Person Identification System Based on a Trapezoid Pyramid Architecture of a Gray-Level Image

Makoto Kosugi and Kouji Yamashita

Dept. of Electrical and Electronical Eng. Musashi Inst. of Tech., Tokyo Japan

Abstract. To realize fully automated face recognition, there must be thorough processing from detection of the face in a scene to recognition. There have been many reports on face recognition, however, studies on detection available for recognition are very few. One of the difficulties comes from many variations of input condition such as illumination and background. As for access control systems such as security or login, input conditions can be rather fixed. Under this condition, fully automated person identification by the facial image is tried and achieved. The face in a scene is first sought by coarse-to-fine processing based on a trapezoid pyramid architecture of a gray-level image, and the result is applied to the recognition. The simple algorithm is implemented by software in a personal computer, and this realizes a series of processing within one second.

1 Introduction

The facial image is one of the most important pieces for personal identification. There have been made a large number of studies on face recognition, aiming at the application to a soft security system to replace the fingerprint or the retinal pattern. To realize fully automated face recognition, there must be thorough processing from the detection of the face in a scene to recognition. For the detection, it is necessary to find a face from the scene and locate its precise address. Most of the past studies, however, are concerned only with the recognition processing, and there are only a few that deal with the thorough processing from detection to recognition[1].

Most studies of face recognition were based on edge detection[2]. In those methods, however, the extraction of the edge itself is often made difficult by weak illumination, defocusing and noise. Especially, in the case of the face, there is little difference in gray-level near the nose or mouth, and this makes very difficult to extract precise edges. As a result, a severe condition is imposed on the focusing illumination and background.

The shape corresponds to higher frequency component of the image. Then the author attempted to recognize a face by using the gray-level mosaic pattern which is the lower frequency component of the image[3]. In the study, a mosaic representation of the central part of a face was used to recognize the face. It was shown that more than 100 persons can be identified. R.Brunelli et al. also compared two methods by geometrical features and gray-level template to recognize the face, and showed the method based on the gray-level made good results in spite of simple algorithm[4].

On the search of a face in a scene, some studies have been also reported[5]. Burt et al. attempted to detect a face by the coarse-to-fine technique[6],[7]. In their studies, the major purpose is the detection of the object, and there left a further study concerning to get precise location sufficient for the recognition of the face. Then the author presented a method to detect and locate a face by the continuous multi-resolution images and recognize the face by the above mosaics[8]. The algorithm could find a face with any size at any position in the scene from the nonuniform background.

One of the targets of person identification by the face is a access control system such as security or login. In this case, input conditions such as illumination, back ground and distance between a camera and a face can be fixed. Therefore, higher recognition rate and shorter executing time can be expected, and restrictions on the face direction can be also relaxed.

This paper describes studies on a person identification system for such an access control, and on investigation of the face direction including a non-front view.

2 Search and Location of a Face in a Scene Based on a Trapezoid-Pyramid Architecture

2.1 Outline of the procedure

For the person identification by the facial image, it is most important to detect and precisely locate a face in a scene. Then, multi-pyramid architecture with continuous resolution was introduced to find a face with any size in a scene[8]. This process is composed of three steps; the coarse search to detect a head portion, the fine search to detect the center part of the face within the head portion, and the decision of the location and size of the face. In the first step, the location and size of the head portion can be roughly obtained by template matching of coarse mosaic. The second step obtains rather exact location of the center part of the face also by template matching of smaller mosaic, and the third step obtains precise location and size which is available to apply for face recognition.

In the head search, it was required to locate the head from the background efficiently and without being affected by the noise. For this purpose, the dictionary of template matching for head search should be a mosaic of as large a block size as possible, as long as the features of the face are retained. For this purpose a much lower resolution of 4 (horizontal) $\times$ 5 (vertical) blocks was used as the mosaic pattern. It must also be an average pattern so that it can cope with any type of face. Then six typical patterns were selected by K-means clustering method from 50 faces. Using these templates an input image is scanned, however, the size of the head part is unknown. Consequently, the block sizes of the input image, that is the resolution of the image, must be varied continuously. This process needs tremendous dynamic steps.

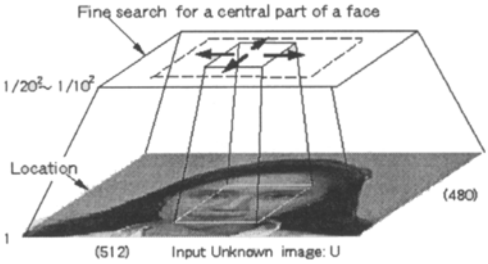


Fig.1. Coarse-to-fine searching strategy for a face by a trapezoid-pyramid architecture.

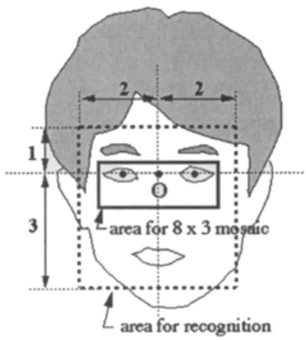


Fig.2. Center of face and area for search and recognition.

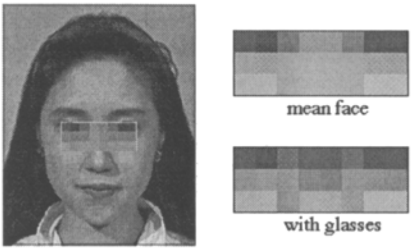
By the way, for the access control system such as security or login, the face can be set almost in the center of a frame by guidance to the user. Especially, if the distance between a camera and a face can be set nearly constant, this can cut the coarse search itself, and cause to much shorten the executing time. In this case, the search process is composed of only two steps, the fine search and the decision of location and size of the face, as shown in Fig.1. This process can be said the face search and location based on a trapezoid-pyramid architecture of gray-level image.

2.2 Face Search for the Central Part of a Face

When the coarse search is executed, the rough location of a head portion is delivered to the fine search. This case, however, starts from the fine search, then the data to be delivered must be given from the input condition, that is, a face is almost set near the center of the frame. Let the center of the face to be the middle point between both eyes O as shown in Fig.2, its x coordinate becomes almost the middle point of the frame width, and the y coordinate also becomes the middle point of the frame height. Now, this point, which is just the center of the frame, is fed to the fine search. For this face search, mosaic template with a finer block size must be used to scan within the head portion.

The mosaic pattern used as the template (dictionary) for the face search must be of a size that is not affected by the hair pattern, and must be an average pattern so that it can fit to any face. From such a viewpoint, the region of the face is defined as the central part of the face as shown in Fig.2. The adequate block size W_b must be determined to represent the features of the eyes and the nose. Smaller block size is suitable to suppress the deviation of the location, however, larger one is better to avoid the computational complexity. From such a viewpoint, the following four kinds of mosaics are compared,(a) Facial part:12 $\times$ 12 blocks, (b) Eyes: 12 $\times$ 1 blocks, (c) Eyes and their vertical neighbors: 16 $\times$ 3 blocks, and (d) Eyes and their lower part: 8 $\times$ 3 blocks.

As a result, 8×3 pattern was selected among four kinds of mosaic patterns as shown in Fig.3. This pattern type contains both eyes and their surroundings. Furthermore, data of this region are extracted from many persons and classified into four kinds of mosaic pattern. Averaged value of each group is used as the templates G_m for the face search as partly shown in Fig.3.



(a) 8 x 3 area (b) template examples

Fig.3. Area and templates for the face search.

Let an input image be U, and the area to detect the head portion be U_a.First the density distribution of the input image is converted by contrast stretching to correct of the density imbalance,if necessary. U_a is represented by the mosaic of block size W_b, and is scanned by the above template G_m to search for the 8×3 area. Namely, let an arbitrary 8×3 mosaic data, among the entire mosaics obtained by dividing U_a by W_b, be C_j. The location of C_j that minimizes the distance between C_j and the template G_m, is defined as the result of search. To absorb the differences of the average gray-level, G_m and C_j are represented by vectors, and its similarity, that is the case when the cosine of the two becomes the maximum, is used to determine the result.

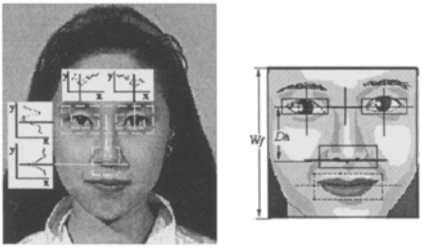
2.3 Precise Location and Determination of Size

The location of the face can be determined by the face search with rather high precision. Still, there can be produced an error corresponding to the size of one block. Consequently, the central part of the face obtained by the face search is projected to the original image level, as is shown in Fig.1, and the highly precise location is determined to be used in the recognition. As an approximate locations of the eyes, nose and mouth are known, a small window is set for each component, as is shown by the dashed line in Fig.4. By examining the gray-level histogram within the window, the precise location of each component is determined.

As to the y coordinate of the eyes, the darkest point is near the center of the eyeball. The gray-level histogram obtained by the image projection to the y axis is examined, and the y coordinates of the eyes (Y_e) is determined from the position of the minimum (the minimum corresponds to black). For each of the eyes, the histogram by the projection to the x axis is examined, and the x coordinate with the minimum value is defined as the center of the eye. The center of the face (X_e) in the horizontal direction is determined as the average of the two.

The next step is to determine the size W_f , which is the central part of the face and is to be used to recognition as shown in Fig.2. For this purpose, the gray-level histogram of the part of the nose or mouth is examined, and the

y coordinate is determined. Using the distance D_h between the eyes and the nose, or the eyes and the mouth, W_f is calculated as being nearly proportional to D_h. The proportional constant is determined as the average of the ratio between W_f and D_h, for several face images. If W_f/D_h differs from the average for some person, the value will result in a slight difference in the size of the facial image, however, it turns out that W_f also gives a cue to the personal identification.



(a) windows and their histogram (b) Wf and Dh

Fig.4. Location of a face by histogram.

3 Search and location of a face with non-front view

For the search and location of the face, an important variation to be considered is the face direction to the camera, when the face is input by TV camera. To fit many kinds of different direction of the face by template matching, there exist two major problems. First, there needs many kinds of template corresponding to the face direction. Second, The size of the face changes according to the angle from the just front direction. At the angle near the front direction, however, the change of face image may be small and there is some possibility of success in search by only the template made from the front view of the face.

Then, allowable direction of the face for the search by only the template generated from the front view of the face is investigated. For this purpose, face images with different direction of every 5 degrees from the front to up/down and right/left are collected. Maximum direction is 20 degrees upward and downward and 30 degrees left and right, respectively. Consequently 117 images per person are collected as shown in Fig.5. The center image is the front view, and the first quadrant has faces with the up and right direction. Totally images of 20 persons are collected and examined.

As a result, face search is found to be able over very large direction, that is up to 25 degrees right and left, and 20 degrees upward and downward, as it is shown in Fig.6. This result is introduced by matching the 8×3 template made from the front view of the face. Let the angle from the front direction to some direction be α, the projection of the image to the front direction, $\cos\alpha$, contributes to the matching. In the up/down cases, $\cos\alpha$ is less than 0.94, and in the right and left cases $\cos\alpha$ is less than 0.91. Accordingly, change of the center part of the face within this region is small enough to lead to the success in the face search. For faces with large directions exceeding the above case, there need templates made from non-font views of faces, however, the number is expected to be not large.

Next, allowable direction of the face for the location by histograms of eyes and nose area is investigated. The region where the the face with non-front view can be located is also shown in Fig.6. The region becomes small compared with the face search, however, very large area is still kept, that is up to 20 degrees right and left, and 15 degrees upward and downward. This is because histograms to get enough information are hold in the area of eyes or nose within the the above face direction.

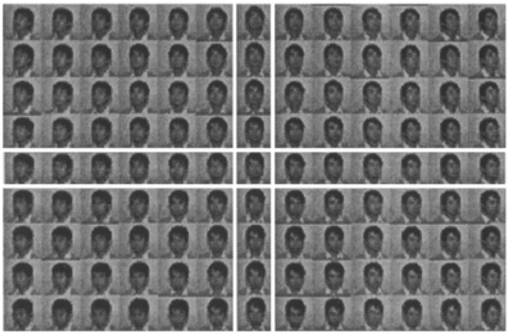


Fig.5. Examples of directional faces (117 direction/person).

4 Recognition of a face with non-front view

The main part of the face obtained by the above search and location is applied to the recognition. It has the width of the face and the vertical length from the eye-brow to the chin. When the facial size containing the main part is measured supposing the midpoint of both eyes O is the center of the face, the size is nearly 2:2 in the hori-zontal direction, and 1:3 in the vertical direction, forming a square, as shown by a dotted line in Fig.2.

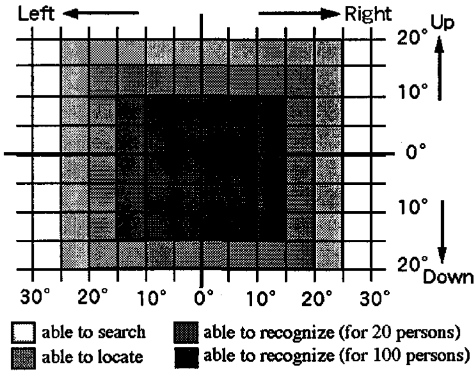


Fig.6. Direction of a face to be detected or recognized.

To represent the main part of the face image by mosaic pattern for recognition, it is divided by adequate block size, and each block is represented by the average value within the block.When the number of mosaic elements is increased, the information concerning the face increases. This makes it possible to represent a large number of faces, while the result becomes more sensitive to the deviation of location. In this study, 12 × 12 blocks are used so that a certain deviation of location can be tolerated, the blocks are then used as the template(dictionary) for recognition[8].

First, the 12 × 12 mosaic templates for the recognition dictionary are constructed after verifying that the location and the size of the face are correctly determined by the previous algorithm. As the dictionary data, only the mosaic from the face of the front view is registered for each person. Next, the main part of the face is extracted from the unknown input image by search and location, and is converted into a 12 × 12 mosaic data. The result is compared with each dictionary data by template matching. Either datum is represented by a 144-dimensional vector, and the case with the smallest angle between the vectors is defined as the result of recognition. This similarity method is used to absorb the gray-level distribution difference, as in the search process.

The recognition rate depends on the number of registered persons, the rate is then investigated by changing the number of persons from 20 to 100. Consequently, in case of the face with a front view, the recognition rate becomes all 100% at any number of person.

Next for the face with a non-front view the recognition rate is examined, and the result is also shown in Fig.6. In the case of the number of registered person is 20, all the persons can be identified over large direction up to 15 degrees right and left, 10 degrees upward and 15 degrees downward. In the case of 100 persons, all the persons can also be identified, however, allowable face direction is reduced to at 10 degrees right and left. From another point of view, this result shows that the person with every direction within 10 degrees difference from the front can be identified even for 100 registered persons by only the templates that is made from the face of the front view. This 10 degree is enough to notice person standing in the non-front direction before the camera.

Furthermore, faces with smile and laugh expression, and with eyeglasses(some with thick frames and some with thin sunglasses), can also be searched and recognized.Fig.7 shows examples succeeded in search and recognition. These results mean that some differences induced by the facial expression or the face directioncan absorbed by using mosaic pattern. Thus necessary information for the face search and recognition is sufficiently kept in spite of changes of facial image,and this produces very robust processing having much tolerance for the input condition. The above algorithm from the search to the recognition of a face is implemented in a personal computer. Fig.8 shows the processing flow of the system, and all the processes are achieved by only program except capturing.A total executing time is within one second by the personal computer with the CPU; Pentium Pro 180 MHz.This executing time is fast enough for a practical use.

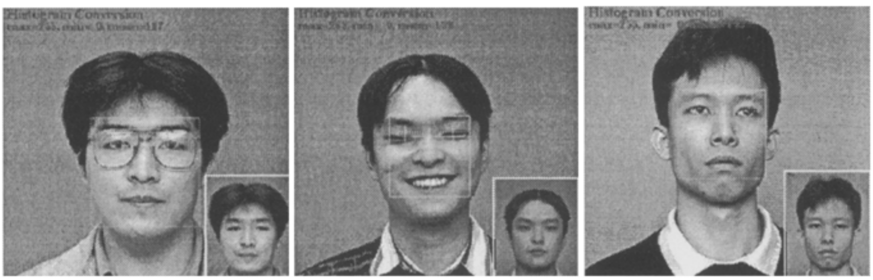


Fig. 7. Resultant examples (right-low: face of front-view for dictionary).

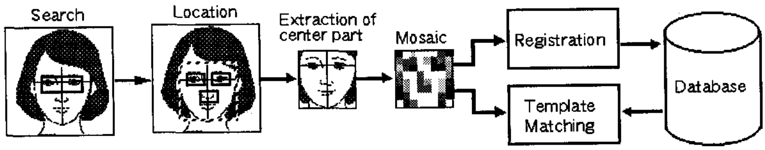


Fig. 8. Processing flow of persons identification by a face.

5 Conclusions

This paper presented an automated method of person identification by the search and recognition of a face in a scene. A face is supposed to be roughly put in a frame, then the face search is executed by matching the template which is a mosaic pattern of the local region containing both eyes. Next, decision of the location and the size of the main part of the face is achieved by histogram of small windows containing eyes and nose. The main part of the face is then represented by a 12 × 12 mosaic and applied to the recognition.

As a result, 100% of the search and recognition rate is obtained for faces with front view. The search and recognition rate for faces with the non-front view is going down as the face direction or the number of registered person becomes larger. Nevertheless, the person with every direction within 10 degrees difference from the front can be identified even for 100 registered persons by only the templates from the face of the front view.

Most of search and recognition processing in the proposed method are are composed of simple template matchings. Then the processing time is within one secondby personal computer without any special hardware, and this is fast enough for a practical use such as access control system.

References

1. M.A.Turk and A.P.Pentland,"Face recognition using eigenfaces",Proc.CVPR'91,pp.586-591(June '91).
2. I.Craw, H.Euis and J.R.Lishman,"Automatic extraction of face-features",Pattern Recognition Letters,5, 2, pp. 183-187 (Feb. '87).
3. M.Kosugi,"Robust identification of human face using mosaic pattern and BPN". In: Neural Networks for Signal Processing II, Proc. of the 1992, IEEE workshop, pp. 296-305 (Sept. '92).
4. R.Brunelli and T.Poggio,"Face recognition: features versus templates",IEEE Trans. Patt. Anal.Machine Intell., 15,10,pp.1042-1052 (Oct. '93).
5. M.S.Lew and N.Huijsmans,"Information theory and face detection",Proc.ICPR'96,pp.601-605(Aug. '96).
6. P.J.Burt,"Smart sensing within a pyramid vision machine",Proc.IEEE,76,8,pp.1006-1015(Aug. '88).
7. P.J.Burt,"Multiresolution techniques for image representation, analysis and 'smart' transmission",SPIE Conf.,Visual Communications and Image Processing IV(Nov.'89).
8. M.Kosugi,"Human-face search and location in a scene by multi-pyramid architecture for personal identification",Systems and Computers in Japan, 26,6, pp.27-38 (June '95).