2007−CVIM−160 (23)
2007／9／4

濃度こう配特徴を用いた全自動正面顔認識

大山 航， 若林 哲史， 木村 文隆
三重大学大学院工学研究科

特徴量に画像の濃度こう配を用いた正面顔自動認識手法を提案する．提案手法は主に以下の３つの処理からなる．(１) 画像探索に基づく顔検出，(２) 検出された顔領域内の顔部品検出に基づく顔外接枠の検出，(３) 顔外接枠内の顔に対する見え方に基づく顔認識．それぞれの処理においては，画像濃度値のかわりに濃度値のこう配を特徴量として用いる．FERET データベースに含まれる正面顔画像を用いた CSU Face Identification Evaluation に従った実験の結果，濃度こう配特徴は，従来の画像特徴よりも低次元の特徴量で高い認識性能を持つことが分かった．

Automatic Frontal Face Recognition Using Gradient Features

Wataru OHYAMA Tetsushi WAKABAYASHI Fumitaka KIMURA
Graduate School of Engineering, Mie University

We propose an automatic frontal face recognition method using gradient features. Our proposed method consists of three main stages: 1) face detection, 2) detection of tight bounding box of face using the result of facial feature extraction and 3) face recognition. In each stage, we use the gradient of image instead of pixel values. Face recognition experiments based on CSU Face identification Evaluation scheme using FERET database suggests that the gradient features has better performance than conventional pixel–based face recognition.

1 Introduction

A fully automatic face recognition or authentication system requires both face detection and recognition[1]. The face detection problem is to segment the input image and isolate the face(s). In particular, it is necessary to determine a tight bounding box around each face that contains just the face (forehead to chin) Of course, the results of the recognition task[2] depend heavily on how well the detection task has been done. For example, when the bounding boxes are not tight, Chen et.al[3] showed that non-face artifacts tend to dominate (and hence corrupt) recognition. Fig. 1 shows face detection results by the SVM–based *holistic* face detector[4]. Holistic approaches take the holistic appearance in the window and binary (two–class) classifiers classify whether the appearance describes a face or not. In the figure, each face is detected by the method, however, the detected face regions are not sufficient for accurate face recognition.

For a human face, there are important features or landmarks that one can exploit for detection purposes. If the position of these facial features is known, then face detection and localization can be done easily and accurately. This approach is called the *feature–based* face detection[5].

In this paper, we propose a novel algorithm which extracts facial features from unoccluded frontal views of human faces in grayscale images. The algorithm utilizes gradient features[6] and distance measures based on a modified quadratic discriminant function (MQDF). Both of these tools have been successfully used for handwritten character recognition. The proposed algorithm also employs a facial shape model to verify and evaluate feature candidates. And we also proposed an appearance–based face recognition using gradient features. In the face recognition.

2 Proposed methodology

2.1 Overview of proposed method

The proposed method consists of three main stages, as shown in Fig. 2. The input is a grayscale image which we assume contains one

Fig. 1 Face detection results by a holistic approach

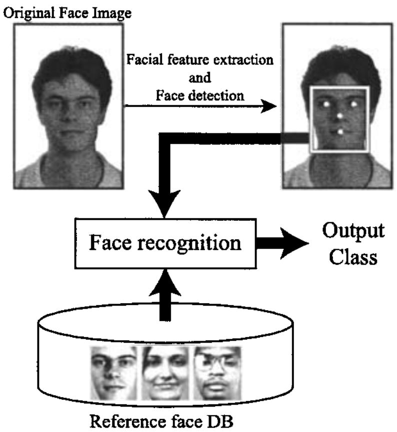

Fig. 2 Main processing stages in the proposed facial feature extraction method.

human face in frontal view and at least 64 × 64 pixels in size. In the first stage, we extract four main facial feature: eyes, nose and mouth, and detect tight bounding box of face using extracted facial features. The second stage is face recognition. In this stage, gradient features are extracted from the tight bounding box detected in the first stage to describe facial appearance. Note that for all matching and distance calculations, we do not use the grayscale pixel data directly; rather, we use gradient features extracted from the pixel data.

2.2 Extraction of gradient feature

Gradient features[6] have been used successfully for handwritten character recognition. In this section, we show how gradient features can be efficiently extracted from grayscale human face images. Given an input image, we first apply a Gaussian smoothing filter with kernel size 5 × 5 and $\sigma = 1.5$. Second, the smoothed grayscale image is normalized so that the mean is 0 and the maximum grayscale value is 1. Finally, the gradient of the normalized grayscale image is obtained by applying the Sobel edge detection operator with a pair of 3 × 3 convolution kernels. The magnitude $|G(x,y)|$ and direction $\theta(x,y)$ of the gradient at point (x,y) of the image is given by

$$|G(x,y)| = \sqrt{Gx^2 + Gy^2}$$
$$\theta(x,y) = \tan^{-1}(\frac{Gy}{Gx}) \quad (1)$$

where Gx and Gy are the horizontal and vertical components of the gradient, respectively.

Fig. 3 shows an example of an input image and the resulting magnitude and direction components of the gradient. These gradient components will form the basis of a feature vector.

After the gradient calculation, a window is placed on the gradient image. The window is partitioned into a grid of $N \times M$ subwindows. Also, we quantize the gradient angle to D different levels. Next, the gradient direction of pixels assigned the same label are added together. So now we have $N \times M \times D$ numbers describing each window.

To further reduce this data, the spatial resolution is reduced from $N \times M$ to $\frac{(N+1)}{2} \times \frac{(M+1)}{2}$ by down–sampling every two horizontal and every two vertical blocks with a Gaussian filter (5 × 5 kernel). Similarly, the directional resolution is reduced from D to $\frac{D}{2}$ by down–sampling with a weight vector $[1\ 4\ 6\ 4\ 1]^T$. After this down-sampling, we obtain a feature vector of size$\frac{(N+1)}{2} \times \frac{(M+1)}{2} \times \frac{D}{2}$ describing each window.

The 5 × 5 Gaussian filter and the weight vector $[1\ 4\ 6\ 4\ 1]^T$ are high cut filters to reduce aliasing from the down–sampling. The filter parameters were determined empirically from trial experiments.

2.3 Facial feature extraction

To obtain tight bounding box for accurate recognition, we employ location of facial features. Each facial feature is extracted using gradient features and MQDF–based matching. Details on this extraction algorithm are described in [7].

As shown in Fig. 4, once the facial features have been located, a tight bounding box can be

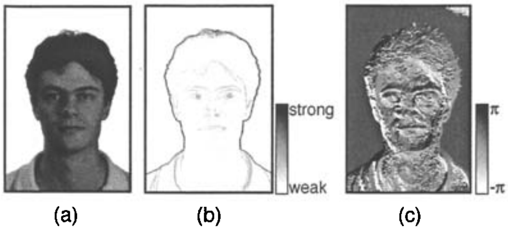

Fig. 3 Extraction of gradient features. (a) smoothed image, (b) magnitude of gradient, (c) direction of gradient

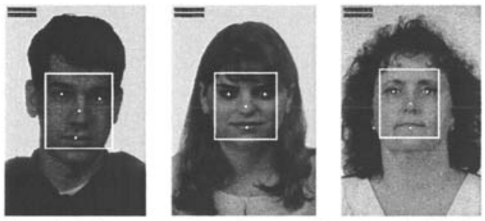
Fig. 4 The examples of correctly extracted facial features.

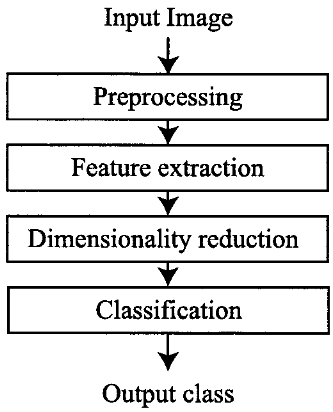

Fig. 5 The 4 stages of the face recognition system


Fig. 6 Example of preprocessed images

found for the face (compare to Fig. 1). The accurate bounding boxes in Fig. 4 can now be used in a recognition system.

2.4 Face recognition

A schematic diagram of the face recognition system to be used here is shown Fig. 5, and consists of four main processing stages: preprocessing, feature vector extraction, dimensionality reduction, and classification.

2.4.1 Preprocessing

For recognition, we are only interested in the face regions from the forehead to the chin and from cheek to cheek. The aim of the preprocessing stage is to normalize the image position, size, and pixel intensity levels so that reasonable distance measures can be made between the input image and each of the stored database images.

The eye coordinates are used to identify and remove any rotation of the input face image. The input is rotated so that the eye coordinates are level (that is, they have the same y-coordinate). In addition, the input is scaled so that the distance between the eyes is 70 pixels. Finally, the input is cropped to a size of 130×150.

Histogram equalization is used to reduce illumination variations, and pixel values are normalized to have a mean of 0 and standard deviation of 1. Fig.6 shows three examples of preprocessed images.

2.4.2 Feature vector extraction

After preprocessing, we have a 130×150 grayscale face image. If we use the pixel data directly, we end up with a feature vector of dimension 19,500. Such a high-dimensional feature vector requires both large computation times and large memory requirements. To get a more efficient representation, we propose extracting gradient feature vectors from the pixel data. The process of extracting the gradient feature vector for face recognition is almost the same as that described in Section 2.2, though slightly different parameters are used. Here, the input image size is 130×150, the image is divided into 11×11 subblocks. 32 gradient angle levels are used. Finally, and similar to what was done in Section 2.2, we reduce the dimensionality of the face gradient feature vector from $11 \times 11 \times 32$ down to $6 \times 6 \times 16$, and hence end up with a 576-dimensional vector of gradient features.

2.4.3 Dimensionality reduction

The dimension of the gradient feature vector can be further reduced using methods like principal component analysis (PCA) and linear discriminant analysis (LDA). In face recognition. PCA and LDA are known as Eigenfaces[8] and Fisherfaces[9], respectively.

2.4.4 Classification

Since we will be using the FERET database[10] for testing, we will assume that each person in the database has 1 image sample. For each image sample, we will compute and store the gradient-based feature vector and perform all of the needed preprocessing, as discussed above. The same feature extraction and preprocessing is done for each input image, as well.

Since we have only have 1 sample per person, we will use a nearest neighbor-based classifier to map the input feature vector to the appropriate class label. The nearest neighbor approach requires a distance measure, and there are several possibilities to choose from: Euclidean distance, Mahalanobis distance, correlation, etc.

3 Experimental results and discussion

3.1 Face detabase

We used the color FERET database[10], which is specifically designed for face verification and identification. From the 1196 images in the gallery, 500 are used for training. Here, training means parameter estimation for dimensionality reduction. We use 4 test sets: fb, fc, Duplicate 1, and Duplicate 2. The duplicate sets contain images of several subjects taken over a period of time.

3.2 Experiments

In this section, we want to examine the efficacy of the proposed face detection and extraction in the context of recognition. To do this, we will compare the recognition results between the semi-automated system, where the face images are cropped manually (and assumed correctly), and the fully automated system, where the face images are processed entirely by our detection/recognition algorithms.

3.2.1 Benchmark algorithms

To compare the proposed gradient features-based recognition algorithm, we will benchmark our results with some other common face recognition algorithms: principal component analysis and Euclidean distance (PCA/Euclidean), the linear discriminant analysis and ldaSoft distance (LDA/ldaSoft), a Bayesian intrapersonal-extrapersonal classifier (BIC). The experiments utilizes the CSU face identification evaluation system (Version 5.0) software package[11].

3.2.2 Performance measures

Let T denote a set of test images and let $|T|$ denote the cardinality of T. One of the advantages of the nearest neighbor classification approach is that we are able to generate an ordered list of best-matching patterns. For each test image, we determine the 10 best matching database images; that is, 10 database images with the smallest distance in the nearest neighbor computation, sorted from lowest to highest. We are interested in knowing where in the list (if at all) the correct person appears. Of all $|T|$ test images, let $C(0)$ denote the number of times the correct person appeared at the top of the list (rank 0). Let $C(1)$ denote the number of times the correct person appears first or second in the list (rank 1). Similarly, let $C(r)$ denote the number of times the correct person appears in the top $r+1$ list of best matching patterns. Performance is then evaluated according to the cumulative match characteristic (CMC) metric presented in citeBolme:2003:CSU and defined as

$$P(r) = \frac{|C(r)|}{|T|} \times 100 \tag{2}$$

3.2.3 Semi-automatic face recognition results

The results of the semi-automatic face recognition experiments are shown in Fig. ??. Here, (a) shows the results using the fb test set, (b) shows the results using the fc test set, (c) shows the results using the Duplicate 1 test set, and (d) shows the results using the Duplicate 2 test set. In 19 each plot, the recognition rate $P(r)$ is plotted against the rank r. The gradient-based features are shown in black solid symbols, whereas

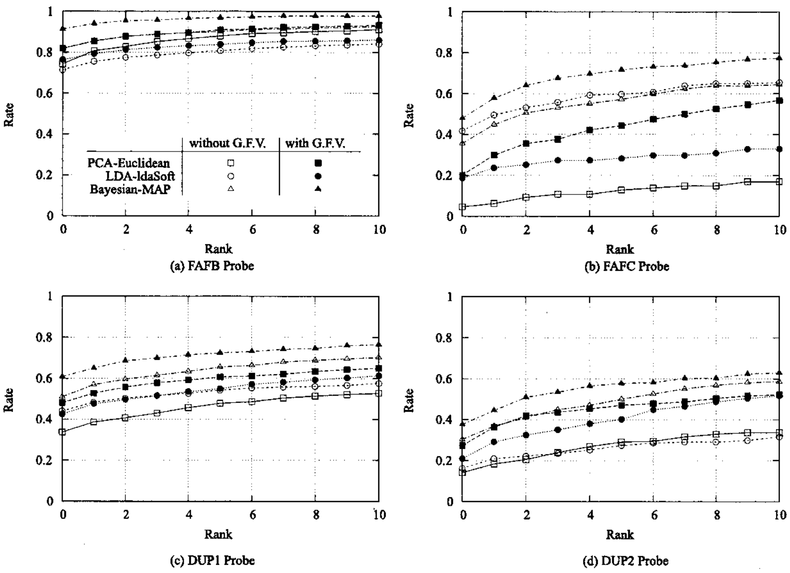

Fig. 7 Results of the semi–automatic system on the (a) fb test set, (b) fc test set (c) duplicate 1 test set and (d) duplicate 2 test set. In each plot, the recognition rate is plotted against the rank of the top winner.

the pixel-based features are shown as open symbols.

In each case, the best results are obtained when the gradient features are used along with the Bayesian MAP classification.

3.2.4 Fully automatic face recognition results

The results on the fully-automated face recognition problem are shown in Fig. 8. As before, the gradient features typically offer better performance over the pixel-based features. As expected, the results here are worse than those in the semi-automated case. On average, there ' s about a 10% reduction in classification performance for the fully automated case.

4 Conclusions

We have shown that the proposed gradient-based features can be successfully used for both facial feature extraction and for face recognition. The advantage of the gradient features is that it gives a compact representation of the image data. In our study, the gradient feature vector was 576-dimensional (whereas the CSU employs 19,500 dimensional pixel value based feature vector). Standard dimensionality reduction and classification approach are easily applicable to each original feature vector. We should also note that computational time required by the classification is significantly reduced by introduction of gradient feature vector.

In future work, we will explore the robustness of the gradient features to other types of image variations; for example, rotation, changes in facial expression, changes in illumination, etc.

5 Acknowledgements

Portions of the research in this paper use the Color FERET database of facial images collected under the FERET program[10).

References

1) Yang, M.-H., Kriegman, D. J. and Ahuja, N.: Detecting Faces in Images: A Survey, *IEEE Trans. on PAMI*, Vol. 24, No. 1, pp. 34–58 (2002).

2) W.Zhao, R.Chellappa, Phillips, P. and A.Rosenfeld : Face Recognition: A Literature Survey, *ACM Computing Surveys*, Vol. 35, No. 4, pp. 399–458 (2003).

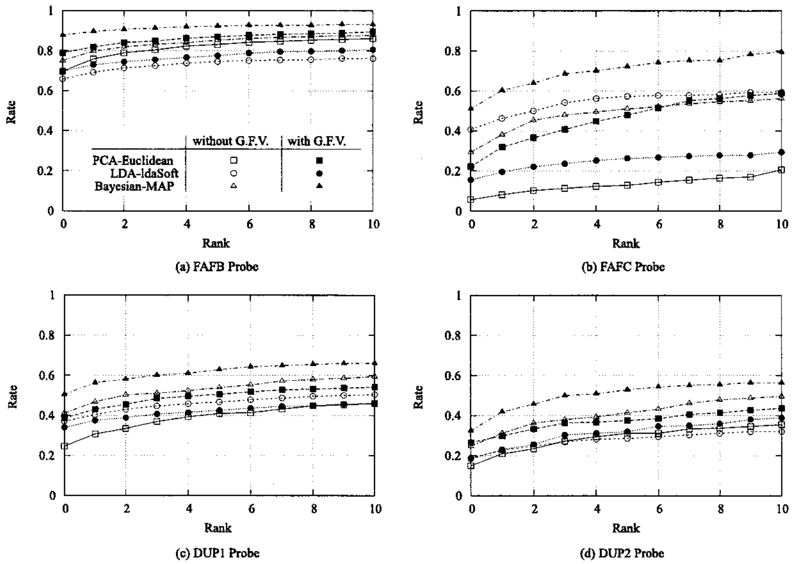

Fig. 8 Results of the fully–automatic system on the (a) fb test set, (b) fc test set (c) duplicate 1 test set and (d) duplicate 2 test set. In each plot, the recognition rate is plotted against the rank of the top winner.

3) Chen, L.-F., Liao, H.-Y. M., Lin, J.-C. and Han, C.-C.: Why recognition in a statistics-based face recognition system should be based on the pure face portion: a probabilistic decision-based proof, *Pattern Recognition*, Vol. 34, No. 7, pp. 1393–1403 (2001).

4) Osuna, E., Freund, R. and Girosi, F.: Training Support Vector Machines: an Application to Face Detection, *Proc. CVPR'97*, p. 130 (1997).

5) Hamouz, M., Kittler, J., Kamarainen, k. J., Paalanen, P., Kalviainen, H. and Matas, J.: Feature-Based Affine-Invariant Localization of Faces, *IEEE Trans. on PAMI*, Vol. 27, No. 9, pp. 1490–1495 (2005).

6) Shi, M., Fujisawa, Y., Wakabayashi, T. and Kimura, F.: Handwritten numeral recognition using gradient and curvature of gray scale image, *Pattern Recognition*, Vol. 25, No. 10, pp. 2051–2059 (2002).

7) Ohyama, W., Shridhar, M. and Watta, P.: Facial Feature Extraction Using Gradient Features and MQDF Matching, in *Proc. of 8-th IASTED International Conference SIP*, pp. 329–334 (2007).

8) Turk, M. and Pentland, A.: Face recognition using eigenfaces, in *Proceedings of IEEE Computer Society Conference on CVPR'91*, pp. 586–591 (1991).

9) Belhumeur, P. N., Hespanha, J. P. and Kriegman, D. J.: Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection, *IEEE Trans. on PAMI*, Vol. 19, No. 7, pp. 711–720 (1997).

10) Phillips, P. J., Moon, H., Rizvi, S. A. and Rauss, P. J.: The FERET Evaluation Methodology for Face-Recognition Algorithms, *IEEE Trans. on PAMI*, Vol. 22, No. 10, pp. 1090–1104 (2000).

11) Bolme, D. S., Beveridge, J. R., Teixeira, M. and Draper, B. A.: The CSU Face Identification Evaluation System: Its Purpose, Features, and Structure, in *Proceedings of Third International Conference Computer Vision Systems*, Vol. 2626/2003 of *Lecture Notes in Computer Science*, pp. 304–313 (2003).