Pergamon

Pattern Recognition, Vol. 29, No. 6, pp. 1007–1017, 1996
Elsevier Science Ltd

Printed in Great Britain.
0031-3203/96 $15.00 + .00

0031-3203(95)00139-5

FACE-TEXTURE MODEL BASED ON SGLD AND ITS APPLICATION IN FACE DETECTION IN A COLOR SCENE

YING DAI and YASUAKI NAKANO*
Department of Information Engineering, Shinshu University, 500 Wakasato, Nagano 380, Japan

(*Received* 4 *April* 1995; *in revised form* 14 *August* 1995; *received for publication* 20 *September* 1995)

Abstract—This paper proposed a new method for the full face detection from complex backgrounds. Based on feature parameters of space gray-level dependence (SGLD) matrix, the face-texture model composed by a set of inequalities was derived. A color information utilization incorporated with the face-texture feature was also investigated. Using the face-texture model, we designed a kind of scanning scheme for face detection in color scenes, in which the orange-like parts including the face areas were enhanced by utilizing the *I* component of the YIQ color system. The experiments showed that this method can locate the face position in the complex backgrounds effectively. Copyright © 1996 Pattern Recognition Society. Published by Elsevier Science Ltd.

Space gray-level dependence matrix Face detection Color information Background
Face-texture model

1. INTRODUCTION

Recently, with the progress of image processing and pattern recognition technology, the study of the face recognition techniques has given rise to more and more interests for researchers. The face recognition and interpretation of human faces have many applications such as security systems, criminal identifications and teleconferences.

Many researchers have reported face recognition techniques.[1–6] However, these researches seem to assume that the facial parts have already been isolated.

It is the key step for face recognition to extract the facial parts from cluttered scenes in a general setting. For example, a security system, where a CCD camera with 512 × 512 resolution covers a whole entrance hall, will show a full-view image of a hall. It is important for this system to detect facial parts at the beginning, if this system is expected to recognize the human beings in the hall automatically. Since this system might only give 20 × 30 pixels to a face, and the focusing of lenses might not be good, it is desirable to extract the facial parts from low quality images. Recent studies have begun to address the problem of face localization.[7–12,15]

In reference (7) a hierarchical knowledge-based method consisting of three levels was utilized to locate unknown human faces. In reference (8) a multi-pyramid architecture is used first and then the parts of eyes and noses were utilized to extract the facial regions. Although these methods can locate the unknown human faces spanning a wide range of sizes in a scene, a requirement for the resolution would be slightly high in order to utilize characteristics of eyes and noses. Moreover, experimental results in the papers seem to restrict to only indoor scenes, so it is ambiguous whether these methods are applicable to outdoor scenes. In reference (9) a model-based approach, where the shape of an object is defined in term of several mini-templates, was adopted to detect the face position. The mini-templates were simply geometric features such as arcs and corners. As it is necessary for this method to extract shapes of edges and lines, the background cannot be complex and the scenes should be with certain constraints. In reference (10) the face space, which is defined by the eigenvectors of a set of faces, was used to detect and identify the facial images. Although it did not necessarily correspond to isolated features such as eyes, ears and noses, it seems that the background would be slightly simple in terms of images shown. In reference (12) the face was localized by coupling a set of local feature detectors with a statistical model of the mutual distances between facial features. However, this algorithm seems to suffer from the same problems mentioned above.

In this paper, we proposed a new method for the detection of full faces in the cluttered color scenes, where the size of faces is comparatively small. Based on the feature parameters of space gray-level dependence matrix (SGLD),[14] a face-texture model composed by a set of inequalities was derived. A face area was defined as such a region where these inequalities hold. The parameters included in the inequalities were decided by experimental basis. Using this face-texture model, we designed a type of scanning scheme for face detection in the color scenes, in which the orange-like parts including the face areas were enhanced by utilizing the *I* component of the YIQ color system.

In order to testify the effectiveness of this method, the experiments were executed and the results were

* Author to whom correspondence should be addressed.

discussed.

2. FACE-TEXTURE MODEL

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 in reference (14). Let $I(i,j)$ be the gray-level value at pixel (i,j) of an image $\in(L_x \times L_y)$, with the brightness values in $[0, L-1]$. Let $P_{ab}(m,n)$ be the number of occurrences in which two neighboring pixels displaced by a vector (m,n) on the image have gray levels a and b, respectively. A matrix composed by $P_{ab}(m,n)$ is called as space gray-level dependence (SGLD) matrix , with a parameter vector (m,n). The formula to obtain $P_{ab}(m,n)$ is defined by:

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

where # denotes the number of elements in the set.

The frequency normalization for the matrix can be computed approximately by the formula:

$$N_{ab}(m,n) = P_{ab}(m,n)/T,$$

where $T = L_x \times L_y$ is the total number of pixels of the image to be analysed.

Based on an SGLD matrix, a set of textural features can be derived. Here, we mainly consider three of them, inertia, inverse difference and correlation features. The equations which define these measures of textural features are:

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

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

$$B_C(m,n) = \frac{1}{\sigma^2} \sum_{a=0}^{L-1} \sum_{b=0}^{L-1} (a-\mu)(b-\mu) N_{ab}(m,n), \quad (3)$$

where L is the number of gray levels, μ and σ are the means and standard deviations of the image. Accordingly, with varying m and n, a set of arrays can be formed, the elements of which are $B_I(m,n)$, $B_D(m,n)$ and $B_C(m,n)$, respectively. Let B_I, B_D and B_C denote these arrays, respectively, then:

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

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

$$B_C = \begin{pmatrix} B_C(0,0) & \cdots & B_C(M,0) \\ B_C(0,N) & \cdots & B_C(M,N) \end{pmatrix} \quad (6)$$

In these arrays, M and N are the upper limits of m and n. In this paper, B_I, B_D and B_C are known as SGLD feature matrices.

2.2. *Face-texture model based on the SGLD feature matrix*

The facial images can be considered as a type 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 almost symmetric with 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 direction paralleling part of the eyes is not larger than that in the direction perpendicular to the part of eyes on the whole, and the difference between the gray-level variation is not so large in the parallel and perpendicular direction of the eyes, because of the special shape of the face, such as the mouth, eyes, facial gradient and so on. Besides, according to the facial shape, the gray levels of neighboring pixels are considered homogeneous in the local region and nonhomogeneous on the whole.

Due to the analysis above, we use the SGLD feature matrices to describe the facial features. The neighbor set of a certain pixel (i,j) is given by

$$N_p = \{(k,l):\ 0 \le k-i \le M \text{ and } 0 \le l-j \le N\},$$

and is represented by a matrix D_p, where the maximum values M and N are selected corresponding to the resolution of facial images. An example illustrated in Fig. 1 uses $M = 2$ and $N = 2$, which are suitable for the facial image shown in Fig. 2, the resolution of which is about 16 × 20 pixels. In this paper, we mainly consider the instance where $M = 2$ and $N = 2$, for the detection of facial images with the low resolution. B_I, B_D and B_C in equations (4), (5) and (6) are calculated on the neighborhood matrix D_p.

In order to simplify the notation, elements of SGLD feature matrices are denoted by the matrix below:

$$D_p = \begin{pmatrix} 1 & 2 & 5 \\ 3 & 4 & 7 \\ 6 & 8 & 9 \end{pmatrix}, \quad (7)$$

where the number of D_p means the order of an element. For example, B_I (3) shows $B_I(0,1)$, and so on.

For inertia matrix B_I, a value of $(a-b)^2$ represents the gray-level variation between two pixels and a value of $\sum_{a=0}^{L-1}\sum_{b=0}^{L-1}(a-b)^2 N_{ab}(m,n)$ can be considered as accumulation of the gray-level variation between all pairs of 2 pixels. According to the facial texture features described above, it is obvious that with the distance between the two pixels in the same direction

(0,0)	(1,0)	(2,0)
(0,1)	(1,1)	(2,1)
(0,2)	(1,2)	(2,2)

Fig. 1. Neighborhood D_p ($M = N = 2$).

$$B_I = \begin{pmatrix} 0.00 & 2.15 & 6.03 \\ 3.38 & 5.10 & 8.51 \\ 10.02 & 11.12 & 13.47 \end{pmatrix}$$

Fig. 2. An example of a facial image and its B_I.

becoming large, B_I becomes great. It is expressed as:

$$B_I(j_1) < B_I(j_2), \tag{8}$$

where $j_1 < j_2$.

Figure 2 shows an example of facial image and the calculated result of B_I. It is shown that the element values of B_I with respect to the neighborhood D_p are increasing all in the direction from the reference point (0, 0).

Furthermore, if the full face appears vertically, $B_I(m, 0)$ represents the gray-level variation in the parallel direction of eyes, while $B_I(0, n)$ represents that in the perpendicular direction of the eyes. Hence, it is established that $B_I(0, n)$ is larger than $B_I(m, 0)$, but it is not much larger than $B_I(m, 0)$ when $m = n$. Although sometimes contrary results are observed because of some disturbance in facial images, the difference of $B_I(m, 0)$ and $B_I(0, n)$ is very small in these cases. If the full face appears obliquely, $B_I(m, 0)$ and $B_I(0, n)$ represent the gray-level variation in the direction inclined to part of eyes, hence values of $B_I(0, n)$ and $B_I(m, 0)$ are close, which is largely determined by the inclined angle to the part of eyes. The larger the angle, the greater the value of B_I.

Generally, $\forall j_1, j_2\,(j_1 < j_2)$, the fact mentioned above is expressed as:

$$B_I(j_1) < (1 + w) B_I(j_2), \tag{9}$$

where w is a small positive constant, which could be determined experimentally.

As an example shown in Fig. 2, $B_I(7)$, which is equal to 8.51, is less than $B_I(8)$, which is equal to 11.12. If the value of w is selected suitably, inequality (9) is expected to serve to discriminate the facial parts from other parts.

Figure 3 shows calculated results of B_D and B_C for the same face as Fig. 2. With the orders increasing, the values of elements in B_D and B_C show the tendency to decrease, because of the homogeneity of faces on the local region and nonhomogeneity on the whole. Moreover, the value of the element in B_D and B_C is only varied in a certain range, because of the similarity among the human faces. Since $B_D(1)$ and $B_C(1)$ are unity, there are positive parameters $d_1(j)$, $d_2(j)$, $c_1(j)$ and $c_2(j)$ less than unity such that:

$$d_1(j) < B_D(j) < d_2(j), \tag{10}$$

$$c_1(j) < B_C(j) < c_2(j), \tag{11}$$

where $1 < j \leq 9$.

$$B_D = \begin{pmatrix} 1.00 & 0.54 & 0.39 \\ 0.48 & 0.42 & 0.34 \\ 0.28 & 0.28 & 0.27 \end{pmatrix}$$

$$B_C = \begin{pmatrix} 1.00 & 0.79 & 0.51 \\ 0.79 & 0.63 & 0.41 \\ 0.51 & 0.39 & 0.24 \end{pmatrix}$$

Fig. 3. B_D and B_C calculated from the face in Fig. 2.

If the values of $d_1(j)$, $d_2(j)$, $c_1(j)$ and $c_2(j)$ are selected suitably, inequalities (10) and (11) are also expected to serve to discriminate the facial parts from other parts.

As a result, the facial region can be estimated as such an area where inequalities (9)–(11) hold.

2.3. *Parameters' determination in the face-texture model*

To construct the face-texture model described above, we have to determine parameters w in equation (9), $d_1(j)$ and $d_2(j)$ in equation (10) and $c_1(j)$ and $c_2(j)$ in equation (11). These parameters are desirable to be selected as small as possible in order to remove non-facial regions, but by doing so, the risk of missing facial parts will become large. In order to make the identification rate high and the false alarm rate low, w, $d_1(j)$, $d_2(j)$, $c_1(j)$ and $c_2(j)$ are determined by experimental procedures described as follows.

Procedure 1

(0) Set the initial values of $w^0 = 0$, $d_1(j)^0 = 1$, $d_2(j)^0 = 0$, $c_1(j)^0 = 1$, $c_2(j)^0 = 0$ and $n = 1$, respectively.

(1) Calculate the B_I, B_D and B_C for a facial image in the test sample set.

(2) **If** the elements of B_I, B_D or B_C meet the inequalities of equations (9), (10) and (11),

then go to next step;

else let $w^n = w^{n-1} + \delta w$, $d_1(j)^n = d_1(j)^{n-1} - \delta d_1$, $d_2(j)^n = d_2(j)^{n-1} + \delta d_2$, $c_1(j)^n = c_1(j)^{n-1} - \delta c_1$ or $c_2(j)^n = c_2(j)^{n-1} + \delta c_2$ and $n = n + 1$.

(3) **If** the values of w, $d_1(j)$, $d_2(j)$, $c_1(j)$ or $c_2(j)$ remain unchangeable for all images,

then end for this unchangeable parameter;

else obtain another facial image, and go to step 1.

Procedure 2

(0) Let $i = 0$ and select the initial values $\delta w = \delta w_0$, $\delta d_1 = \delta d_{10}$, $\delta d_2 = \delta d_{20}$, $\delta c_1 = \delta c_{10}$ and $\delta c_2 = \delta c_{20}$.
(1) Execute Procedure 1 and obtain the parameters w_0, $d_1(j)_0$, $d_2(j)_0$, $c_1(j)_0$ and $c_2(j)_0$.
(2) Let $i = i + 1$ and select the new initial values $\delta w = \delta w_1$, $\delta d_i = \delta d_{1i}$, $\delta d_2 = \delta d_{2i}$, $\delta c_1 = \delta c_{1i}$ or $\delta c_2 = \delta c_{2i}$ ($\delta w_i < \delta w_{i-1}$ and so on).
(3) Execute Procedure 1 again and obtain the new parameters w_i, $d_1(j)_i$, $d_2(j)_i$, $c_1(j)_i$ or $c_2(j)_i$.
(4) **If** $w_i > w_{i-1}$, $d_1(j)_i < d_1(j)_{(i-1)}$, $d_2(j)_i > d_2(j)_{(i-1)}$, $c_1(j)_i < c_1(j)_{(i-1)}$ or $c_2(j)_i > c_2(j)_{(i-1)}$, **then** let $w = w_{i-1}$, $d_1(j) = d_1(j)_{(i-1)}$, $d_2(j) = d_2(j)_{(i-1)}$, $c_1(j) = c_1(j)_{(i-1)}$ or $c_2(j) = c_2(j)_{(i-1)}$, end; **else** goto step 2.

2.4. *Examination for face-texture model*

In order to determine the parameters w, $d_1(j)$, $d_2(j)$, $c_1(j)$ or $c_2(j)$, we used 50 facial images from 20 different full faces including the vertical and inclined faces under the different light condition for experiments. The size of the facial images was about 16 × 20 pixels and as stated above, $M = 2$ and $N = 2$. The experiments testified that the parameters converge to $w = 0.2$. Moreover, for $B_d(2)$ and $B_c(2)$, $d_1(2) = 0.5$, $d_2(2) = 0.8$, $c_1(2) = 0.55$ and $c_2(2) = 0.85$.

Next, we performed an experiment with *ca* 310 test samples to examine the performance of the face-feature model described above. There are full facial images including the vertical and inclined faces, nonfacial images and partially facial images in the sample set. An example of four types of samples is shown in Fig. 4. The size of images was varied from 16 × 20 to 20 × 26. By calculating the SGLD feature matrices B_I, B_D and B_C, the facial images were determined, if B_I, B_D and B_C met the inequalities (9)–(11). The experimental results are shown in Table 1.

As shown in Table 1, the identification rate for the full facial images in the sample set is 100% and the false alarm rate, i.e. the rate of misidentification, for non-

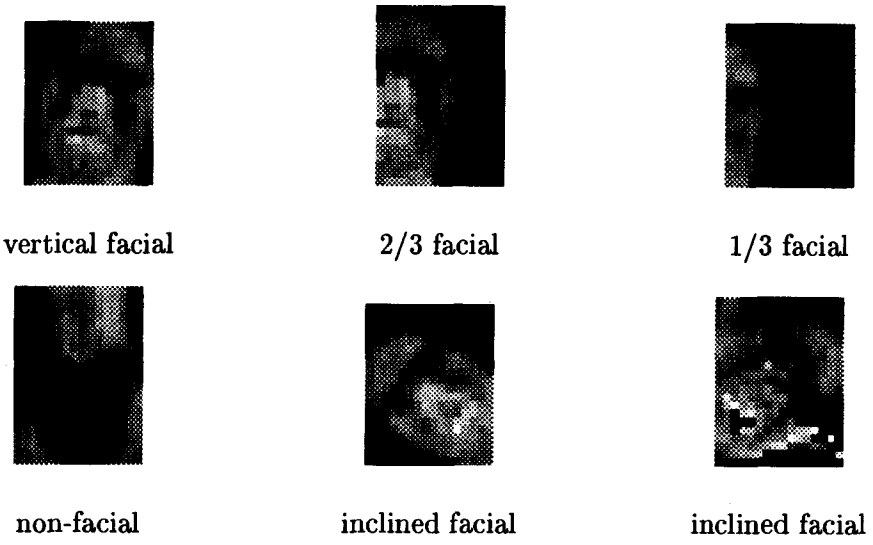


Fig. 4. Variety of test samples used in the experiments of Table 1.

Table 1. Experimental results for face identification in small regions

test samples		rate determined as faces
case	number	
verti. facial img.	35	100%
incli. facial img.	10	100%
2/3 facial img.	35	80%
1/2 facial img.	40	40%
1/3 facial img.	40	6%
nonfacial img.	150	4%

facial images is 4%. For the partially facial images, the more the facial part is included, the higher the identification rate is.

It is important that the rate of faces' missing is 0%, although some false alarms exist, because we can expect that the false faces might be rejected in the recognition stage. In this sense, we can conclude that the facial texture model described above is effective for the face detection and can be used to express the facial characteristics.

3. LOCATION OF FACIAL REGIONS IN THE COLOR SCENES

3.1. *Utilization of color information*

The face-texture model described in the previous section showed good performances, but some false alarms emerge for tough samples. In order to overcome the problem, we develop an approach to utilize the color information.

Most people in East Asia have almost same skin color. Especially, since Japanese (and many people in East Asia) have yellow skins. The feature can be utilized as follows. Skin regions might be enhanced by converting RGB color images to YIQ representation and using the I component, which includes color components from orange to cyan.

As is well known, a conversion formula from RGB to YIQ is given as:

$$\begin{pmatrix} Y \\ I \\ Q \end{pmatrix} = \begin{pmatrix} 0.30 & 0.59 & 0.11 \\ 0.60 & -0.27 & -0.32 \\ 0.21 & -0.52 & 0.31 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}. \quad (12)$$

For the values of the I component, the maximum is *ca* 150 in terms of the conversion formula (12), when values of R, G and B components are varied in the interval [0, 255]. 0 represents gray part in the color images. Hues of I component are changed from orange to cyan. The smaller the value of the I component, the more it contains the component of orange hues and the less it contains that of cyan hues. Especially, the values of orange hues which include a skin color are changed from 1 to 50.

Accordingly, only the orange hues which include the skin color in the images are kept, if the values of the I component in the interval [0, 50] are adopted and those larger than 50 are set to 0. In the following, we term the images the gray level of which is expressed by the value of the I component filtered as above, as I component images.

An example of color images (printed in monochrome) and their I components are shown in Figs 5(a) and (b), respectively. From Fig. 5(a), we observe that the facial part of the I component image corresponds to the original facial image approximately, but is enhanced. The eyes and eyebrows appear as the original shapes in black color, but the mouth appears in white color, because the values of the mouth's I component are larger than those of the eyes, eyebrows and skin. However, if the values of the mouth's I component are larger than the threshold 50, they are set to 0. In this time, the corresponding part of the mouth appears as black in the I component image. Generally speaking, the distribution of light and shade in the facial part has a similar tendency to that in the original image. It is indicated that the face-texture model described in Section 2 is also applicable for the facial parts in the I component images.

From Fig. 5(b), it is clear that the parts of orange hues including the facial part are enhanced by highlights in the I component image and other parts which give other hues are removed. Namely, the nonfacial parts are removed moderately and the facial parts are in good condition for the face extraction in the I component image.

3.2. *Scanning scheme*

The face extraction from the color scenes is executed in the I component images obtained above. In this scheme, the image is scanned by a window for locating the facial position on the basis of the face-texture model. The choice of the window size depends mainly on two factors: (1) how large are the localized faces; (2) whether the model represents the facial features in a window of such size. Since this paper involves the extraction of facial parts with the low resolution in the complex backgrounds, it is desirable to set the size of window as small. From the conclusion on the recognition bound given in reference (13), the window size was set to be 16×20 pixels in advance.

The method of detecting facial areas consists of five steps.

(1) RGB to YIQ conversion

- I component images are calculated.

(2) Noise reduction

- In order to reduce scattered noise in the I component images, a smoothing filter is applied.

(3) Screening

- Regions having intensities in a band are selected.

(4) Extraction of facial regions from I component images.

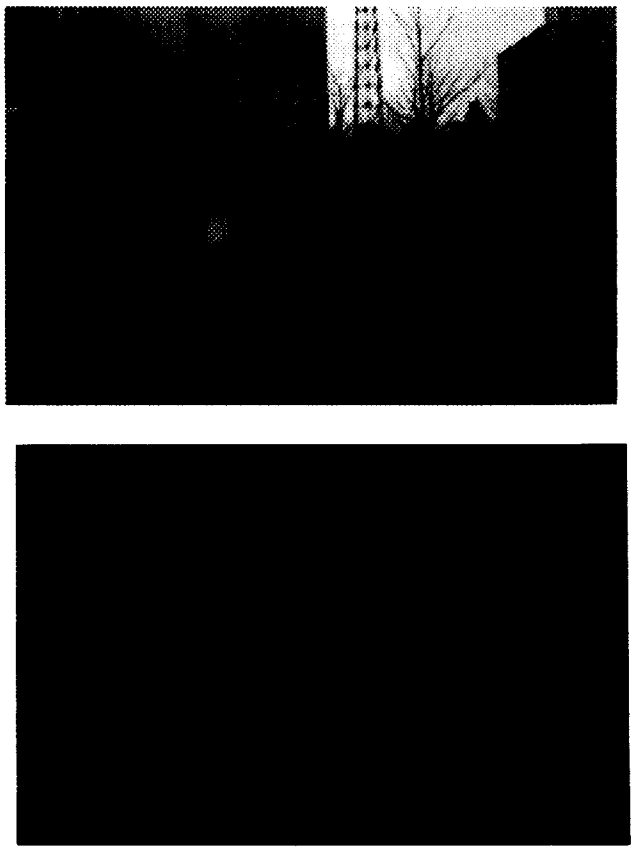
(5) Postprocessing.

The first and second steps will be obvious.

At the third step, a window scans the whole picture and the average intensity in the window is calculated for the purpose of screening. If the average is less than a threshold, the area covered by the window is not considered as the facial region and is rejected immediately. If the value is greater, the fourth step is tested.

At the fourth step, calculating the B_I, B_D and B_C of the images in the scanning window area, it is decided that the facial part exists at the position of the window, if the elements of matrices meet the inequalities (9)–(11), continuing until the whole image is scanned.

(a)



(b)

Fig. 5. Examples of color images and I component images. Note: upper images are color images, but printed in monochrome here.

Although the small step of scanning is desirable to obtain the high rate of detection, the time for scanning a whole image will become long. According to the experimental results shown in Table 1, we found that the rate of identification was 80% for the samples which included 2/3 facial area and only 6% for those which included 1/3 facial area. Thus, the horizontal and vertical step lengths for searching the face position are set as 1/3 of window's width and height, respectively.

The fifth and last step check if the extracted candidate regions overlap. When the overlap is detected, it is necessary to process further to make the overlapped regions sole. The overlap conditions include: (1) the regions are overlapped in the horizontal direction; (2) the regions are overlapped in the vertical direction; (3) the regions are overlapped in horizontal and vertical directions. For the first case, the number of overlapped regions is generally two or three. If two regions are overlapped, such regions merge into one as the facial area; if three regions are overlapped, the middle region is considered as the facial area and other regions are discarded. For the second case, the overlapped regions are processed in the same way. For the third case, the

region which is overlapped most frequently is considered as the facial area and other regions are discarded. Here, it is assumed that more than one faces do not exist closely.

3.3. *Experimental results*

To testify the effectiveness of the algorithm proposed above for face detection, we executed the experiments with two sets of test samples. One was the face database of Olivetti Research Laboratory (ORL) taken via ftp and the other was the scenery images containing faces, which were obtained by scanning pictures into the computer.

The experimental results of face detection are explained below.

Experiment 1

The face database of ORL is used as the test samples to examine the performance of the face-texture model further.

In this experiment, 10 different persons were selected randomly from the face database and a set of five or six test views per person was adopted. The test views included face's rotation, inclination, expression and so on. In order to apply to our face-texture model proposed above, the resolution of the test images was reduced to an interval of $[16 \times 20, 20 \times 26]$. Some examples are shown in Fig. 6.

For these test samples, a face identification experiment was executed using the face-texture model. Although the images in the face database of ORL are monochrome images, the experiment is executed as if they are color (I component) images, because the distribution of the gray level in the monochrome face images has the similar tendency to that in the I component images.

Table 2 summarizes our experimental face identification results. It can be seen that except for an instance of person 24, to whom the face identification failed for one view, all the test samples were identified as face successfully. The overall face identification rate was

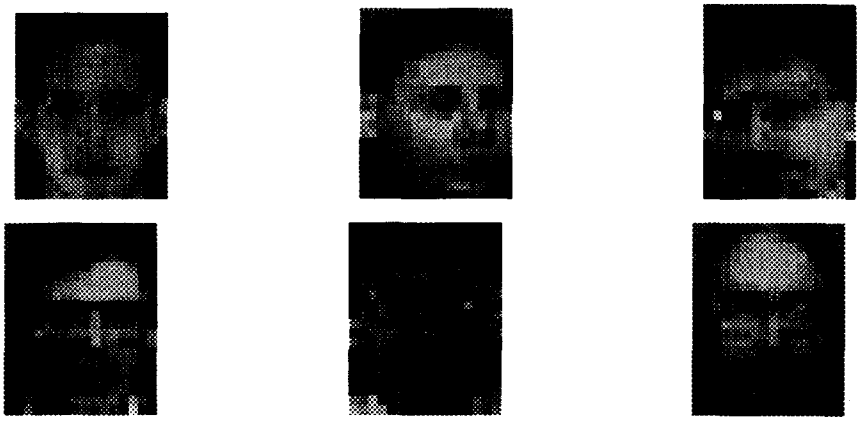


Fig. 6. Examples of ORL face database.

Table 2. Experimental results of face identification using ORL face database

test samples		test samples' condition					number of
ID	number	rotation	inclination	expression	glasses	beard	face identification
1	5	o					5
4	5	o		o	o		5
17	5	o			o	o	5
18	5	o					5
20	5	o		o	o		5
24	5	o		o			4
28	5		o		o	o	5
29	6	o					6
34	6	o	o		o		6
38	5		o	o			5

Table 3. Experimental results of face detection using scenery pictures

test samples			rate of face detection	number of false faces
ID	average size (pixels)	number of regions covered by windows		
1–27	500 × 400	625	100%	0
28	524 × 416	681	100%	3
29	1142 × 264	942	100%	2
30	346 × 388	420	100%	2

98%. It is indicated that the face identification is reliable by using the face-texture model based on the SGLD matrix.

Experiment 2

We executed a face extraction experiment using 30 images containing *ca* 60 faces, which were obtained by scanning the different types of scenery pictures into the computer. These scenes included the indoor scenes such as desks, shelves, wall and so on, and the outdoor scenes such as buildings, trees, grass and so on. The pictures were taken without any special constraints. The position of the facial regions was arbitrary in these test pictures and size was varied from 16 × 20 pixels to 20 × 26 pixels.

Table 3 summarizes our experimental face detection results. As a result, all 60 facial regions in 30 images were extracted from the complex backgrounds correctly, that is, missing rate was 0%. However, there were three pictures in which besides the correctly located faces, false faces appeared. The number of false faces was seven. For the images in which the false alarms happened, the number of regions covered by windows was 2043. Although the rate of which false alarms happened for images was 10% (3 from 30), in the image where the false alarms happened, the average rate of false alarm was 0.34% (7 from 2043), which is very small. On the other hand, such as analysed in Section 2, some false alarms should be tolerated rather than the missed faces, because the false faces might be rejected in the recognition stage. In this sense, the experimental results were satisfactory.

Figures 7 and 8 show some examples of the facial parts extracted from the complex backgrounds done in our experiments. The white rectangles in images represent the face area extracted. Figures 7(a) and (b) are the examples with the outdoor scenes. In Fig. 7(a), the original image of which is shown in Fig. 5(a), there is only one face, and the background is wide relatively. The face was located successfully. In Fig. 7(b) three facial areas were also located successfully. Figure 8 is an example of an indoor scene. In Fig. 8 there are three full faces, including two vertical faces and one inclined face. These facial regions were extracted correctly without any false alarms.

3.4. *Effectiveness of color information*

Although the face-texture model proposed above is also applicable for the face extraction in a black and white image, the utilization of color information ensure face detection is more effective, because the non-facial regions which were needed to detect were reduced in the I component image by means of discarding the nonorange hues. To evaluate the effectiveness of color information the RGB image of Fig. 7(a) was converted to monochromatic by averaging RGB values. Using the pseudomonochrome image the previous method was tested, a condition that the face image should be identified. The result is that there were 14 false alarms detected. Since no false alarm was detected in the I component, the effectiveness of color information is obvious.

CPU time for calculation in the I component image of Fig. 7(a) was 5.5 times smaller than in the pseudomonochrome image. This speedup is considered to be based on the screening. By skipping the regions whose average intensity is too small in the I component images, the burden of calculating SGLD matrices was reduced remarkably.

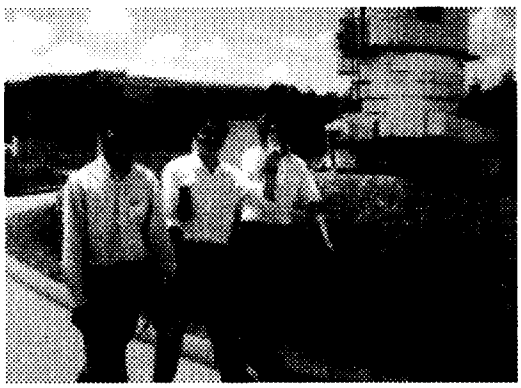
4. CONCLUSION

The full face detection in the color scenes using the face-texture model based on the SGLD matrix was newly proposed and tested. Our method has three contributions. First, the face-texture model based on the SGLD matrix we presented is effective to facial images with low resolution. Second, the utilization of incorporating the color information with the texture feature improves the performance of face detection. Third, the face-texture model is applied to extracting facial parts in the wide, complex backgrounds successfully, where the size of facial areas is relatively small.

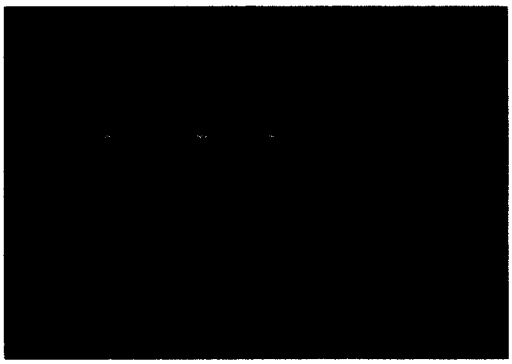
This paper showed that the utilization of the I component color information in face extraction was applicable to the Asian. As a future issue, it is necessary to

Extracted face
Resolution of image is 1142 × 738 pixels

(a)



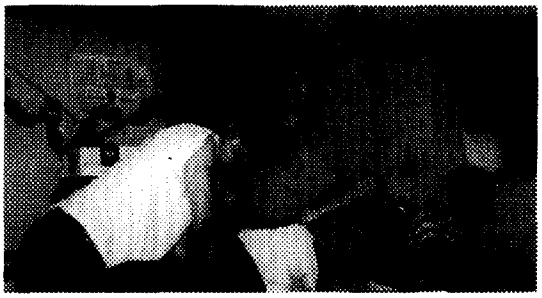
Inputted image (color in original)
Resolution of image is 236 × 164 pixles



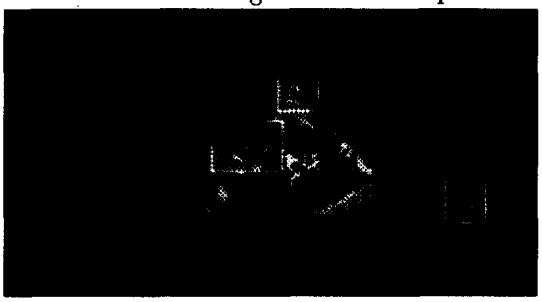
Extracted faces

(b)

Fig. 7. Detected facial regions.



Inputted image (color in original)
Resolution of image is 213 × 113 pixels



Extracted faces

Fig. 8. Detected facial regions.

examine the effectiveness of color information for other races' extraction.

Acknowledgements—This work was partly supported by a grant No. 95-3224 from the Japan Society for the Promotion of Science for Young Scientists. The authors acknowledge the contribution from all members of Nakano Laboratory of the Department of Information Engineering, Shinshu University, in providing with their portraits and helps in experiments. They also acknowledge Olivetti Research Laboratory who supplies the face database via ftp.

REFERENCES

1. L. D. Harmon, Automatic recognition of human face profiles, *Comput. Graphics Image Process.* **6**, 135–156 (1977).
2. T. Kurita, N. Otsu and T. Sato, A face recognition method using higher order local autocorrelation and multivariate analysis, *Proc. 11th Intl. Conf. Pattern Recognition* 213–216 (1992).
3. S. Akamatsu, T. Sasaki and H. Fukamachi, An accurate robust face identification scheme, *Proc. 11th Intl. Conf. Pattern Recognition* 217–220 (1992).
4. K. Matsuno, C. Lee, S. Kimura and S. Tsuji, Automatic recognition of human facial expressions, *Proc. ICCV'95* 352–359 (1995).
5. I. A. Essa and A. P. Pentland, Facial expression using a dynamic model and motion energy, *Proc. ICCV'95* 360–367 (1995).
6. D. Beymer and T. Poggio, Face recognition from one example view, *Proc. ICCV'95* 500–507 (1995).
7. G. Yang and T. S. Huang, Human face detection in a complex background, *Pattern Recognition* **27**(1), 53–63 (1994).
8. M. Kosugi, Human-face search and location in a scene by multi-pyramid architecture for personal identification, *Trans. Inst. EICE* **J77-D-II**(4) (in Japanese) 672–681 (1994).
9. V. Govindaraju, S. N. Srihari, and D. B. Sher, A computational model for face location, *Proc. ICCV'90* 718–721 (1990).
10. M. Turk and A. P. Pentland, Face recognition using eigenface, *Proc. CVPR* 586–591 (1991).
11. Q. Chen, H. Wu and M. Yachida, Face detection by fuzzy pattern matching, *Proc. ICCV'95* 591–596 (1995).
12. T. K. Leung, M. C. Burl and P. Perona, Finding faces in cluttered scenes using random labeled graph matching, *Proc. ICCV'95* 637–644 (1995).
13. Y. Dai, Y. Nakano and H. Miyao, A study of face recognition with low quality images, *Proc. ICARCV'94* 1442–1446 (1994).
14. R. M. Haralick, Textural features for image classification, *IEEE Trans. SMC* **SMC-3**(6), 610–621 (1973).
15. Y. Dai, Extraction of facial image from a complex background using SGLD matrices, *Proc. 12th Intl Conf. Pattern Recognition* **I**, 137–141 (1994).

SUMMARY

It is the key step for face recognition systems to extract the facial parts in the complex backgrounds, if such systems are expected to recognize the human beings automatically. On the other hand, such systems might only give 20 × 30 pixels to a face. Despite the importance of the problem, there are little discussion on the face extraction in the complex backgrounds.

In this paper, we proposed a new method for face extraction in the complex backgrounds using a texture model incorporating with color information utilization. Based on the space gray-level dependence matrix, the facial texture model composed by a set of condition inequalities was derived. The coefficients in the inequalities were decided with experimental procedures. By transforming color images from RGB color representation to YIQ color one, the orange-like parts including the face areas were enhanced in the original images, if the I component of YIQ color system was only used. Using the face-texture model, we designed a kind of scanning scheme for face detection in the I component images. After scanning, some regions, which were considered as face area, but might be overlapped, were obtained. By using a set of relational judgement conditions, the facial parts were determined finally.

The experiments showed that this method can detect the face regions in the color complex backgrounds effectively. Using thirty images containing sixty faces, the facial parts in the images were all extracted correctly, but, false alarms, which indicate unfacial parts were extracted as faces, appeared in three images. Total number of false alarms was 7, which means the false alarm rate was 0.34%, when false alarm rate was defined as the ratio of total number of false alarms to that of all scanned windows in the images. In fact, a little of false alarms should be tolerated rather than the face missing, because the false faces might be rejected in the recognition step. In this sense, the experimental results were satisfactory.

About the Author—YING DAI was born in Xian, China. She received the B.S. and M.S. degrees in information and control from Xian Jiaotong University, China, in 1985 and 1988, respectively. From 1988 to 1991, she was an assistant and then a lecturer in Xian Jiaotong University, China. From 1991 to 1992, she was a lecturer in Huanan Technology University, China. In 1992, she joined Shinshu University as a research student where she is currently a postgraduate student for a doctor degree. Her research interests include computer vision, pattern recognition and image processing.

About the Author— YASUAKI NAKANO received B.E., M.E. and Ph.D. degrees in mathematical engineering from the University of Tokyo in 1961, 1963 and 1975, respectively. From 1963 to 1989, he was with the Central Research Laboratory, Hitachi Ltd, where he carried out research on speech synthesis, speech recognition, character recognition and document understanding. In 1989 he was appointed as a professor of the Faculty of Engineering at Shinshu University. Currently he is responsible for research and education in pattern recognition and artificial intelligence. He wrote many papers in these studies related to the areas stated above and his current concern is pattern recognition, including handwritten character and cursive word, neural networks and natural language processing. Professor Nakano is a senior member of IEEE and belongs to many academic societies in Japan. He has served as the Program Chair at ICDAR'93, the Publicity Co-Chair at ICDAR'95 and the Program Chair at ICDAR'97. He is also an associate editor of the Pattern Recognition Journal.