face identification system. Section 6 demonstrates the performance of the algorithm. Section 7 concludes the article.

2. Directional edge-based feature representation algorithm

The feature representations utilized in the present work are described in this section. They are based on the spatial distribution of directional edges extracted from an original image. As a region of interests (ROI) for generating a feature vector, we employ a fixed-size window of 64×64-pixel sites, which is compatible with the hardware organization of the VLSI vector generator chip (Yamasaki & Shibata (2007)). A 64-dimension feature vector is generated from the pixel intensities within the window.

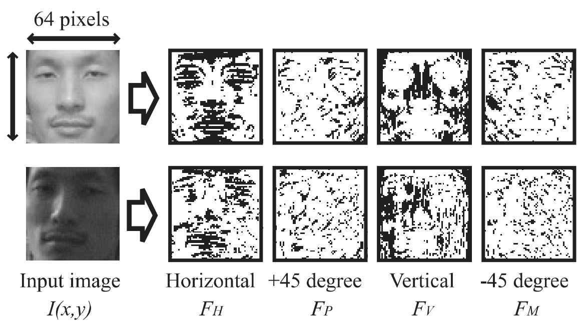

Fig. 1. Directional edge-based feature maps generated from bright and dark facial images. Since edges are detected taking local luminance variance into account, similar edge maps are obtained independent of the illumination condition.

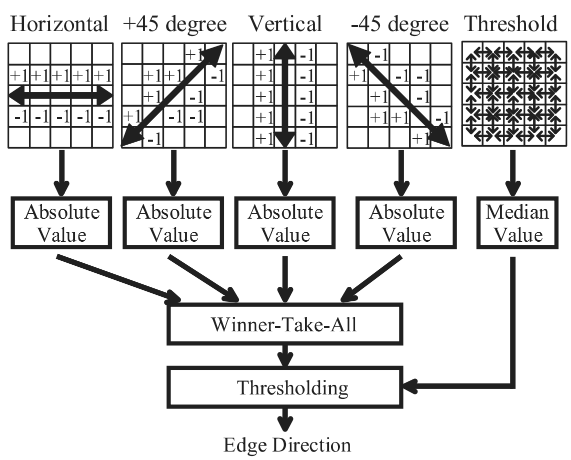

Fig. 2. Procedure of directional edge extraction.

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. Fig. 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 degree, vertical, and -45 degree. These four directional-edge-based feature maps are regarded as representing the most fundamental features extracted from the original image

$$\underset{K_H}{\overset{\text{Horizontal}}{\begin{array}{|c|c|c|c|c|}\hline 0&0&0&0&0\\\hline 1&1&1&1&1\\\hline 0&0&0&0&0\\\hline -1&-1&-1&-1&-1\\\hline 0&0&0&0&0\\\hline\end{array}}}\quad \underset{K_P}{\overset{\text{+45 degree}}{\begin{array}{|c|c|c|c|c|}\hline 0&0&0&1&0\\\hline 0&1&1&0&-1\\\hline 0&1&0&-1&0\\\hline 1&0&-1&-1&0\\\hline 0&-1&0&0&0\\\hline\end{array}}}\quad \underset{K_V}{\overset{\text{Vertical}}{\begin{array}{|c|c|c|c|c|}\hline 0&1&0&-1&0\\\hline 0&1&0&-1&0\\\hline 0&1&0&-1&0\\\hline 0&1&0&-1&0\\\hline 0&1&0&-1&0\\\hline\end{array}}}\quad \underset{K_M}{\overset{\text{-45 degree}}{\begin{array}{|c|c|c|c|c|}\hline 0&-1&0&0&0\\\hline 1&0&-1&-1&0\\\hline 0&1&0&-1&0\\\hline 0&1&1&0&-1\\\hline 0&0&0&1&0\\\hline\end{array}}}$$

Fig. 3. Filtering kernels utilized for extracting directional edges.

and all feature representations utilized in this work are derived from the feature maps. The procedure of feature-map generation is illustrated in Fig. 2. The input image is firstly subjected to pixel-by-pixel spatial filtering operations using kernels of 5×5-pixel size as in the following:

$$I_d(x,y) = \left| \sum_{i=-2}^{2} \sum_{j=-2}^{2} K_d(i,j) \cdot I(x+i, y+j) \right| \tag{1}$$

$$d \in \{H, P, V, M\}, \tag{2}$$

where $I(x,y)$ is the pixel intensity at the location (x,y), and $K_d(i,j)$ is the filtering kernel shown in Fig. 3. The kernel which gives the maximum value of $I_d(x,y)$ determines the direction of the edge at the pixel site. Namely, the preliminary edge flag $F_d^*(x,y)$ is determined as follows:

$$F_d^*(x,y) = \begin{cases} 1, & \text{if } I_d(x,y) = \max\limits_{d^* \in \{H,P,V,M\}} I_{d^*}(x,y) \\ 0, & \text{otherwise.} \end{cases} \tag{3}$$

This assigns one edge flag at every 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 difference between two neighboring pixels in the horizontal direction $H_{nm}(x,y)$ and that in the vertical direction $V_{nm}(x,y)$ are obtained as

$$\begin{aligned} H_{nm}(x,y) &= |I(x+n+1, y+m) \\ &\quad - I(x+n, y+m)| \end{aligned} \tag{4}$$

$$\begin{aligned} V_{nm}(x,y) &= |I(x+m, y+n+1) \\ &\quad - I(x+m, y+n)| \end{aligned} \tag{5}$$

respectively, where $n = -2, -1, 0, 1$, and $m = -2, -1, 0, 1, 2$. The threshold value $TH(x,y)$ is calculated as

$$TH(x,y) = M_{ed}(x,y) \times 5, \tag{6}$$

where $M_{ed}(x,y)$ is the median of the 40 values of $H_{ij}(x,y)$ and $V_{ij}(x,y)$. For each direction $d \in \{H, P, V, M\}$, the directional edge map $F_d(x,y)$ is obtained as

$$F_d(x,y) = \begin{cases} F_d^*(x,y), & \text{if } I_d(x,y) > TH(x,y) \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

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.

2.2 Projected principal-edge distribution

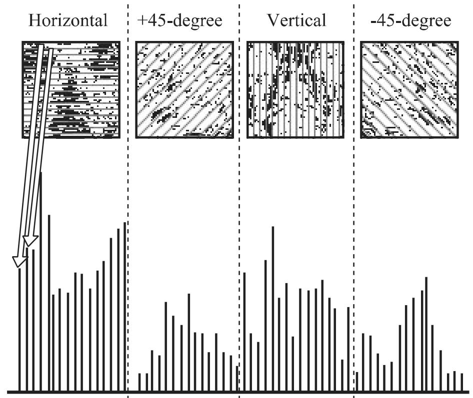

Fig. 4. Partitions of feature maps for vector generation based on projected principal-edge distribution (PPED).

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. Fig. 4 illustrates the procedure of feature-vector generation in the Projected Principal-Edge Distribution (PPED) (Yagi & Shibata (2003)). In the horizontal edge map, for example, edge flags in every four rows are accumulated and spatial distribution of edge flags are represented by a histogram. Similar procedures are applied to other three directions. Finally, a 64-dimension vector is formed by concatenating the four histograms. Details of the PPED feature representation are given in Yagi & Shibata (2003).

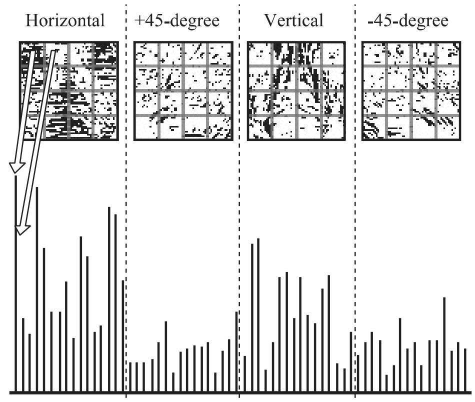

Fig. 5. Partitions of feature maps for vector generation based on averaged principal-edge distribution (APED).

2.3 Averaged principal-edge distribution

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, other feature-vector generation schemes have been developed (Suzuki & Shibata (2004)). In the Averaged Principal-Edge Distribution scheme (which was originally named Cell Edge Distribution (CED) in Suzuki & Shibata (2004)), every feature map is divided into 4×4 cells each containing 16×16-pixel sites as shown in Fig. 5. The number of edge flags in each cell is counted and the number constitute a single element in the vector representation by the Averaged Principal-Edge Distribution (APED) scheme. From the four directional edge feature maps F_H, F_P, F_V, and F_M, four 16-dimension vectors are obtained as in the following:

$$\begin{aligned}
\mathbf{H}(a+4b) &= \sum_{i=0}^{15}\sum_{j=0}^{15} F_H(16a+i, 16b+j) \\
\mathbf{P}(a+4b) &= \sum_{i=0}^{15}\sum_{j=0}^{15} F_P(16a+i, 16b+j) \\
\mathbf{V}(a+4b) &= \sum_{i=0}^{15}\sum_{j=0}^{15} F_V(16a+i, 16b+j) \\
\mathbf{M}(a+4b) &= \sum_{i=0}^{15}\sum_{j=0}^{15} F_M(16a+i, 16b+j) \\
& a = 0,1,2,3, \; b = 0,1,2,3.
\end{aligned} \tag{8}$$

A 64-dimension feature vector **X** in the Averaged Principal-Edge Distribution scheme is obtained by concatenating these four vectors as

$$\mathbf{X} = [\mathbf{H}, \mathbf{P}, \mathbf{V}, \mathbf{M}]. \tag{9}$$

Two types of 64-dimension feature vectors, the PPED (Projected Principal-Edge Distribution) vector and the APED (Averaged Principal-Edge Distribution) vector, are utilized for face detection in the present work.

3. System organization of multiple-clue face detection

In this section, the organization of the face detection system developed in the present work is presented.

3.1 Overview of the system

In order to detect all faces in a target image, a window-scanning technique is employed for face detection as illustrated in Fig. 6. A partial image in the fixed-size window of 64×64 pixels is taken from the input image and a 64-dimension feature vector is generated according to the procedure described in Section 2. Then, the feature vector is matched with all template vectors of both face samples and non-face samples stored in the system and classified as a face or a non-face according to the category of the best-matched template vector. Namely, when the best-matched template is found in the group of face samples, the partial image is determined as a face. If the best-matched template is in the group of non-face samples, it is determined as a non-face. In this work, no threshold value is employed in the template matching. The template matching is carried out utilizing the Manhattan distance as the dissimilarity measure, which is given by

$$d(\mathbf{X}, \mathbf{T}) = \sum_{i=1}^{64} |\mathbf{X}(i) - \mathbf{T}(i)|, \tag{10}$$