Journal of the Faculty of Engineering, Shinshu University, **No.74,1994**
信州大学工学部紀要 第 74 号

The Extraction and Recognition of Facial Images in the Complex Background

Dai Ying *, Yasuaki Nakano ** and Hidetoshi Miyao **

複雑背景中の顔画像の抽出と識別

戴 瑩 *　　中野 康明 **　　宮尾 秀俊 **

In this paper, we proposed a new method for face extraction in the complex backgrounds. Based on the space gray level dependence matrix, a facial texture model composed by a set of inequalities was derived. The weight coefficients in the inequalities were decided through Perceptron type learning. According to this facial texture model, the face positions in the complex backgrounds were detected and located effectively.

Next, the experimental evidences for determinating the resolution bound of facial images at which the faces can be identified were given on the basis of subjective tests. Machine recognition based on the pattern matching approach was also executed. The experimental results suggest that the bound for the correct recognition by pattern matching method coincides with that of perceptional tests, and the pattern matching is also effective to the face recognition with low quality images.

⋮

1. Introduction

Recently, the study of the face recognition techniques has given rise to the more interests for researchers from the viewpoint of many applications such as security systems, criminal identifications and teleconferences.

Many researchers have reported face recognition techniques[1], [2], [3] [4]. However, these researches assume that the facial parts have already been isolated.

* Doctral Program, Faculty of Engineering, Shinshu University, Wakasato, Nagano, 380 JAPAN
** Department of Information Engineering, Faculty of Engineering, Shinshu University, Wakasato, Nagano, 380 JAPAN

It is the key step for face recognition to extract the facial parts from the low quality images. From a practical viewpoint the extraction of facial images must be done automatically. Furthermore, the quality might be very low in many situations. For example, a security system, where a CCD camera with 512×512 resolution covers a whole entrace hall, it might give only 20×30 pixels to a face, and the focusing of lenses might not be so good.

There has been little discussion on the face extraction from complex backgrounds and the lower bound of the resolution to give the sufficient performance in face recognition.

In this paper, we proposed a new method for the face extraction in the complex background. Based on the space gray level matrix (SGLD) [5], a textural model composed by a set of face judgement conditions was derived as a set of inequalities. By means of this textural model, we designed a kind of scanning scheme for face extraction in the complex backgrounds.

After extracting the facial parts from complex backgrounds, we tried to have a test on the recognition of facial images with low quality based on the human perception. In the perception test, facial images with variety of resolution were given, and human subjects familiar with these faces were asked to identify them. By this test, an approximate resolution bound of face recognition based on the human perception was obtained. The computational recognition experiments based on a pattern matching method were also done and the results gave good correspondence with the perception test.

2. Textural Feature Model of the Facial Images

2.1 Space Gray Level Dependence Matrix

The performance of the space gray level dependence (SGLD) matrix which is used in textural feature analysis was presented by Haralick [5]. Let us denote the size of an image to be analyzed by $X \times Y$ pixels. The gray level appearing in each pixel is quantized to L levels. Let $L_x = \{0, 1, ..., X-1\}$, and $L_y = \{0, 1, ..., Y-1\}$ be the sets of horizontal and vertical indices respectively, and $G = \{0, 1, ..., L-1\}$ that of quantized gray levels. Let $I(i,j)$ be the gray level value at pixel$(i,j) \in (L_x \times Ly)$. Suppose $P_{ab}(m,n)$ to be the number of occurrence in which two neighboring pixels displaced by a vector (m,n) on the image have gray level a and b, respectively. A matrix composed by $P_{ab}(m,n)$ is called as space gray level dependence matrix (SGLD) $P(m,n)$, with a parameter vector (m,n). The formula to obtain $P_{ab}(m,n)$ is defined by

$$P_{ab}(m,n) = \sharp\{((i,j),(i+m,j+n)) \in (L_x \times L_y), I(i,j) = a, I(i+m,j+n) = b\} \quad (1)$$

where $\sharp$ denotes the number of elements in the set.

The frequency normalization for the matrix can be computed approximately by the formula $N_{ab}(m,n) = P_{ab}(m,n)/M$, where $M = X \times Y$ is the total number of pixels of the image to be analyzed.

By means of SGLD matrix, a set of textural features can be derived. Especially, we consider two of them, which will be used in the next section. They are covariance, inertia, and inverse difference features. The equations which define these measures of textural features are

$$B_I(m,n) = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1}(a-b)^2 N_{ab}(m,n) \tag{2}$$

$$B_D(m,n) = \sum_{a=0}^{L-1}\sum_{b=0}^{L-1}\frac{N_{ab}(m,n)}{1+(a-b)^2} \tag{3}$$

where L is number of gray levels, and μ denotes average level. Accordingly, with varying in m and n, a set of array can be formed, the elements of which are $B_I(m,n)$, and $B_D(m,n)$ respectively. Let B_I, and B_D denote these array respectively, then

$$B_I = \begin{pmatrix} B_I(0,0) & \ldots & B_I(M,0) \\ \\ B_I(0,N) & \ldots & B_I(M,N) \end{pmatrix} \tag{4}$$

$$B_D = \begin{pmatrix} B_D(0,0) & \ldots & B_D(M,0) \\ \\ B_D(0,N) & \ldots & B_D(M,N) \end{pmatrix} \tag{5}$$

In these arrays, M and N are the upper limits of displacement to be determined by a preliminary experiment.

2.2 Texture Model of Faces Based on the SGLD Matrix

The facial images can be considered as a kind of texture with some special textural characteristics. First, the texture of facial images is not of orientation and duplication, but appears as a whole. Second, it is nearly symmetric respect to the medial axis. On the other hand, from the view of gray level distribution of facial images, we know that the gray level variation on the local region in the horizontal direction is not larger than the one in the vertical direction on the whole, because of the influence of eyes, mouths, and facial gradient. Besides, according to the facial shape, the gray levels of neighboring pixels are considered homogeneous in the local region. Due to the analysis above, we can use the feature parameter matrix of SGLD to describe the facial features.

The neighbor set of a certain pixel (i, j) can be given by

$$N_p = \{\mathbf{v} = (l, k) : 0 \leq l - i \leq N_1,\ and\ 0 \leq k - j \leq N_2\}$$

The feature matrices B_I, and B_D in eq. (4) and (5) are calculated on the neighborhood N_p.

Observing B_I's calculated for many human faces, we can get

$$B_I(m_1, n_1) < B_I(m_2, n_1),\ if\ m_1 < m_2, \tag{6}$$

$$B_I(m_1, n_1) < B_I(m_1, n_2),\ if\ n_1 < n_2; \tag{7}$$

$$B_I(m_1, n_1)(1 + w) < B_I(m_2, n_2), \tag{8}$$

if $0 \leq \| \mathbf{r_2} \| - \| \mathbf{r_1} \| < \sqrt{2} - 1$, and $| \theta(\mathbf{r_2}) | > | \theta(\mathbf{r_1}) |$, where $\mathbf{r_i} = (m_i, n_i)$.

Here w is a positive coefficient.

Moreover, because of the homogeneity on the local region for facial part, it is observed that

$$B_D(m, n) < c, \tag{9}$$

where $0 < c < 1$, and$(m, n) \in D_p$, $(m, n) \neq (0, 0)$.

If w is large enough and c is small enough, and if they are stable for facial images, we can use ineq. (6) and (9) as the characteristics of them. In other words, these two coefficient will serve to discriminate the facial parts from other parts. Thus, the facial region can be estimated as such area where inequalities (6)–(9) hold. The weight w and c can be determined by sample learning.

2.3 Examination for Facial Feature Model

In order to obtain the weight coefficients w and c, we use 40 facial image samples with 20 different faces under the different light condition for learning. The experiments testified that the coefficients converge to $w = 0.3$ and $c = 0.6$.

Next, we performed an experiment with about 150 test samples to examine the performance of the facial feature model described above. The samples included full facial images, nonfacial images, and the images which only have part of face. Examples of test samples are shown in Fig.1. The size of images were varied from 12×16 to 20×26. By calculating the parameter matrix B_I and B_D of SGLD, the facial images were determined, if the B_I and B_D met the inequalities (6), (7), (8), (9). The experimental results are shown in Table 1.

According to the experimental result, we can conclude that the facial feature model described above is effective. We can expect that it works for extracting the facial parts in the complex backgrounds. In the next section, we will discuss the face extraction using the above model.

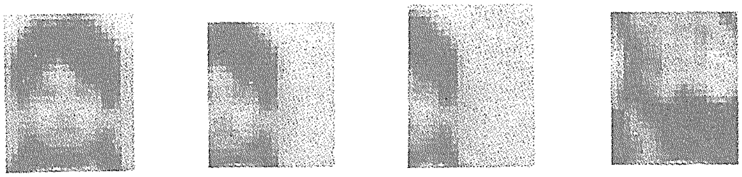

Figure 1 Test samples with varied portions of facial parts

Table 1 Experiment result

test samples		rate of samples determined as faces
case	number	
full facial img.	20	100%
2/3 facial img.	25	80%
1/2 facial img.	30	40%
1/4 facial img.	25	6%
nonfacial img.	50	2%

3. Extraction of the Facial Images

Based on the facial feature model described above, a method to extract facial parts from full image is proposed. In this scheme, the original image is scanned by window for locating the facial position on the basis of facial feature model. The choice of the window size depends mainly on two factors: 1) how large are the localized faces; 2)whether does the model represent the facial features in a window of such size.

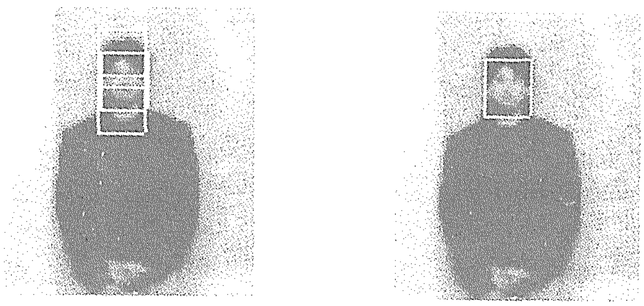

Figure 2 Overlaped regions and post-processing result

According to the examination for the facial feature model performance, the part of image scanned by window will be considered as face, if the window contains "face" pixels more than a half. So, the horizontal and vertical step length of scan for searching the face position is set as 1/3 of window's width and height respectively. Calculating the inertia and inverse difference matrix B_I and B_D of image in the scanning window, it is considered that the facial part exists in this window, if the elements of matrixs meet the inequalities (6)–(9). This procedure continues until whole image is scanned.

The regions extracted as facial parts may be overlapped, because the window moves with rather small step. Using an intelligent post-processing algorithm, overlapped candidates can be unified. An example of the processed results is shown in Fig.2.

4. Subjective Tests for Face Recognition

The samples for subjective test consisted of 14 facial images taken from 11 persons. Among them, a couple of images were taken for 3 persons with and without glasses. The width of facial images were varied from 12 pixels to 20 pixels. Fig.3 shows a set of samples on this test.

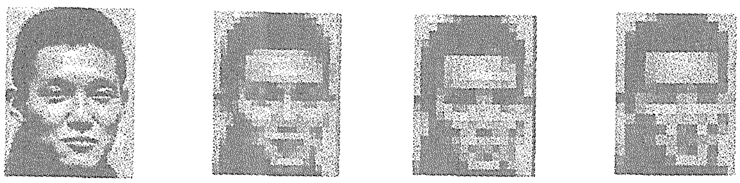

Figure 3 Facial images with varied resolution
From left to right, original, 20×26, 16×20 and 12×15

Nine subjects who are familiar with the inputed faces were asked to recognize these facial images, and give the answer of belonging to someone or being unrecognizable.

The experiments showed that at the resolution of 12×16 pixels, the images become almost unrecognizable. At the resolution of 16×20 pixels, the number of unrecognized images decreased, and all the test images can be identified by three subjects who are considered to have the powerful identification ability. It indicates the information of the images with this resolution are reserved essentially. At the resolution of 20×26 pixels, all facial images can be recognized correctly. The conclusion roughly coincides that of preliminary results suggested by Harmon[1].

5. Face Recognition Based on Pattern Matching

5.1 Pattern Matching Based on Similarities

Along with the subjective tests, we executed computer recognition experiments. As the simplest and most basic method, we adopted a pattern matching approach, where the similarity between the template and test images with a certain resolution is used to identify the test images. The recognition performance will be compared at the various resolution. The similarity is defined by the formula

$$m = \frac{\int\int(f(x,y)-\bar{f})(t(x,y)-\bar{t})dxdy}{\sqrt{\int\int(f(x,y)-\bar{f})^2dxdy\int\int(t(x,y)-\bar{t})^2dxdy}}$$

where

$f(x,y), t(x,y)$: template and test images, respectively.

$\bar{f}, \bar{t}$: the average of gray values of $f(x,y)$ and $t(x,y)$, respectively.

It is well known that the similarity is in the range of $[-1,1]$. The larger the value is, the more resemble the template and test images are.

5.2 Experimental Results by the Pattern Matching Method

The samples of facial images which were used by subjective tests set the subjects with or without glasses as the two set of templates respectively. These template samples were normalized and the resolution was varied from 8×10 pixels to 20×26 pixels. Two kinds of similarity matrices between the test and template samples were computed. One of them was the similarity matrix in which the template set included the subjects with glasses, and the test samples were the subjects without glasses. The results are shown in Table 2. The other was just contrary, but the results are not shown because of the lack of space.

5.3 Experimental Results under Temporal Variations

In the experiments at the preceding sections, the photographs used were taken in a very short era. As to investigate the influence of temporal variations, we executed another series of recognition. Photographs of five subjects were taken about one year later than the former experiments. They were subjects 2,6,7,8,10. Subject 7 put glasses on, and photographs with and without glasses were taken.

In these images, the blurring was done by optical means rather than by computer simulation. In other words, the photographs were taken at a long distance, and the enlargement ratio was changed in the development of the photo-prints. The whole scene of a photograph from which a face was extracted is shown in Fig.4.

The sampling pitch was adjusted so as to represent the facial part at 8×10, 12×16, 16×20, and 20×26 resolution. A set of samples are shown in Fig.5. Using these

Table 2 Similarity Matrices

resolution: 8×10 pixels											
test	similarity										
	1_g	2_g	3	4	5	6	7_g	8	9	10	11
1	0.84	0.41	0.31	0.63	0.61	0.42	0.79	0.59	0.64	0.55	0.66
2	0.34	0.76	0.72	0.72	0.82	0.74	0.52	0.78	0.74	0.73	0.46
7	0.70	0.47	0.41	0.62	0.73	0.47	0.95	0.53	0.66	0.52	0.58

resolution: 12×16 pixels											
test	similarity										
	1_g	2_g	3	4	5	6	7_g	8	9	10	11
1	0.78	0.45	0.33	0.63	0.50	0.34	0.56	0.58	0.53	0.44	0.59
2	0.34	0.76	0.59	0.60	0.62	0.55	0.48	0.72	0.64	0.67	0.40
7	0.65	0.48	0.45	0.56	0.57	0.42	0.89	0.57	0.56	0.38	0.59

resolution: 16×20 pixels											
test	similarity										
	1_g	2_g	3	4	5	6	7_g	8	9	10	11
1	0.73	0.37	0.36	0.60	0.50	0.38	0.62	0.44	0.57	0.38	0.51
2	0.40	0.73	0.62	0.60	0.68	0.66	0.54	0.68	0.60	0.69	0.49
7	0.64	0.52	0.43	0.63	0.64	0.38	0.87	0.46	0.58	0.38	0.53

resolution: 20×26 pixels											
test	similarity										
	1_g	2_g	3	4	5	6	7_g	8	9	10	11
1	0.71	0.36	0.38	0.52	0.53	0.24	0.55	0.49	0.47	0.45	0.56
2	0.21	0.70	0.44	0.56	0.58	0.28	0.38	0.67	0.53	0.68	0.30
7	0.59	0.39	0.41	0.48	0.44	0.30	0.87	0.47	0.55	0.40	0.46

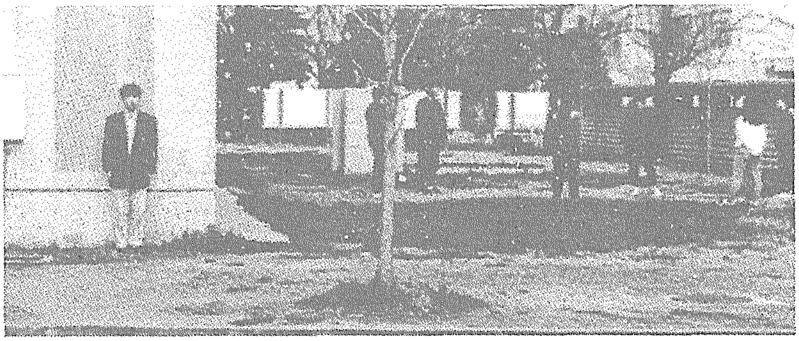

Figure 4 Whole scene from which the facial image in Fig.5 is extracted

images, the recognition experiments based on pattern matching were executed. The results using 20×26 resolution are shown in Table 3, but other results are not shown in order to save the space.

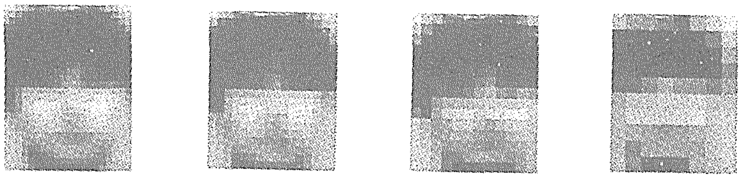

Figure 5 Test samples with different resolution varied from 8×10 to 20×26

5.4 Analysis of the Results

According to the recognition results in Table2, one of test samples without glasses can not be identified with the resolution corresponding to 8×10 pixels, although all test samples with glasses are identified with a series of resolution based on a min-distance criterion between test and template samples. On the other hand, we can observe that the distances between the different subject's templates each other keep large enough not to be disturbed by the degradation of facial image's quality beyond resolution corresponding to 16×20 pixels, but below that resolution, the distances between the candidate and other samples become small, and the similarity distance rank are varied. In Table 3, we find that except one of them, all test samples beyond

Table 3 Similarity between test and template samples

resolution: 20x26 pixels											
test	similarity										
	1	2	3	4	5	6	7	8	9	10	11
2	0.42	0.78	0.32	0.69	0.56	0.19	0.61	0.68	0.70	0.57	0.23
6	0.34	0.65	0.70	0.49	0.51	0.53	0.50	0.54	0.38	0.62	0.40
7	0.56	0.55	0.17	0.62	0.31	0.01	0.76	0.53	0.54	0.35	0.33
7_g	0.59	0.47	0.24	0.51	0.31	0.00	0.72	0.43	0.40	0.35	0.42
8	0.52	0.60	0.35	0.66	0.50	0.25	0.64	0.69	0.67	0.56	0.34
10	0.41	0.77	0.62	0.59	0.64	0.46	0.51	0.72	0.45	0.81	0.44

resolution corresponding to 16×20 pixels were identified correctly based on the min-distance criterion between test and template samples. The reason that one of test samples could not be identified is that there are too much difference between the test and its template patterns to be recognized even by human. But below the resolution corresponding to 16×20 pixels, the error of identification becomes larger based on the min-distance criterion.

Considering the results above, we can conclude that the optimal resolution of face images for machine recognition is about 16×20 pixels. This resolution corresponds with the results of subjective tests given in the previous section.

6. Conclusion

The face extraction in the complex backgrounds using facial textural model based on SGLD matrix was newly proposed and tested.

We conclude that the facial texture model based on SGLD matrix we presented is effective to facial images with low quality. Using the method, facial/nonfacial discrimination experiments were conducted and the recognition rate of 100% and the false alarm rate of 2% were obtained. Also, the facial texture model is robust, and is applied in extracting facial parts in the complex backgrounds successfully. In the face extraction experiments using 30 images containing 70 faces in the complex backgrounds, all facial parts were extracted correctly, with the false alarm rate of 5%.

On the other hand, the experiments based on the human perception and recognition processing showed that the bound of the low quality of the facial images which can be identified is corresponding to 16×20 pixels. A series of computer recognition experiments was also executed, changing the resolution from 8×10 to 20×26.

Recognition scheme was based on the template matching method. From eleven candidates, the scheme succeeded to identify the correct person, when the resolution corresponded 16×20 or better.

The recognition scheme was tested under the more severe condition, where the photographs taken from long distance were used so as to give low quality images of faces. Furthermore, the test images were taken about one year after those of the templates. The result of computer recognition also showed correspondence with the subjective test explained above.

Acknowledgement

The authors acknowledge the contribution from all members of Nakano Laboratory of Dept. of Information Eng., Shinshu University for their providing with portraits and helps in experiments.

Reference

1) L.D.Harmon et al. "Automatic Recognition of Human Face Profiles", Computer Graphics and Image Precessing 6, pp. 135-156 (1977).

2) T.Kurita, et al. "A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analysis", The Proceedings of 11th International Conference on Pattern Recognition, pp.213-216, (1992).

3) S. Akamatsu, et al. "An Accurate Robust Face Identification Scheme", The Proceedings of 11th International Conference on Pattern Recognition, pp.217-220, (1992).

4) M. Kosugi, "Human-Face Recognition Using Mosaic Pattern and Neural Networks", The Trans. of the Institute of EICE, Vol.J76 -D-II No.6 (1993), pp.1132-1139.(in Japanese)

5) R. M. Haralick, "Textural Features for Image Classification" , IEEE Trans. on SMC, Vol. SMC-3, No.6, pp. 610-621 (1973)