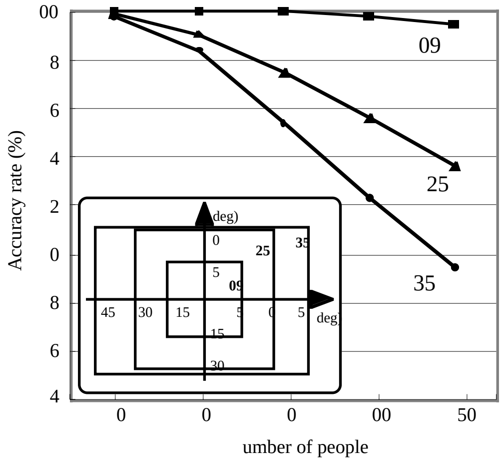


Figure 7. Accuracy rates of face recognition

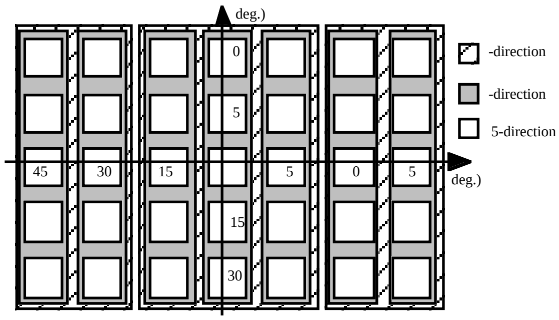


Figure 8. Classification for face direction

4.2. Exp-2: Face direction estimation

To evaluate the precision of our face direction estimation method, we constructed three kinds of discriminant spaces as follows:

3-dir.: 35 directions are divided into three columns.

7-dir.: 35 directions are divided into seven columns.

35-dir.: each of the 35 directions occupies a column.

Figure 8 shows the divided areas for face direction estimation respectively. In this experiment, we divided the areas vertically, because we assumed that there are fewer different features between faces vertically than horizontally. The experimental results are shown in Table 1. The accuracy rates of face direction estimation are shown for “specified persons” and “unspecified persons”. “Specified persons” signifies that images of a given member of persons are used for the discriminant space as training. The number of specified persons was 50. We constructed three kinds of discriminant spaces from 17,500 training images. For the test data, we used 17,500 images for 50 specified persons, and 17,500 images for the unspecified persons that consisted of 50 persons were not included in the training data.

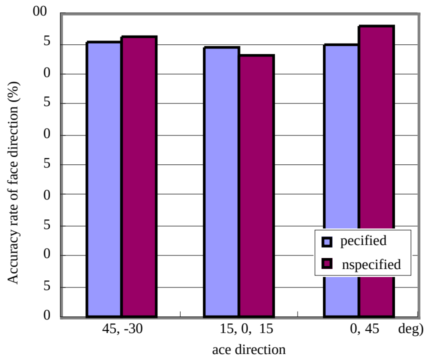
Table 1. Accuracy rate of the face direction estimation (%)

Classification	3-dir.	7-dir.	35-dir.
Specified persons	95.0	86.0	73.6
Unspecified persons	95.8	84.0	65.3

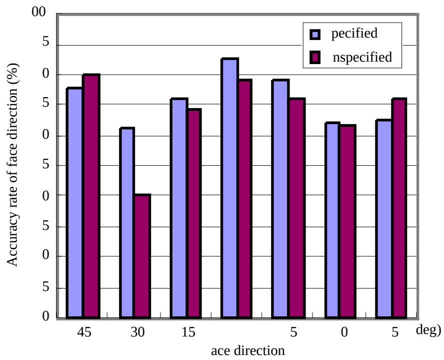
Experiments show that there are some differences in average accuracy rates between the specified persons and unspecified persons as shown in Table 1. Figure 9 shows the accuracy rates by use of the three discriminant spaces.

The classification of 3-directions has the highest accuracy rate in this experiment, as shown in Figure 9(a). However, it could only estimate 3 directions. In the case of 7-directions, the frontal views of faces and both end views of faces were distinguished with a high accuracy, as shown in Figure 9(b). We assume that there is a large difference between persons for the 30 degrees view of faces. In the case of 35-directions, the accuracy rates of the unspecified persons decreased, as shown in Figure 9(c-2). In particular, the accuracy rates decreased when the subjects looked down. We assume that the features of these faces are lost dynamically.

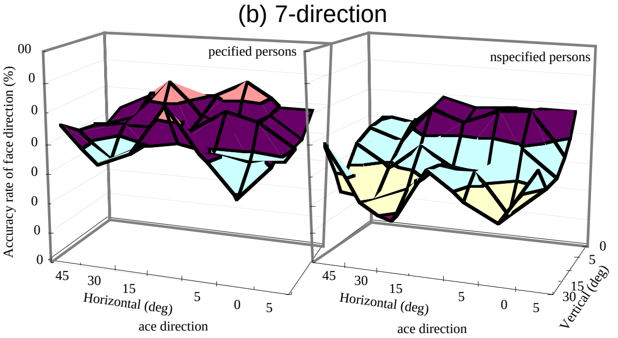
From the results, we assume that the discriminant space which was constructed from 50 persons has an ability of face direction estimation for unspecified persons,



(a) 3-direction



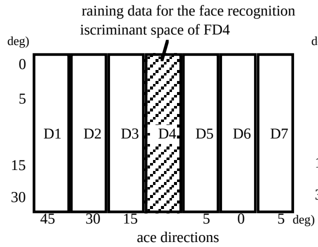
(b) 7-direction



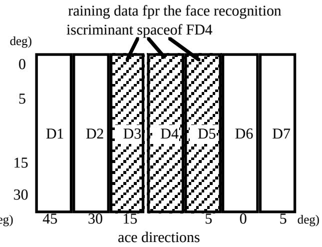
(c-1) 35-dir. specified persons, (c-2) unspecified persons

Figure 9. Accuracy rates of face direction estimation

and we also assume that it was difficult to estimate the face direction exactly with our method.



(a) Regular area



(b) Extended area

Figure 10. Area of training data for face recognition discriminant space

Table 2. Accuracy rate of face recognition

Discriminant space \ Classification	3-directions / (simulation)	7-directions / (simulation)
Regular Discr. Sp.	95.6 / (97.7)	93.6 / (98.6)
Extended Discr. Sp.	95.6 / (96.5)	97.6 / (98.3)

4.3. Exp-3: Combination of face direction estimation and face recognition

From the experiments 1 and 2, we expect that the combination of the face direction estimation and the face recognition is effective. We propose a hierarchical method that estimates the face direction and then recognizes faces in the limited face directions. From experiment 1, we can expect that the discriminant space constructed from the limited variations of face directions has high performance. However, our method could not estimate the face direction exactly. Thus, there is a trade-off between the accuracy rate of face direction estimation R_{dir} and the accuracy rate of face recognition R_{face}. The overall accuracy rate R_{total} is given by equation (1).

$$R_{total} = R_{dir} \cdot R_{face} \qquad (1)$$

Ordinarily, the hierarchical method intends to decrease the accuracy rate through each process. In other words, the overall accuracy rate can not exceed the accuracy rate of face direction estimation. To solve this problem, we construct the face recognition discriminant space by making the wider than the range estimated by the face direction estimation method. As shown in Figure 10, in the case of 7-directions, the area FD4 represents 0 degree. From the viewpoint of efficiency, face recognition discriminant space is constructed within FD4. On the other hand, the extended face recognition discriminant space for FD4 covers FD3, FD4 and FD5.

We experimented on face recognition with 3-directions and 7-directions. We compared the performance for the regular and extended face recognition discriminant spaces. Table 2 shows that the hierarchical method achieved over 93 % accuracy rate. This is higher than in experiment 1. In total, 7-directions achieved the highest accuracy rate, 97.6 %, even if the accuracy rate of the face direction estimation was 86 %. We also show the accuracy rates by a simulation supposing that the estimation of the face direction is correct. The difference between both accuracy rates are less than 1 % in 7-directions. We assume that our method performs well with the combination of face direction estimation and face recognition.

5. Conclusions

We proposed a new method that identifies people using a combination of face direction estimation and face recognition. Our method is a hierarchical appearance-based method that uses a linear discriminant analysis (LDA) on Four-Directional Features (FDF). Using the estimation of face direction, our method utilizes a face recognition discriminant space that is constructed from the limited variations of face directions. Combining the face direction estimation and the face recognition, our method achieved an accuracy rate of 97.6 % for a facial database containing 105,000 images: 150 people and 35 face directions.

Since the face direction estimation and face recognition use the same framework, the available range of face directions can be extended and that would be easy to implement. We expect that our method could maintain a high accuracy rate even when the available range of face directions is extended.

The future work is to optimize the combination of face direction estimation and face recognition, and then to construct the optimum combination automatically.

Reference

[1] M. Weiser, “The Computer for the 21st Century,” Scientific American, Vol. 256, No. 3, pp.66-76, 1991.

[2] A. Pentland, “Smart Rooms, Smart Clothes,” Scientific American, Vol. 274, No. 4, pp.68-76, 1996.

[3] M. Torrance, “Advances in Human-Computer Interaction: The Intelligent Room,” CHI’95 Research Symp., 1995.

[4] A. Pentland, “Looking at People: Sensing for Ubiquitous and Wearable Computing,” IEEE Trans. PAMI, Vol.22, No.1, pp.107-119, 2000.

[5] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, Patrick J. Rauss, “The FERET Evaluation Methodology for Face-Recognition Algorithms,” IEEE Trans. PAMI, Vol. 22, No. 10, pp.1090-1104, Oct. 2000.

[6] T. Watanabe and M. Yachida, “Real Time Gesture Recognition Using Eigenspace from Multiple Input Image Sequences,” Proc. of FG’98, pp.428-433, 1998.

[7] Y. Matsumoto, A. Zelinsky, “An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement,” FG2000, pp.499-504, 2000.

[8] M. Thonnat, N. Rota, “Image Understanding for Visual Surveillance Applications,” 3rd International Workshop on Cooperative Distributed Vision, 1999.

[9] M.Turk and A.Pentland, “Face recognition using eigenfaces,” Journal of Cognitive Nuroscience, vol.3, no.1, pp.71-86, 1991.

[10] T. Kurita, N. Otsu, and T. Sato, “A face recognition method using higher order local autocorrelation and multivariate analysis,” 11th ICPR, pp.213-216, 1992.

[11] B.Moghaddam, and A.Pentland, “Probabilistic Visual Learning for Object Detection,” ICCV’95, pp.786-793, 1995.

[12] W. Zhao, R. Chellappa and A. Krishnaswamy, “Discriminant Analysis of Principal Components for Face Recognition,” Proc. of FG’98, pp.336-341, 1998.

[13] A.Georghiades, P.Belhumeur, and D.Kriegman, “From Few to Many: Generative Models for Recognition Under Variable Pose and Illumination,” FG2000, pp.277-284, 2000.

[14] H.Murase and S.Nayar, “Illumination Planning for object recognition using parametric eigenspace,” IEEE Trans. PAMI, Vol. 16, No. 12, pp.1219-1227, 1994.

[15] T.Kurita, H.Shimai, T.Mishima, and T.Takahashi, “Self-organization of viewpoint dependent face representation by the self-supervised learning and viewpoint independent face recognition by the mixture of classifiers,” IAPR Workshop MVA2000, pp.319-322, 2000.

[16] Fukunaga, K. “Introduction to Statistical Pattern Recognition (2nd ed.),” Academic Press, 1990.

[17] H. Hongo, M. Ohya, M. Yasumoto, Y. Niwa, and K. Yamamoto, Focus of Attention for Face and Hand Gesture Recognition Using Multiple Cameras, Proc. FG2000, pp.156-161, 2000.

[18] H. Hongo, H. Watanabe, M. Yasumoto, Y. Niwa, and K. Yamamoto, “Eye Region Extraction for omni-directional gaze detection using multiple-cameras,” IASTED Int. Conf. SPPRA, pp.241-246, 2001.

[19] J.-C.Terrillon, M. David and S. Akamatsu, “Automatic Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and of Invariant Moments”, Proc. of FG’98, pp.112-117, 1998.

[20] Y. Raja, S. J. McKenna and S. Gong, “Tracking and Segmenting People in Varying Lighting Conditions using Colour”, Proc. of FG’98, pp.228-233, 1998.

[21] M. Yasumoto, H. Hongo, H. Watanabe, K. Yamamoto, “Face Direction Estimation using Multiple Cameras for Human Computer Interaction,” 3rd Int. Conf. on Multimodal Interfaces (ICMI2000), pp.222-229, 2000.