14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

ILLUMINATION-INVARIANT FACE IDENTIFICATION USING EDGE-BASED FEATURE VECTORS IN PSEUDO-2D HIDDEN MARKOV MODELS

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

A pseudo-2D Hidden Markov Model-based face identification system employing the edge-based feature representation has been developed. In the HMM-based face recognition algorithms, 2D discrete cosine transform (DCT) is often used for generating feature vectors. However, DCT-based feature representations are not robust against the variation in illumination changes. In order to enhance the robustness against illumination conditions, the edge-based feature representation has been employed. This edge-based feature representation has already been applied to robust face detection in our previous work and is compatible to processing in the dedicated VLSI hardware system which we have developed for real-time performance. The robustness against illumination change of the pseudo-2D HMM-based face identification system has been demonstrated using both the AT&T face database and the Yale face database B.

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, robot vision, and so forth. Real-time human face recognition, in particular, plays an important role in establishing user-friendly interfaces between humans and computers. In order to realize real-time-responding low-power embedded systems, the development of hardware-friendly algorithms compatible to dedicated VLSI chips is quite essential. For this purpose, we have already developed the VLSI chips dedicated for image feature extraction [1, 2] and template matching [3]. An automatic face recognition system consists of two stages, the face detection and the face identification [4]. In the face detection, facial images are localized in an input image. In the face identification, the localized faces are identified as individuals registered in the system. Therefore, developing both face detection algorithms and face identification algorithms is quite important. We have already developed a robust face detection algorithm and the robust nature of the system under various conditions such as illumination, scale, and rotation has been demonstrated in our previous work [5–7].

In this paper, we focused on the face identification stage. In this stage, facial images are already detected and localized in the scene. Therefore, more computational time is allowed for identifying each face. Namely, we can employ more sophisticated classification algorithms. Hidden Markov Models (HMM) are one of the statistical classifiers successfully applied to the speech recognition. HMM has already been utilized for face identification. The original form of HMM is based on a simple one dimensional (1D) Markov chain.

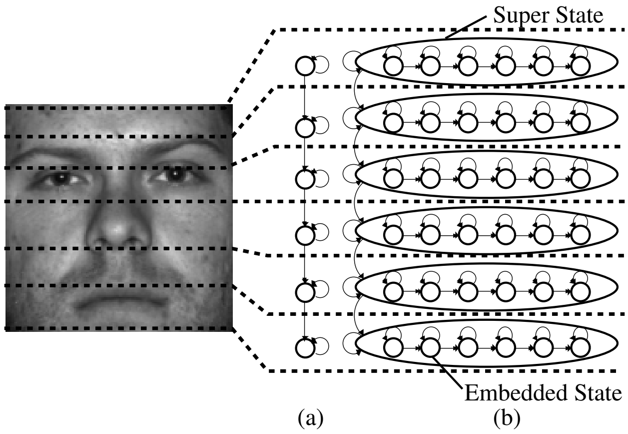

Figure 1: 1D Hidden Markov Model (a) and pseudo-2D Hidden Markov Model (b).

This traditional 1D HMM was adapted for the face identification in Ref. [8], where a facial image was divided into several regions, such as the forehead, eyes, nose, and mouth, and then each of these regions was assigned to a state in the 1D HMM as illustrated in Fig. 1 (a). This dimensionality reduction of the image from two dimension (2D) to 1D implies the loss of 2D structural information in the facial image along the horizontal direction. In order to process images as 2D data, pseudo-2D HMM has been introduced in Ref. [9] and applied to face recognition in Ref. [10]. The pseudo-2D HMM consists of a set of super states which contain a 1D HMM within themselves. The pseudo-2D HMM is utilized for modeling facial images in a hierarchical manners as in the following. Several super states correspond to the vertical facial features, such as forehead, eyes, nose, and mouth as illustrated in Fig. 1 (b). Each state within the super state is utilized for modeling the horizontal sequence of the localized feature. The pseudo-2D HMM face identification has often utilized the coefficients of 2D discrete cosine transform (DCT) for feature representation [11, 12], and good identification performance has been proven. However, the representation is sensitive to the change in illumination conditions.

The purpose of this paper is to develop an illumination-invariant pseudo-2D Hidden-Markov-Model-based face identification system in which the edge-based feature representation is employed. The robustness of the face identification system against the variation in illumination conditions has been demonstrated for the AT&T face database [8] and the Yale face database B [13].

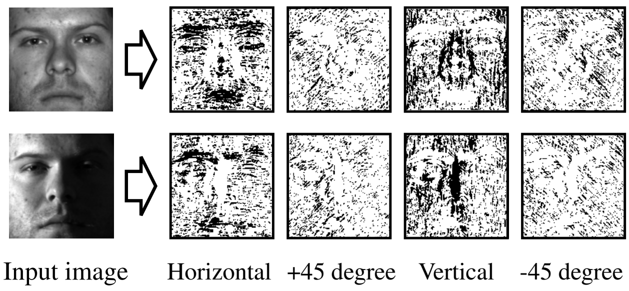

Figure 2: Edge-based feature maps generated from different illumination conditions.

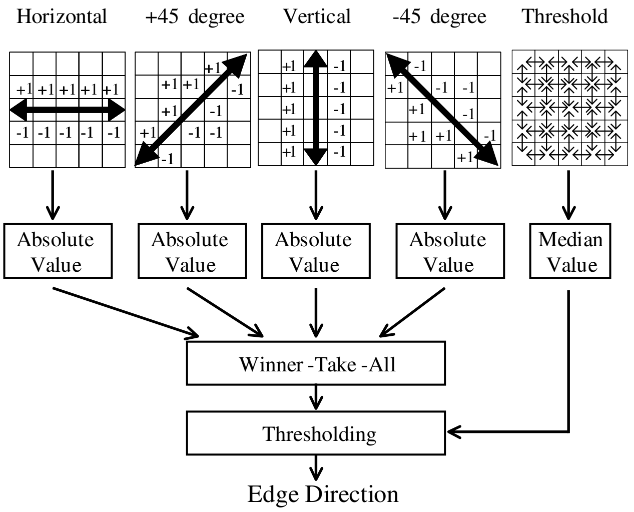

Figure 3: Four directional edges are extracted by 5×5-pixel kernel filtering for each pixel.

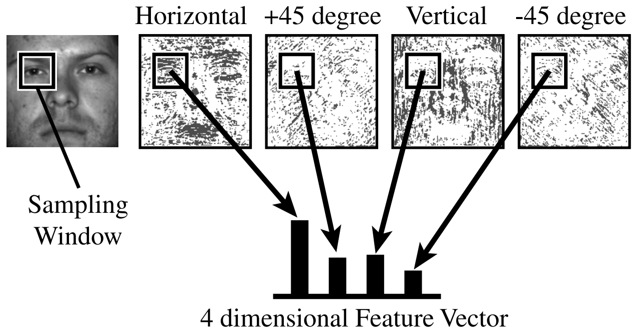

Figure 4: Feature-vector-generation scheme from edge-based feature maps: each element represents the number of edge flags within sampling window.

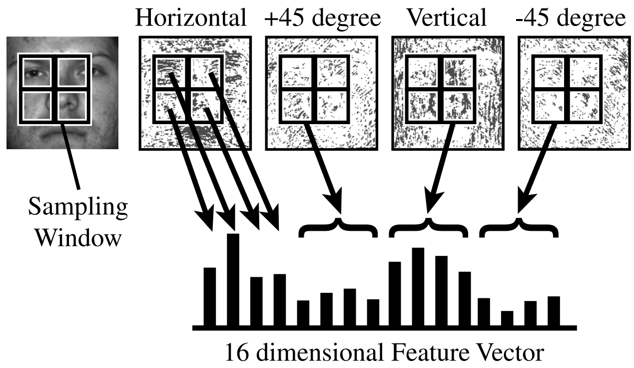

Figure 5: Another feature-vector-generation scheme: four times larger size sampling window is divided into 2×2 cells and each element represents the number of edge flags within a cell.

2. EDGE-BASED FEATURE REPRESENTATION

2.1 Edge-Based Feature Maps

Edge-based feature maps are the bases of our image representation algorithm [14]. The feature maps represent the distribution of edge flags corresponding to four-directional edges extracted from an input image as shown in Fig. 2. Figure 3 illustrates how to extract edges from an input image. The input image is firstly subjected to pixel-by-pixel spatial filtering operations using kernels of 5×5-pixel size to extract edges in four directions, i.e. horizontal, +45 degree, vertical, or -45 degree. The kernel yielding the maximum gradient determines the direction of the edge. The threshold value for edge extraction is determined taking the local variance of luminance data into account. Namely, the median of the 40 values of neighboring pixel intensity differences in a 5×5-pixel kernel is adopted as the threshold. This is quite important to retain all essential features in an input image in the feature maps. All the pertinent parameters have been optimized in the medical radiograph recognition as a test vehicle. But the robust nature has been proven in more general applications [14]. Therefore, the parameters determined in Ref. [14] are utilized as they are in the present work. Figure 2 shows an example of feature maps generated from the same person under different illumination conditions from the Yale face database B. The edge information is very well extracted from both bright and dark images.

2.2 Feature Representation

The proposed feature representation vectors are generated from the edge-based feature maps by counting the number of edge flags within the sampling window as illustrated in Fig. 4. Each element of the vector represents the number of edge flags within the window of the corresponding edge direction. Since four directional edges are available in the feature maps, four elements of the feature representation vector are generated from a sampling window. In order to represent the 2D structural information within the sampling window, another type of feature-vector-generation scheme illustrated in Fig. 5 was also employed. In this scheme, a larger size sampling window is employed and then divided into 2×2 cells. A 16-dimension feature vector is generated by counting the number of edge flags in each cell.

The sequence of observation vectors is acquired as follows. The $w \times w$-pixel window scans the image from left to right and makes the sequence of feature vectors as illustrated in Fig. 6 (a). For each sampling, the sampling window shifts s pixels to right. The horizontal scanning is repeated with the vertical pitch of s pixels from the top to bottom as shown in Fig. 6 (b). In this manner, a series of feature vec-

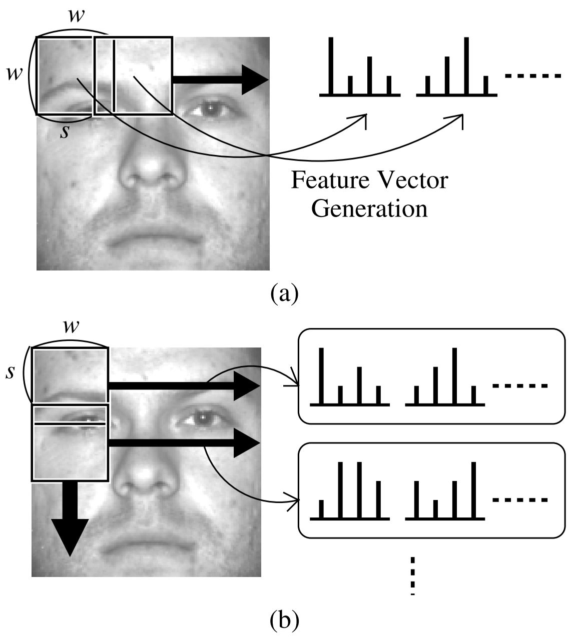

Figure 6: Sampling window scans face image; horizontal scanning (a) and vertical scrolling (b).

tors is generated. These feature vectors are utilized as observation vectors for the pseudo-2D Hidden Markov Models on both training and identification stages. The algorithms of the feature-vector generation and image scanning scheme are compatible to the processing in our VLSI chip [1].

3. FACE IDENTIFICATION USING PSEUDO-2D HIDDEN MARKOV MODELS

The identification of the target facial image is carried out as in the following. The sequence of edge-based feature-vectors is classified by the pseudo-2D Hidden Markov Models. The 6×6-state left-right model illustrated in Fig. 1 (b) is utilized in this work. The model consists of six super states each of which contains an one-dimension Hidden Markov Model with six embedded states. The sum of three Gaussian mixtures is employed for the probability function of the embedded state. These parameters of the HMM are optimized by the experimental results on the AT&T face database.

In the training, one HMM is generated for each person using the Baum-Welch algorithm. For avoiding false local minima, the common initial face model described in Ref. [11] is utilized . Namely, the common initial face model is firstly trained on all faces in the training set. The face model for each person is obtained from the common model by refining it on the training faces of the person with the Baum-Welch algorithm. In the identification, the target facial image is evaluated by each face model using the Viterbi algorithm. The face model which gives the maximum probability determines the identity of the input image.

4. EXPERIMENTAL RESULTS AND DISCUSSION

In order to evaluate the performance of the proposed system, the AT&T face database [8] was used for both training and test sets. This database contains 10 different images for each

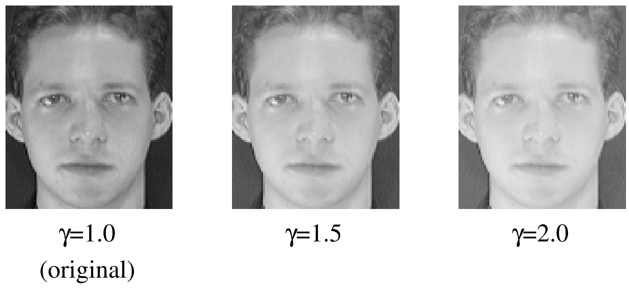

Figure 7: Test images under different illumination conditions emulated by gamma correction.

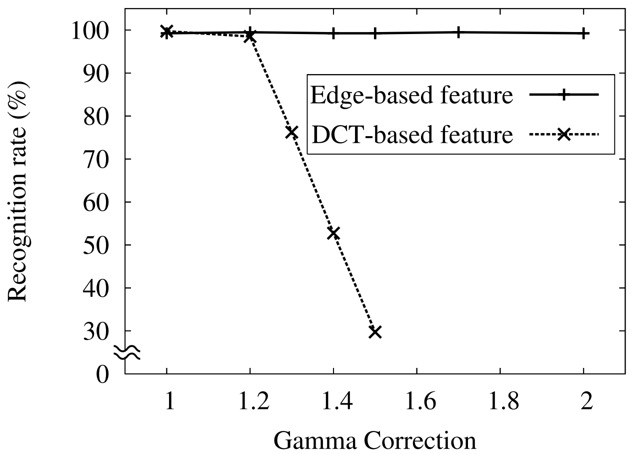

Figure 8: Recognition rate on AT&T database with different illumination conditions.

of 40 people. These face images are taken at similar illumination conditions. As a preliminary evaluation of the robustness against illumination variations, various illumination conditions were emulated by the gamma correction. Examples of the test images utilized for the experiments are shown in Fig. 7. The recognition rate was evaluated by the cross validation. Namely, each person's model was trained on all face images except for one image and the classification was carried out for the image excluded in the training. This procedure was repeated for all images in the database.

The recognition rates on various illumination conditions employing the edge-based and the DCT-based feature representations are presented in Fig. 8. In this experiment, first nine coefficients of discrete cosine transform on the sampling window are utilized for the DCT-based feature vectors, and the four dimensional feature vector of Fig. 4 was used as the edge-based feature vectors. $w = 8$ and $s = 4$ were employed for both feature vectors as the parameters of window scanning. Figure 8 shows that the recognition rate of the DCT-based feature vectors falls rapidly at $\gamma = 1.3$ while the edge-based feature vectors performed over 99% recognition rate at the entire range. Figure 9 shows the results of the Viterbi segmentation on the two test images ($\gamma = 1.0$ and $\gamma = 2.0$). Each block separated by the white lines corresponds to each state of the pseudo-2D HMM illustrated in Fig. 1. The segmentation results using the edge-based features are almost the same independent of illumination conditions as shown in

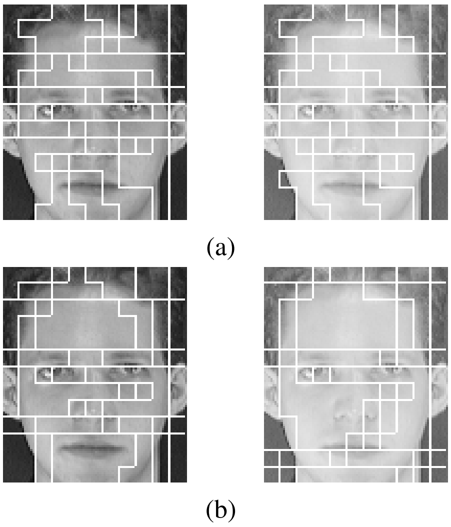

Figure 9: Viterbi segmentation of face images under different illumination conditions using edge-based feature vectors (a) and DCT-based feature vectors (b); each block corresponds to each state of pseudo-2D HMM.

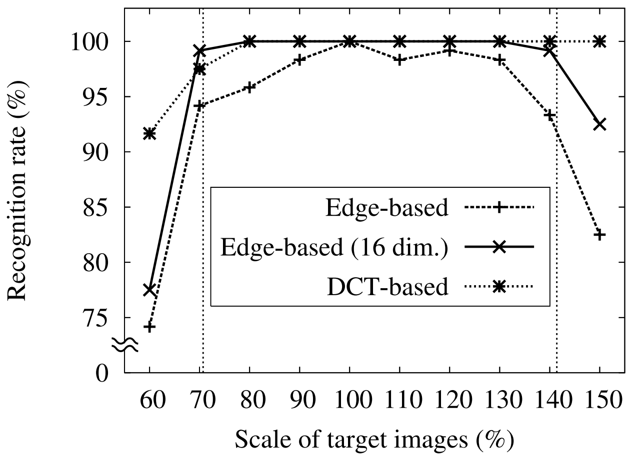

Figure 10: Recognition rate on AT&T database with different scales of target facial images.

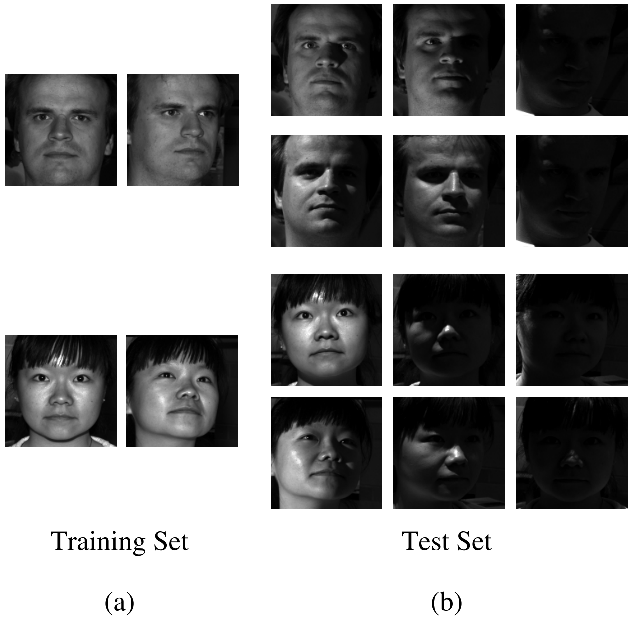

Figure 11: Examples of facial images from Yale face database B used for training (a) and test (b).

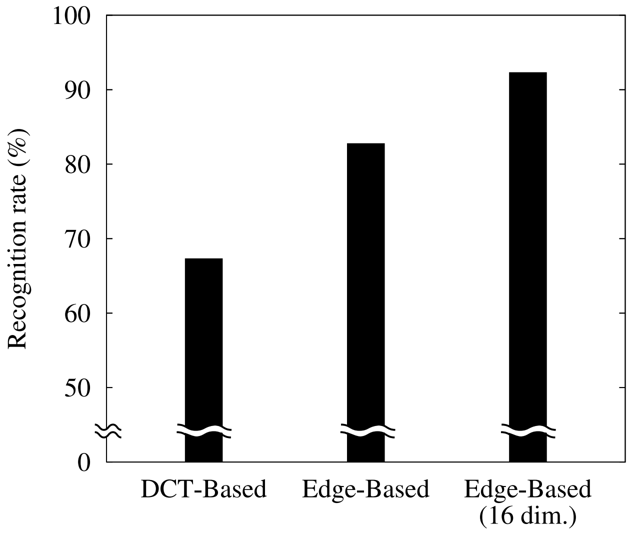

Figure 12: Recognition rate on Yale face database B.

Fig. 9 (a). On the other hand, Fig. 9 (b) shows that the states to which the nose and the mouth belong do not match between the dark and bright images in case of the DCT-based feature vectors.

Figure 10 shows the recognition rate on the various size of the target faces. Although the recognition rate using the edge-based feature vectors (4 dimensions) is a little degraded as compared to that using the DCT-based ones, the recognition rate using the 16 dimensional edge-based feature vectors is almost same with that using the DCT-based ones. Over 99% recognition rate was obtained with the 16 dimensional edge-based vectors on the scale range between 70% $\left(\approx \frac{1}{\sqrt{2}}\right)$ and 140% $\left(\approx \sqrt{2}\right)$. The scale-invariant face detection system developed in our earlier work [6] is capable of enclosing facial images of any sizes in a frame in which the scale variation of the face image is limited within the range between $\frac{1}{\sqrt{2}}$ and $\sqrt{2}$. Therefore, it would be possible to build a scale-invariant face identification system using our already-developed face detection system as a preprocessing stage in the present identification system.

The proposed system demonstrates a good performance. However, the results are obtained from the AT&T face database which consists of relatively simple images for identification. We also evaluated the performance on the Yale face database B [13]. The Yale face database B contains 5760 single light source images of 10 people each seen under 576 viewing conditions (9 poses $\times$ 64 illumination conditions). The face model of each person was learned from the nine

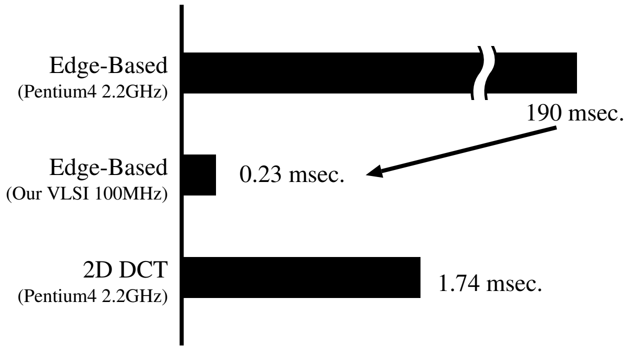

Figure 13: Comparison of computational time to generate all feature vectors from a 92×112-pixel image.

poses with frontal illumination, and then the face identification is carried out for the images under various illumination conditions. Examples of facial images are shown in Fig. 11. Figure 12 illustrates the results of face identification. The recognition rate obtained with the proposed 16 dimensional edge-based feature vectors was 92% and better than that of the DCT-based feature vectors.

The proposed feature vectors demonstrated the good performance independent of illumination conditions. However, the computational costs of generating edge-based feature vectors are very expensive. Figure 13 illustrates the computational time to generate all feature vectors required for the identification of a 92×112-pixel image. 190ms are required for generating the edge-based feature vectors using the Intel Pentium 4 processor operating at 2.2GHz. This is too much as compared to the DCT-based feature vector generation. However, it takes only 0.23ms when the dedicated feature-extraction VLSI chip developed in our group [1] is used. This is much faster than the DCT-based vector generation using a Pentium 4 processor.

5. CONCLUSION

An illumination-invariant pseudo-2D HMM-based face identification system has been developed. In order to enhance the robustness against the variation in illumination conditions, the edge-based feature representation has been employed. This edge-based feature representation is compatible to the processing in our dedicated VLSI chip, making it possible to build a real-time responding system. As a result, the recognition rate over 99% has been obtained for the AT&T face database and 92% for the Yale face database B independent of the illumination conditions, thus verifying the robustness of the system against illumination condition variations.

6. ACKNOWLEDGMENT

The authors would like to thank Takeshi Kanamori of The University of Tokyo for his contribution in carrying out the computer simulation.

REFERENCES

[1] H. Yamasaki and T. Shibata, "A real-time image-feature-extraction and vector-generation VLSI employing arrayed-shift-register architecture," in *Proc. of 31st European Solid-State Circuits Conf. (ESSCIRC)*, Grenoble, France, Sept. 2005, pp. 121–124.

[2] M. Yagi, H. Yamasaki, and T. Shibata, "A mixed-signal VLSI for real-time generation of edge-based image vectors," in *Advances in Neural Information Processing Systems 16 (Proc. of NIPS 2003)*, Vancouver, Canada, Dec. 2003, pp. 1035–1042.

[3] M. Ogawa, K. Ito, and T. Shibata, "A general-purpose vector-quantization processor employing two-dimensional bit-propagating winner-take-all," in *IEEE Symp. on VLSI Circuits Dig. Tech. Papers*, June 2002, pp. 244–247.

[4] T. Kondo and H. Yan, "Automatic human face detection and recognition under non-uniform illumination," *Pattern Recognition*, vol. 32, pp. 1707–1718, 1999.

[5] Y. Suzuki and T. Shibata, "Multiple-resolution edge-based feature representations for robust face segmentation and verification," in *Proc. EUSIPCO 2005*, Antalya, Turkey, Sept. 2005.

[6] Y. Suzuki and T. Shibata, "An edge-based face detection algorithm robust against illumination, focus, and scale variations," in *Proc. EUSIPCO 2004*, Vienna, Austria, Sept. 2004, vol. 3, pp. 2279–2282.

[7] Y. Suzuki and T. Shibata, "Multiple-clue face detection algorithm using edge-based feature vectors," in *Proc. Int'l Conf. Accoustics, Speech and Signal Processing (ICASSP 2004)*, Montreal, Canada, May 2004, vol. 5, pp. 737–740.

[8] F. Samaria and A. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota, FL, Dec. 1994, pp. 138–142.

[9] S. Kuo and O. Agazzi, "Keyword spotting in poorly printed documents using pseudo 2-D hidden Markov models," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 16, no. 8, pp. 842–848, Aug. 1994.

[10] A. Nefian and M. Hayes, "Hidden Markov models for face recognition," in *Proc. Int'l Conf. Accoustics, Speech and Signal Processing (ICASSP 1998)*, May 1998, vol. 5, pp. 2721–2724.

[11] S. Eickeler, S. Muller, and G. Rigoll, "Recognition of JPEG compressed face images based on statistical methods," *Image and Vision Computing*, vol. 18, pp. 279–287, Mar. 2000.

[12] A. Nefian and M. Hayes, "An embedded HMM-based approach for face detection and recognition," in *Proc. Int'l Conf. Accoustics, Speech and Signal Processing (ICASSP 1999)*, Mar. 1999, vol. 6, pp. 3553–3556.

[13] A. Georghiades, P. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643–660, 2001.

[14] M. Yagi and T. Shibata, "An image representation algorithm compatible with neural-associative-processor-based hardware recognition systems," *IEEE Trans. Neural Networks*, vol. 14, no. 5, pp. 1144–1161, Sept. 2003.