IEICE TRANS. INF. & SYST., VOL.E88–D, NO.10 OCTOBER 2005

PAPER Special Section on Image Recognition and Understanding

An Efficient Search Method Based on Dynamic Attention Map by Ising Model

Kazuhiro HOTTA[†a)], **Masaru TANAKA**[††], **Takio KURITA**[†††], *and* **Taketoshi MISHIMA**[††], *Members*

SUMMARY This paper presents Dynamic Attention Map by Ising model for face detection. In general, a face detector can not know where faces there are and how many faces there are in advance. Therefore, the face detector must search the whole regions on the image and requires much computational time. To speed up the search, the information obtained at previous search points should be used effectively. In order to use the likelihood of face obtained at previous search points effectively, Ising model is adopted to face detection. Ising model has the two-state spins; "up" and "down". The state of a spin is updated by depending on the neighboring spins and an external magnetic field. Ising spins are assigned to "face" and "non-face" states of face detection. In addition, the measured likelihood of face is integrated into the energy function of Ising model as the external magnetic field. It is confirmed that face candidates would be reduced effectively by spin flip dynamics. To improve the search performance further, the single level Ising search method is extended to the multilevel Ising search. The interactions between two layers which are characterized by the renormalization group method is used to reduce the face candidates. The effectiveness of the multilevel Ising search method is also confirmed by the comparison with the single level Ising search method.

key words: *Ising model, dynamic attention map, renormalization group, efficient search, face detection*

1. Introduction

Face detection is the first essential step for automatic face recognition. Since automatic face recognition has many potential applications [1]–[6], face detection becomes an active research area [7], [8]. Some frontal face detection methods give high detection rate under the restricted environment [7]–[12]. However, in general, the face detector must search the whole regions on the input image because the system can not know where faces there are and how many faces there are in advance. Therefore, face detection requires much computational time. The efficient search algorithm without decreasing the detection accuracy is required.

The face detection methods based on color information of faces do not require much computational cost [13], [14]. However, it is not easy to detect the correct position of faces from only color information. Color information of faces is effective to reduce the candidates of faces. Rowley et al. [11] make the face detector be invariant to translations about 25%. The search speed is improved by using coarse search. However, in general, there is the trade-off between the search speed and the false detection rate. On the other hand, we applied the random search method, in which the search point is selected randomly, to face detection method [15]. Although the average speed of the random search is improved, the speed of the random search depends on the random number and is unstable. The reasons are as follows. (1) Each matching in random search is performed independently. (2) The random search method does not utilize the information (the likelihood of face) obtained at previous search points. In this paper, in order to use the likelihood of face obtained at previous search points effectively, Ising model [16], [17] is adopted to face detection [18].

Ising model is the simplest model of magnetization. It can take only one state of "up" and "down". The state of the spin depends on both the state of the neighboring spins and an external magnetic field. Since face detection problem has only two states; "face" and "non-face", we can assign Ising spins to "face" and "non-face" states of face detection. In our face detection method, the proximity to the mean vector of face class in discriminant space represents the likelihood of face. The neighboring spins of previous search point (the selected spin) are expected to have similar likelihood of face as that of previous search points. If the measured likelihood of face is integrated into the energy function of Ising model as the external magnetic field, then the states of the spins in neighboring region of previous search point can be estimated through spin flip dynamics. This paper demonstrates that the search space for face detection is narrowed down effectively by spin flip dynamics which use the state of neighboring spins and the measured likelihood of face. Furthermore, we extend Ising search to multilevel Ising search by taking the renormalization group method into consideration [19], [20]. In the multilevel Ising search, Ising model is adopted to the different layers and the interaction between the different layers is used to reduce the face candidates. The effectiveness of the multilevel Ising search method is also confirmed by the comparison with single level Ising search method.

In Sect. 2, we explain a scale and rotation invariant face detection method used in this paper. The performances of that method are also shown in Sect. 2. Ising model, which is the simplest model of magnetization, is explained in Sect. 3. How to adopt Ising model to face detection is explained in

Manuscript received October 8, 2004.
Manuscript revised February 1, 2005.

[†]The author is with The University of Electro-Communications, Chofu-shi, 182–8585 Japan.
[††]The authors are with Saitama University, Saitama-shi, 338–8570 Japan.
[†††]The author is with National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba-shi, 305–8568 Japan.
a) E-mail: hotta@ice.uec.ac.jp

DOI: 10.1093/ietisy/e88–d.10.2286

Sect. 4. Section 5 is for the experimental results of single level Ising search method. In Sect. 6, Ising search method is extended to multilevel Ising search method. The effectiveness of the multilevel Ising search method is shown in Sect. 7. Finally, conclusion is described in Sect. 8.

2. Scale and Rotation Invariant Face Detection Method

This section gives a review of the scale and (2D) rotation invariant face detection method using Higher-order Local AutoCorrelation (HLAC) features extracted from Log-Polar image [15]. That method consists of the following three steps.

(1) The center point of Log-Polar transformation is set to a certain position on the input image and a Log-Polar image is constructed.
(2) HLAC features are extracted from the Log-Polar image.
(3) The extracted features are projected into the discriminant space for face and non-face classification. Then measure the proximity to mean of face class (likelihood of face) and decide face or non-face.

By applying this process to all positions on the input image, the face detector can find faces in the image. Figure 1 shows the flow of face detection. In the training of the face class, the center point of Log-Polar transform is set to the top of one's nose. Therefore, that detection method is invariant to scalings and rotations in terms of top of one's nose. First, we explain Log-Polar transformation [21]–[24] and HLAC features [25], [26].

2.1 HLAC Features Extracted from Log-Polar Image

Input image is generally represented as a collection of pixel points on the Cartesian coordinate in which the origin is at the middle pixels in the height and width of the image. Log-Polar image can be constructed by the following transformations of the coordinates. At first, the point (x, y) on the Cartesian coordinate is transformed into the point $(\rho = \sqrt{(x^2 + y^2)}, \theta = \arctan(y/x))$ on the Polar coordinate. The point on the Polar coordinate is transformed into the point $(z = \log(\rho), \theta)$ on the Log-Polar coordinate by taking the logarithm of the scale ρ. Figure 2 (a) and (b) show Cartesian coordinate (input image) and Log-Polar coordinate (image).

In this paper, we use the re-sampling method by the inverse transformation to obtain the Log-Polar image from the input image. To obtain the pixel value at the point (z_i, θ_j) on the Log-Polar image, the point is inversely transformed into the point $(\exp(z_i)\cos(\theta_j), \exp(z_i)\sin(\theta_j))$ on the Cartesian coordinate. Then the value of the point (z_i, θ_j) is estimated as the mean intensity value of the neighboring points of the back-projected point $(\exp(z_i)\cos(\theta_j), \exp(z_i)\sin(\theta_j))$ on the input image. We can obtain a Log-Polar image by performing this estimation for all points on the Log-Polar coordinate. Figure 3 (a)–(c) show an input image, the sampling points used to construct the Log-Polar image, and its Log-Polar image. Note that the sampling density decreases from the center to the periphery. This means that the extracted features contain much information of the target on the central region than that on the peripheral regions such as background.

Log-Polar image has a good property for scale and rotation invariant feature extraction. Scalings of a target are represented as the shifts along z (= $\log(\rho)$) axis on the Log-Polar image. Rotations of a target are also represented as the shifts along θ axis. Figure 4 (a)–(f) show Log-Polar image of a simple 2D shape with different scales and rotations. Both scalings and rotations of the target are represented as the shifts in Log-Polar image.

To obtain the scale and rotation invariant features, we have to extract shift invariant features from the Log-Polar image because the scalings and rotations are represented as the shifts in Log-Polar image. In this paper, HLAC features [25], [26] are utilized as shift invariant features.

It is well known that autocorrelation function is shift-invariant. Its extension to higher orders is higher-order autocorrelation function. The Nth-order autocorrelation functions with N displacements $(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N)$ from the refer-

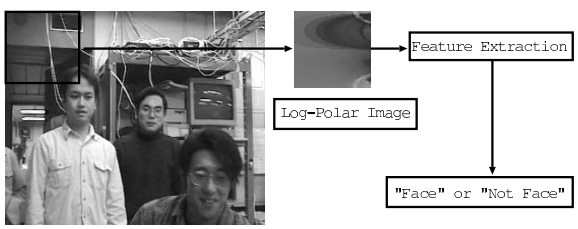


Fig. 1 The flow of the scale and rotation invariant face detection method.

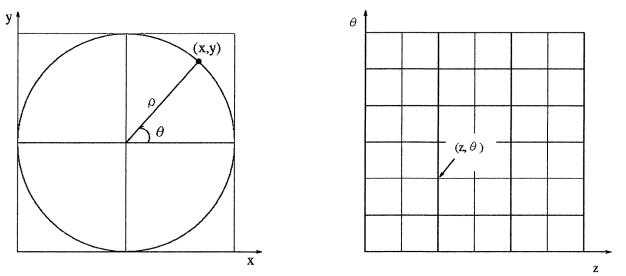


Fig. 2 Cartesian coordinate and Log-Polar coordinate.

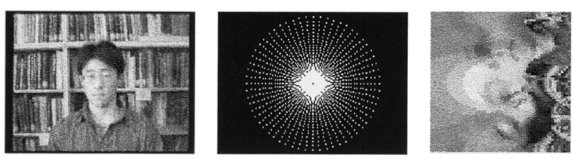


Fig. 3 Log-Polar transformation. (a) Input image (160 × 120 pixels). (b) Sampling points used to construct the Log-Polar image. (c) Log-Polar transformed image (60 × 60 pixels).

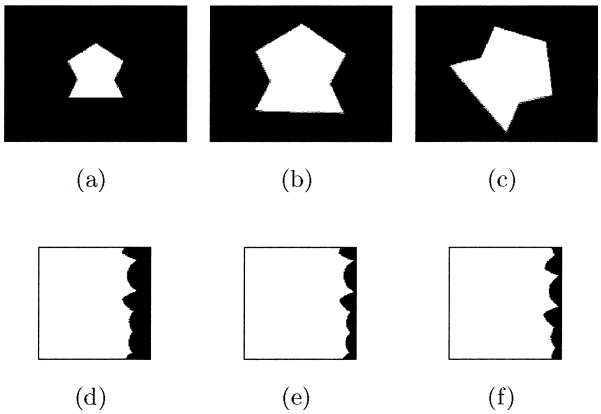


Fig. 4 Examples of Log-Polar image of 2D shapes. (a) Small size. (b) Normal size. (e) 45° rotated image of (b). (d) Log-Polar image of (a). (e) Log-Polar image of (b). (f) Log-Polar image of (c).

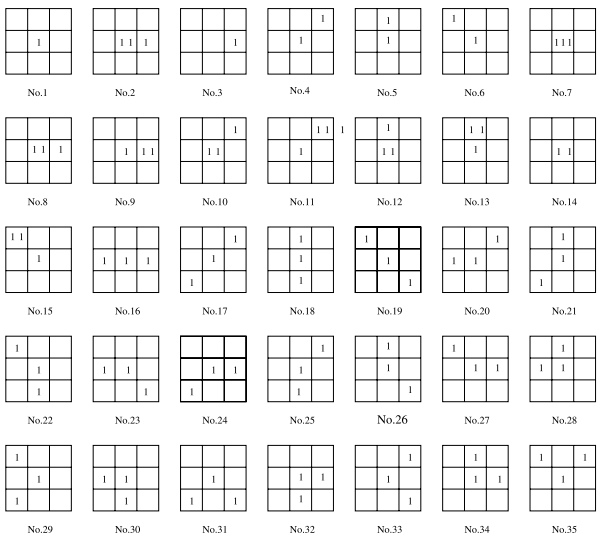


Fig. 5 The 35 local mask patterns.

ence point $\boldsymbol{r}$ are defined by

$$x^N(\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_N) \int I(\boldsymbol{r})I(\boldsymbol{r}+\boldsymbol{a}_1)\cdots I(\boldsymbol{r}+\boldsymbol{a}_N)d\boldsymbol{r}, \quad (1)$$

where function $I(\boldsymbol{r})$ denotes a intensity value of a Log-Polar image and $\boldsymbol{r} = (z = \log(\rho), \theta)$. Since the number of these autocorrelation functions obtained by the combination of the displacements over the image are enormous, we must reduce them for practical application. At first, we restrict the order N up to the second ($N = 0, 1, 2$). Then, we also restrict the range of displacements within a local 3×3 window, because the correlation within local region is much higher than the correlation between far points. In other words, we consider the autocorrelations up to the three points within 3×3 window. By eliminating the displacements to which are equivalent by shift, the number of the patterns of the displacements is reduced to 35. Figure 5 shows 35 mask patterns. The features are obtained by scanning the Log-Polar image with the 35 local 3×3 mask patterns and by computing the sums of the products of the corresponding pixels to "1" in the mask patterns. The "11" and "111" in the mask patterns represent the square and the cube of the same pixel value. Since these features are obviously invariant to the shift, HLAC features extracted from the Log-Polar image become robust to linear scalings and rotations of a target in the input image.

In following experiments, the color informations are used to improve the accuracy of face and non-face classification. Since RGB informations correlate each other, we utilize the color representation (R+G+B)/3, R-B, (2G-R-B)/2, which are obtained by Principal Component Analysis of colors [27], to obtain the independent color information. HLAC features of Log-Polar image are extracted from each color independently. Namely, the number of features is 105 (= 35 features × 3 colors).

2.2 Face and Non-face Classification Based on Linear Discriminant Analysis

HLAC features extracted from a Log-Polar image are general, primitive, and independent of the recognition task. These features have enough information to discriminate faces from non-faces. To get new effective features for the given recognition task (face and non-face classification), it is necessary to combine these features. For this purpose, we use Linear Discriminant Analysis (LDA).

For face detection, we have to design a classifier which can classify face and non-face. It is expected that face class includes only face images, but non-face class includes many kinds of images except face images. It is difficult to deal with non-face class as a single cluster in the feature space. Thus we modified the discriminant criterion such that the covariance of face class is minimized while the covariance between face class and each sample in non-face class is maximized.

In this paper, face class and non-face samples are represented as

$$C_F = \{\boldsymbol{x}_{Fi} \mid i = 1, \cdots, N_F\},$$
$$C_{NF} = \{\boldsymbol{x}_{NFk} \mid k = 1, \cdots, N_{NF}\}, \quad (2)$$

where N_F is the number of face samples and N_{NF} is the number of non-face samples. The mean vector of face class, the covariance matrix (Σ_F) of face class, and the covariance matrix (Σ_C) between the mean vector of face class and each sample of non-face class are given by

$$\bar{\boldsymbol{x}}_F = \frac{1}{N_F}\sum_{i=1}^{N_F}\boldsymbol{x}_{Fi},$$
$$\Sigma_F = \frac{1}{N_F}\sum_{i=1}^{N_F}\boldsymbol{x}_{Fi}\boldsymbol{x}_{Fi}^T - \bar{\boldsymbol{x}}_F\bar{\boldsymbol{x}}_F^T,$$
$$\Sigma_C = \frac{1}{N_{NF}}\sum_{k=1}^{N_{NF}}(\boldsymbol{x}_{NFk} - \bar{\boldsymbol{x}}_F)(\boldsymbol{x}_{NFk} - \bar{\boldsymbol{x}}_F)^T, \quad (3)$$

where the symbol ^{T} denotes the transpose.

New features $\boldsymbol{y}$ are obtained by linear combination of primitive features $\boldsymbol{x}$ as $\boldsymbol{y} = A^T\boldsymbol{x}$, where $A = [a_{ij}]$ is a coefficients matrix.

To construct the discriminant space in which the covariance of face class is minimized and the covariance between the mean vector of face class and each sample of non-face class is maximized, we use the discriminant criterion $J = \boldsymbol{tr}\left(\hat{\Sigma}_F^{-1}\hat{\Sigma}_C\right)$, where $\hat{\Sigma}_F$ and $\hat{\Sigma}_C$ are the covariance matrix of face class and the covariance matrix between the mean vector of face class and each sample of non-face class in the discriminant space, respectively. The optimal coefficient matrix A, which maximizes this discriminant criterion J, is obtained by solving the eigen-value problem

$$\Sigma_C A = \Sigma_F A \Lambda \qquad \left(A^T\Sigma_F A = I\right). \tag{4}$$

The dimension L of discriminant space is given as the lowest value that satisfies

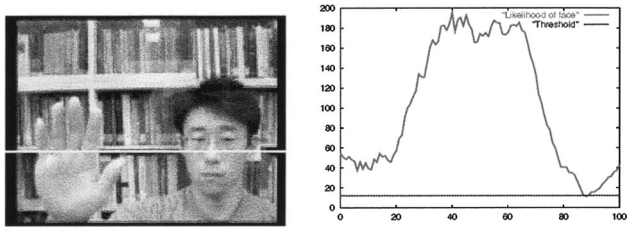
$$98\% \leqq \frac{\sum_{i=1}^{L}\lambda_i}{\sum_{j=1}^{M}\lambda_j}, \tag{5}$$

where M is the dimension of primitive features.

To investigate whether the unknown input is face or non-face, the proximity to the mean vector of face class in discriminant space is evaluated. If the measured distance is below a certain threshold, then the input is classified as face class. In this classification, the performance of face detection depends on the value of the threshold. The optimal threshold is experimentally determined by using the following two probabilities. The first probability is $P_F = 1 - n_F/N_F$ in which the samples of face class are miss-classified as non-face, where n_F is the number of the samples of face class which has a value less than the threshold and N_F is the total number of samples of face class. The second probability is $P_{NF} = n_{NF}/N_{NF}$ in which the samples of non-face class are miss-classified as face, where n_{NF} is the number of non-face samples which has a value less than the threshold and N_{NF} is the number of non-face samples. We define the threshold as a distance from the mean of the face class. As the threshold is changed from zero to infinity, two probabilities vary depending on the threshold. Since these two probabilities are error probabilities, we would like to minimize both error probabilities. Thus we can select the optimal threshold in which the sum of the two probabilities is minimized.

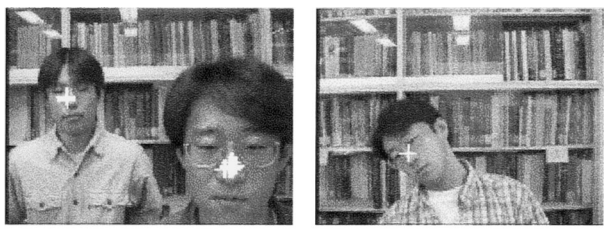
2.3 Performance of Face Detection

First, we investigated how the proximity to the mean of face class changes, when the searched point is moved along the white line in Fig. 6 (a). For good face detection, only the point whose center corresponds to the center of a face (one's nose) must be below the threshold and other points must be above the threshold. The results are shown in Fig. 6 (b). The horizontal dotted line in Fig. 6 (b) shows the value of optimal threshold determined by the experiment. The image in Fig. 6 (a) includes face and hand with same size and same color. Although we find the local minimum at hand regions, their distances are bigger than the threshold. On the other hand, only the region in which a face locates at center gives the distances below threshold. These results show that the points, which have the distance below threshold, contain certainly face.



(a) (b)

Fig. 6 The proximity to the mean vector of face class in discriminant space.



(a) (b)

Fig. 7 Examples of the scale and rotation invariant face detection. (a) Two persons with different scales. (b) One person inclined his face.

Examples of the scale and rotation invariant face detection are shown in Fig. 7. The white cross is plotted on the center of the detected region as face. If the white crosses are on one's nose, then the detection is correct. Figure 7 (a) represents the example of two persons with different scales. In spite of the different scales, two faces are detected correctly without changing the size of the image. Figure 7 (b) is the example of the image in which human inclines his face. His face is also detected correctly. This is because HLAC features extracted from Log-Polar image are robust to scalings and (2D) rotations of a face.

These results are obtained by searching the whole regions on the image. In general, the system must search the whole regions on the image because it is difficult to know where are faces and how many faces there are in advance. However, the exhaustive search method is not practical because of its computational cost. Previously we applied the random search method, in which the search point is selected randomly, to the face detection [15]. The average speed of the search is improved by the random search method. However, the speed of random search depends on the random numbers and is unstable because the random search method does not make use of the information obtained at previous search points. In order to use of the likelihood of face obtained at previous search points effectively, Ising model is adopted to face detection.

3. Ising Model

Ising model is the simplest model of magnetization [16], [17]. It consists of two state Ising spins; "up" and "down". The state of the spin depends on both the state of neighboring spins and the external magnetic field. Originally, Ising model was proposed as a simplified version of Heisenberg model, which consists of two state spins and interactions between all spins. Now this Ising model is quite famous for its usefulness in the fields of physics and neural networks. The energy of the spin s_i is given by

$$E_i = -J \sum_{j \in nn(i)} s_i s_j - H s_i, \tag{6}$$

where E_i is the energy of the ith spin s_i, J is a coupling constant of the spins, H is an external magnetic field, $nn(i)$ represents the nearest neighboring spins of the spin s_i, and the spin s_i called Ising spin takes 1 ("up") or -1 ("down"). The state of the spin is updated according to the probability which is proportional to $\exp(-\beta \Delta E_i)$, where β is a reciprocal of the temperature and ΔE_i is the energy change caused by flipping the spin s_i. The energy change ΔE_i caused by flipping the spin s_i is given by

$$\Delta E_i = 2J \sum_{j \in nn(i)} s_i s_j + 2H s_i. \tag{7}$$

The dynamics of Ising model work to minimize the total energy E.

4. Dynamic Attention Map by Ising Model

From Fig. 6 (b), we understand that the proximity to the mean vector of face class in discriminant space represents the likelihood of face. The shorter distance in the discriminant space means the higher likelihood of face. The spin flip dynamics works to minimize the energy function which includes the state of neighboring spins and an external magnetic field. If we integrate the measured likelihood of face into the energy function of Ising model as an external magnetic field, then the state of neighboring spins of the selected spin can be estimated through spin flip dynamics. This can be used to reduce the search space dynamically by introducing the information obtained at previous searched points.

In face detection, there are also two states; "face" and "non-face". Here we set "face" to -1 ("down") and "non-face" to 1 ("up"). The direction of an external magnetic field (H) is assumed to be "up" basically, because the non-face regions in the images are wider than that of face. However, the direction and magnitude of an external magnetic field should be changed adaptively by the measured likelihood of face. For example, if the measured likelihood of face is high, then the strong external magnetic field should be given toward "down" (the direction of "face"). On the other hand, if the measured likelihood of face is low, then the strong external magnetic field should be given toward "up" (the direction of "non-face"). To do this, $H_d(m_d(a) - \theta_d)$ is used as a external magnetic filed, where H_d is the coefficient of the external magnetic field, $m_d(a)$ is the measured likelihood of face (the proximity to the mean vector of face class in discriminant space) of the spin s_a, and θ_d is the threshold to classify face and non-face. In this formulation, the magnitude of the external magnetic field is changed according to the measured likelihood of face. In addition, the direction of the external magnetic field changes by depending on both threshold and measured likelihood of face. The energy function reflecting the measured likelihood of face ($m_d(a)$) of a selected spin s_a is given by

$$E_i = -J \sum_{j \in nn(i)} s_i s_j - H_d \left(m_d(a) - \theta_d\right) s_i, \quad i \in NN(a), \tag{8}$$

where $NN(a)$ means the neighboring spins of the spin s_a for performing the spin flip dynamics, E_i is the energy of the spin s_i, and $nn(i)$ represents the nearest neighboring spins of the spin s_i.

The color information of faces is effective to reduce the search space for face detection. We can easily integrate the likelihood of face in terms of the color information into the energy function of Ising model. The discriminant space in terms of color information is constructed by using color signals (RGB data) extracted from face and non-face images. The proximity to the mean of face class in the discriminant space in terms of color information represents the likelihood of face in terms of color information. The optimal threshold for discrimination using color information is determined by an experiment.

The likelihood of face in terms of color information is integrated into the energy of the spin s_i as

$$E_i = -J \sum_{j \in nn(i)} s_i s_j - H_d \left(m_d(a) - \theta_d\right) s_i - H_c \left(m_c(a) - \theta_c\right) s_i, \tag{9}$$

where H_c represents the coefficient of the external magnetic field in terms of color informations, $m_c(a)$ is the measured likelihood of face using color information, and θ_c is the threshold for discrimination using color information.

Then the state of each spin is updated according to the probability which is proportional to $\exp(-\beta \Delta E_i)$, where ΔE_i is the energy change caused by flipping the spin s_i, that is,

$$\Delta E_i = 2J \sum_{j \in nn(i)} s_i s_j + 2H_d \left(m_d(a) - \theta_d\right) s_i + 2H_c \left(m_c(a) - \theta_c\right) s_i. \tag{10}$$

We call the spin map obtained after spin flip dynamics iteration "Dynamics Attention Map". Example of Dynamic Attention Map is shown in Fig. 8. White pixels in the image represent the spins remained as face candidates. The state of spins are changed dynamically through spin flip dynamics and face candidates are narrowed down effectively.

In the following, the meta algorithm for Ising search

Fig. 8 Example of dynamic attention map.

method using Dynamic Attention Map is shown.

1. Set all spins to −1 ("face" state) and make the face list for search. The list consists of the spins whose state is "face".
2. Select one spin s_a randomly from the face list.
3. Extract the HLAC features from Log-Polar image of the region centered at the selected spin s_a. Then measure the likelihood of face in the discriminant space and that in the discriminant space in terms of color information of the selected spin s_a.
4. Apply the spin flip dynamics for suitable number of iterations within the neighboring regions of the selected spin s_a.
5. Remove the spins flipped from "face" to "non-face" from the face list and add the spins flipped from "non-face" to "face" to the face list. Note that, if the selected spin to measure the likelihood of face is found to be "non-face", that spin is removed from the face list and never flipped again. (The spin, which is found to be "non-face", works as the magnetic field. Its direction is "up".)
6. By repeating from 2 to 5, face candidates are narrowed down effectively.

5. Evaluation of Ising Search Method

To investigate the effectiveness of Ising search method, Ising search method is compared with the random search method. In this experiment, the number of search is evaluated when one spin below threshold is found. If face candidates are narrowed down effectively, face is detected with the small number of search.

The parameters of Ising search method are set to $\beta = 0.2$, $J = 1.0$, $H_d = 0.25$, $H_c = 4.0$, and Monte Carlo steps (MCS) are performed 5 times in the neighboring 5×5 lattice centered the selected spin. MCS is the process which includes the spin selection, energy computation, and spin flip dynamics. The likelihood of face in terms of color information is much smaller than that in terms of HLAC features of Log-Polar image. In order to compensate for that gap, we use large H_c.

Experiments are performed for four cases. Figure 9 (a)–(d) shows the face detection results obtained by exhaustive search method. The white crosses in this Figure represent the center of the detected region. The size of these

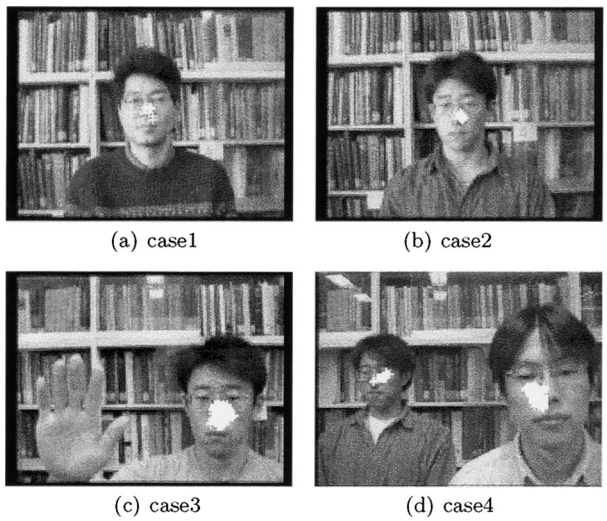


Fig. 9 The results obtained by using exhaustive search method.

Table 1 The comparison of the performance among the random search method and Ising search method.

	mean		median	
	Ising	random	Ising	random
case 1	213.32	580.26	217	474
case 2	211.29	551.03	217	361
case 3	46.35	58.70	34	42
case 4	54.25	71.06	45	50

images is 160×120 pixels. In this paper, the window size for Log-Polar transformation is set to 60×60 pixels, and the peripheral 32 pixels of the input image are not used as the center points of Log-Polar transformation. Therefore, the number of face candidates in the input image is 5,376. The case 1 is one person in an image with face to the non-face pixel-ratio 8/5,368. The case 2 is one person in an image with face to the non-face pixel-ratio 9/5,367. The case 3 is one person and his hand with face to the non-face pixel-ratio 91/5,285. The case 4 is two person at the different size with face to the non-face pixel-ratio 89/5,287. Ising and random search were repeated 100 times for each case with changing the random number. The search was continued until face is detected. In this experiment, there is no failure. The failure means that the face list becomes empty before finding a face.

The mean and median number of search are shown in Table 1. From Table 1, we understand that the number of Ising search is about the half of random search. This result shows that Ising search can narrow down face candidates effectively. Figure 10 shows how the number of face candidates decreases. The random search makes the number of the face candidates decrease gently. On the other hand, Ising search makes the number of the face candidates decrease steeply. Note that Ising search method makes the search space narrower effectively than the random search method.

When an input image does not include faces, the search time of the random search method corresponds to the exhaustive search method. On the other hand, Ising search

method does not require so much time because Ising search method can make the number of face candidates decrease through spin flip dynamics. In the upper experiments, the search is finished if only one face is detected. This search method can not detect multiple faces. If Ising search is performed until the number of face candidates becomes zero, multiple faces can be detected. Since the face candidates are already decreased through spin flip dynamics, the additional search does not require much computational time. Figure 11 (a) and (b) show the result obtained by exhaustive search method. Dynamic Attention Maps obtained through spin flip dynamics are shown in Fig. 11 (c) and (d). The white pixels in Fig. 11 (c) and (d) represent the spins classified as face. The obtained Dynamic Attention Maps are nearly equal to the results obtained by exhaustive search method. Note that two person's faces are detected correctly. The search maps obtained in this experiment are shown in Fig. 11 (e) and (f). The white pixels represent the spins which are selected and classified. From Fig. 11 (e) and (f), the regions around faces are searched finely. On the other hand, the non-face regions are searched coarsely. This is because the state of neighboring spins are changed to non-face though spin flip dynamics. From these results, the effectiveness of Ising search method is demonstrated.

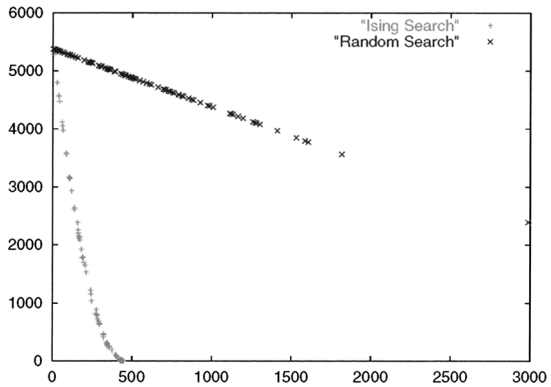


Fig. 10 How the number of face candidates decrease. (case 1 of Fig. 9)

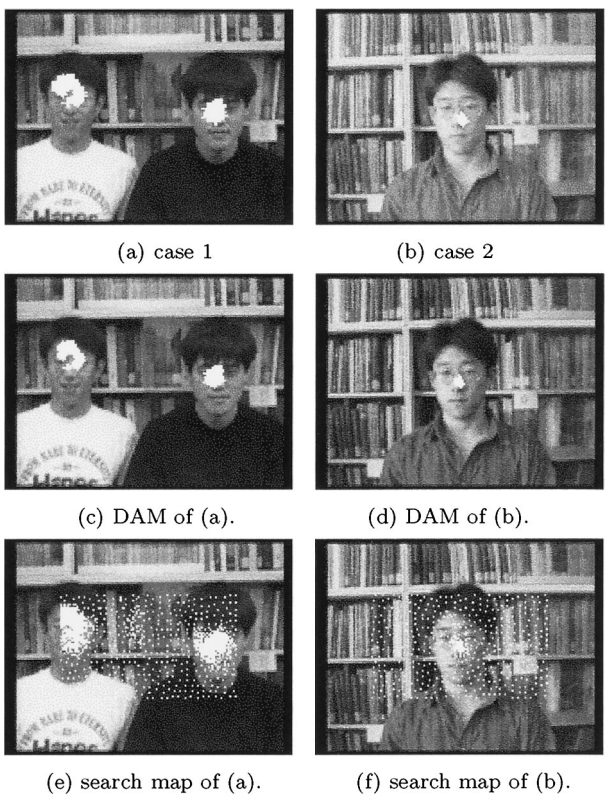


Fig. 11 The face detection result obtained by using exhaustive search method, Dynamic Attention Map, and search map obtained by using Ising search method.

6. Multilevel Dynamic Attention Map

To improve the search performance further, the single level Ising search method is extended to the multilevel Ising search. The interactions between upper (coarse scaled) and lower (fine scaled) layer are used to construct the multilevel Dynamic Attention Map. The interactions between two layers are characterized by the renormalization group method such that the state of the spin on the upper layer is determined by the states of the corresponding spins on the lower layer and the couplings of the corresponding spins on the lower layer is used in the spin flip dynamics on the upper layer. Figure 12 shows the multilevel structure. When a spin s_a^l on the lower layer, which is in the "face" state and one of the component spins of the spin s_a^u on the upper layer, is selected to evaluate a likelihood of face on the lower layer, the energy of the spin s_i^u on the upper layer layer can be given as

$$\begin{aligned} E_i^{upper} = &-J \sum_{j \in nn(i)} s_i^u s_j^u - J_{ul} \sum_{k \in com(i)} s_i^u s_k^l \\ &- H_d^u (m_d^l(a) - \theta_d) s_i^u \\ &- H_c^u (m_c^l(a) - \theta_c) s_i^u, \ i \in NNU(a), \end{aligned} \tag{11}$$

where the super scripts "u" and "l" stand for the quantities on the upper layer and those on the lower layer, J_{ul} is a coupling constant for the interactions between the layers, $com(i)$ means the component spins s_j^l on the lower layer of the spin s_i^u, and $NNU(a)$ means the neighboring spins to the spin s_a^u for performing the spin flip dynamics. On the other hand, the energy of the spin s_i^l on the lower layer can be given as

$$E_i^{lower} = -J \sum_{j \in nn(i)} s_i^l s_j^l - H_d^l (m_d^l(a) - \theta_d) s_i^l$$

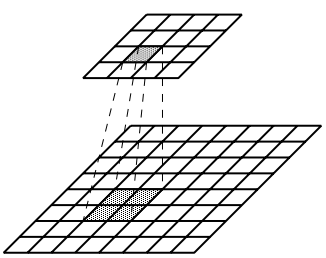


Fig. 12 Multilevel spin map.

$$- H_c^l(m_c^l(a) - \theta_c)s_i^l, \quad i \in NNL(a), \tag{12}$$

where $NNL(a)$ means the neighboring spins to the spin s_a^l for performing the spin flip dynamics.

Then the state of every spin $s_i^{u,\,l}$ on each layer is updated according to the probability which is proportional to $\exp\left(-\beta\Delta E_i^{upper,\,lower}\right)$, where

$$\begin{aligned}\Delta E_i^{upper} &= 2J\sum_{j\in nn(i)} s_i^u s_j^u + 2J_{ul}\sum_{k\in com(i)} s_i^u s_k^l \\ &\quad + 2H_d^u(m_d^l(a) - \theta_d)s_i^u \\ &\quad + 2H_c^u(m_c^l(a) - \theta_c)s_i^u, \\ \Delta E_i^{lower} &= 2J\sum_{j\in nn(i)} s_i^l s_j^l + 2H_d^l(m_d^l(a) - \theta_d)s_i^l \\ &\quad + 2H_c^l(m_c^l(a) - \theta_c)s_i^l. \end{aligned} \tag{13}$$

This spin flip dynamics in Ising model on each layer creates the multilevel Dynamic Attention Map.

In the following, we show the meta algorithm for the multilevel Ising search.

1. Set all the spins on each layer to −1 ("face") and make the face list for search on the upper layer. The face list consists of the spins in the "face" state on the upper layer.
2. Select one spin (s_a^u) randomly from the face list on the upper layer. Then select one spin $\left(s_a^l\right)$ in "face" state randomly from the lower layer, which must be a component spin of the selected spin on the upper layer. If such all component spins on the lower layer are in the "non-face" state, repeat this step until such the spin in the "face" state on the lower layer is selected.
3. Measure the likelihood of face in the discriminant space and that in the discriminant space in terms of color information of the selected spin on the lower layer.
4. Apply the spin flip dynamics for suitable number of iterations on the lower layer.
5. Update the state of the spin on the upper layer. The state of the spin on the upper layer depends on the number of spins in the "face" state on the lower layer. This is a kind of the renormalization group method.
6. Apply the spin flip dynamics for suitable number of iterations on the upper layer using the interactions between the layers and the likelihood of face obtained on lower layer.
7. Remove the spins flipped from the "face" to the "non-face" from the face list and add the spins flipped from the "non-face" to the "face" to the face list. Note that, if the selected spin to measure the likelihood of face is found to be in the "non-face" state, that spin is removed from the face list and never flipped again.
8. By repeating from 2 to 7, face candidates are narrowed down effectively.

7. Evaluation of Multilevel Ising Search Method

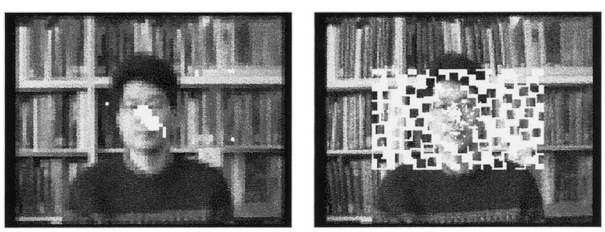
This section shows the effectiveness of multilevel Ising search method. In the following experiments, we use the same images used in the experiments for the single level Ising search method. The number of the component spins on the lower layer of the single spin on the upper layer is 2×2. The other parameters of the multilevel Ising search are set to $\beta = 0.2$, $H_d^u = H_d^l = 0.2$, $H_c^u = H_c^l = 6.0$, $J = 1.0$, and $J_{ul} = 2.0$. Monte Carlo steps (MCS) are performed 5 times in the neighboring 5×5 lattices centered the selected spin.

The face detection experiment was repeated 100 times for each case. Table 2 shows the means and the medians over 100 trials for each case. Comparing this result with that of the single level Ising search (See Tables 1 and 2), we understand that the performance of the multilevel Ising search is better than that of the single level Ising search. Examples of Dynamic Attention Maps obtained after detecting a face are shown in Fig. 13. The white regions in Fig. 13 (a) and (b) represent the spins in the "face" state on the upper and lower layer respectively. Dynamic Attention Map on the upper layer shows where the face is. On the other hand, the many spins in "face" state are remained in Dynamic Attention Map on the lower layer. The remained spins in "face" state on non-face regions represent the component spins of the spin which is removed from face list on the upper layer after only one search. It is considered that the likelihood of face of these remained spins on non-face regions are low and the corresponding spins on the upper layer of these spins on the lower layer are flipped from "face" to "non-face" by the spin flip dynamics on the upper layer. The spin flip dynamics on the upper layer makes the search on non-face regions coarse. The number of search is decreased by this process. In contrast to it, face regions are searched finely.

To investigate the effectiveness of the multilevel Ising

Table 2 The performance of multilevel Ising search method.

	mean	median
case 1	109.15	108
case 2	105.77	111
case 3	35.37	32
case 4	34.99	29



(a) (b)

Fig. 13 Dynamic Attention Map obtained by multilevel Ising search method. (a) Dynamic Attention Map on upper layer. (b) Dynamic Attention Map on lower layer.

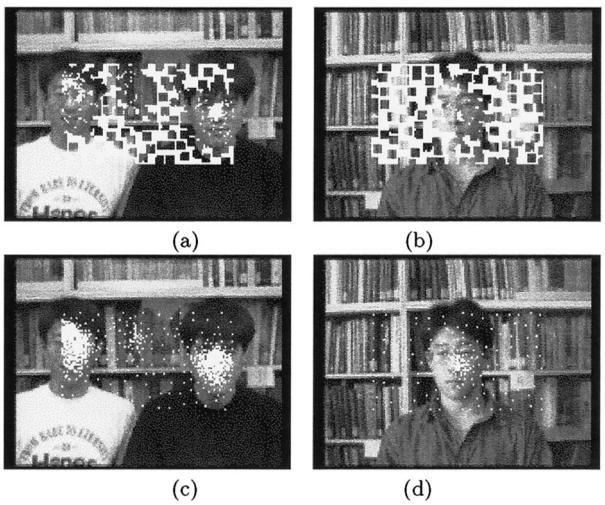


Fig. 14 Dynamic Attention Map on lower layer and search map on lower layer obtained by multilevel Ising search method. (a) and (b) Dynamic Attention Map on lower layer (c) and (d) Search map on lower layer.

search, the search is continued until the face list becomes empty. Figure 14 shows the Dynamic Attention Map and search map on the lower layer. Figure 14 (c) and (d) show the search map on the lower layer, in which the white pixels represent the searched spins. Comparing these search maps with those of the single level Ising search (See Fig. 11), we understand that the density of the search on face regions are nearly equal to that of the single level Ising search. On the other hand, the density of the search on non-face regions is coarser than that of the single level Ising search. This is because face candidates are narrowed down effectively through the spin flip dynamics on the upper layer. The remained white blocks in Fig. 14 (a) and (b) represent the component spins of the spins whose state are flipped to "non-face" by the spin flip dynamics on the upper layer. In the case of single level Ising search, all remained spins (white blocks in Fig. 14 (a) and (b)) are searched until the face list becomes empty. Therefore, the number of the search of multilevel Ising search is smaller than that of the single level Ising search. On the other hand, the spins in "face" state around face regions in Dynamic Attention Map on lower layer represent the spins detected as face.

From these results, the effectiveness of the multilevel Ising search is demonstrated.

8. Conclusion

The efficient search is realized by adopting Ising model to face detection. To improve the search performance further, the single level Ising search is extended to multilevel Ising search by taking the renormalization group method into consideration. The effectiveness of the multilevel Ising search method is confirmed by the comparison with the single level Ising search method. The multilevel Ising search method requires about 0.3 seconds to find faces in the case of Fig. 9 (a) and (b) on PC with Xeon CPU 2 GHz, when the search is continued until the face list becomes empty. On the other hand, the single level Ising search method requires about 0.6 seconds to find faces on same PC. This result also shows the effectiveness of the multilevel Ising search.

In general, there is the trade-off between the search speed and the false detection rate. It is difficult to detect faces efficiently and accurately. One of them is sacrificed [11], [15]. The proposed search method does not guarantee to obtain optimal positions of faces. However, the proposed method can narrow down the search space effectively. Therefore, it is expected that the optimal positions of faces are obtained with high probability and with low computational cost.

Acknowledgements

We would like to thank the anonymous reviewers for valuable comments.

References

[1] M. Doi, K. Sato, and K. Chihara, "A robust face identification against lighting fluctuation for lock control," Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition, pp.42–47, 1998.
[2] O. Hasegawa, K. Itou, T. Kurita, S. Hayamizu, K. Tanaka, K. Yamamoto, and N. Otsu, "Active gent oriented mutimodal interface system," Proc. International Joint Conference on Artificial Intelligence, pp.82–87, 1995.
[3] I. Hara, A. Zelinsky, T. Matui, H. Asoh, T. Kurita, M. Tanaka, and K. Hotta, "Communicative functions to support human robot cooperation," Proc. International Conference on Intelligent Robots and Systems, pp.683–688, 1999.
[4] S. Satoh and T. Kanade, "Name-it: Association of face and name in video," Tech. Rep., CMU-CS-96-205, 1996.
[5] R. Chellappa, C.L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," Proc. IEEE., vol.83, no.5, pp.705–740, 1995.
[6] W. Zhao, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Comput. Surv., vol.35, no.4, pp.399–458, 2003.
[7] E. Hjelmas and B.K. Low, "Face detection: A survey," Computer Vision and Image Understanding, vol.83, no.2, pp.236–274, 2001.
[8] M.-H.Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," IEEE Trans. Pattern Anal. Mach. Intell., vol.24, no.1, pp.34–58, 2002.
[9] E. Osuna, R. Freund, and F. Girosi, "Training support vector machines: An application to face detection," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.130–136, 1997.
[10] K. Sung and T. Poggio, "Example-based learning for view-based human face detection," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.1, pp.39–51, 1998.
[11] H.A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," IEEE Trans. Pattern Anal. Mach. Intell., vol.20, no.1, pp.23–38, 1998.
[12] B. Heisele, T. Serre, M. Pontil, and T. Poggio, "Component-based face detection," Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp.657–662, 2001.
[13] G. Xu and T. Sugimoto, "Rits eye: A software-based system for realtime face detection and tracking using pan-tilt-zoom controllable camera," Proc. 14th International Conference on Pattern Recognition, pp.1194–1197, 1998.

[14] H. Wu, Q. Chen, and M. Yachida, "Face detection from color images using a fuzzy pattern matching method," IEEE Trans. Pattern Anal. Mach. Intell., vol.21, no.6, pp.557–563, 1999.
[15] K. Hotta, T. Kurita, and T. Mishima, "Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image," Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition, pp.70–75, 1998.
[16] M. Mezard and G. Parisi, Spin Glass Theory and Beyond, World Scientific, 1987.
[17] H. Gould and J. Tobochnik, An Introduction to Computer Simulation Methods — Applications to Physical Systems, 2nd ed., Addison Wesley, 1996.
[18] M. Tanaka, K. Hotta, T. Kurita, and T. Mishima, "Dynamic attention map by Ising model for human face detection," Proc. 14th International Conference on Pattern Recognition, pp.1044–1046, 1998.
[19] M. Tanaka, K. Hotta, T. Kurita, and T. Mishima, "Multilevel dynamic attention map for human face detection," Vision Geometry VII, Proc. SPIE, vol.3454, pp.274–285, 1998.
[20] K. Hotta, M. Tanaka, T. Kurita, and T. Mishima, "Multilevel ising search for human face detection," Applications of Digital Image Processing XXI, Proc. SPIE, vol.3460, pp.202–213, 1998.
[21] G. Sandini and V. Tagliasco, "An anthropomorphic retina-like structure for scene analysis," Computer Graphics and Image Processing, vol.14, pp.365–372, 1980.
[22] L. Massone, G. Sandini, and V. Tagliasco, "Form-invariant: Topological mapping strategy for 2d shape recognition," Comput. Vis. Graph. Image Process., vol.30, pp.169–188, 1985.
[23] J. der Spiegel, G. Kreider, C. Claeys, I. Debusschere, G. Sandini, P. Dario, F. Fantini, P. Bellutti, and G. Soncini, "A foveated retina-like sensor using ccd technology," Analog VLSI and Neural Network Implementations, Kluwer, 1989.
[24] M. Tistarelli and G. Sandini, "On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow," IEEE Trans. Pattern Anal. Mach. Intell., vol.15, no.4, pp.401–410, 1993.
[25] N. Otsu and T. Kurita, "A new scheme for practical flexible and intelligent vision systems," Proc. IAPR Workshop on Computer Vision, pp.431–435, 1988.
[26] T. Kurita, N. Otsu, and T. Sato, "A face recognition method using higher order local autocorrelation and multivariate analysis," Proc. 11th IAPR International Conference on Pattern Recognition, pp.213–216, 1992.
[27] Y. Ohta, T. Kanade, and T. Sakai, "Color information for region segmentation," Computer Graphics and Image Processing, vol.13, pp.222–241, 1980.

Kazuhiro Hotta received B.Eng., M.Eng., and Dr.Eng. degrees from Saitama University in 1997, 1999, and 2002, respectively. From 1999 to 2002, he was a research fellow of Japan Society for the Promotion of Science. Since 2002, he is a research associate at the Department of Information and Communication Engineering, The University of Electro-Communications. His current research interests are pattern recognition and computer vision. He is a member of IEEE computer society, IPSJ, and Japanese Academy of Facial Studies.

Masaru Tanaka received B.Sc. degree in 1986, M.Sc. degree in 1988 and Dr.Sc. degree in 1991 from Kyushu University. From 1991 to 1999, he was with Electrotechnical Laboratory, AIST, MITI, Japan. From 1995 to 1996, he was a visiting research scientist at Institute for Information Technology, NRC, Ottawa, Canada. Since 2000, he is an Associate Professor at the Department of Information and Computer Sciences, Saitama University. His current research interests are pattern recognition, biomimetic vision and information geometry.

Takio Kurita received the B.Eng. degree from Nagoya Institute of Technology and the Dr.Eng. degree from the University of Tsukuba, in 1981 and in 1993, respectively. He joined the Electrotechnical Laboratory, AIST, MITI, Japan in 1981. From 1990 to 1991 he was a visiting research scientist at Institute for Information Technology, NRC, Ottawa, Canada. He is currently Deputy Director of Neuroscience Research Institute, National Institute of Advanced Industrial Science and Technology (AIST). His current research interests include statistical pattern recognition and neural networks. He is a member of IEEE Computer Society, IPSJ, JNNS, the Behaviormetric Society of Japan, and Japanese Academy of Facial Studies.

Taketoshi Mishima Professor of Information and Computer Science at Saitama University, was born in Kagoshima City, Kagoshima. He received his BE, ME and Ph.D. in electrical engineering from Meiji University in 1968, 1970 and 1973 respectively. He had worked for ETL (Electrotechnical Laboratory, AIST, MITI) as a Research Scientist and a Senior Research Scientist from 1974 to 1992. Since 1995, he has been a faculty member of the Department of Information and Computer Sciences, Saitama University, after he worked for Josai International University as a professor of the Faculty of Management and Information Sciences, and for the College of Liberal Arts, Saitama University, as a professor of Computer Science. Now, he also works as a Visiting Scientist at RIKEN (Institute of Physical and Chemical Research) and a Visiting Researcher at RACE (Research into Artifacts, Center for Engineering), University of Tokyo. His current research interests are the security and ethics in computer network society, mathematical pattern recognition, DARS (Distributed Autonomous Robotic System), computational biomechanics, and especially the crystallization process and state of proteins. He is a member of MSJ, IEEE, IPSJ and JSSM.