MVA '92 IAPR Workshop on Machine Vision Applications Dec. 7-9,1992, Tokyo

TARGET IMAGE EXTRACTION FOR FACE RECOGNITION USING THE SUB-SPACE CLASSIFICATION METHOD

Shigeru Akamatsu * , Tsutomu Sasaki ** ,
Hideo Fukamachi *** , and Yasuhito Suenaga **

* ATR Human Information Processing Research Laboratories
2 - 2 Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02, Japan
** NTT Human Interface Laboratories
1 - 2356 Take, Yokosuka-shi, Kanagawa 238-03, Japan
*** NTT Software Corp.
1 - 2356 Take, Yokosuka-shi, Kanagawa 238-03, Japan

ABSTRACT

This paper proposes a scheme that offers robust extraction of the target face image in standard view, which is defined using internal facial features as steady reference points. The scheme is characterized by two steps: facial feature detection using color image segmentation, and target image selection from among the candidates using the sub-space classification method. The scheme's flexibility has been confirmed in experiments under a wide range of image acquisition conditions.

INTRODUCTION

Computer recognition of the human face is expected to improve future human-computer interaction. Most face recognition tasks such as personal identification and facial expression understanding require automatic and robust extraction of the target face in standard view from input images. Aimed at automatic detection and identification of human faces, some systems were reported to show fair recognition of a very limited number of subjects under restricted conditions [1][2]. However, they were not flexible enough to achieve accurate and robust identification under more practical conditions. One reason is that the process of standardizing the target images was inadequate for human faces and therefore, the features used for identification are strongly subject to distortion by imaging conditions.

One promising approach to defining consistent target images of faces under various imaging conditions is the use of internal facial features as the steady reference points. We proposed to extract the target images to be matched through affine transformation defined by the two eyes and the mouth [3]. As shown in Fig.1, the original image is first translated, scaled and rotated so that the three reference points satisfy a standard spatial arrangement. Then a standard window enclosing these reference points is superimposed on the affine transformed face image specified by the five empirically determined parameters shown in the figure. The standard window is sampled, for example at 128 by 128 pixels, to form a target image. The target images obtained by this method are rather robust to changes in location, size and orientation of the face and less sensitive to variations of hair style and background. The drawback to this approach, however, is that reliable detection of the exact set of the facial features as the reference points is not easily achieved by image processing alone. There is usually a danger of missing the correct facial features when we attempt to prevent false extraction.

Therefore, a joint solution to defining the target image is proposed by introducing a classification framework: first, all facial features with the potential for use as reference points are extracted by an image segmentation algorithm to yield possible target image candidates; then, the target image is selected from the candidates with regard to overall resemblance to the ensemble of correct target images using a statistical method of pattern classification.

The proposed scheme was used as a preprocessing module for image standardization in a face recognition system, by which the assessment was made based on the accuracy of identification [4]. In this paper, however, there will be a more detailed discussion of the algorithm and assessment of the target image extraction procedure.

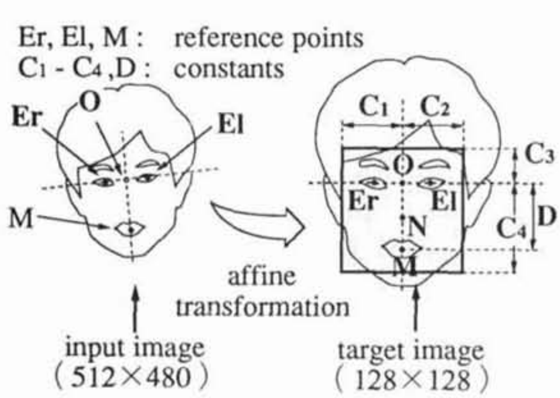


Fig.1 Target image extraction from front view face

EXTRACTION OF POSSIBLE FACIAL FEATURES

It is empirically known that, for the average Japanese, the facial features in question show the following color characteristics in face images represented in the HSV and YIQ color coordinate systems:

- the lips have higher Q-component values than the surrounding area.
- the skin area exhibits clear peaks in the histograms of both the I and H components.

Accordingly, we have developed the following algorithm to extract all possible regions that correspond to facial features [5]: the image is first transformed from the RGB color coordinate system to the HSV and the YIQ systems and then thresholding, hole filling and image subtraction operations are applied to the specified components and the resultant segmented regions are combined to yield regions that potentially represent the facial features.

A preliminary experiment carried out on more than 100 face images showed nearly 100% success in placing the true three facial features within the list of candidates for the reference points. Simultaneously, however, the facial feature extraction algorithm usually yields several false regions that do not correspond to the correct reference points. The number of candidates for the reference point might be reduced by using local information on color, size, and relative position of the regions. It is difficult, however, to determine the correct set for the reference points by such ad hoc criteria alone because there often remain some regions that closely resemble each other, for example an eye and an eyebrow.

SELECTION FROM TARGET IMAGE CANDIDATES

Because of the false reference points extracted by color image segmentation, we usually have several target image candidates for a single input image. Fig.2 shows an example of the set of target image candidates. The problem here is to select the one that nearly fits the face view in the standard window described in Fig.1.

Our solution to this problem is to apply a classification framework so that the selection is made by measuring the degree of fitness of an image to the ensemble of correctly standardized face images. Intensity images of size N x N pixels can be represented by points scattered in a huge vector space of N x N dimensions. Target image candidates obtained from the face images of many individuals are not spread randomly over this space, but converge to several clusters characterized by the facial compositions peculiar to the set of facial features frequently used as reference points. Each cluster representing its own type of standardization among target image candidates spans a lower dimensional sub-space. Thus, target image selection is performed in the form of sub-space classification by examining the distance between the point corresponding to the target image candidate and each sub-space.

Fig.2 Target image candidates obtained from a single input image

Assume that K target image standardization classes are defined by observing target image candidates obtained from face images of a large number of individuals. A mathematical description of the sub-space for standardization class k is obtained by applying the Karhunen-Loeve (K-L) expansion to the training samples of the target images assigned to the class k, and is given by mean vector μ_k and eigenvectors ψ_{kl} ($l = 1,2,...L$) which are associated with the L largest eigenvalues for the sample covariance matrix R of the training set. It is beyond our computational power to determine the eigenvectors and eigenvalues of R for a typical image size of N=128, because the matrix R would be N^2 by N^2 in size. As a practical matter, however, the number of samples in the training set for class k, M_k, is less than the dimension of matrix R, which makes the matrix singular. In that case, the matrix R has, at most, M_k non-zero eigenvalues and the corresponding eigenvectors can be calculated through a smaller M_k by M_k matrix problem by applying singular value decomposition technique [6]. Since μ_k and ψ_{kl}'s are the vectors of the same dimensions as the training samples and are therefore representative of some images, they are called mean face and eigenfaces [7].

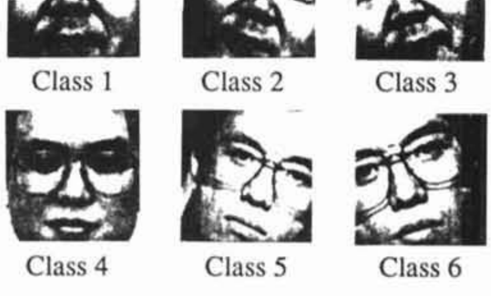


Fig.3 Examples of target image candidates for the six classes

Target image selection for the input image is carried out as follows. Suppose that there are I target image candidates obtained from the input image. Let $\mathbf{X}_i$ denote a N x N dimensional vector representing the intensity image of the i-th target image candidate. For each $\mathbf{X}_i$, the sub-space classification method [6] is applied using the distance d_{ik} defined between the data point and the sub-space of class k.

$$d_{ik}^2 = \| X_i - \mu_k \|^2 - \sum_{l=1}^{L} | (X_i - \mu_k)^t \psi_{kl} |^2$$

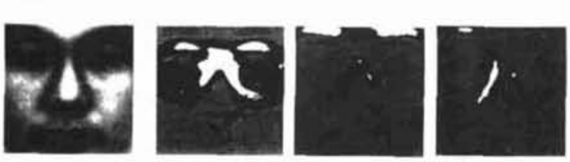
And after executing K-class discrimination by the minimum distance decision rule for all the I target image candidates, the simple algorithm is applied to select the correct target image. Assume that class one is assigned to the cluster for correct standardization. The algorithm is: if any one of the target image candidates, say X_j, was classified as class one, then select that candidate as the target image; otherwise reject all the given candidates with no target image selected.

EXPERIMENTS

We examined the color segmentation algorithm using face images of 241 individuals to obtain tentative sets of reference points for the standardization and obtained a total of 1,462 target image candidates in various standardization conditions (a set of intensity images of 128 by 128 pixels). Through observation of this set, six major classes of facial composition were defined. The examples of the target image candidates for the six classes are shown in Fig.3. Class 1 corresponds to the standardization that we judge correct and the other classes correspond to irregular standardization, for example, patterns leaning to the right are assigned to Class 2.

By setting L to 40, the same number of eigenfaces describing the sub-space of each class were obtained from the training set of 241 individuals. The mean face and the top three eigenfaces representing the sub-space for Class 1 are shown in Fig.4. The eigenfaces shown in the figure are those quantized into four intensity levels.

As shown in Table 1, four sets of face images were prepared as test samples with the combination of the following conditions; "familiar face" vs. "unfamiliar face" depending on whether the face is of a person included in the training set of 241 individuals, and "front-view" vs. "side-on" for face imaging conditions. Classification regarding the imaging condition was made subjectively after sampling frame images from a motion picture sequence taken while the subject was shaking or nodding his/her head.



mean face　　the top three eigenfaces with the largest eigenvalues

Fig.4 A description of the sub-space for Class 1

Table 1 Four types of test samples

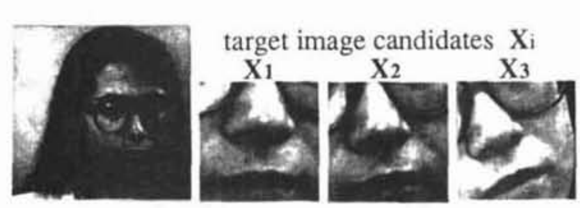
Set	Familiarity of person	Imaging view	# of samples	Notes
A	familiar	front-view	165	6 males, 2 females 4 wearing glasses
B	familiar	side-on	126	2 males, 1 female 2 wearing glasses
C	unfamiliar	front-view	77	4 males,17 females no glasses
D	unfamiliar	side-on	41	1 female no glasses

The target image extraction performance achieved on the test samples is shown in Table 2. Whether the extracted target image is correct or not grounded on the standard face view defined in Fig.1, was given by subjective observations.

For the input images that roughly match the front view, a nearly 100% correct extraction rate was achieved not only for familiar faces but also for unfamiliar faces. For the

Table 2 Evaluation of the scheme

Set	Number of samples	Correctly extracted	Falsely extracted	Rejected
A	165	165 (100 %)	0 (0 %)	0 (0 %)
B	126	109 (86.5 %)	2 (1.6 %)	15 (11.9 %)
C	77	76 (98.7 %)	0 (0 %)	1 (1.3 %)
D	41	38 (92.7 %)	0 (0 %)	3 (7.3 %)



input image

	X1	X2	X3
k=1	4641	4660	4699
k=2	4804	4897	4478
k=3	4728	4491	***4362***
k=4	4679	4730	4621
k=5	4759	4961	4519
k=6	***4584***	***4488***	4421

distance d_{ik} from the sub-space

Fig.5 A successful example of the target image selection

side-on facial images, the success rate of the target image extraction was observed to fall. However, the worst case, extracting false target images, rarely occurred, because most of the input images for which the correct target image was not extracted were properly rejected with no target images extracted. In Fig.5, a successful example of the procedure in avoiding false target image extraction for an input image from set B is shown. In this case, extraction of the correct reference points for standardization accidentally failed in the facial features detection stage so that there was no appropriate target image among the candidates. However, all the candidates were successfully rejected by the proposed algorithm, because none of them were classified as class one. These cases of rejection could be treated properly in the total operation of the target image extraction system by repeating the facial feature extraction process with different thresholding parameters.

Fig.6 shows some examples of input face images from which the correct target images were successfully extracted by the proposed scheme. The three plus marks are the reference points finally positioned after the target image selection procedure. Shown at the upper left of each image is the target image extracted from the original input color image. The scheme worked well when the subject assumed various poses, when the subject wore glasses and when the background included various objects.

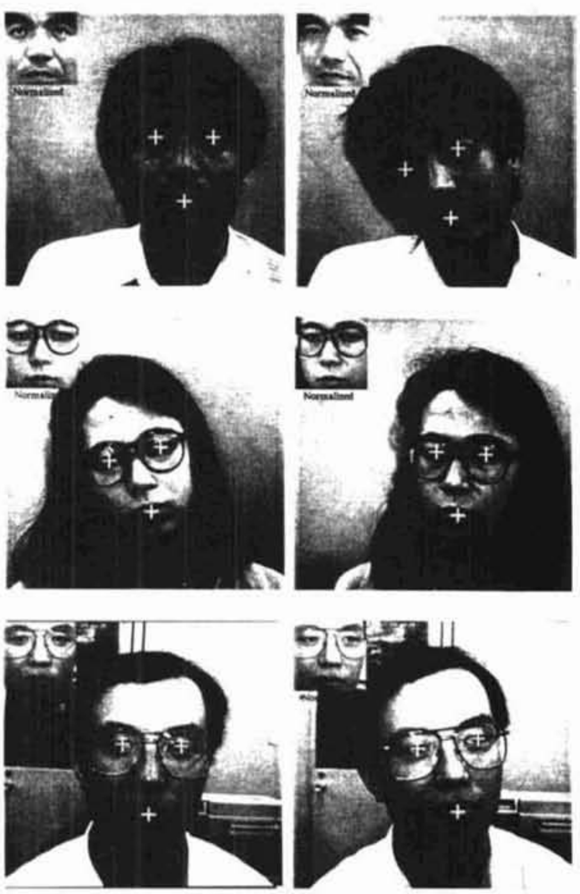


Fig.6 Examples of successfully extracted target images

CONCLUSION

We have proposed a scheme for automatic extraction of the target image in standard view from input facial images. To achieve higher robustness, we combined a locally sensitive operation of facial feature extraction with global inspection of the target image in the framework of classification by the sub-space method. Experiments indicated that consistent target images can be extracted from unfamiliar faces under a wide range of imaging conditions. Although real-time operation is not yet feasible, we believe this approach will be applicable to the preprocessing stage for most automatic face recognition tasks such as personal identification, lip reading and facial expression recognition.

ACKNOWLEDGEMENTS

This paper is based on experiments carried out while the first author was affiliated with NTT Human Interface Laboratories. The authors thank Drs. Takahiko Kamae, and Takaya Endou for their encouragement and advice. Discussions with colleagues and the assistance of Keiko Isobe in arranging the experimental data are gratefully acknowledged. Encouragement and support by Drs. Kouhei Habara and Yoh'ichi Tohkura, ATR Human Information Processing Research Laboratories, are also gratefully acknowledged.

REFERENCES

[1] P.Burt, "Smart sensing within a Pyramid Vision Machine," Proc.IEEE, Vol.76, No.8, pp.1006-1015, Aug. 1988

[2] M.A.Turk and A.P.Pentland, "Face Recognition Using Eigenface," Proc.CVPR, Maui, June 1991

[3] S.Akamatsu, T.Sasaki, H.Fukamachi and Y.Suenaga, "Robust face identification scheme: K-L expansion of an invariant feature space," SPIE Intelligent Robots and Computer Vision, 1607-007, Nov. 1991

[4] S.Akamatsu, T.Sasaki, H.Fukamachi, N.Masui and Y.Suenaga, "An Accurate and Robust Face Identification Scheme," Proc.ICPR'92, Hague, Aug. 1992

[5] T.Sasaki, S.Akamatsu, H.Fukamachi and Y.Suenaga, "Color segmentation method for image registration in automatic face recognition," OSA Advances in Color Vision, Jan. 1992

[6] Erkki Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press, England, 1983

[7] L.Sirovich and M.Kirby, "Low-dimensional procedure for the characterization of human faces," J.Opt.Soc.Am.A, Vol.4, No.3, pp.519-524, Mar. 1987