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

MULTI-VIEW FACE DETECTION AND POSE ESTIMATION EMPLOYING EDGE-BASED FEATURE VECTORS

Daisuke Moriya, Yasufumi Suzuki, and Tadashi Shibata

Masakazu Yagi and Kenji Takada**

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: moriya@else.k.u-tokyo.ac.jp, yasufumi@else.k.u-tokyo.ac.jp, shibata@ee.t.u-tokyo.ac.jp

*Department of Orthodontics and Dentofacial Orthopedics, School of Dentistry, Osaka University
1-8 Yamadaoka, Suita, Osaka, 565-0871, Japan
email: mgoat@dent.osaka-u.ac.jp, ktakada@dent.osaka-u.ac.jp

ABSTRACT

A multi-view face detection and pose estimation system has been developed employing the edge-based feature representations. Using the posed face images at four angles: 0°, 30°, 60°, and 90° as templates, the performance of pose estimation of about 80% has been achieved for test images in the entire angle range of 0° - 90°. In order to further enhance the performance, the concept of focus-of-attention (FOA) has been introduced in the vector generation. Namely, edge-based feature vectors are generated from the restricted area mostly containing essential information of facial images. As a result, the detection rate has been enhanced to more than 90% for profile images, which is difficult to achieve when original edge-based vectors are used.

1. INTRODUCTION

Human face detection is an important processing in a variety of applications such as security cameras, video surveillance systems, robot vision and so forth. A number of algorithms have been proposed so far. However, face detection is still a challenging task since it is difficult to achieve robustness under various situations such as different illumination, focus, scales, and poses, while achieving real-time performance. The face detection algorithms are classified into two categories: image-based approaches and feature-based approaches [1]. The feature-based approaches make explicit use of facial features such as skin color information [2], geometrical relationship among facial parts (i.e. the eyes, the mouth, and the nose) [3] and so on. Such features, however, are not immune to variations in environmental conditions. For example, skin color information is sensitive to the change in illumination conditions and geometrical relationship is likely to be affected by beards, mustaches, hairstyles, and so on. In image-based approaches, on the other hand, face images are handled as a whole, and statistical learning algorithms such as neural networks [4], eigenfaces [5], etc. are employed for detection. Usually the image-based approaches show better performance than the feature-based approaches. However, computationally the former is much more expensive than the latter.

A robust face detection algorithm has been developed, employing edge-based feature vectors in conjunction with the template matching technique [6][7][8]. It has been shown that the algorithm is very robust against illumination variations, defocussing, scale variations, and image rotation. Such a performance can be attributed to the robust nature of edge-based feature representation originally developed for medical radio-graph analysis and handwritten pattern recognition [9][10]. The face detection systems presented in Ref. [6][7][8], however, have been developed only for frontal face detection.

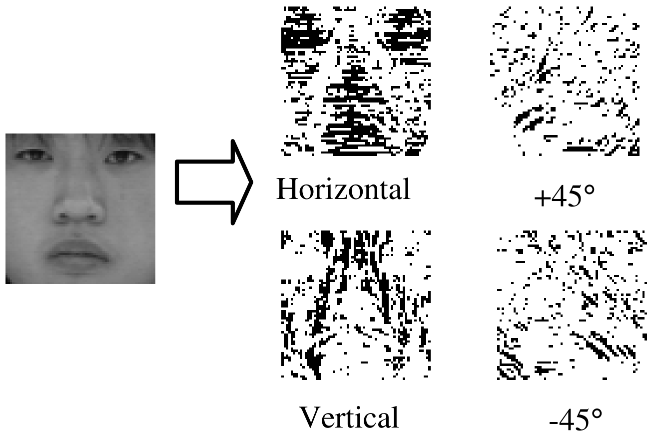

Figure 1: Edge-based feature maps.

Then the purpose of this paper is to further extend the capability of the system and make it applicable to multi-view face detection. The ultimate goal of this research is to develop a medical analysis system extracting structural characteristics of a human face from multi-view angles and to apply the multi-dimensional facial soft tissue analysis for the treatment planning in orthodontics. In order to detect omni-directional posed faces robustly, the concept of focus-of-attention (FOA) has been introduced in the vector generation. Namely, edge-based feature vectors are generated from the restricted area mostly containing the essential information of facial images. As a result, more than 90% detection rate has been achieved for profile faces which is difficult with the original edge-based vectors.

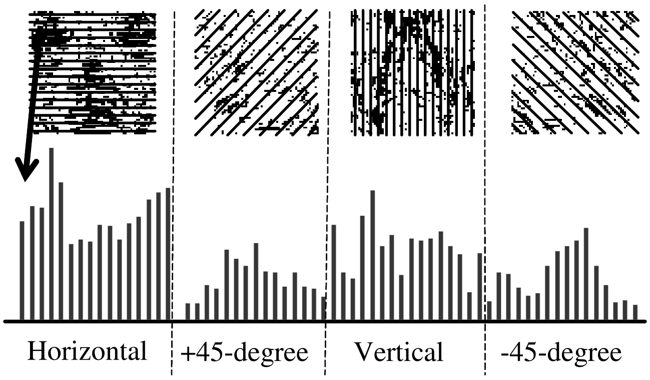

Figure 2: Feature vector based on Projected Principal-Edge Distribution (PPED); PPED vector is performed by taking a histogram of edge flags along same direction with edges.

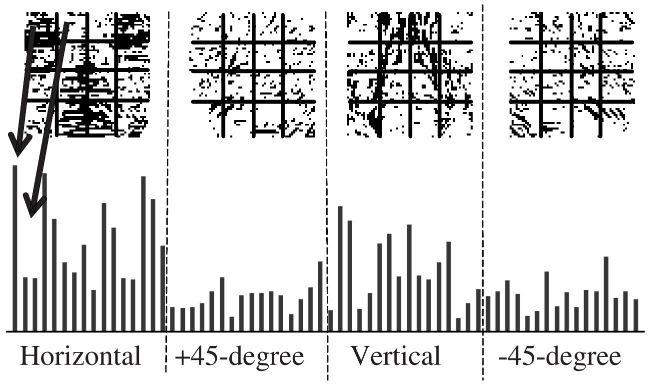

Figure 3: Feature vector based on Cell Edge Distribution (CED); CED vector is performed by counting number of edge flags within each cell.

2. IMAGE PERCEPTION USING EDGE-BASED FEATURE VECTORS

2.1 Vector generation algorithms

The first step of the perception is the feature map generation which extracts edge information from a 64×64-pixel input images. Figure 1 shows an input image and its four-directional feature maps. Each feature map represents the distribution of edges corresponding to each direction, i.e. horizontal, $+45°$, vertical, and $-45°$ in the 64×64-pixel image. The feature maps are the bases of our feature representations and all of the feature vectors utilized in this paper are generated from the feature maps. 64-dimension vectors are generated as spatial distribution histograms of edge flags in the feature maps. We have already developed two types of feature representation, the Projected Principal-Edge Distribution (PPED) [11] and the Cell Edge Distribution (CED) [6]. Figure 2 illustrates the feature-vector-generation procedure of the Projected Principal-Edge Distribution (PPED). 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. Another scheme of vector generation is shown in Fig. 3. In the Cell Edge Distribution (CED) vector generation, each feature map is divided into 4×4 cells and each cell contains 16×16-pixel sites. Each element in the CED vector indicates the number of edge flags within the corresponding cell.

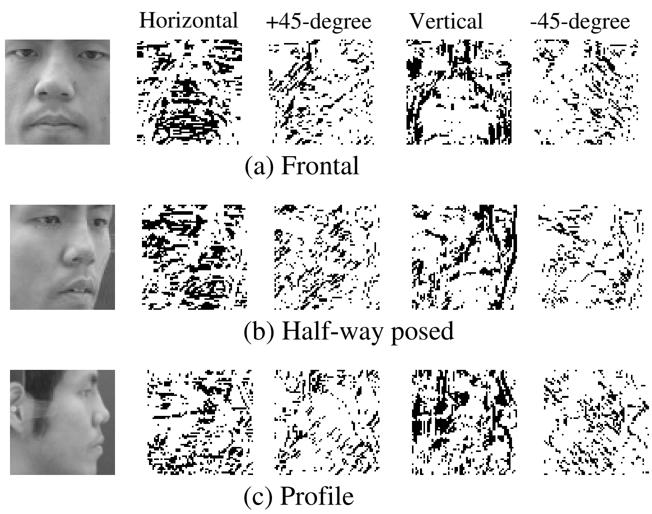

Figure 4: Feature maps generated from frontal face (a), half-way posed face (b), and profile (c) images .

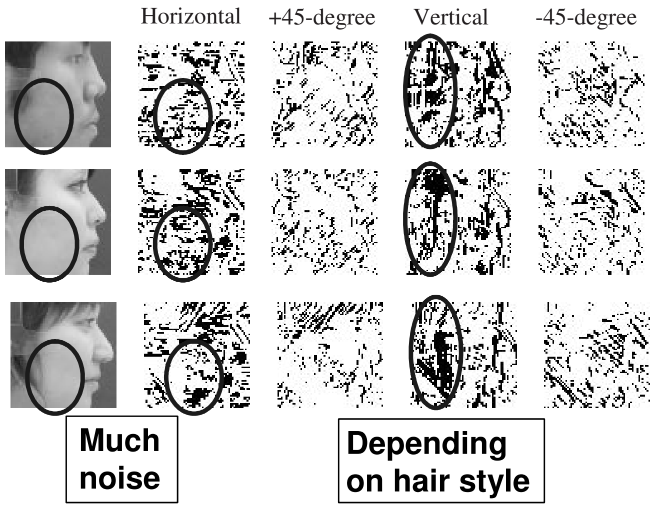

Figure 5: Feature maps generated from profile images.

2.2 Profile-specific feature representation

Figure 4 shows the feature maps generated from frontal face, half-way posed face, and profile images of one person. Facial parts such as the eyes, the nose and the mouth can be easily recognized from the horizontal and vertical feature maps generated from the frontal face images. However, it is difficult to identify the patterns of facial parts from the feature maps of the profile. Figure 5 illustrates the comparison with the feature maps from three profile images. Many edges are detected on the cheek area where no specific features are visible in original images. Moreover, feature maps of back side of the profile vary depending on the hairstyle of each person. In addition, a number of characteristic edge distribution are

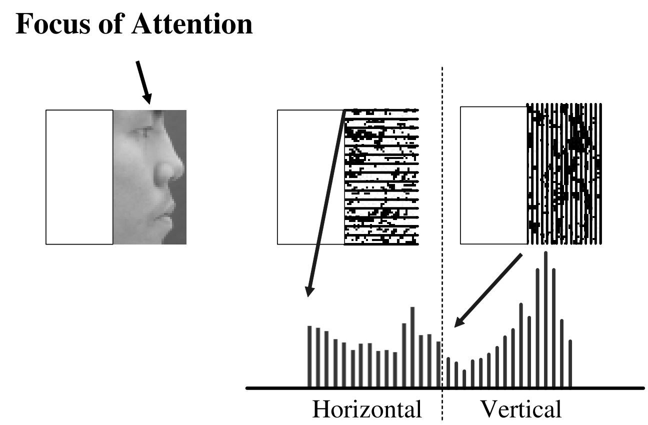

Figure 6: Feature vector based on Projected Principal-Edge Distribution (PPED) generated from the focus-of-attention (FOA) area.

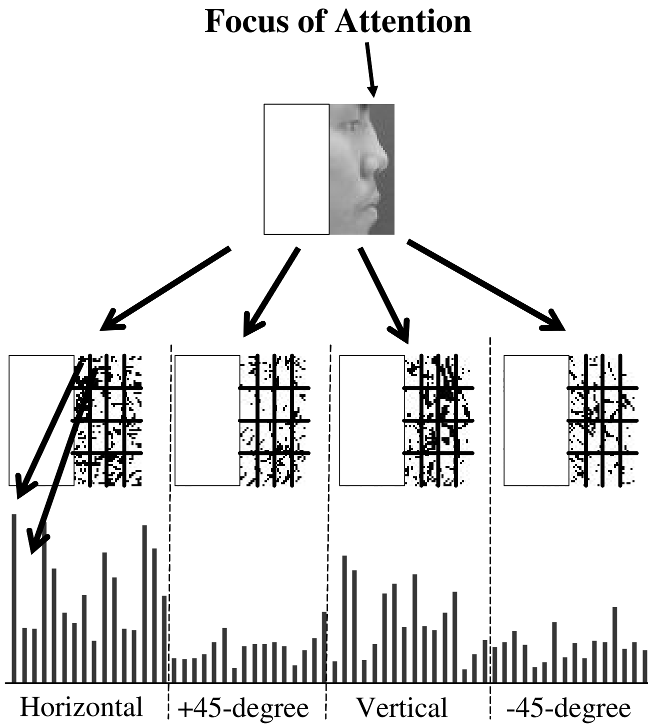

Figure 7: Feature vector based on Cell Edge Distribution (CED) generated from the focus-of-attention (FOA) area.

present at the rear side of the face in the vertical edge map. However, they are due to the hairstyle and vary a lot from person to person. Therefore, the front side of profile image is utilized for focus-of attention (FOA) area of the profile. Two kinds of feature vectors are generated from the focus-of-attention (FOA) area. One is the PPED-like vector (we call this PPED* for short) shown in Fig. 6. Namely, the feature vector is generated by taking the histogram of edge flags in every four rows for the horizontal edge map and in every two columns for the vertical edge map. The other is similar to the CED vector-generation scheme (we call this CED*

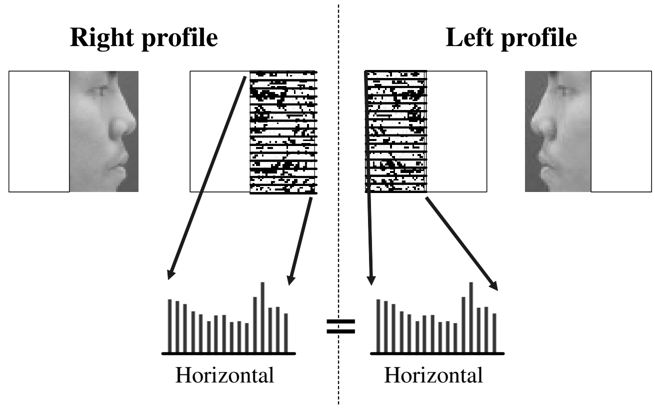

Figure 8: Horizontal part of PPED vector of right and left profile; they are same due to horizontal projection.

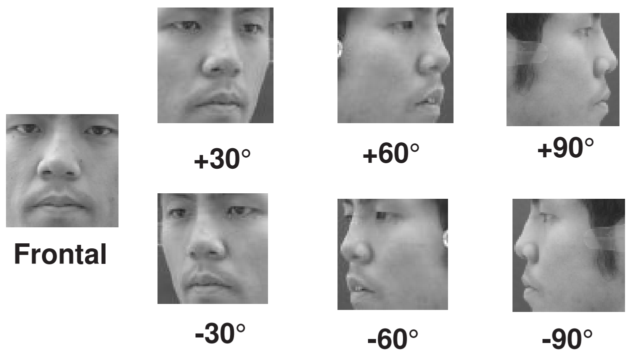

Figure 9: Template images of posed faces.

for short) illustrated in Fig. 7. The FOA area is divided into 4×4 cells each of which contains 8×16-pixel sites and each element of the 64-dimension vector represents the number of edge flags within the corresponding cell.

Figure 8 illustrates PPED* feature vectors from both the profile and its mirror images. The horizontal projection makes the same vectors from both horizontal edge maps. This makes it difficult to distinguish the PPED vector of right profile from one of left profile. Therefore, CED* feature vector is employed for the profile-specific feature representation.

3. EXPERIMENTAL RESULTS AND DISCUSSION

The samples used in this work were 130 face images of 13 people taken from 10 directions between the range of 0° to 90° with an increment of 10°. The sample photos were prepared as a preliminary database to use in the multi-dimensional facial soft tissue analysis for the treatment planning in orthodontics. Four direction posed faces at 0°, 30°, 60° and 90° were utilized as templates. In addition, −30°, −60° and −90° direction faces were generated by taking the mirror images of original samples as shown in Fig. 9. Face detection was carried out on face images angled at 0°, 20°, 40°, 60°, 80°, and 90° for 13 people as shown in Fig. 10. The detection rate was evaluated by the cross validation. Namely, all face images except for one person were utilized as templates and the face detection carried out for the face images of the person excluded from the templates. This procedure was

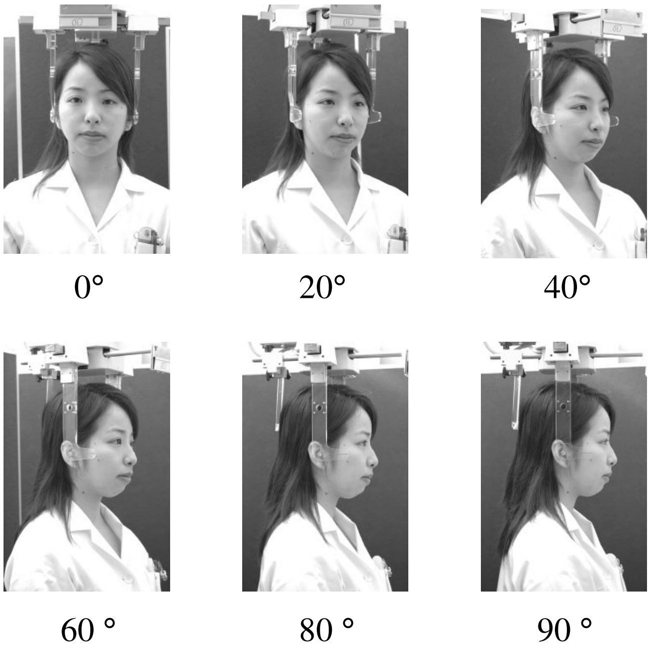

Figure 10: Target images of posed faces.

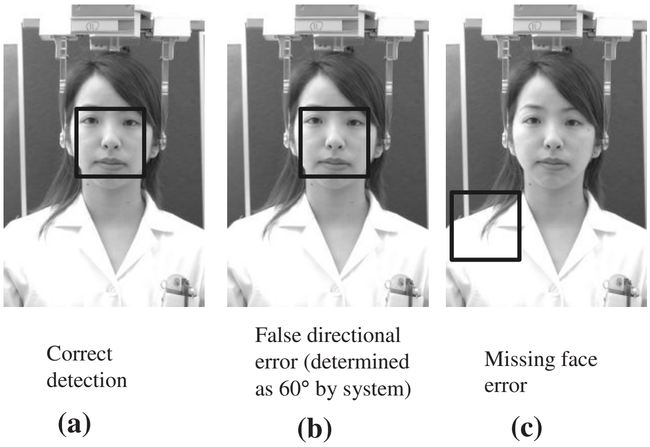

Figure 11: Correct detection (a) and two kinds of errors (b)(c) of system.

repeated for all images in the database. In this experiments, based on the prior knowledge that only one face image is existing in the target image, the location where the feature vector gives the minimum distance to the best matched template was determined as a face. And the direction of the template which gave the minimum distance among the all templates determines the pose of the target face. There are two kinds of errors in the detection system as illustrated in Fig. 11. Figure 11 (b) shows a "false directional error" which means that the face is localized correctly but the pose estimation of the detected face is failed. The other type of errors is a "missing face error" in which the system cannot localize the face correctly.

Figure 12 and Fig. 13 shows the detection rates of six direction faces using PPED and CED feature vectors, respectively. Although almost over 80% detection rates are obtained in frontal and 20° faces for both feature vectors, the detection rates degrade to less than 60% in profile images. In order to improve the performance, the multiple-clue criteria [6] have been utilized. Namely, the distance was evaluated using both PPED and CED vectors, and the detection was carried out using the sum of both distances and searching for the minimum. Figure 14 illustrates the correct detection rates using the multiple-clue criteria. The detection rates all of angeled faces were improved and approximate 80% detection rates were obtained.

Figure 15 shows the detection rates of profile images for three types of feature vectors: PPED, CED, and CED*. In addition, the detection rates using the multiple clue derived from all three feature vectors are also shown in Figure 15. Only about 50% of the target profile images were detected with the original feature vectors (PPED and CED). However, over 90% detection rate was obtained with the proposed

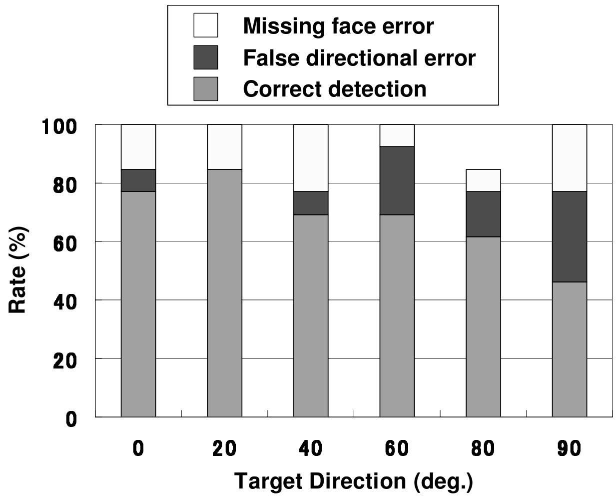

Figure 12: Face detection rate of omni-directional faces using PPED feature vector.

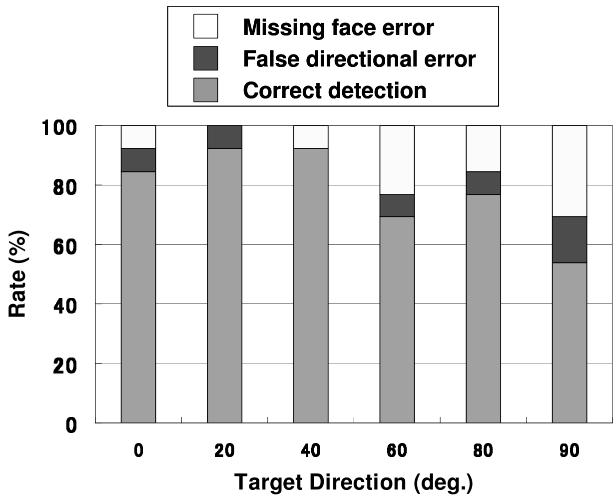

Figure 13: Face detection rate of omni-directional faces using CED feature vector.

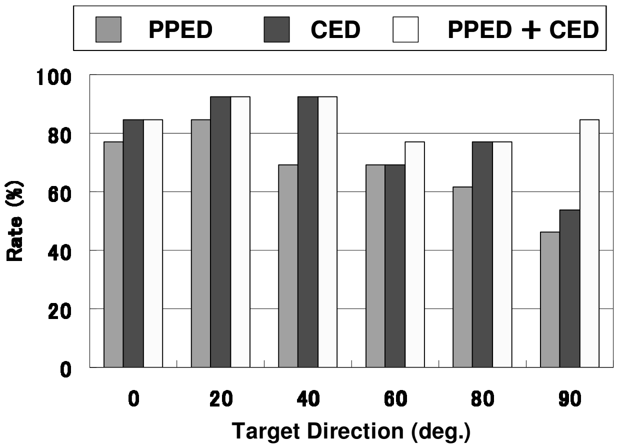

Figure 14: Face detection rate of omni-directional faces employing multiple-clue scheme.

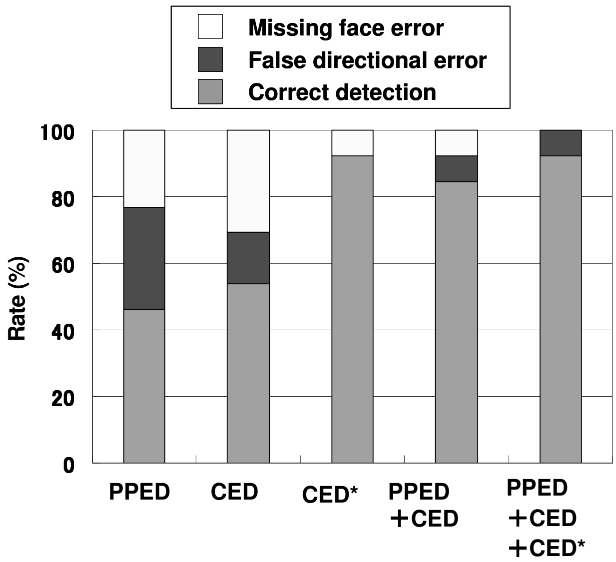

Figure 15: Profile detection rate using PPED, CED, CED* and multiple clue of them.

CED* vector.

4. CONCLUSION

A multi-view face detection and pose estimation system employing the edge-based feature representations has been developed. Approximate 80% detection rates were obtained by emplyoing the multiple clue derived from our original feature vectors. In order to enhance the performance of the system, the concept of focus-of-attention (FOA) has been introduced in the feature-vector generation. As a result, more than 90% detection rate has been achived for profile and the peformance of the system has been sucessfully improved.

REFERENCES

[1] E.Hjelmas and B.K.Low, "Face detection: A survey," in *Computer Vision and Image Ungerstanding*, vol. 83, pp.236-274, 2001.

[2] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," in *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 9, no.4, pp.551-564, 1999.

[3] C. Lin and K. C. Fan, "Human Face Detection Using Geometric Triangle Relationship," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol.2, pp.945-948, 2000.

[4] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," in *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 23-38, Jan. 1998.

[5] C. Liu and H. Wechsler, "Gabor Feature Based Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," in *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467-476, Apr. 2002.

[6] Y. Suzuki and T. Shibata, "Multiple-clue face detection algorithm using edge-based feature vectors," in *Proc. ICASSP 2004*, Montreal, Canada, May 17-21. 2004, vol. 5, pp. 737–740.

[7] Y. Suzuki and T. Shibata, "An edge-based face detection algorithm robust against illumination, focus, and scale variations," in *Proc. EUSIPCO 2004*, Vienna, Austria, September 6-10. 2004, pp. 2279–2282.

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

[9] 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.

[10] 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.

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