MVA2000 IAPR Workshop on Machine Vision Applications, Nov.28-30, 2000, The University of Tokyo, Japan, pp.211-214

Feature Ordering by Cross Validation for Face Detection

Takio KURITA *
Electrotechnical Laboratory

Kazuhiro HOTTA and Taketoshi MISHIMA †
Saitama University

Abstract

This paper presents the method to determine the order of feature (attention) points for face detection. The order of feature points is determined in terms of the classification ability to face and non-face images (cross validation). The matching is performed in the order of the selected feature points. It is confirmed that high recognition rate is obtained by using only a small number of feature points. By ordering of the feature points, we can reject the input image as non-face by systematically checking the feature points. This can speed up the face detection.

1 Introduction

In order to recognize an object, human pays attention to the characteristic points of the object. Our eyes always move on such characteristic points. By integrating the information obtained from such attention points, we can easily recognize the object (face) under various environments. Previously, we have proposed a recognition method which introduces the information processing in biological vision system and takes the idea of attention mechanism into consideration [1]. The contrast filter is used to simulate the information processing at the retina. The profiles of the contrast filter resembles the receptive field profiles of retinal ganglion cells [2, 3, 4]. Gabor filter [5] is also used to simulate the information processing at V1 area. In addition, information measure of Gabor features is used as saliency for attention. By using the saliency value (information of Gabor features) as weight in matching, the feature (attention) points with rich information have much influence to the matching result. It is confirmed that high recognition rate is obtained when only the feature points with high saliency are used in matching. Although it is expected that the feature points with rich information give much influence to the face recognition task, the relation between the feature points with rich information and classification ability is not so clear. In this paper, the feature points are directly selected (ordered) in terms of the classification ability to face and non-face images (cross validation). The matching is performed in the order of the selected feature points. It is confirmed that high recognition rate is obtained by using only a small number of feature points. By ordering of the feature points, we can reject the input image as non-face by systematically checking the feature points. This can speed up the face detection.

In section 2, a face detection method used in the following experiments is explained. The selection method of feature points are shown in section 3. Experimental results are shown in section 4.

2 Face Detection Method

In the face detection method [1], we apply the contrast filter to simulate the information processing at the retina. It is known that the profile of such filter resembles the profile of retinal ganglion cells[2, 3, 4]. Then Gabor filters are applied to the contrast filtered images. It is well known that Gabor filters resemble the receptive field profiles of the simple cells in V1 area [5]. Thus this process corresponds to the information processing at V1 area. The information of the outputs of Gabor filters extracted from the contrast filtered image is used as the weight for matching. This means that the feature points with rich information have much influence to the matching result. The output probabilities of Gabor filters were estimated from a set of general images. In this paper, 1500 different TV images were obtained by using the cut detection method based on the robust correlation [6]. Then a set of local regions with 9×9 pixels were randomly selected from the TV images and the probability distributions of the outputs of Gabor filters were estimated.

Since the outputs of each Gabor filters can be regarded as independent [7], the output probability density $p(\mathbf{x})$ to the Gabor feature vector $\mathbf{x} = (x_1, \ldots, x_N)$ can be calculated by

$$p(\mathbf{x}) = \prod_{j=1}^{N} p(x_j), \tag{1}$$

where N denotes the number of Gabor filters which

* Address: 1-1-4, Umezono, Tsukuba, Ibaraki, 305-8568 Japan. E-mail: kurita@etl.go.jp
† Address: 255, Shimo-Okubo, Urawa, Saitama, 338-8570 Japan. E-mail: {hotta, mishima}@me.ics.saitama-u.ac.jp

have different orientations and $p(x_j)$ is the output probability density of jth Gabor filter. Then the information of Gabor feature vector $\mathbf{x}$ can be defined by

$$I(\mathbf{x}) = -\log p(\mathbf{x}) = -\sum\nolimits_{j=1}^{N} \log p(x_j). \quad (2)$$

In this paper, the probability density of an output of a Gabor filter is approximated by using normal distribution with mean zero. Thus the probability density $p(\mathbf{x})$ and the information $I(\mathbf{x})$ to Gabor features $\mathbf{x}$ are computed as

$$\begin{aligned} p(\mathbf{x}) &= \prod_{j=1}^{N} \frac{1}{\sqrt{2\pi}\sigma_j} \exp\left(-\frac{{x_j}^2}{2{\sigma_j}^2}\right), \\ I(\mathbf{x}) &= \sum\nolimits_{j=1}^{N} \{-\log\left(\frac{1}{\sqrt{2\pi}\sigma_j}\right) + \frac{{x_j}^2}{2{\sigma_j}^2}\}, \quad (3) \end{aligned}$$

where ${\sigma_j}^2$ denotes the variance of output to jth Gabor filter.

The information values of all points on the model face image are calculated and those values are used as weights for matching. The matching similarity is defined as the weighted sum of the correlations of Gabor features extracted from contrast filtered image. The similarity measure is defined by

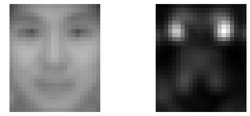
$$\text{Similarity} = \sum_{i=0}^{M} \frac{I(\mathbf{x}_i)}{\sum_{k=0}^{M} I(\mathbf{x}_k)} \operatorname{Cor}(i), \quad (4)$$

where M is the number of pixels on the model face image, $I(\mathbf{x}_i)$ is the information of ith pixel on the model image, and $\operatorname{Cor}(i)$ is the correlation between Gabor features of ith pixel on the model image and the input image.

The model face image (26×22 pixels) and its information map are shown in Figure 1. The model image is the mean image of 104 person's faces. Original face images are obtained from Web. It is noticed that eyes, nose, and mouth have high information values. To detect faces from a given input image, all pixels on the input image are checked as candidates of a face center. Namely, similarities of all candidates are computed and the locations whose similarities are larger than a given threshold are classified as face. To cope with the scale changes of a face on the image, several images of different scales are generated by interpolation from the input image and these images are used as the matching candidates.

3 Feature Ordering by Cross Validation

To select the feature points, the classification ability is evaluated on samples of face and non-face images. The ordered feature points are systematically used in the matching.



(a) (b)

Figure 1: Model face and Information map. (a) Model face for face detection (26×22 pixels). (b) Information map of (a).

In the following experiments, the mean face image of 104 persons (Figure 1(a)) is used as model face. The weighted sum of the correlations of Gabor features (4) is used as the similarity measure for face and non-face classification. The correct recognition rates to the samples of face and non-face images are evaluated for the combinations of the feature points. For simplicity, we use the same strategy with *forward stepping* of stepwise regression [8]. At first, the feature point with the largest possible value of the correct recognition rates is selected as the initial set of the feature points. Then one feature point is added at a time so as to produce at each step the largest possible value of the correct recognition rates by adding just a single feature point to the set of the feature points. This procedure produces an increasing sequence of nested subsets of the feature points. In this paper, 246 face images and 413 non-face images are used as the images to determine the ordering of the feature points. Examples of face and non-face images are shown in Figure 2. The size of these images is 26×22 pixels. Face images from MIT face database and the face images on the Web are used. The images obtained from Web are used as non-face images. The non-face images misclassified as face image by the face detection method described in section 2 are also included in the set of the non-face images.

The algorithm to determine the ordering of the feature points is given as follows.

1 Extract Gabor features of contrast filtered image from data set.

2 Select one feature point with the largest value of the correct recognition rates as the initial set of the feature points.

3 Add a single feature point with the largest possible value of the correct recognition rates to the set of the feature points.

4 Continue 3 until all points are ordered (selected).

In the ordering process of the feature points, the correct recognition rate is selected by computing the optimal threshold for face and non-face classification. The order of feature points and the optimal threshold at each step are systematically used in the matching for face detection.

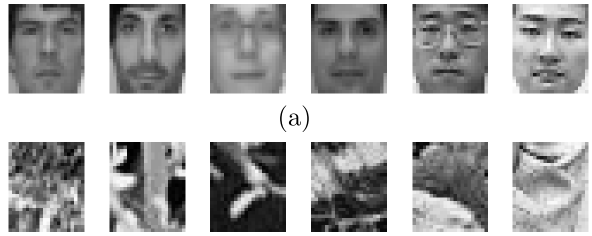
4 Experimental Results

For evaluation of the proposed method, the set of face and non-face images are divided into two data sets. The data set 1 (123 face images and 207 non-face images) is used to select the feature points. The data set 2 (123 face images and 206 non-face images) is for evaluation.

Figure 3 (a) (black line) shows the correct recognition rate to data set 1 when the different number of feature points are used in the order of the selected feature points. The horizontal axis represents the number of feature points. The vertical axis represents the recognition rate for the sets of feature points. The high recognition rates were obtained at which the number of feature points is between 65 and 245. The recognition rates of these features are higher than the case that all feature points are used in the matching. In Figure 4, the first 7-th feature points are shown. These feature points are located in different positions of the face. This means that information from different positions is necessary to classify faces and non-faces using small number of feature points. To investigate the effect of the ordering of feature points, we have compared the proposed feature sets with the feature sets in which feature points are ordered at random. The recognition rates obtained by using the order of feature points selected at random is shown in Figure 3 (a) (gray line). The optimal thresholds are also computed at each step of the ordering process. In this case the recognition rates are always lower than the proposed ordering method. This result shows the effectiveness of the order of feature points.

To investigate the generalization ability of the proposed ordering method, the experiments using unknown samples (data set 2) have been performed. Figure 3 (b) (black line) shows the recognition rates to data set 2 when the matching is performed in the order of the selected feature points. The threshold values selected to data set 1 are used to discriminate faces with non-faces. For comparison, the recognition rates are computed for the case that feature points are randomly ordered. The recognition rates at this feature sets are shown in Figure 3 (b) (gray line). The recognition rate of the proposed method is always higher than the case that the feature points are randomly ordered.

The proposed method is applied to face detection task. The number of feature points is set to 150 because the high recognition rate is obtained in the previous experiments (Figure 3).



(a)

(b)

Figure 2: Examples of face and non-face images. (a) Examples of face images. (b) Examples of non-face images.

To detect faces from a given input image, all pixels on the input image are checked as candidates of a face center. Then the locations whose similarities (4) are larger than a given threshold are classified as face. By using the ordering of the feature points, we can reject the candidates as non-face by checking only a small number of the feature points. The rejection condition in the case of the first k-th feature points is given by

$$\sum_{i=1}^{k} w(i)\,(1-\mathrm{Cor}(i)) > \left(\sum_{j=1}^{ep} w(j) - \theta\right), \qquad (5)$$

where $w(i)$, $\mathrm{Cor}(i)$, ep, θ is the weight at i-th point, the correlation value of Gabor features at i-th point, the maximum number of feature points used in matching (we set to 150 in the experiments), the threshold value, respectively.

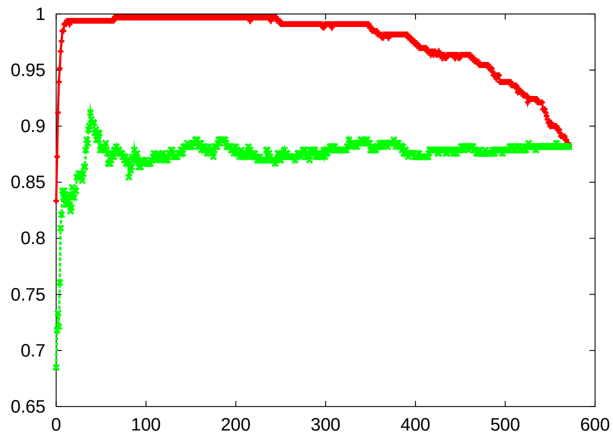
To cope with the scale changes of a face on the input image, several images of different scales are generated by interpolation from the input image and these images are used as the matching candidates. The results of face detection are shown in Figure 5. White squares represent the regions detected as faces. It is noticed that some faces with different scales are correctly detected by the proposed method.

References

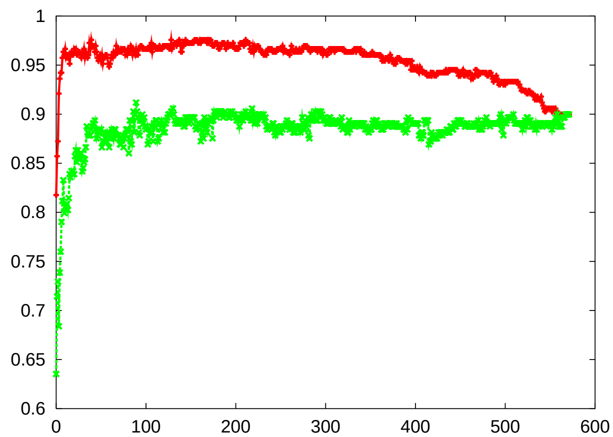
[1] K.Hotta, T.Mishima, T.Kurita, and S.Umeyama, "Face matching through information theoretical attention points and its applications to face detection and classification," in *Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 34–39, 2000.

[2] J.J.Atick and A.N.Redlich, "What does the retina know about natural scenes?," *Neural Computation* **4**(2), pp. 196–210, 1992.

[3] B.A.Olshausen and D.J.Field, "Sparse coding with an overcomplete basis set: A strategy employed by v1?," *Vision Research* **37**, pp. 3311–3325, 1997.

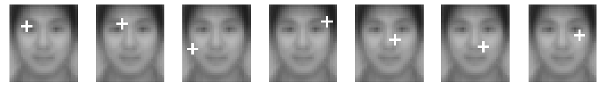


(a) Data set 1. (selection of feature points)



(b) Data set 2. (evaluation of ordered feature points)

Figure 3: Recognition rates when the different number of feature points are used in the order of the selected feature points.



(a) (b) (c) (d) (e) (f) (g)

Figure 4: The selected first 7-th feature points ((a)-(g): 1-th point to 7-th point)

Figure 5: Results of face detection.

[4] R. P. N. Rao and D. H. Ballard, "Efficient encoding of natural time varying images produces oriented space-time receptive fields," tech. rep., 97.4, Dept of Comp Sci, Univ of Rochester, 1997.

[5] J. Daugman, "Entropy reduction and decorrelation in visual coding by oriented neural receptive fields," *Trans. on Biomedical Engineering* **36**(1), pp. 107–114, 1989.

[6] T.Kurita, "Robust template matching and its application to cut detection," *Proc. of 1997 IEICE General Conference* (D-12-61), p. 268, 1997 (in Japanese).

[7] A. J. Bell and T. J. Sejnowski, "Edes are the 'independent components' of natural scenes.," *Vision Research* **37**(23), pp. 3327–3338, 1997.

[8] R. A. Thisted, *Elements of Statistical Computing Numerical computation*, Chapman and Hall, 1988.