A Transformation-Based Mechanism for Face Recognition

Yea-Shuan Huang and Yao-Hong Tsai

Advanced Technology Center
Computer & Communications Research Laboratories
Industrial Technology Research Institute, Chutung, Hsinchu, Taiwan
yeashuan@itri.org.tw

Abstract. This paper proposes a novel mechanism to seamlessly integrate face detection and face recognition. After extracting a human face x from an input image, not only x but also its various kinds of transformations are performed recognition. The final decision is then derived from aggregating the accumulated recognition results of each transformed pattern. From experiments, the proposed method has shown a significantly improved recognition performance compared with the traditional method on recognizing human faces.

1 Introduction

Due to the rapid advance of computer hardware and the continuous progress of computer software, we are looking forward to developing more powerful and friendly computer use models so that computers can serve people in a more active and intelligent way. The concept of "*computer will be more human*" is not required just in the scientific fiction but also is in our daily life. To this end, the computer basically needs to have a surveillance ability, which enables it to detect, track, and recognize its surrounding people so that the computer can offer various kinds of user oriented services automatically. This results in the situation that researches on face processing (including detection [1-2], tracking [3], and recognition [4-7]) are very prosperous in the last two decades.

Many promising algorithms have been proposed to deal with the basic face processing problems, such as (1) how to detect the human faces from an input image? and (2) how to recognize the people identity based on a detected face? High accuracy has already been individually reported for face detection (FD) and face recognition (FR). However, when integrating FD and FR together (as shown in Figure 1), it often results in a considerably degraded performance. For example, both of the chosen algorithms of FD and FR may have over a 90% individual correction rate, but the integrated system actually just has only 60% in correctly recognizing a face from an input image. This phenomenon mainly comes from three factors: (1) the criterion of deciding successful face detection is too rough for face recognition so that many detected face

T. Caelli et al. (Eds.): SSPR&SPR 2002, LNCS 2396, pp. 636-643, 2002.

images considered to be correct in the FD stage are not good enough in the FR stage due to either partially lost important face information or contain extra non-face image; (2) the chosen FR has little generalization ability so that it is easy to obtain a wrong recognition when a face image is not perfectly detected; and (3) the training samples of FR are extracted manually so that the automatically detected face images from a FD having different properties from the manually extracted ones are prone to be mis-recognized. In this paper, a novel mechanism has been devised which can seamlessly integrate FD and FR and improve the accuracy of the whole integrated system. The basic concept of this approach is that not only the detected face image x but also its various transformations are performed the face recognition operation, and the final decision is derived from the accumulated recognition results such as the decision is the class with the highest similarity among all individual recognition or the class with the highest average similarity.

This paper consists of four sections. Section 2 describes the newly proposed face recognition integration mechanism which contains four main processing steps: face detection, image transformation, image matching, and result accumulation. Section 3 then performs experiments with and without multiple transformation and investigates their performance difference on the ITRI (Industrial Technology and Research Institute) face databases. Finally, Section 4 draws our conclusions and point out the future research directions.

2 The Proposed Integration Mechanism

Figure 2 shows the newly proposed transformation-based mechanism to integrate a face recognition system. The main concept of this mechanism is that not only an extracted face image x but also its various transformations are performed recognition, and the final decision is derived from the accumulated recognition results. In this section, face detection and face matching is discussed first, then a recognition by accumulating multiple transformation is formally introduced.

2.1 Face Detection

In a face recognition system, it is essential that a face image is extracted from a processed image by containing only the portion of face image which is useful for recognition and excluding the image portions (such as background, clothes, and hair) invalid to recognition. Chen et al. [8] have shown that it is incorrect to recognize a person by using the image including face, hair, shoulder, and background as used in [9] because the trained statistical classifiers learns too much irrelevant information in identifying a person. Instead, they proposed to extract the face-only image for training and testing a face recognition system. A face-only image as shown in Figure 3 ideally is the minimal rectangle containing eyebrows. eyes, nose, and mouth of a face. Here, we adopted Han's method [10] to extract the face-only image which consists of three main steps. In the first step, a morphology-based technique is devised to perform eye-analogue segmentation. Morphological operations are applied to locate eye-analogue pixels in the original image. Then a labeling process is executed to generate the eye-analogue segments. Each eye-analogue segment is considered a candidate of one of the eyes. In the second step, the previously located eye-analogue segments are used to find meaningful eye pairs by using four geometrically matching rules to guide the merging of the eye-analogue elements into pairs. Each meaningful eye pair is further directed to specify a corresponding face region. The last step is to verify each specified face region to be a face or an non-face by a neural network. This method performs rather fast and accurate when dealing with uniformly well lit face images.

2.2 Face Matching

A three-layer feed-forward network with a Generalized Probabilistic Descent (GPD) learning rule is served as the face classifier. GPD is originally proposed by Juang [11] to train a speech classifier, which is reported to have a much better recognition performance than the well-known Back-Propagation (BP) training. However, to our best knowledge, GPD is rarely or even never used in the computer-vision community. Because GPD is based on minimizing a classification related error function, it theoretically can produce a better classification performance than the classifiers (such as BP) based on minimizing a least-mean-square error. Because the space limit, this GPD face learning approach is not introduced in this article, interested readers can find detail information in [12].

2.3 Recognition by Accumulating Multiple Transformations

Assume K subjects of people in the concerned pattern domain, $C_1,\cdots,C_K$, x denotes an inputted face-only image, and $S_k(x)$ is the possibility that *x* belongs to subject *k*. Let $F_n(x)$ be the feature of the *n*-th transformed image of *x*, where $0 \le n \le N$ and N is the total number of transformations, and $F_0(x)$ be the feature of the original *x*. Traditionally, x is recognized to subject *k* if $S_j(x)$ has the largest value among all $S_k(x)$ for $1 \le k \le K$. That is

$$D(x) = j, \qquad \text{if } S_j(x) = \arg\max_{1\le k\le K} S_k(x).$$

The proposed decision by accumulating multiple transformation is

$$\begin{aligned} D(x) = j, \quad \text{if } & G_j(S_j(F_0(x)),\cdots,S_j(F_N(x))) \\ & = \arg\max_{1\le k\le K} G_k(S_k(F_0(x)),\cdots,S_k(F_N(x))) \end{aligned}$$

where $G_k(\cdots)$ is an aggregation function which specifies the appropriate way to derive the accumulated score that *x* belongs to subject *k*. Two possible selections of G_k have been proposed, they are

$$G_k(S_k(F_0(x)),\cdots,S_k(F_N(x))) = \max_{0\le n\le N} S_k(F_n(x)) \tag{1}$$

and,

$$G_k(S_k(F_0(x)),\cdots,S_k(F_N(x))) = \frac{1}{N}\sum_{n=0}^{N} S_k(F_n(x)) \tag{2}$$

Of course, there are many other possible choices of G_k. As a matter of fact, this concept can be applied to deduce a lot of variational recognition processes, such as

(1) not only the best detected object but also many other possible detected objects are performed recognition and the final decision is derived from the proposed G_k, and
(2) not only many possibly detected object but also their various transformations are applied recognition, and the final decision is derived from these tentative recognitions.

In reality, there exists various kinds of image transformations such as image rotation, affine transform, boundary shifting, lighting compensation and so on. In this paper, two kinds of commonly-used transformations (image rotation and boundary shifting) are described as follows: Let *I* be an image with *M* horizontal and *N* vertical pixels, *I(m,n)* be one image pixel locating at the *m*-th horizontal and the *n*-th vertical pixels ($1 \le m \le M$ and $1 \le n \le N$). Let *T* denote a segmented target image, *T(x,y)* be one pixel of *T*, and T' be a transform of *T*. The rotation transform is

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$

where θ is the rotation angle. One more generalized transformation form is the affine transform and its matrix notation is

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

which covers translation, scaling, rotation and slant.

Another possible transformation is boundary shifting which shifts any of *T'*s four boundaries (i.e. top, bottom, left and right) in and out. Suppose *T(x,y)* be one pixel of *T* where $x_{min} \le x \le x_{max}$ and $y_{min} \le y \le y_{max}$, if T' is achieved by shifting the *T'*s left boundary *p* pixels out (i.e. the left boundary is shifted *p* pixels more left but keep the top, the bottom and the right boundaries the same), then $T'(x,y) = T(x,y)$ with $x_{min} - p \le x \le x_{max}$ and $y_{min} \le y \le y_{max}$. T' could also be achieved by shifting both *T*'s left boundary *p* pixels out and *T*'s bottom boundary *q* pixels in, then $T'(x,y) = T(x,y)$ with $x_{min} - p \le x \le x_{max}$ and $y_{min} \le y \le y_{max} - q$.

3 Experiment Results

To investigate the performance of this method, two sets of face databases were constructed and face recognition experiments were performed by training and testing a face classifier (GPD with 625-100-26 nodes) with and without adopting this proposed method respectively. The first face dataset was taken by asking 26 persons to slightly rotate their faces or change their face expressions when they stood about 1.5 meters away from the camera. The second face dataset was taken by allowing the same 26 persons to approach from 6 meters to 1 meter away from the camera. Figure 4 shows some examples of the two face datasets which reveals that (1) the face images are about the same size in the first dataset, but they have quite different sizes (more than 10 times) in the second dataset, and (2) the second dataset consists of much image difference in brightness gain and intensity uniformity.

In this experiment, face-only images are manually selected from the first face database, which are then used to train the GPD face classifier; but face-only images are automatically extracted from the second face database, which are further used to test the face recognition performance. A threshold T is defined to be $0.125 \times D$, where D is the distance between the centers of two eyes. If any location distance of the eye on the automatically extracted face-only image to its corresponding manually selected eye is larger than T, then it is counted as an invalid face detection; otherwise, it is counted as a valid one. Since a face-only image is extracted directly based on its right and left eyes, shifting the eye's locations can correspondingly generate transformed images having the rotation and boundary shifting effects to the original one. In this experiment, the locations of each right and left eyes are individually shift S pixels (here S is $0.06 \times D$) left and right. This results in totally 81 $(3 \times 3 \times 3 \times 3)$ transformed images from each detected face-only image.

After the training procedure, the test images are inputted to the trained recognizer, and the recognition decision is made respectively by using the tradition decision rule and the proposed decision rule with the first aggregating rule

$$G_k(S_k(F_0(x)), \cdots, S_k(F_N(x))) = \max_{0 \le n \le N} S_k(F_n(x)).$$

In order to analyze the effectiveness of the proposed method, the face recognition rate is computed by two different situations: recognition of validly detected faces, and recognition of invalidly detected faces. Table 1 shows respectively the recognition performance of the tradition decision rule and the proposed transformation-based decision accumulation rule. Obviously, the proposed method performs much better than the traditional one by improving the recognition accuracy from 65% to 85% in the first (valid) situation, and from 40% to 60% in the second (invalid) situation. Since the two recognition indexes are based on the same face detection result, it is clear that the better performance of the proposed method comes from the better generalization ability of the proposed method. This robustness enables the integration system to have the ability in producing the correct recognition decision even when the detected face image is not complete or good enough to the adopted face classifier.

4 Conclusions

This paper proposes a novel mechanism to seamlessly integrate face detection and face recognition. After extracting a human face x from an input image, not only x but also its various kinds of transformations are performed recognition. The final decision is then derived from aggregating the accumulated recognition results of each trans-

formed pattern. From experiments, the proposed method has shown a much better performance compared with the traditional face recognition system.
One drawback of this method is that it takes a much longer period of FD+FR (0.85 seconds/image) than the traditional one (0.2 second/image). Many speed-up methods can be easily applied to reduce the processing time, such as (1) to reduce the number of transformations, and (2) to reduce the scale in each transformation. However, it is more interesting if an appropriate way can be designed to guide the suitable transformation types and transformation scales automatically based on the extracted face information. We are currently working on this direction.

References

1. K.K. sung and T. Poggio, "*Example-Based Learning for view-Based Human Face Detection*," IEEE Trans. Patt. Anal. Machine Intell., Vol. 20, pp. 39-51, 1998.
2. H. A. Rowley, S. Baluja, and T. Kanade, "Neural network-based face detection," *IEEE transactions on PAMI.*, vol. 20, no. 1, pp. 22-38, Jan. 1998.
3. D. M. Gavrila , "The visual analysis of human movement: a survey," Computer Vision and Image Understanding, vol. 73, pp. 82-98, 1999.
4. M. Turk and A. Pentland, "*Eihenfaces for Recognition*", Journal of Cognitive Neuroscience, March, 1991.
5. R. Brunelli and T. Poggio, "*Face Recognition: Features Versus Templates*", IEEE Trans. Patt. Anal. Machine Intell. Vol. 15, No. 10, October, pp 1042-1052, 1993.
6. R. Chellappa, C. Wilson and S. Sirohey, "*Human and Machine Recognition of Faces: A Survey*", Proc. Of IEEE, Vol. 83, No. 5, May, pp 705-740, 1995.
7. A.K. Jain, R. Bolle and S. Pankanti, *Biometrics: Personal Identification in Networked Society*, Kluwer Academic Publishers, 1999.
8. L.-F. Chen, C.-C. Han, and J.-C. Lin, "Why Recognition in a Statistics-Based Face Recognition System Should be Based on the Pure Face Portion: a Probabilistic Decision-Based Proof," to appear in *Pattern Recognition*, 2001.
9. F. Goudail, E. Lange, T. Iwamoto, K. Kyuma, and N. Otsu, "Face Recognition System Using Local Autocorrelations and Multiscale Integrations," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 18, pp. 1024-1028, Oct., 1996.
10. C.C. Han, H.Y. Mark, K.C. Yu, and L.H. Chen, "Fast Face Detection via Morphology-Based Pre-Processing," Pattern Recognition 33, pp. 1701-1712, 2000.
11. B.H. Juang and S. Katagiri, "*Discriminative Learning for Minimum Error Classification*", IEEE Trans. On Signal Processing, Vol. 40, No. 12, December, pp 3043-3054, 1992.
12. Yea-Shuan Huang, Yao-Hong Tsai, Jun-Wei Shieh, "Robust Face Recognition with Light Compensation," to appear in The Second IEEE Pacific-Rim Conference on Multimedia, 2001.

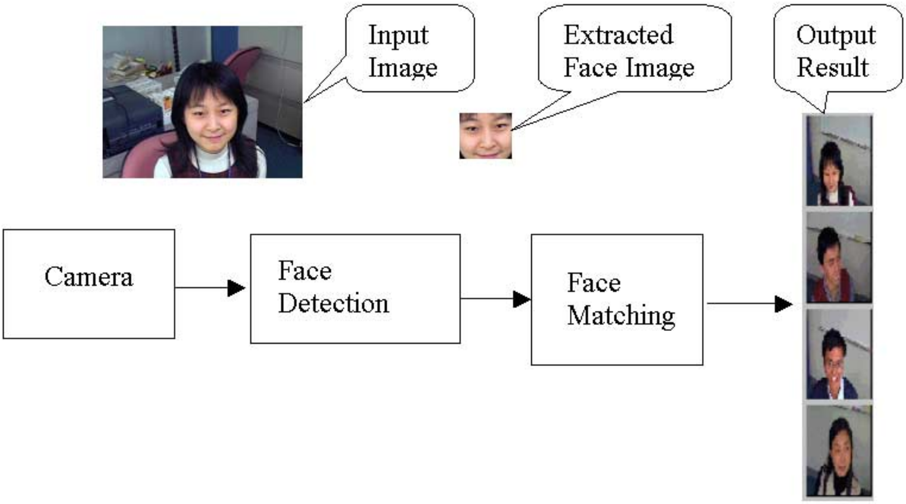

Fig. 1. A traditional face recognition system

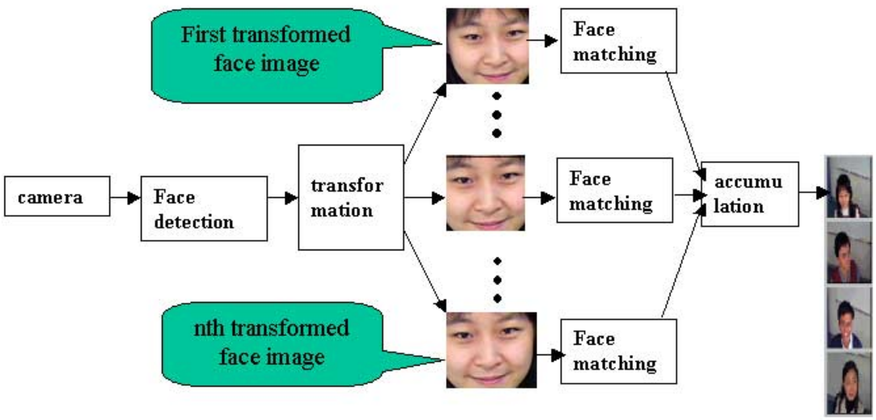

Fig. 2. The proposed transformation-based face recognition system

Fig. 3. One example of the approximate face-only image delimitated by a black-line rectangle which is extracted based on the two eyes marked in white circles

Fig. 4. Examples of two face datasets. The first row shows images of the first dataset images which were taken at a constant distance between camera and the subject, and the second row displays the images of the second dataset which were taken during a subject approached close to the camera

Table 1. This table shows (1) the performance of face detection, and (2) the recognition performance by the tradition decision rule and the corresponding recognition rate by the proposed decision rule

	Valid Faces	Invalid Faces
Sample Number	2456	345
Traditional FR	65%	40%
Proposed FR	85%	60%