Detection and Recognition Technologies

Advances in Face Detection and Recognition Technologies

By Atsushi SATO,* Hitoshi IMAOKA,* Tetsuaki SUZUKI* and Toshinori HOSOI*

ABSTRACT This paper describes advances in the authors' face detection and recognition technologies. For face detection, a hierarchical scheme for combined face and eye detection has been developed based on the Generalized Learning Vector Quantization method to achieve precise face alignment. For face recognition, the perturbation space method has been improved to reduce the adverse effects of illumination changes as well as pose changes by using a standard face model. Experimental results have revealed that the proposed method outperforms our previously employed method.

KEYWORDS **Face detection, Face recognition, Face authentication, Biometric authentication**

1. INTRODUCTION

In recent years there have been great expectations of biometric authentication in view of increasing vicious crimes and terrorist threats. Biometric authentication is the automatic identification or identity verification of an individual based on physiological or behavioral characteristics such as fingerprint, iris, face, vein, voice, and so on. These kinds of authentications are most commonly used to safeguard international borders, control access to facilities, and enhance computer network security.

Face authentication has interesting characteristics that other biometrics do not have; facial images can be captured from a distance, any special actions are not always required for authentication, and a crime-deterrent effect can be expected because the captured images can be recorded and we can see who the person is at a glance. Due to such characteristics, the face recognition technique is expected to be applied widely not only to security applications such as video surveillance but also to image indexing, image retrievals and natural user interfaces.

The authors have previously developed face detection and recognition technologies[1]. Subsequently, several improvements have been accomplished on the previous method. This paper describes such advances in the authors' face detection and recognition technologies. For face detection, a hierarchical scheme for combined face and eye detection has been developed based on the Generalized Learning Vector Quantization method (GLVQ). For face recognition, the perturbation space method has been improved to reduce the adverse effects of illumination changes as well as pose changes. This paper is organized as follows: Face detection technology and face recognition technology are described in Section 2 and Section 3, respectively. Experimental results are given in Section 4, and conclusions are presented in Section 5.

2. FACE DETECTION AND ALIGNMENT

Face detection has two important tasks; one is to determine facial regions in an image against various backgrounds, and the other is to determine alignment of each face such as position, size and rotation to obtain better performance in face recognition. Assuming that the face is rigid, its pose is completely defined by six parameters; three coordinates and three rotation angles in three-dimensional space. If the face is a frontal view, its pose can be defined by four parameters, because the degree of rotational freedom decreases from three to one; that is, in-plane rotation. Therefore, most face detection algorithms focus on determining the position of both eyes in the image, which has four free parameters. To take account of its out-of-plane rotation, another facial feature point should be detected; the center of the mouth for example, or face recognition algorithms should be designed to be robust with respect to such variations.

Much work has been done in face detection so far[2]. Skin colors are often used to determine facial

*Media and Information Research Laboratories

regions because of ease of implementation, but its performance is easily degraded by illumination changes. View-based methods achieve very high performance for detecting faces against complicated backgrounds without using skin colors[3,4]. However, these methods consume much time for searching facial regions exhaustively over the image. Moreover, face alignment is not so precise in general, because such methods ignore high-frequency components of the image to speed up searching.

The authors have developed a hierarchical scheme for face detection and alignment in order to improve performance. **Figure 1** shows a block diagram of the proposed method. In face detection, the face position is roughly determined using low-frequency components by searching over multi-scale images taking account of in-plane rotation. And then in eye detection, the position of both eyes is determined precisely by a coarse-to-fine search using high-frequency components of the image. View-based method has been employed for these detectors in which templates are trained by Generalized Learning Vector Quantization (GLVQ) as shown next.

2.1 Generalized Learning Vector Quantization

GLVQ[5,6] is a learning method of templates, as used in nearest neighbor classifiers, based on Minimum Classification Error (MCE) criterion. MCE minimizes the smoothed empirical risk defined by

$$R_e(\theta) = \frac{1}{N}\sum_{n=1}^{N}\sum_{k=1}^{K}\ell(\rho_k(\mathbf{x}_n;\theta))1(\mathbf{x}_n \in \omega_k), \quad (1)$$

where $\mathbf{x}_n$ $(n=1,\cdots,N)$ and ω_k $(k=1,\cdots,K)$ denote training samples and classes, respectively, and $1(\cdot)$ is an indicator function such that 1(true)=1 and 1(false)=0. $\ell(\cdot)$ is a smoothed loss function defined by

$$\ell(\rho) = \frac{1}{1+\exp(-\xi\rho)}, \quad (2)$$

where ξ (>0) controls the slant of the sigmoid function, and when goes to infinity, Eq. (1) becomes identical to the empirical loss in Bayes decision theory. $\rho_k(\mathbf{x}_n;\theta)$ is called a misclassification measure as explained later. The classifier parameter θ can be updated for a given $\mathbf{x}_n$ to minimize the smoothed empirical risk as follows in an online learning form called probabilistic descent:

$$\theta \leftarrow \theta - \varepsilon\frac{\partial R_e(\theta)}{\partial\theta}, \quad (3)$$

$$\frac{\partial R_e(\theta)}{\partial\theta} = \sum_{k=1}^{K}\frac{\partial\ell(\rho_k(\mathbf{x}_n;\theta))}{\partial\theta}1(\mathbf{x}_n \in \omega_k). \quad (4)$$

For nearest neighbor classifiers, the classifier parameter consists of reference vectors called templates; that is, $\theta = \{\mathbf{m}_{ki} \mid k=1,\cdots,K; i=1,\cdots,N_k\}$ where N_k is the number of reference vectors in class ω_k. In GLVQ, the misclassification measure is defined as follows to ensure convergence of reference vectors:

$$\rho_k(\mathbf{x}_n;\theta) = \frac{d_k(\mathbf{x}_n;\theta) - d_l(\mathbf{x}_n;\theta)}{d_k(\mathbf{x}_n;\theta) + d_l(\mathbf{x}_n;\theta)}, \quad (5)$$

where $d_k(\mathbf{x}_n;\theta)$ is the squared Euclidian distance between $\mathbf{x}_n$ and the nearest reference vector $\mathbf{m}_{ki}$ of class ω_k to which $\mathbf{x}_n$ belongs, and likewise $d_l(\mathbf{x}_n;\theta)$ is the squared Euclidian distance between $\mathbf{x}_n$ and the nearest reference vector $\mathbf{m}_{lj}$ of the other classes. Then we can obtain GLVQ learning rule for these two reference vectors as follows:

$$\mathbf{m}_{ki} \leftarrow \mathbf{m}_{ki} + \varepsilon w(\rho_k(\mathbf{x}_n;\theta)) \times \frac{d_k(\mathbf{x}_n;\theta)}{\{d_k(\mathbf{x}_n;\theta) + d_l(\mathbf{x}_n;\theta)\}^2}(\mathbf{x}_n - \mathbf{m}_{ki}) \quad (6)$$

$$\mathbf{m}_{lj} \leftarrow \mathbf{m}_{lj} - \varepsilon w(\rho_k(\mathbf{x}_n;\theta)) \times \frac{d_l(\mathbf{x}_n;\theta)}{\{d_k(\mathbf{x}_n;\theta) + d_l(\mathbf{x}_n;\theta)\}^2}(\mathbf{x}_n - \mathbf{m}_{lj}) \quad (7)$$

where $w(\rho_k(\mathbf{x};\theta)) = 4\ell(\rho_k(\mathbf{x};\theta))\{1-\ell(\rho_k(\mathbf{x};\theta))\}$. In addition, GLVQ employs a simulated annealing technique to avoid getting trapped in local minima, in

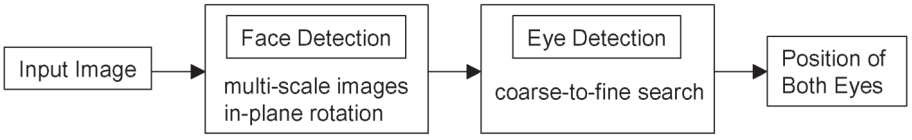

Fig. 1 Block diagram of the proposed face detection and alignment method.

which the slant parameter ξ is set to a small positive number at the beginning and is increased during learning.

2.2 Face Detection

Figure 2 shows the flow of the proposed face detection system. First, multi-scale images are generated from an input image, and then reliability maps are generated by GLVQ. Finally, these maps are merged through interpolation to obtain final results. The reliability maps are generated by scanning the multi-scale images with templates as shown in **Fig. 3**. The reliability is calculated by

$$\rho(\mathbf{x}) = (d_{NF}(\mathbf{x}) - d_F(\mathbf{x}))/(d_{NF}(\mathbf{x}) + d_F(\mathbf{x})), \quad (8)$$

where $d_F(\mathbf{x})$ and $d_{NF}(\mathbf{x})$ denote the minimum distances between a query sub-image $\mathbf{x}$ and templates which belong to face and non-face categories, respectively. The value of $\rho(\mathbf{x})$ ranges in [-1, 1], and if the value is positive, the query sub-image is regarded as face. **Figure 4** shows examples of face images and non-face images used in GLVQ training. The size of face was normalized on the basis of eye positions. To take account of in-plane rotation, rotated face images were also added to the training samples as shown in **Fig. 5**.

Since searching speed depends on the size of templates, reducing the image resolution is effective for speeding up searching. However, doing this degrades face detection performance, because high-frequency components of the image vanish. To solve this problem, high-frequency components are extracted before reducing the image size in our method. **Figure 6** shows an example of extracting high-frequency features from the image; that is, extracting edge strength for each direction. The size of these edge images is reduced by summing them up in local areas. Since the size of each multi-scale image can be reduced as well, the searching speed can be improved without noticeably degrading the performance.

2.3 Eye Detection

Eye detection determines the precise position of

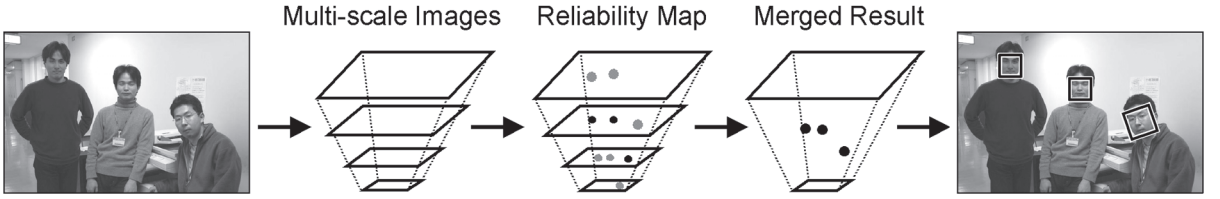

Fig. 2 Processing flow of the face detection.

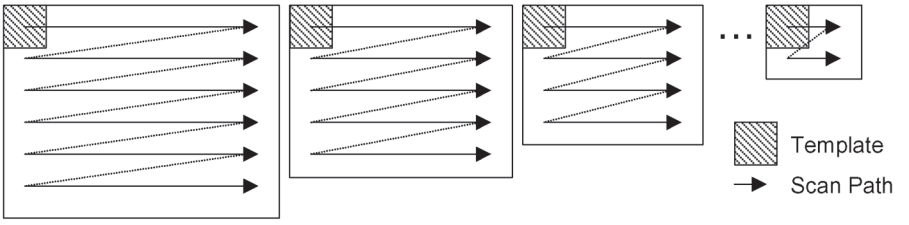

Fig. 3 Scanning multi-scale images with templates.

Fig. 4 Examples of face images and non-face images used in GLVQ training for face detection.

both eyes by a course-to-fine search as shown in **Fig. 7**. First, a grid is assumed on the image centered at the initial position of each eye, which was determined by face detection. Next, the reliability of the eye is calculated for each grid point in the same manner as in the face detection. The most likely point for each grid is selected for the candidate for the eye and then the size of the grid is reduced for the next search. Templates were trained by GLVQ with eye images and non-eye images as in the face detection. However, in order to preserve high-frequency components of the image, training samples were zoomed in to the eye regions as shown in **Fig. 8**.

3. FACE RECOGNITION

Face recognition is to calculate similarities or scores between a query facial image and enrolled facial images. Several methods have been proposed thus far; extracting effective features from whole facial images based on principal component analysis[7], extracting local features based on local feature analysis[8], and extracting relative positions of facial landmarks[9]. The performance of face recognition degrades by two factors: one is global image variations caused by pose and illumination changes, and the other is local image variations caused by facial expression, aging, wearing glasses and so on. However, such variations are not taken into account explicitly in the above previously discussed methods.

To reduce such adverse effects, two key technologies have been developed; the perturbation space method for global image variations, and adaptive regional blend matching for local image variations.

3.1 Perturbation Space Method

For reducing the adverse effect of global image variations, a model based approach is very effective, because such variations are caused by physical properties. As mentioned, the position and the rotation of the face in three-dimensional space can be described by six degrees of freedom. Moreover, it has been shown that illumination changes can be described no more than ten degrees of freedom[10]. Thus, face recognition using a three-dimensional facial model improves the recognition performance greatly even if a pose or illumination changes. However, this method requires the individual's own facial shape and reflectance on the surface, thus it cannot be applied to ordinary face recognition systems based on image matching.

We have developed a standard face model to

Fig. 5 Examples of face images taking account of in-plane rotation.

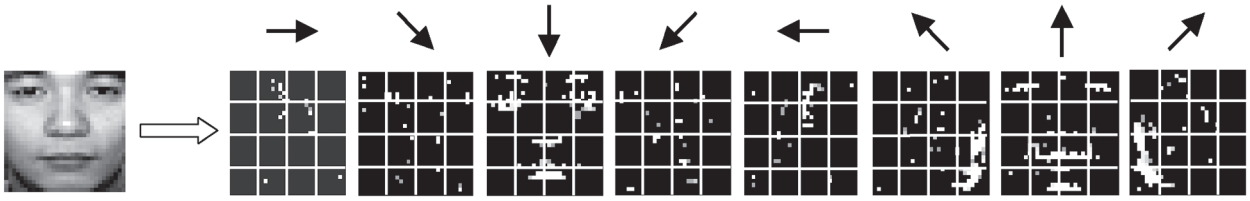

Fig. 6 Example of extracting edge strength for each direction.

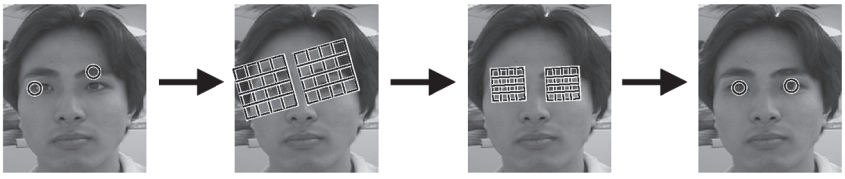

Fig. 7 Processing flow of course-to-fine search in eye detection.

Fig. 8 Examples of eye images and non-eye images used in GLVQ training for eye detection.

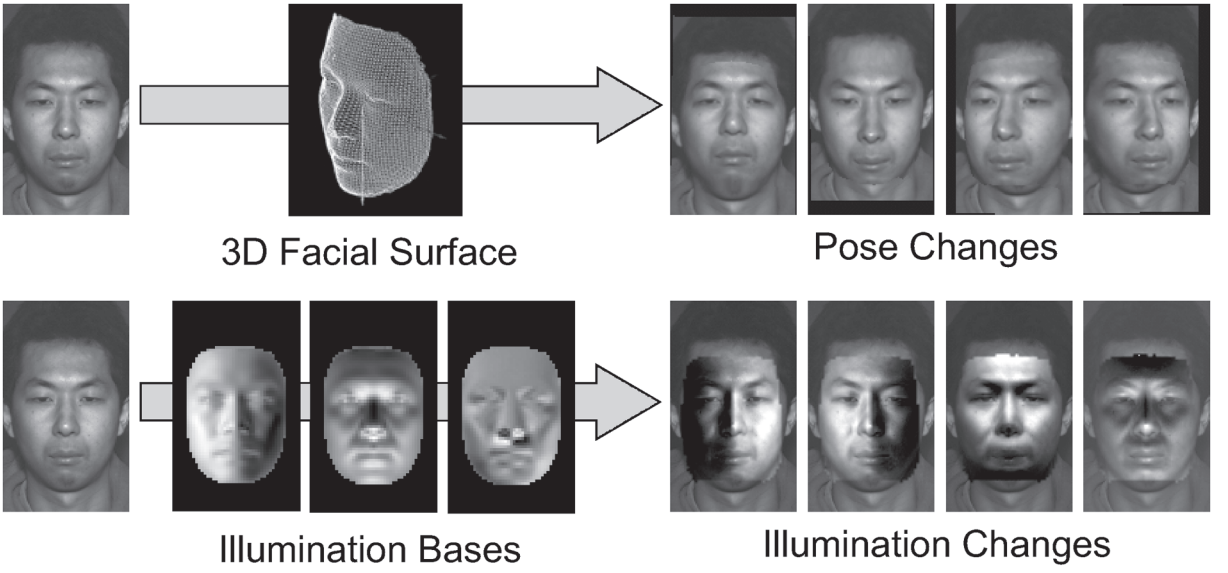

Fig. 9 Conceptual figure of generation of various facial images using standard face model.

generate various facial appearances. This face model consists of a three dimensional facial surface and a set of illumination bases as shown in **Fig. 9**. For pose changes, the enrolled image is mapped onto the three dimensional facial surface, and is rotated in three dimensional space to simulate out-of-plane pose changes. For illumination changes, various shaded images are generated from the enrolled image using the illumination bases which were obtained by executing principal component analysis (PCA) on various facial images.

If all of these generated images are enrolled to the database, it may take long to compare between a query and the database in the recognition process. To solve this problem, these images are compressed using PCA, and the comparison is executed using the following equation:

$$D^2(\mathbf{x}) = \| \mathbf{x} - \mu \|^2 - \sum_{j=1}^{M} \{\phi_j^T (\mathbf{x} - \mu)\}^2 \qquad (9)$$

where $\mathbf{x}$ is the query image, μ is the mean of all generated images, and ϕ_j is the j-th eigenvector obtained by PCA. Eq. (9) denotes the squared distance between $\mathbf{x}$ and the subspace spanned by the eigenvectors. This method, which is called the perturbation space method, achieves fast comparisons in the

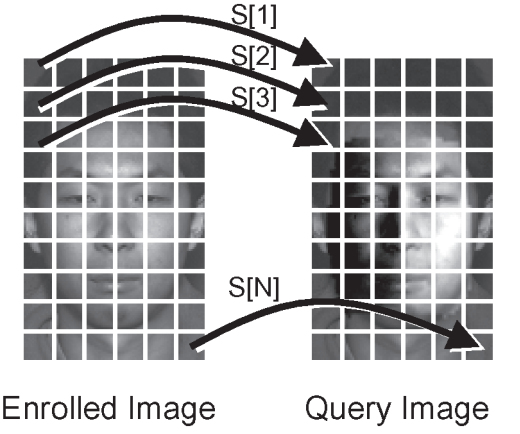

Fig. 10 Conceptual figure of adaptive regional blend matching.

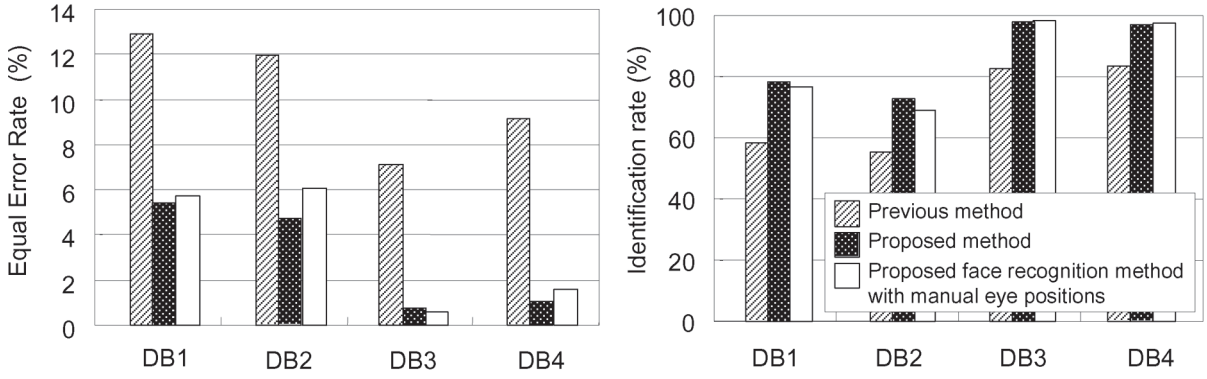

Fig. 11 Experimental results of face recognition for several databases.

recognition process, taking account of pose and illumination changes.

3.2 Adaptive Regional Blend Matching

For local image variations, robust image matching is desirable because the modeling of facial expressions change and aging effects, or wearing glasses is quite a difficult problem. To reduce the adverse effects of such local changes, adaptive regional blend matching has been developed.

As shown in **Fig. 10**, the enrolled image and the query image are divided into N segments, and a score S[i] is calculated for each pair of segments by the perturbation space method. Some scores having larger values are taken into account for calculating the final score, instead of using all of them. This means that segments in the query image which are quite different from the corresponding segments in the enrolled image are neglected for calculating the final score, so that this method achieves robust matching with respect to local image variations.

4. EXPERIMENTS

Face recognition experiments were conducted to demonstrate the effectiveness of the proposed method for several databases collected in our laboratories. DB1 and DB2 contain aging effects, DB3 contains facial expression changes, and DB4 contains illumination changes. The number of persons represented in each database is from 200 to 1,000.

Figure 11 shows experimental results by the proposed method compared with our previous method. Equal error rate (EER) is the performance measure of verification in which the threshold for scores is tuned so that its false acceptance rate is equal to its false rejection rate. The identification rate is the probability that the 1st candidate having the highest score is the genuine person. The performance of the proposed method is drastically improved compared with the previous method, and its performance is almost the same as the case that eye positions were given by hand instead of using face detection. It can be said that the accuracy of eye positions by the proposed face detection method is almost the same as human estimation.

5. CONCLUSION

This paper has described advances in the authors' face detection and recognition technologies. For face detection, a hierarchical scheme for combined face and eye detection has been developed, based on the Generalized Learning Vector Quantization method. For face recognition, the perturbation space method has been improved to reduce the adverse effects of illumination changes as well as of pose changes. Experimental results reveal that the proposed method achieves much higher performances for several databases than by our previous method. It is also revealed that the accuracy of eye positions by the proposed face detection method is almost the same as for human estimation. This would be greatly helpful for putting automatic face recognition into practice. In future works, estimating three-dimensional facial shape from an image should be developed to improve the recognition performance even further.

REFERENCES

[1] A. Sato, A. Inoue, et al., "NeoFace - Development of Face Detection and Recognition Engine," *NEC Res. & Develop.*, **44**, 3, pp.302-306, July 2003.
[2] E. Hjelman and B. K. Low, "Face Detection: A Survey," *Computer Vision and Image Understanding*, **83**, 3, pp.236-274, 2001.
[3] H. A. Rowley, S. Baluja and T. Kanade, "Neural NetWork-Based Face Detection," *IEEE Trans. PAMI*, **20**,

1, pp.23-38, 1998.
[4] K. K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection," *IEEE Trans. PAMI*, **20**, 1, pp.39-51, 1998.
[5] A. Sato and K. Yamada, "Generalized Learning Vector Quantization," *In Advances in Neural Information Processing Systems*, **8**, pp.423-429, MIT Press, 1996.
[6] A. Sato, "Discriminative Dimensionality Reduction Based on Generalized LVQ," *In Artificial Neural Networks - ICANN2001*, pp.65-72, Springer LNCS 2130, 2001.
[7] M. Turk and A. Pentland, "Eigenfaces for Recognition," J. of Cognitive Neuroscience, **3**, 1, pp.71-86, 1991.
[8] P. Penev and J. Atick, "Local Feature Analysis: A General Statistical Theory for Object Representation," *Network: Computation in Neural Systems*, pp.477-500, 1996.
[9] L. Wiskott, J. M. Fellous, et al., "Face Recognition by Elastic Bunch Graph Matching," *IEEE Trans. PAMI*, **19**, 7, pp.775-779, 1997.
[10] R. Ishiyama and S. Sakamoto, "Geodesic Illumination Basis: Compensating for Illumination Variations in Any Pose for Face Recognition," *In Proc. of the Int. Conf. on Pattern Recognition*, **4**, pp.297-301, 2002.

Received January 17, 2005

* * * * * * * * * * * * * * * *

Atsushi SATO received his D.S degree in physics from Tohoku University 1989. He joined NEC Corporation in 1989, and is now a principal researcher at the Media and Information Research Laboratories. His research interests include image processing, pattern recognition and artificial neural networks. During 1994-1995, he was a visiting researcher in Department of Electrical Engineering at University of Washington.

Dr. Sato is a member of the Institute of Electronics, Information and Communication Engineers (IEICE) and the Physical Society of Japan (JPS). He received the Convention Award from the Information Processing Society of Japan (IPSJ) and from IEICE in 1991 and 1994, respectively.

Hitoshi IMAOKA received his Dr. Eng. degree in statistical mechanics from Osaka University in 1997. He joined NEC Corporation in 1997, and is now a senior researcher at the Media and Information Research Laboratories. His research interests include pattern recognition, and image processing.

Dr. Imaoka is a member of the Japanese Neural Network Society and the Institute of Electronics, Information and Communication Engineers (IEICE).

Tetsuaki SUZUKI received his M.E. degree in computer science from the Tokyo Institute of Technology in 2000. He joined NEC Corporation in 2000, and is now a researcher at the Media and Information Research Laboratories. His research interests include pattern recognition and color image processing.

Mr. Suzuki is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

Toshinori HOSOI received his M.E. degree in mechanical science from Osaka University in 2001. He joined NEC Corporation in 2001, and is now a researcher at the Media and Information Research Laboratories. His research interests include pattern recognition and image processing.

Mr. Hosoi is a member of the Institute of Electronics, Information and Communication Engineers (IEICE).

* * * * * * * * * * * * * * * *