VALIDATING DIRECTIONAL EDGE-BASED IMAGE FEATURE REPRESENTATIONS IN FACE RECOGNITION BY SPATIAL CORRELATION-BASED CLUSTERING

Yasufumi Suzuki and Tadashi Shibata

Department of Frontier Informatics, School of Frontier Sciences, The University of Tokyo
701 Kiban Bldg. 5-1-5 Kashiwanoha, Kashiwa, Chiba, 277-8561, Japan
email: yasufumi@else.k.u-tokyo.ac.jp, shibata@ee.t.u-tokyo.ac.jp

ABSTRACT

Directional edge-based image feature representations have already been developed and applied to medical radiograph analysis, face detection, and face identification. The scheme of the feature-vector generation is considered as a dimensionality reduction from a high-dimension edge feature map space to a low-dimension edge histogram space. However, the validity of such schemes in dimensionality reduction has not yet been studied on the sound basis of statistics. In this paper, in order to validate the directional edge-based feature representations, the scheme of dimensionality reduction employing the hierarchical clustering based on the spatial correlations of edges has been investigated. In this study, the feature representations of images have been evaluated using facial images as the test vehicle. As a result, the validity of using the directional edge-based feature vectors in image recognition has been verified by the similarity between the results of the hierarchical clustering and the schemes employed in the feature representations.

1. INTRODUCTION

The development of robust image recognition systems is quite essential in a variety of applications such as intelligent human-computer interfaces, security systems, and so forth. For human-like robust image perception, an image feature representation which extracts only essential features from images and represents them in a vector format is of critical importance. A number of feature representations have been developed so far. In the visual cortex of animals, it is known that directional edge information in visual inputs plays an essential role in early visual processing. Namely, biological systems rely on the spatial relationship among edges in various directions for image perception. Being inspired by such a biological principle, a directional edge-based feature representation called Projected Principal-Edge Distribution (PPED for short) [1] has been developed and has been successfully applied to hand-written pattern recognition and medical radiograph analysis [2, 3]. It has also been applied successfully to the face detection system [4] in conjunction with another directional edge-based feature representation called Averaged Principal-Edge Distribution (APED for short) [5]. In addition, APED-like feature representations have been applied to face identification [6]. Both PPED and APED feature vectors are generated from the spatial distributions of the four directional edges (horizontal, $+45°$, vertical, and $-45°$). Since the amount of data in the directional edge maps is massive, the dimensionality reduction is carried out by producing the spatial distribution histograms from the four directional edge maps. The feature-vector-generation schemes in PPED and APED are compatible with the dedicated feature-extraction VLSI chip [7], making the feature representation algorithms compatible to building real-time responding systems. However, the schemes of dimensionality reduction in PPED and APED have not yet been supported by statistical analysis.

The principal components analysis (PCA) is often used for the dimensionality reduction of feature representations. The PCA calculates the orthogonal linear transformation which maximizes the variances among feature vectors. Such a linear transformation is derived from the eigenvectors of the spatial correlations in the image data. Although the PCA achieves the dimensionality reduction efficiently preserving the characteristics of images, the linear transformation of the PCA is not suitable for direct hardware implementations due to the computational cost.

The purpose of this paper to validate the directional edge-based feature representations by the statistical analysis. The dimensionality reduction in both PPED and APED vector representations is carried out by taking the spatial histograms employing their own partitions of the directional edge maps. The dedicated VLSI chip has already been developed [7] which is compatible to any types of edge-histogram generation. Therefore, it is crucial to find how to divide the directional edge maps for generating histograms in order to maximize the performance in image classification. In this work, as like in the PCA, it is considered to maximize the variance among feature vectors. For each directional edge map, grouping the pixel sites having the high correlations together into the same element of the feature vector increases the variance as described in Section 3. In order to determine such partitions, the technique of hierarchical clustering based on the spatial correlation of the directional edges has been introduced. In this work, the evaluation has been carried out using a set of frontal facial images as a test vehicle. The results of the hierarchical clustering are similar to the partition employed in PPED or APED vector-generation scheme, thus validating the directional edge-based feature representations.

2. DIRECTIONAL EDGE-BASED FEATURE REPRESENTATIONS

This section describes the directional edge-based feature representations utilized for the face detection system in the previous works [4, 5]. Firstly, the directional edge-based feature maps which represent the distribution of four directional edges are generated. Then, the 64-dimension feature vectors are formed by taking the spatial histograms of the directional

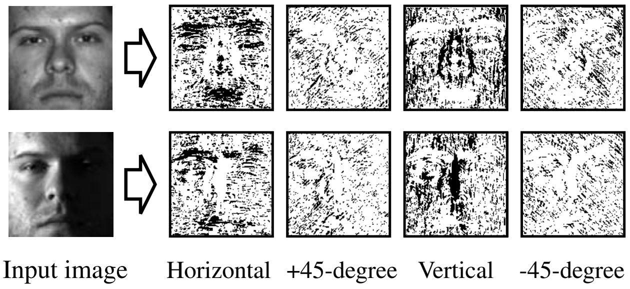

Figure 1: Directional edge-based feature maps generated from bright and dark facial images.

edge-based feature maps.

2.1 Directional Edge-Based Feature Maps

The first step in forming feature vectors is the generation of four feature maps in which edges are extracted from a 64×64-pixel input image in four directions. Figure 1 shows the relationship between an input image and four feature maps. Each feature map represents the distribution of edge flags corresponding to one of the four directions, i.e. horizontal, $+45°$, vertical, and $-45°$. In the feature maps, each black pixel which we call an edge flag represents the location where an edge is identified in the corresponding direction. These four directional edge-based feature maps are regarded as representing the most fundamental features extracted from the original image. The procedure of feature-map generation is as in the following. The input image is firstly subjected to pixel-by-pixel spatial filtering operations using kernels of 5×5-pixel size. The kernel which gives the maximum value determines the direction of the edge at the pixel site. In order to retain only edges of significance in feature maps, thresholding operation is introduced. The threshold value is determined by taking the local variance of luminance data into account. The intensity differences between two neighboring pixels in the horizontal and vertical directions in a 5×5-pixel block are obtained and the median of the 40 values of them is utilized as the threshold value for edge identification. Thanks to such a thresholding operation, essential edges representing facial features are well extracted from both bright and dark facial images as shown in Fig. 1, thus making directional edge-based representations robust against illumination conditions. Details of the directional edge-based feature maps are given in [1]. In the following discussion, for each direction $d \in \{H,P,V,M\}$, the directional edge map is expressed as $F_d(x,y) \in \{0,1\}$, where the directional edge exists at the location (x,y) if $F_d(x,y) = 1$. Here, H, P, V, and M represent "horizontal," "$+45°$," "vertical," and "$-45°$," respectively.

2.2 Directional Edge-Based Feature Representations

Although the directional edge-based feature maps retain essential feature information in the original image, the amount of data is still massive and dimensionality reduction is essential for efficient processing of classification. Figure 2 illustrates the procedure of feature-vector generation in the Projected Principal-Edge Distribution (PPED) [1]. The feature vector in PPED is formed by projecting the feature maps along the direction of the edges. In the horizontal edge map, for example, edge flags in every four rows are accumulated,

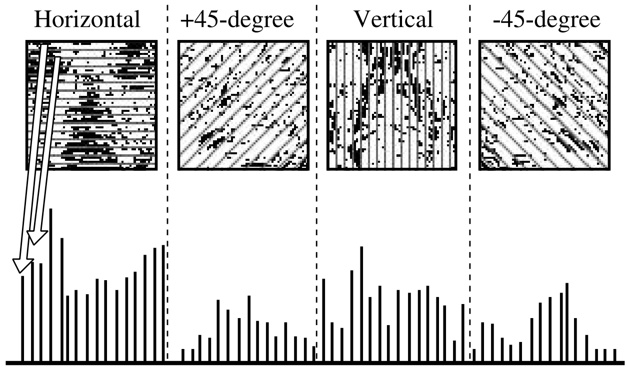

Figure 2: Partitions of feature maps for vector generation based on projected principal-edge distribution (PPED).

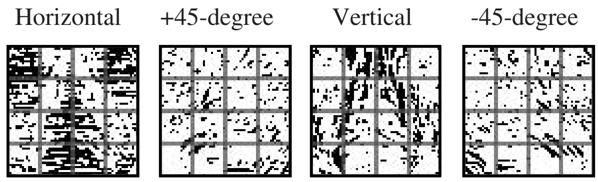

Figure 3: Partitions of feature maps for vector generation based on averaged principal-edge distribution (APED).

and a spatial distribution of edge flags is represented by a histogram. Similar procedures are applied to other three directions. Finally, a 64-dimension vector is formed by concatenating the four histograms.

In the PPED feature representation, the information of edge distribution along the direction identical to the directional edge (e.g. the horizontal distribution of horizontal edge flags) is lost during the accumulation. In order to complement such loss in PPED vectors, another feature-vector generation scheme has been developed [5]. In the Averaged Principal-Edge Distribution scheme (which was originally named Cell Edge Distribution in [5]), every feature map is divided into 4×4 cells each of which contains 16×16-pixel sites as shown in Fig. 3. The number of edge flags in each cell is counted and the number constitute a single element in the vector representation by the APED scheme. A 64-dimension feature vector in the APED scheme is obtained by concatenating these four vectors.

3. SPATIAL CORRELATION-BASED CLUSTERING

The process of the feature-vector generation is considered as a dimensionality reduction in the space of four directional edge maps onto the lower-dimension spaces. In the feature-vector-generation schemes, each directional edge-based 64×64-pixel feature map of F_H, F_P, F_V, or F_M is divided into 16 groups as shown in Figs. 2 and 3. The number of directional edge flags within each group is counted and constitutes a single element in the feature vector, thus making a 16-dimensional feature vector. The dedicated VLSI chip [7] is capable of generating feature vectors using such a scheme with any form of feature-map partitions. Therefore, in the present study, it is considered that how the feature maps should be divided for maximizing the variance among the feature vectors generated from facial images. In the following discussion, 64×64-pixel two dimensional image data of a feature map are treated as a 4096-dimension binary vector. Namely, each feature map F_d ($d \in \{H,P,V,M\}$) is converted to a 4096-dimension vector $E_d(i)$ which is expressed as

$$E_d\,(x+64y) = F_d(x,y), \tag{1}$$

where $x = 0,1,2,\ldots,63$ and $y = 0,1,2,\ldots,63$. For each direction $d \in \{H,P,V,M\}$, the partition $P_d(i)$ of the feature map $E_d(i)$ is given as

$$P_d(i) \in \{j \mid j \in \mathbb{N}, 1 \le j \le 16\},$$

where $i = 0,1,2,\ldots,4095$. This means that the pixel site corresponding to the feature map $E_d(i)$ belongs to the $P_d(i)$-th element in the feature vector. A mask function $M_d(i,j)$ is defined from $P_d(i)$ as

$$M_d(i,j) = \begin{cases} 1, & \text{if } P_d(i) = j \\ 0, & \text{otherwise,} \end{cases}$$

where $i = 0,1,2,\ldots,4095$ and $j = 1,2,3,\ldots,16$. Note that the mask function $M_d(i,j)$ satisfies the following equation:

$$\sum_{j=1}^{16} M_d(i,j) = \sum_{j=1}^{16} \Big\{M_d(i,j)\Big\}^2 = 1. \tag{2}$$

From the directional edge map $E_d(i)$ defined in Eq. (1), a 16-dimensional feature vector $\mathbf{V}_d$ is generated using the mask function M_d as

$$V_d(j) = \sum_{i=0}^{4095} M_d(i,j)E_d(i), \tag{3}$$

where $j = 1,2,3,\ldots,16$. When n sample images are given, the mean $\overline{E_d(i)}$ and the variance $\mathrm{Var}(E_d(i))$ of the directional edge-based feature maps $E_d(i)$ are obtained as

$$\overline{E_d(i)} = \frac{1}{n}\sum_{k=1}^{n} E_d^k(i), \tag{4}$$

$$\mathrm{Var}(E_d(i)) = \frac{1}{n}\sum_{k=1}^{n} \Big\{E_d^k(i)\Big\}^2 - \Big\{\overline{E_d(i)}\Big\}^2, \tag{5}$$

respectively, where $E_d^k(i)$ is the edge map of the direction d obtained from the k-th face sample. The covariance $C_d(i,j)$ between two different pixel sites $E_d(i)$ and $E_d(j)$ in the feature map is also obtained as follows:

$$C_d(i,j) = \frac{1}{n}\sum_{k=1}^{n} E_d^k(i)E_d^k(j) - \overline{E_d(i)}\;\overline{E_d(j)}. \tag{6}$$

The feature vector $V_d(j)$ is generated from the directional edge map $E_d(i)$ using Eq. (3). The mean $\overline{V_d(j)}$ and the variance $\mathrm{Var}(V_d(j))$ of the j-th element in the feature vector $\mathbf{V}_d$ are expressed as

$$\overline{V_d(j)} = \frac{1}{n}\sum_{k=1}^{n} V_d^k(j), \tag{7}$$

$$\mathrm{Var}\,(V_d(j)) = \frac{1}{n}\sum_{k=1}^{n} \Big\{V_d^k(j)\Big\}^2 - \Big\{\overline{V_d(j)}\Big\}^2, \tag{8}$$

respectively, where $V_d^k(j)$ is the j-th element in the feature vector in the direction d generated from the k-th sample. From Eqs. (3), (4), and (7), the values of $\{V_d^k(j)\}^2$ and $\Big\{\overline{V_d(j)}\Big\}^2$ are obtained as

$$\begin{aligned} \Big\{V_d^k(j)\Big\}^2 &= 2\sum_{l=0}^{4094}\sum_{m=l+1}^{4095} M_d(l,j)E_d^k(l)M_d(m,j)E_d^k(m) \\ &\quad + \sum_{i=0}^{4095}\Big\{M_d(i,j)E_d^k(i)\Big\}^2, \end{aligned} \tag{9}$$

$$\begin{aligned} \Big\{\overline{V_d(j)}\Big\}^2 &= \left\{\sum_{i=0}^{4095} M_d(i,j)\overline{E_d(i)}\right\}^2 \\ &= 2\sum_{l=0}^{4094}\sum_{m=l+1}^{4095} M_d(l,j)\overline{E_d(l)}M_d(m,j)\overline{E_d(m)} \\ &\quad + \sum_{i=0}^{4095}\Big\{M_d(i,j)\overline{E_d(i)}\Big\}^2, \end{aligned} \tag{10}$$

respectively. The value of $\mathrm{Var}(V_d(j))$ is simplified using Eqs. (5), (6), (8), (9), and (10) as

$$\begin{aligned} \mathrm{Var}(V_d(J)) &= \sum_{i=0}^{4095}\Big\{M_d(i,j)\Big\}^2 \mathrm{Var}(E_d(i)) \\ &\quad + 2\sum_{l=0}^{4094}\sum_{m=l+1}^{4095} M_d(l,j)M_d(m,j)C_d(l,m). \end{aligned} \tag{11}$$

Here, we consider the summation of variances of feature-vector elements W_d which is given as

$$W_d = \sum_{j=1}^{16} \mathrm{Var}\,(V_d(j)).$$

From Eqs. (2) and (11), W_d is expressed as in the following:

$$\begin{aligned} W_d &= 2\sum_{j=1}^{16}\sum_{l=0}^{4094}\sum_{m=l+1}^{4095} M_d(l,j)M_d(m,j)C_d(l,m) \\ &\quad + \sum_{i=0}^{4095} \mathrm{Var}(E_d(j)). \end{aligned} \tag{12}$$

In Eq. (12), since $\mathrm{Var}(E_d(j))$ is a constant independent of the mask function $M_d(i,j)$, maximizing W_d is equivalent to maximize α_d defined as

$$\alpha_d = \sum_{j=1}^{16}\sum_{l=1}^{4095}\sum_{m=l+1}^{4096} M_d(l,j)M_d(m,j)C_d(l,m). \tag{13}$$

α_d in Eq. (13) means the summation of all covariances between any pair of pixels which belong to the same element in the feature vector.

In order to determine the division of four feature maps which maximizes the sum of variances of feature-vector elements (this is identical to maximizing the value of α_d), the hierarchical clustering technique has been introduced. All sites within 64×64-pixel window are divided into 16 clusters using the agglomerative hierarchical clustering algorithm [8] where the covariances between different pixel sites are utilized as the dissimilarity measure.

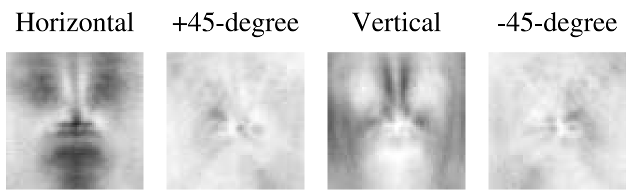

Figure 4: Mean images of directional edge-based feature maps generated from 900 frontal facial images.

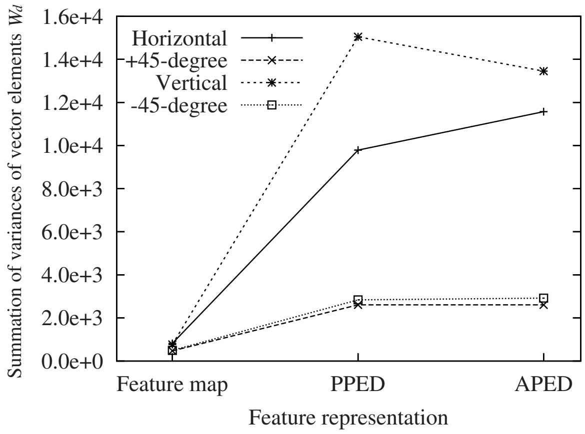

Figure 5: Summation of variances of all feature-vector elements for PPED and APED.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In order to evaluate the variances of feature-vector elements in PPED and APED, a statistical analysis has been done as in the following. In this work, 900 frontal face samples (100%, 80%, and 60% scaled faces) from HOIP face database [11] were utilized for the evaluation. For each of facial images, the directional edge-based feature vector was generated. The mean images $\overline{E_d(i)}$ of 900 directional edge-based feature maps are presented in Fig. 4. Then, for each direction (horizontal (H), +45-degree (P), vertical (V), and -45-degree (M)), the value of W_d ($d \in \{H,P,V,M\}$) which is the summation of the variances of feature-vector elements was calculated. The values of W_d in PPED and APED are shown in Fig. 5. In Fig. 5, the values of "feature map" represent the summation of variances of all pixel sites in the feature map, and are equivalent to the value $\sum_{i=0}^{4095} \mathrm{Var}(E_d(j))$ in Eq. (12). The difference between the value of PPED and that of feature map indicates the increment of variances α_d caused by the dimensionality reduction during the PPED-vector generation.

The hierarchical clustering was carried out on the covariance matrices derived from the 900 frontal face samples to determine the partitions of edge maps maximizing the variance. Figure 6 demonstrates the results of hierarchical clustering employing the complete linkage algorithm [9] and the average linkage algorithm [10]. The clustering result by the complete linkage algorithm, which has the tendency of making clusters compact, shows that the neighboring pixel sites along the direction of edge flags (e.g. horizontal neighboring sites of horizontal edge flags) are likely to be classified into the same cluster. Such clustering is very similar to the feature

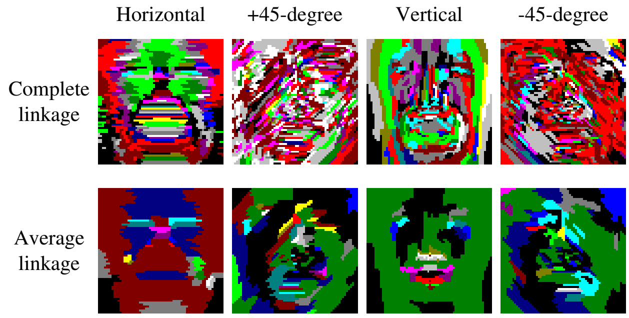

Figure 6: Results of agglomerative hierarchical clustering of all pixel sites in directional edge-based feature maps. Each feature map is divided into 16 clusters which are indicated by 16 different colors.

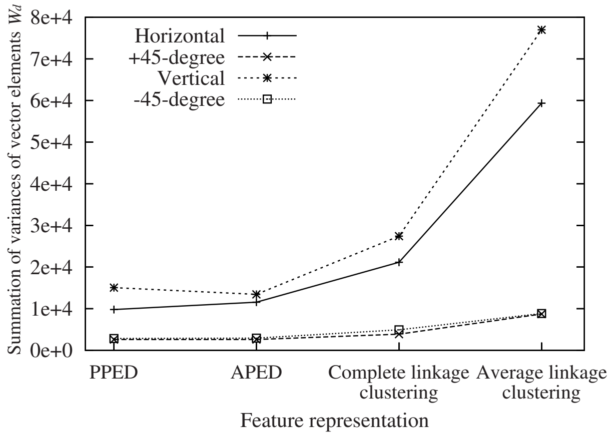

Figure 7: Summation of variances of all feature-vector elements for each vector-generation scheme.

map division scheme employed in the PPED vector generation shown in Fig. 2. On the other hand, it is interesting to see the clustering result by the average linkage algorithm. It displays more connectivity than that by the complete linkage algorithm, in which each facial part such as the eyes, the nose, and the outline of a face is assigned to the same cluster. This implies that the feature map division scheme employed in the APED vector generation (Fig. 3) is essential for facial image recognition.

The results of the clustering suggest that PPED and APED are essential for facial image classification. However, PPED and APED have been developed as general-purpose feature representations. Therefore, we can expect that employing the results of the hierarchical clustering as the partitions can improve the performance of facial image recognition. Figure 7 shows the values of W_d in Eq. (12) employing the results of hierarchical clustering illustrated in Fig. 6 as the partitions of the edge maps are demonstrated in comparison with those of PPED and APED. The clustering-based feature representations achieve larger variances than PPED and APED. The preliminary experiments of face detection were carried out using an image in CMU test set C [12]. In the template matching-based face detection, a 64×64-pixel image was taken from an input image and a feature vector was generated from the taken image. Then, the feature vector was

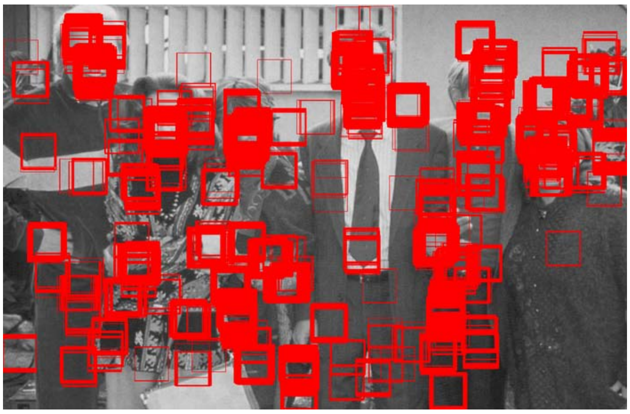

Figure 8: Result of face detection from an image in CMU test set C [12] using PPED and APED feature vectors.

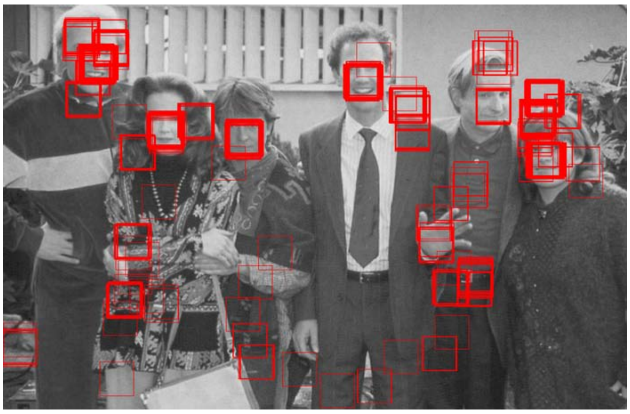

Figure 9: Result of face detection using hierarchical clustering-based feature vectors.

matched with all template vectors of face and nonface samples, and classified as a face or a nonface according to the category of the best-matched template vectors. This classification was carried out by pixel-by-pixel scanning the entire image. The 900 facial images and 2000 randomly chosen background images were utilized for face samples and nonface samples, respectively. In the experiment, the concept of multiple-clue criterion [5] has been introduced. Namely, only pixel sites where both feature vectors classified as a face were determined as a true face. The results of face detection using PPED and APED feature vectors are shown in Fig. 8. Although all faces are detected, many false positives also occur. Figure 9 demonstrates the result of face detection using the feature vectors utilizing the results of the complete linkage clustering and the average linkage clustering shown in Fig. 6. The number of false positives has been reduced without missing the true faces. (These false positives can be eliminated by the facial parts verification algorithm described in [4].) This result implies that the directional edge-based feature representations using the hierarchical clustering improve the performance of image recognition.

5. CONCLUSIONS

The directional edge-based image feature representations have been studied on the basis of statistic analysis. Using 900 frontal facial images as test samples, four directional edge maps were analyzed employing the hierarchical clustering based on the spatial correlations of edge flags. As a result, the validity of using PPED and APED feature vectors for facial image recognition has been verified by the similarity between the results of the hierarchical clustering and the partition employed in PPED or APED vector-generation scheme. In addition, it has also been demonstrated that the hierarchical clustering-based feature representations improve the performance of face detection.

REFERENCES

[1] M. Yagi and T. Shibata, "An image representation algorithm compatible with neural-associative-processor-based hardware recognition systems," *IEEE Trans. Neural Netw.*, vol. 14, pp. 1144-1161, Sept. 2003.

[2] M. Yagi, M. Adachi, and T. Shibata, "A hardware-friendly soft-computing algorithm for image recognition," in *Proc. EUSIPCO 2000*, Tampere, Finland, September 5-8. 2000, pp. 729-732.

[3] M. Yagi and T. Shibata, "Human-perception-like image recognition system based on the associative processor architecture," in *Proc. EUSIPCO 2002*, Toulouse, France, September 3-6. 2002, pp. 103-106.

[4] Y. Suzuki and T. Shibata, "Multiple-Resolution Edge-Based Feature Representations for Robust Face Segmentation and Verification," in *Proc. EUSIPCO 2005*, Antalya, Turkey, Sept. 4-8. 2005.

[5] Y. Suzuki and T. Shibata, "Multiple-clue face detection algorithm using edge-based feature vectors," in *Porc. of Intl. Conf. on Accoustics, Speech and Sig. Proc. (ICASSP)*, Montreal, Canada, May 17-21. 2004, pp. 737-740.

[6] Y. Suzuki and T. Shibata, "Illumination-invariant face identification using edge-based feature vectors in pseudo-2D hidden Markov models," in *Proc. EUSIPCO 2006*, Florence, Italy, Sept. 4-8. 2006.

[7] H. Yamasaki and T. Shibata, "A real-time image-feature-extraction and vector-generation VLSI employing arrayed-shift-register architecture," in *Proc of European Solid-State Circuits Conf. (ESSIRC)*, Grenoble, France, Sept. 12-16. 2005, pp. 121-124.

[8] W. H. E. Day and H. Edelsbrunner, "Efficient algorithms for agglomerative hierarchical clustering methods," *Journal of Classification*, vol. 1, pp. 7-24, Dec. 1984.

[9] D. Defays, "An efficient algorithm for a complete link method," *Computer Journal*, vol. 20, pp. 364-366, 1977.

[10] E. M. Voorhees, "Implementing agglomerative hierarchic clustering algorithms for use in document retrieval," *Information Processing and Management*, vol. 22, pp. 465-476, 1986.

[11] Softpia Japan, Research and Development Division, HOIP Laboratory, "HOIP facial image database," http://www.hoip.jp/web_catalog/top.html

[12] H. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection", *IEEE Trans. Pattern Anal. Machine Intell.*, vol.20, pp. 23-38, Jan. 1998.