MULTIPLE-CLUE FACE DETECTION ALGORITHM USING EDGE-BASED FEATURE VECTORS

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 multiple-clue image perception has been developed aiming at direct implementation in a VLSI hardware accelerator, and was applied to the problem of face detection. In the algorithm, feature vectors are generated utilizing distribution of edges in an input image, thus achieving dimensionality reduction for efficient processing. In addition, multiple clues in the edge distribution are utilized to enhance the accuracy of face detection. For this purpose, several feature vector generation schemes have been developed, which are all compatible to the direct hardware implementation. In software simulation, over 91% of human faces have been detected correctly with only 20 face templates, and the number of false positives has been reduced drastically by the multiple-clue scheme.

1. INTRODUCTION

The development of human-like robust image recognition systems is quite essential in a variety of applications such as intelligent human-computer interfaces, robotics, and so forth. Real-time face recognition, in particular, plays an important role to establish a user-friendly human-computer interfacing. In automatic face recognition systems, face detection is one of the important pre-processing steps to identify the location of each face in input images. A number of face detection algorithms using such as eigenfaces [1] and neural networks [2] have been developed. In these algorithms, however, a large amount of numerical computation is required, making the processing computationally very expensive. Therefore, it is not feasible to build real-time responding systems by software running on general-purpose computers. In addition, since these algorithms are based on complex floating-point operations, direct system-on-chip (SOC) implementation is not an efficient solution, either. In this regard, the development of hardware-friendly algorithms is a critical issue. For this purpose, we have developed Projected Principal-Edge Distribution (PPED) algorithm for robust image representation [3][4], and highly-parallel vector quantization VLSI chips in both digital [5] and analog [6] technologies. In the PPED algorithm, four-directional edges are detected to form edge maps, which we call feature maps. A feature vector is generated by forming spatial distribution histograms of detected edges, which is then subjected to template matching. The robust nature of the PPED representation has been demonstrated in medical radiograph analysis [3], hand-written character recognition [6], and so on. However, the original PPED cannot be applied to face detection as it is, since a lot of false positives result. This is because some essential features specific to human faces are lost during the transformation from feature maps to a feature vector.

The purpose of this paper is to present a robust face detection algorithm using multiple clues obtained from the feature maps. In this scheme, several feature vector generation methods are employed in addition to PPED. The vector generation algorithms are all compatible to the VLSI chips developed in Refs. [4]-[6]. The locations of the faces are identified by combining all the results of the face detection conducted in each vector space. The software simulation has been carried out in order to demonstrate the validity of the algorithm. As a result, the number of false positives has been greatly reduced in the multiple-clue method as compared to that in original PPED, while robustly detecting human faces using a very small number of face samples as templates.

2. MULTIPLE-CLUE IMAGE PERCEPTION USING EDGE-BASED FEATURE VECTORS

2.1. Edge-based feature maps

The first step of the system is the feature map generation which extracts edge information from 64x64 pixel input images. Fig. 1 shows an input image and its four-directional feature maps. Each feature map represents the distribution of edge flags corresponding to each direction, i.e. horizontal, +45 degree, vertical, or -45 degree in the 64x64 pixel image. The feature maps are the most fundamental features extracted from the original image and all the feature vector generation algorithms we propose in this paper are based on the feature maps.

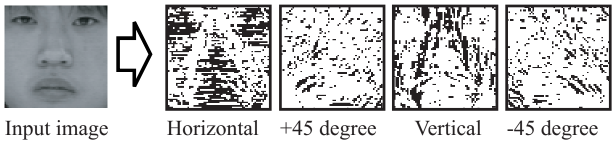


Fig. 1. Feature map images.

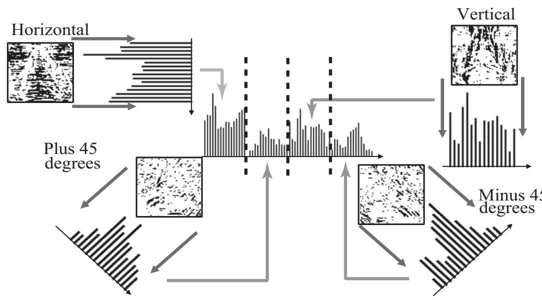


Fig. 2. Feature vector based on Projected Principal-Edge Distribution (PPED).

2.2. Feature vectors

Although feature maps very well represent the image features, the amount of data is still massive and dimensionality reduction is essential for efficient processing. 64-dimension vectors are generated as spatial distribution histograms of edge flags in feature maps. We have introduced two new vector representations, namely the general-purpose representation and the target specific representation in addition to PPED.

Fig. 2 illustrates the feature vector generation procedure in the original Projected Principal-Edge Distribution (PPED) [3]. 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 general-purpose vector generation is shown in Fig. 3. In Cell Edge Distribution (CED) method, each feature map is divided into 4x4 cells and each cell has 16x16 pixel sites. The edge flags in the cell are counted and each element in the feature vector indicates the number of edge flags within the corresponding cell.

In order to perform high-accuracy detection of a certain object, it is helpful to develop a vector representation specific to the target. Among the four-direction feature maps of a human face, as shown in Fig. 1, the horizontal edge map retains the most important features of a human face. A lot of edge flags are detected around the areas of eyes and a mouth. Fig. 4 illustrates the face specific vector generation scheme based on this characteristic. At first, two 16-pixel-height rows which are assumed to be the location of eyes and a mouth if the 64x64 pixel window fist a human face are cut from the horizontal feature map. Then both eyes and mouth bands are divided into two-pixel-width columns and the number of horizontal edge flags within each column is counted to yield a single vector element. A face specific 64-dimension feature vector, which we call Eyes and Mouth Extraction (EME), is thus generated.

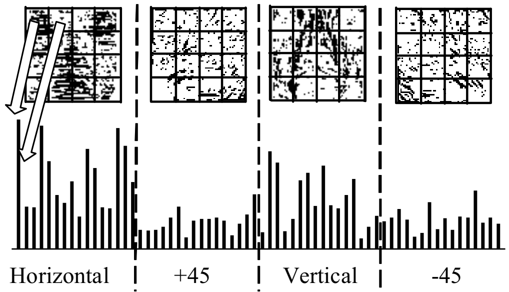


Fig. 3. Feature vector based on Cell Edge Distribution (CED).

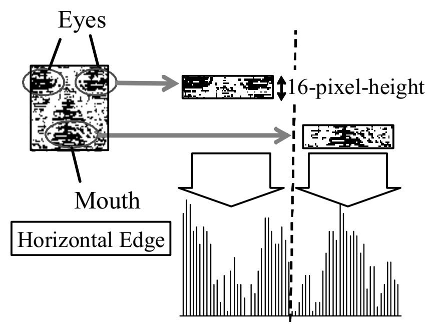


Fig. 4. Feature vector based on Eyes and Mouth Extraction (EME).

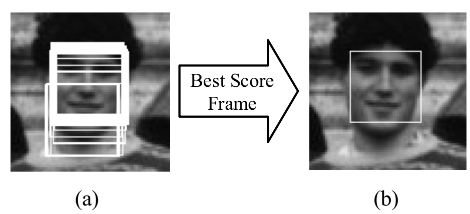


Fig. 5. Face localization.

2.3. Template matching and localization

A 64x64-pixel area was taken from the target image as an input to the system and a feature vector is generated. Two types of feature vectors of both face and non-face images are stored as templates in the system in advance. The feature vector of the input image is compared with all the template vectors. The template vector yielding the minimum distance (the Manhattan distance is utilized in this work) is detected as the maximum-likelihood case to the input, and then the input vector is classified as face or non-face depending on the matching result.

Due to the robust nature of the edge-based histogram representations, multiple pixel sites are detected as candidates of a face when the window comes closer to a face location as show in Fig. 5(a). In order to localize the window at the best location enclosing the detected face, the flowing algorithm is utilized. Assuming that two faces does not overlap in the 64x64-pixel window, only one candidate should remain as the detected face within the window. Therefore, the matching distance is employed as a measure for decision. At each point where a face is detected, the matching distance is compared with those of the

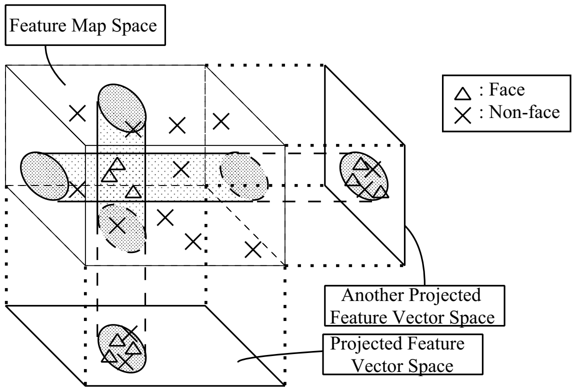


Fig. 6. Concept of multiple-clue perception.

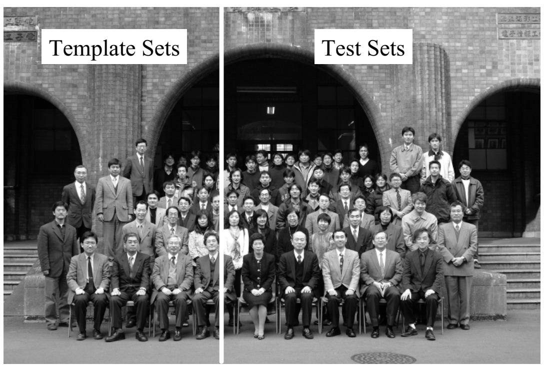


Fig. 7. Input picture utilized as template sets (left) and test sets (right).

other already detected points within the current 64x64-pixel window. If the current distance is smaller than the others, then all other points are eliminated. If the current distance is larger, the point is classified as a non-face site. In this manner, multiple detection is reduced to a single window as show in Fig. 5(b).

2.4. Multiple-clue perception

The process of the feature vector generation is considered as dimensionality reduction of the feature space for data compression. In other word, the feature vectors are generated by projecting the data in the feature map space to a lower-dimension space as illustrated in Fig. 6. In this process, however, some important information that separates faces from non-faces is lost. As a result, a number of false-positives are inevitably detected. Therefore, we have developed a multiple-clue perception scheme to enhance the accuracy of face detection as shown in Fig. 6. In this scheme, two or more vector generation algorithms as discussed in Section 2.2 are employed and face detection is carried out by combining all the face detection results in each vector space. Employing other feature vector generation algorithms corresponds to projecting the data in the feature map space to other feature vector spaces. False-positives detected in a certain feature vector space are unlikely to appear in another feature vector space since different data reduction approaches are applied to each projection algorithm as illustrated in Fig. 6.

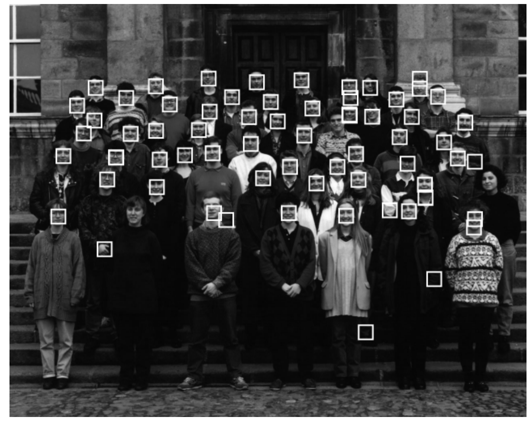
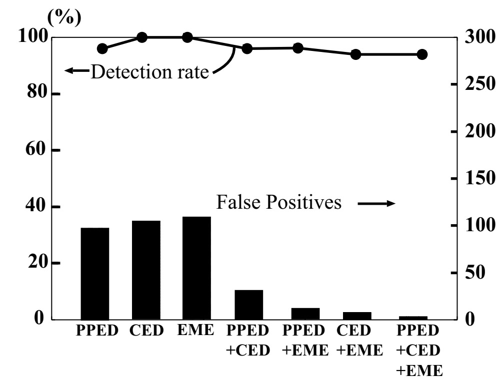
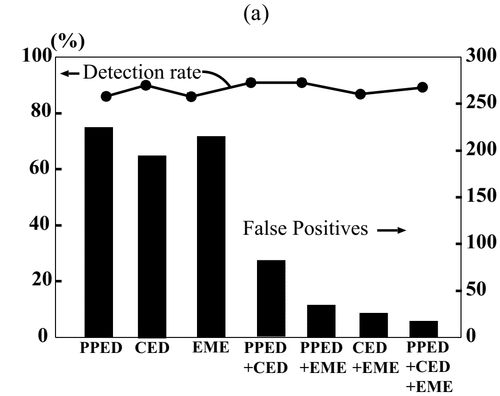


Fig. 8. Output obtained from multiple-clue perception.



(a)

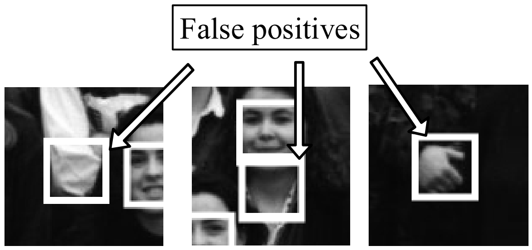


(b)

Fig. 9. Face detection rate and false positives: (a) results using right-hand side of picture in Fig. 7; (b) results using picture in Ref. [2] (Fig. 8).

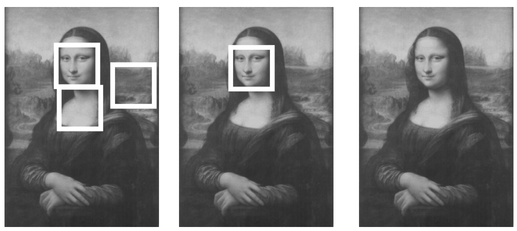
3. EXPERIMENTAL RESULTS AND DISCUSSIONS

Template image of faces and non-faces were selected from the picture shown in Fig. 7. Twenty human faces on the left-hand side of the picture were utilized as face templates, and 2000 non-face templates were chosen randomly from the background scenery. The face detection was carried out using the right-hand



(a) (b) (c)

Fig. 10. Examples of false positives.



PPED (a) PPED + EME (b) PPED + EME + CED (c)

Fig. 11. Face detection using realistic face drawing. The PPED + EME + CED fails to detect faces (c).

side of the picture in Fig. 7 and the picture of a different human race utilized in Ref. [2] (Fig. 8).

Fig. 9 demonstrates the results of face detection. If the face templates are generated from the same picture of the test faces, almost 100% of human faces are detected correctly and about 100 false positives are detected by using only a single clue as shown in Fig. 9(a). As the number of clues increases, the number of false positives is decreased drastically while preserving a high face detection rate (~93%). Even with test images taken under the different condition (e.g. brightness, contrast, human race, and so on), 91% of faces (52 faces out of 57) are detected and the number of false positives is suppressed to 16 by the multiple-clue method as shown in Fig. 9(b). In the experiment, even when the number of face templates was increased, no improvement in the face detection performance was observed. Therefore, it is concluded that 20 templates are sufficient for the multiple-clue algorithm.

The locations which are identified as faces are illustrated in Fig. 8, and some false positives are enlarged in Fig. 10. As shown in Fig. 10, false positives are detected at the location of a crease of a shirt or at the location where the distribution of edges is coincident with that of a human face. However, it is expected that these errors can be eliminated by a simple confirmation process.

Fig. 11 demonstrates the face detection result using an realistic face drawing. Although false positives are detected in the PPED vector representation, they are successfully eliminated in the multiple-clue employing PPED and EME vector representations. However, in the case of PPED + CED + EME, no face is detected. This is because the location of the detected face in CED method is slightly different from those of the others. This issue can be solved by merging the detected locations.

The operational time for vector generation and face detection in a VGA image is estimated to be within 0.1 sec by using the next version of the hardware developed in Ref. [3], which is feasible to real-time applications.

4. CONCLUSION

A hardware-friendly multiple-clue image perception algorithm has been developed for robust image representation, and successfully applied to human face detection. Two edge-based feature vector generation schemes have been proposed for this purpose in addition to PPED feature representation. As a result, robust face detection has been achieved with only a small amount of face templates, while reducing the number of false positives drastically.

ACKNOWLEDGMENT

The facial data utilized in Fig. 1 is used by permission of Softopia Japan, Research and Development Division, HOIP Laboratory. It is strictly prohibited to copy, use, or distribute the facial data without permission.

REFERENCES

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

[2] 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.

[3] 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. 1141-1161, Sep. 2003.

[4] M. Yagi, H. Yamasaki, and T. Shibata, "A Mixed-Signal VLSI for Real-Time Generation of Edge-Based Image Vectors," to be published in Advances in Neural Information Processing Systems 16, Dec. 2003.

[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-Matching Classifier Using Floating-Gate MOS Technology," Advances in Neural Information Processing Systems 14, Vol. II, pp. 1131-1138, Dec. 2001.