

AN EDGE-BASED FACE DETECTION ALGORITHM ROBUST AGAINST ILLUMINATION, FOCUS, AND SCALE VARIATIONS

Yasufumi Suzuki and Tadashi Shibata

Department of Frontier Informatics, School of Frontier Science, The University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan
yasufumi@else.k.u-tokyo.ac.jp, shibata@ee.t.u-tokyo.ac.jp

ABSTRACT

A face detection algorithm very robust against illumination, focus and scale variations in input images has been developed based on the edge-based image representation. The multiple-clue face detection algorithm developed in our previous work [1] has been employed in conjunction with a new decision criterion called “density rule,” where only high density clusters of detected face candidates are retained as faces. As a result, the occurrence of false negatives has been greatly reduced. The robustness of the algorithm against circumstance variations has been demonstrated.

1. INTRODUCTION

The development of robust image recognition systems is quite essential in a variety of applications such as intelligent human-computer interfaces, robotics, security systems, and so forth. Real-time face recognition, in particular, plays an important role in establishing user-friendly human-computer interfacing. In realizing real-time automatic face recognition systems, the issues are twofold [2]. In the first place, locations of faces in an input image must be detected robustly independent of circumstance conditions such as illumination, focus, size of faces, and so on. The most important point at this stage is to detect all faces without loss due to false negatives despite some false positives included in the detection results. The decision that a partial image belongs to faces or non faces needs be done very fast because such decision must be repeated for all locations in the entire image.

The detected face candidates are then identified as individuals by matching with stored samples in the system. The computational cost at this stage is quite reduced as compared to that in the first stage since only a small number of candidates are involved in the recognition process. This allows us to execute more complex computation for verification and identification. Therefore, false positives detected in the first stage can be eliminated by this operation.

A number of face detection algorithms such as those using eigenfaces [3] and neural networks [4], for instance, have been developed. In these algorithms, however, a large amount of numerical computation is required, making the processing extremely time-consuming. Therefore, it is not feasible to build real-time responding systems by software running on general-purpose computers. In this regard, the development of hardware-friendly algorithms compatible to dedicated VLSI chips is quite essential.

For this purpose, we have developed search engine VLSI chips in both CMOS digital [5] and analog [6] technologies. The projected principal-edge distribution (PPED) algorithm has been developed as an image representation scheme compatible to the search engine VLSI's and its robust nature has

been proven in applications to hand-written pattern recognition and medical X-ray analysis [7][8][9]. The multiple-clue pattern classification algorithm has been developed as an extension of the PPED and successfully applied to face detection [1]. However, we found a severe degradation in the performance when the number of face templates is increased aiming at enhancing its robustness against the variations in input images.

The purpose of this paper is to analyze the cause of the degradation and to develop a face detection algorithm that is very robust against illumination, focus, and scale variations. We have introduced a new decision criterion called “density rule,” where only high density clusters of detected face candidates are retained as faces. And the enhanced robustness of the multiple-clue face detection algorithm has been demonstrated for a number of sample images.

2. FACE DETECTION ALGORITHM

2.1 Edge-Based Feature Maps

Edge-based feature maps are the very bases of our image representation algorithm [1][9]. The feature maps represent the distribution of four-direction edges extracted from a 64×64 -pixel input image. The input image is first subjected to pixel-by-pixel spatial filtering operations using kernels of 5×5 -pixel size to detect edges in four directions, i.e. horizontal, +45 degree, vertical, or -45 degree. The threshold for edge detection 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. Fig. 1 shows an example of feature maps generated from the same person under different illumination conditions. The edge information is very well extracted from both bright and dark images.

2.2 Feature Vectors

64-dimension feature vectors are generated from feature maps by taking the spatial distribution histograms of edge flags. In this work, three types of feature vectors, two general-purpose vectors and a face-specific vector generated from the same set of feature maps were employed to perform multiple-clue face detection algorithm [1]. Fig. 2 illustrates the feature vector generation procedure in the projected principal-edge distribution (PPED) [9]. This provides a general purpose vector. In the horizontal edge map, for example, edge flags in every four rows are accumulated and the spatial distribution of edge flags along the vertical axis is represented by a histogram. Similar histograms are generated

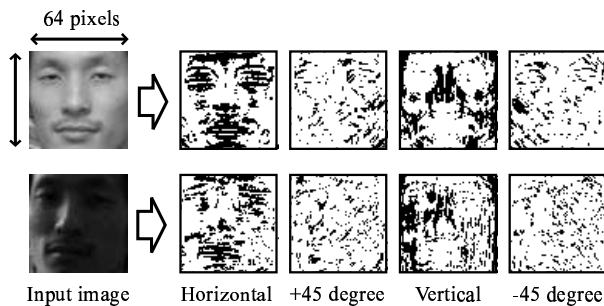


Figure 1: Feature maps generated from bright and dark face images.

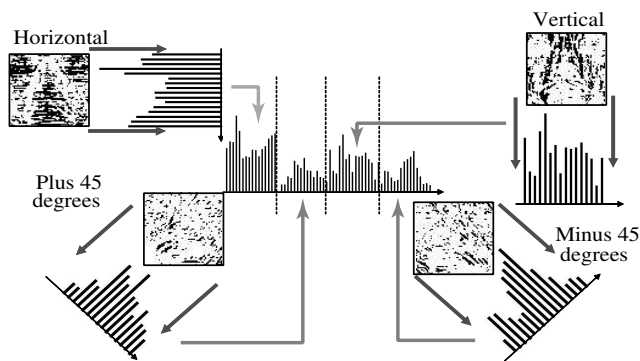


Figure 2: Feature vector generation based on projected principal-edge distribution (PPED).

from three other feature maps, and a 64-dimension vector is formed by concatenating the four histograms.

Generation of the other general-purpose vector called the cell edge distribution (CED) vector is illustrated in Fig. 3. Each feature map is divided into 4×4 cells. Each element in a CED vector indicates the number of edge flags in the corresponding cell.

A face specific feature vector generation scheme called the eyes-and-mouth (EM) extraction is shown in Fig. 4. Two 16-pixel-high bands of rows are cut from the horizontal feature map. They correspond to the location of eyes and a mouth when the 64×64 -pixel window encloses a human face. Then the number of edge flags in two neighboring columns is counted to yield a single vector element in a 64-dimension EM vector.

2.3 Face Detection by Template Matching

Face detection is carried out in two steps : the coarse selection step and the fine selection step. In the coarse selection step, a 64×64 -pixel area is taken from the input image and a feature vector is generated. Then, the feature vector is matched with all template vectors of 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. The matching is carried out using the Manhattan distance as the dissimilarity measure.

As the criterion in the coarse selection step, the multiple-clue method is utilized [1]. Template matching is performed using three feature vector generation algorithms: PPED,

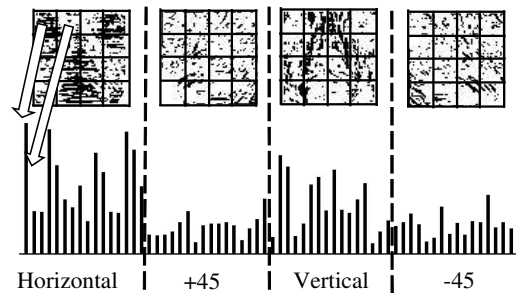


Figure 3: Feature vector generation based on cell edge distribution (CED) algorithm.

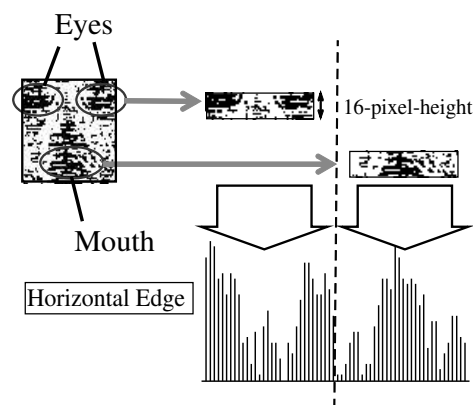


Figure 4: Feature vector generated by the eyes-and-mouth (EM) extraction.

CED and EM. If a local image is classified as a face by all the three vector representations, then it is adopted as a face candidate. This classification is carried out by pixel-by-pixel scanning of the 64×64 -pixel window over the entire image. An example of such coarse selection is demonstrated in Fig. 5 (b).

Due to the robustness of the edge-based feature representations, multiple pixel sites are detected as face candidates in the vicinity of a real face. In our previous work [1], the “regional minimum selection rule” was adopted to leave only one face candidate at the location of a real face in the fine selection step. We assumed that two faces do not overlap in a 64×64 -pixel window, and that only one candidate should remain within the window. Therefore, the face candidate having the minimum matching distance within each 64×64 -pixel window is retained as a face in the fine selection step. Although the regional minimum selection rule works quite well with a small number of face templates, increasing the number of face templates degrades the performance. Fig. 6 (a) demonstrates the example of failure in regional minimum selection where 1500 templates were used as face samples. Although the face was detected in the coarse selection stage as show in Fig. 5 (b), the candidates detected from the true face were removed due to candidates around the face which have better matching scores. Increasing the number of face templates raises the probability of non-face images accidentally matching to face templates with very small distances.

In order to solve this problem, we introduced a new selec-

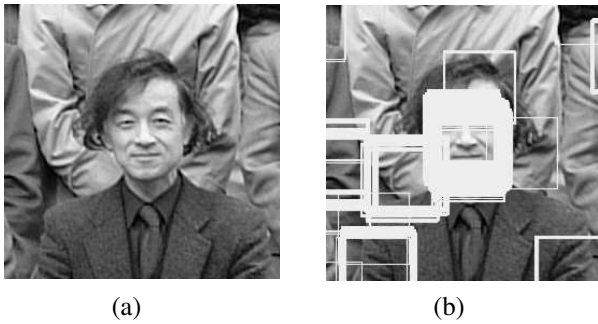


Figure 5: Results of coarse selection in multiple-clue face detection: (a) input image; (b) after coarse selection step.

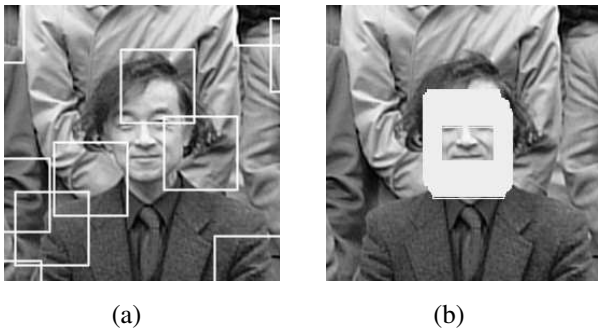


Figure 6: Results of fine selection: using regional minimum selection rule (a) and using density rule (b).

tion criterion called the density rule in the fine selection step as an alternative to the regional minimum selection rule. The density rule is based on the fact that a cluster of high-density face candidates is formed near the real face location as show in Fig. 5 (b). In addition, false positives arising from non-face input images are likely to appear very sparsely. Therefore, we have introduced a threshold for the density of face candidates. Namely, if the local density of face candidates is higher than the threshold, the location is identified as a face, or discarded otherwise. As the result of examinations in various conditions, we adopted 80 candidates within 15×30 -pixels rectangle as the threshold value. Fig. 6 (b) shows the result of the fine selection employing the density rule.

2.4 Scaling

In the present system, the fixed-size 64×64 -pixel window is utilized in the feature vector generation and the template vectors were generated so that the window just fits the size of faces. However, various size faces can appear in an input image and they will not necessarily fit to the window size. In order to solve the problem, we introduced the concept of scaling to the template images as well as to the input image. The size of the input image is reduced by factors of $1/2$, $1/4$, $1/8 \dots$ and face detection is carried out for each reduction. In addition, the size of the template images are also varied as 100%, 90%, 80%, 70% and 60%. In this manner, nearly continuous scaling has been realized in the matching operation. Due to the robust nature of the edge-based representation, the discontinuity of 10% in scaling does not affect the matching performance.

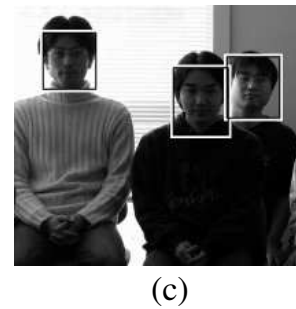
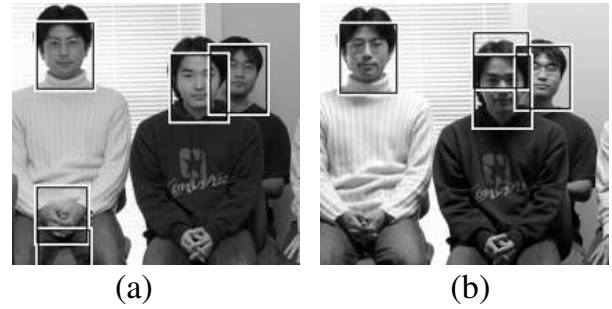


Figure 7: Face detection results under varying illumination conditions.

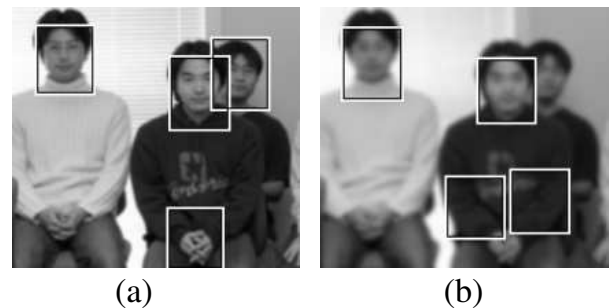


Figure 8: Results of face detection using blurred images: (a) 5×5 Gaussian filter; (b) 11×11 Gaussian filter.

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this work, 1500 face images and 2000 non-face images are utilized as templates. Template images of faces are all frontal faces of 300 people [11]. As scale variations of face templates, 90%, 80%, 70%, 5% and 60% sized face images of them are used in addition to the original images. Template images of non-faces were chosen randomly from the background scenery. The results of face detection experiment using pictures taken under the three different illumination conditions and the results are shown in Fig. 7. All frontal faces were detected correctly even under the back-light dim condition.

Fig. 8 illustrates the result of face detection using blurred images. The defocusing of the images was performed by Gaussian blur filter. In order to generate the images shown in Fig. 8 (a) and Fig. 8 (b), 5×5 and 11×11 Gaussian filter were utilized to the picture shown in Fig. 7 (a), respectively. All faces are detected correctly in the image with 5×5 Gaussian filter. However, some faces failed to be detected in the 11×11 Gaussian filtered image since edge flags of sharp

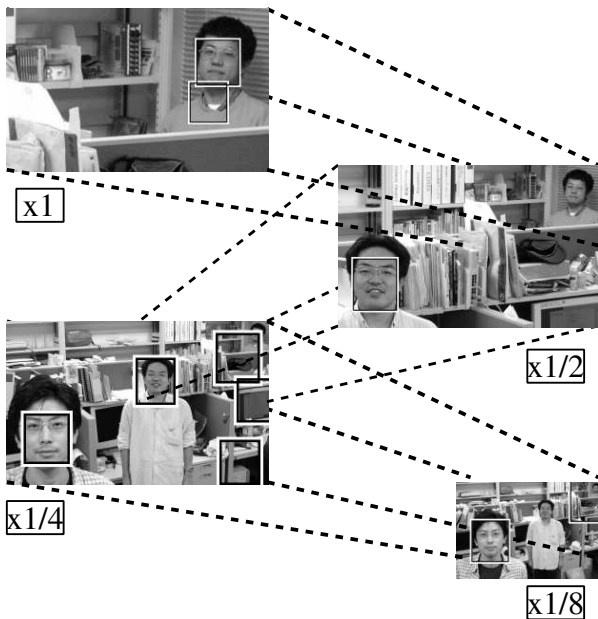


Figure 9: Results of scaling face detection using original, 1/2, 1/4, and 1/8 scale images.

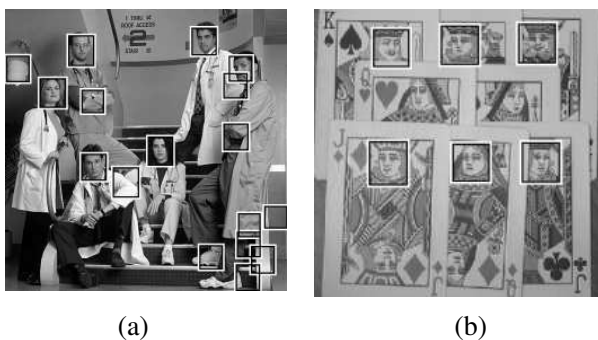


Figure 10: Detection results using images in Ref. [10] (a) and in Ref. [4] (b).

edges were enlarged in the feature maps by the blur operation and flags of weak edges were lost.

The results of face detection using the image that contains different size of faces are illustrated in Fig. 9. All faces were detected correctly on one or two scale factors of the image. This indicates that continuous scaling has been realized.

Fig. 10 illustrates the results of detecting faces on the picture in Ref. [10] and the face drawing in Ref. [4]. All real faces in the picture were detected correctly. However, at first, no face drawing in Fig. 10 (b) was detected because the density of face candidates of drawings was smaller than that of real faces. Therefore, we adopted 20 candidates as the threshold value for the image in Fig. 10 (b), then 6 face drawings out of 8 were detected.

4. CONCLUSION

A robust face detection algorithm based on the edge-based image representation has been developed. The multiple-clue face detection algorithm has been employed in conjunction with the density rule where only high density clusters of de-

tected face candidates are retained as faces. As a result, the occurrence of false negatives has been greatly reduced. And the robustness of the algorithm against the variance of illumination, focus and scales has been demonstrated for a number of sample images.

REFERENCES

- [1] Y. Suzuki and T. Shibata, "Multiple-Clue Face Detection Algorithm Using Edge-Based Feature Vectors," accepted for presentation in IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2004), May 2004.
- [2] T. Kondo and H. Yan, "Automatic human face detection and recognition under non-uniform illumination," *Pattern Recognition Letters*, vol. 32, pp.1707-1718, 1999.
- [3] C. Liu and H. Wechsler, "Gabor Feature Base Classification Using the Enhanced Fisher Linear Discriminant Model for Face Recognition," *IEEE Trans. Image Processing*, Vol. 11, no. 4, pp. 467-476, Apr. 2002.
- [4] H. Rowley, S. Baluja, and T. Kanade, "Neural Network-Based Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 20, no. 1, pp. 23-38, Jan. 1998.
- [5] 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, pp. 244-247, Jun. 2002.
- [6] T. Yamasaki and T. Shibata, "Analog Soft-Pattern-Matching Classifier Using Floating-Gate MOS Technology," *IEEE Trans. Neural Networks*, Vol. 14, no. 5, pp.1257-1265, Sep. 2003.
- [7] M. Yagi, M Adachi and T. Shibata, "A Hardware-Friendly Soft-Computing Algorithm for Image Recognition," Proc. of 10th European Signal Processing Conference (EUSIPCO 2000), pp. 729-732, Sep. 2000.
- [8] M. Yagi and T. Shibata, "Human-Perception-Like Image Recognition System Based on the Associative Processor Architecture," in the Proc. of 11th European Signal Processing Conference (EUSIPCO 2002), pp. I-103 - I-106, Sep. 2002
- [9] 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, Sep. 2003.
- [10] K. Sung and T. Poggio, "Example-Based Learning for View-Based Human Face Detection", *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no.1, pp. 39-51, Jan. 1998.
- [11] Softpia Japan, Research and Development Division, HOIP Laboratory, "HOIP facial image database," http://www.hoip.jp/web_catalog/top.html