

Facial Feature Point Extraction Method Based on Combination of Shape Extraction and Pattern Matching

Kazuhiro Fukui and Osamu Yamaguchi

Toshiba Kansai Research Laboratories, Kobe, Japan 658

SUMMARY

In this paper, we propose a method for fast and accurate extraction of feature points such as pupils, nostrils, mouth edges, and the like from dynamic images with the purpose of face recognition. Accuracy of face extraction with these feature points used as criteria greatly affects the capabilities of face recognition methods based on pattern matching. This processing must be performed rapidly, considering the large number of recognition trials required for dynamic images and the requirements of real-time registration. In various extraction methods proposed in the past, steady extraction was difficult due to influences such as individual differences, expression variations, face direction or illumination variations, and so forth. These methods are far from satisfactory in terms of extraction accuracy and processing speed. The proposed method achieves high position accuracy at a low computing cost by combining shape extraction with pattern matching. In particular, it uses a separability filter to extract feature point candidates for pupils, nostrils, mouth edges, and the like. Next, it uses pattern matching based on the subspace method to select the correct feature points from the candidates. Results of testing facial images under various conditions using an evaluation system demonstrated that for 1700 static images the feature point extraction rate was 99%, and in the case of dynamic images the extraction rate for 9880 frames was 98% at a speed of 10 trials/s, without using hardware. © 1998 Scripta Technica, Syst Comp Jpn, 29(6): 49–58, 1998

Key words: Facial feature point extraction; face recognition; subspace method; separability.

1. Introduction

This paper describes a novel method for fast and accurate extraction of facial feature points such as pupils, nostrils, mouth edges, and the like from dynamic images, with the purpose of face recognition. The method is robust to variations of face direction and illumination. Steady extraction of feature points is a difficult task due to variations resulting from the fact that the shape and brightness vary for different individuals, as well as from changes in expression, head movements, illumination, and other influences.

Accuracy of facial feature point extraction also greatly affects the capabilities of face recognition methods which are based on template matching [1, 2, 11–13]. These methods involve cropping facial images after normalizing them in size and position by a two-dimensional affine transform with feature points such as pupils, nostrils, and mouth edges used as criteria, and measuring the pattern similarity degree by comparison against previously registered reference images, in other words, dictionary images. Therefore, the cropping accuracy of the normalized image, and thus the extraction accuracy of the feature points which were used as normalization criteria, is a factor with strong influence on recognition capability.

As for the speed of facial point extraction, it should be fast, considering the large trial number of the face cropping and recognition actions in the case of dynamic

images, and also the practical requirements of taking numerous crops of sample facial images for real-time generation of facial dictionaries. For example, when face recognition of an unsuspecting recognition object is needed (such as for security purposes), recognizable frames are selected for processing from a dynamic image sequence, and therefore, the higher the number of facial crops the better. As for dictionary generation, it is known that in conventional recognition of handwritten characters, dictionaries are generated from hundreds of character sample images; thus, the case of faces, which are three-dimensional objects subject to greater variations of shape and brightness than written characters, would require even more numerous learning sample images [3]. Thus, conventional facial cropping techniques [1, 2], or automatic crops performed under restricted conditions such as full face, do not meet practical requirements.

Known facial feature point extraction methods can be roughly subdivided into “shape information-based” and “pattern information-based” methods.

Typical representatives of the shape information-based methods are edge information-based techniques [9, 16]. For extracting feature points, these methods involve fitting the edges within a region into a structural model by energy minimization. However, problems arise because it is not easy to properly set up the initial parameters, and also because steady extraction is difficult in the cases when sharp edges cannot be obtained due to resolution and illumination conditions.

On the other hand, the pattern information-based methods use brightness and color information instead of edge information. For instance, it has been shown that the eigenfeature method using the eigenspace [12] is more robust to noise, as well as to illumination and shape variations, than the edge-based methods, and thus is effective for real images. However, problems are caused by the fact that it requires multiple calculations of the inner product between images, which entails extensive processing; in addition, the position detection accuracy is lower than in the edge-based techniques. Methods based on color information [4] are more robust to variations of illumination conditions than those based on image intensity information, but color information is often lost in some illumination positions and face directions, so they are not always effective.

As for the processing speed, one method has been proposed [14] for real-time tracking of feature points by simple feature point extraction processing, but it is impossible to know which facial feature point is represented by the tracked point, and therefore these points cannot be used as normalization criteria.

In this paper, we propose a method which combines shape information and pattern information in order to solve the abovementioned problems of conventional

pattern information-based techniques. Its basic approach consists in determining the feature point candidates based on the shape information with high position precision, and verifying them by pattern matching. In particular, a region-based separability filter is used to obtain partial shapes of the feature points corresponding to circles or edge points, and they are verified using the subspace method [10]. This method can be expected to produce high position accuracy, considering that positioning is based on the shape information. In addition, since pattern matching based on the subspace method is used to select the correct feature points from the candidate group, the method is comparatively robust to feature point shape variations and brightness variations.

As for the processing speed, pattern matching is performed only for the candidates narrowed down by the separability filter with low computation cost, which means a substantial reduction in computation volume as compared to the methods employing full pattern matching.

In addition, we shall discuss a method for verifying the correctness of the geometrical relationship between the extracted feature points. Usually, multiple learning patterns which are accumulated within a short time contain extraction errors caused by blinking or similar actions. Therefore, the newly detected feature points are verified for correctness based on the two-dimensional coordinates of correctly extracted facial feature points obtained in the course of several frames.

We shall describe a method of steady extraction of facial feature points such as pupils, mouth edges, nostrils, and others from dynamic images, which is robust to illumination variations and face direction variations. In addition, evaluation results are described for static and dynamic images taken under various conditions.

2. Proposed Feature Point Extraction Method

2.1. Combination of shape information and pattern information

Reference 8 describes an efficient approach based on extracting those feature point parts which are not easily influenced by individual differences, face expression or direction. Usually, these extracted partial features include many erroneously identified features, and extraction therefore involves selecting the correct candidate by applying empirically obtained geometrical constraints [8, 17]. However, using only geometrical constraints is not enough to detect minor position errors that lead, for example, to the eye corners being mistaken for the pupils due to changes in face direction or illumination conditions.

To solve this problem, we have used pattern matching in addition to the geometrical constraints. As shown in Fig. 1, based on information regarding the partial shape of a feature point (for instance, position and diameter of the pupil in the eye), the partial images in the vicinity of the feature point are cropped after being normalized by position and size. If the partial shapes were extracted correctly, then the partial images in the vicinity of the feature point which are normalized and extracted based on these partial shapes (case f3 in Fig. 1) must exhibit high similarity to the dictionary images which have been previously cropped based on the correct shape information (referred to hereafter as d1). Conversely, in the case of erroneous shape information, the extracted images exhibit low similarity with the dictionary image d1 (cases f1, f2, f4 in the figure). In this paper, we shall discuss the case of pupils approximated by circles; it seems that normalization accuracy and pattern matching efficiency are improved if they are approximated precisely by ellipses.

2.2. Shape information extraction using a separability filter

A drawback of conventional edge-based shape extraction methods is their sensitivity to noise. To solve this problem, we have proposed region-based methods for edge [15] and contour [5, 6] extraction. In this paper, a “separability filter” is proposed for extracting feature points such as pupils, nostrils, and mouth edges. According to Fig. 2, let us assume a template mask comprised of two concentric circle regions: region 1 (r_1-r) and region 2 ($r-r_2$) (r defines the mask radius). We shall introduce from the linear discrimination method a “separability” quantity which in-

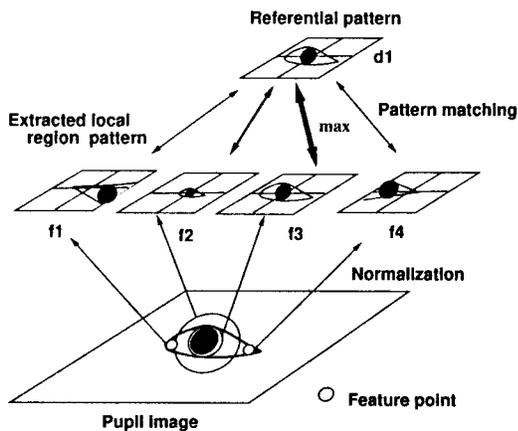


Fig. 1. Combination of shape extraction and pattern matching.

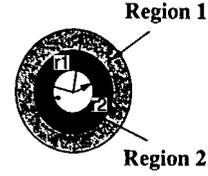


Fig. 2. Separability filter.

dicates the degree of separability of the region information (for instance, image intensity) characterizing regions 1 and 2. When the information of regions 1 and 2 can be completely separated in normalized values, this separability degree reaches its maximum value of 1.0. The separability degree is resistant to noise influence, and can be obtained regardless of brightness differences in the feature points (image intensity differences between regions 1 and 2) [15].

The separability filter outputs the separability measure η which can be obtained from the following equations:

$$\eta = \frac{\sigma_b^2}{\sigma_T^2} \quad (1)$$

$$\sigma_b^2 = n_1(\overline{P_1} - \overline{P_m})^2 + n_2(\overline{P_2} - \overline{P_m})^2 \quad (2)$$

$$\sigma_T^2 = \sum_{i=1}^N (P_i - \overline{P_m})^2 \quad (3)$$

where N is the total number of pixels in the entire region, n_1 is the number of pixels in region 1, n_2 is the number of pixels in region 2, σ_T is the total dispersion in the entire region, P_i is the image intensity level at location i , P_1 is the mean brightness level in region 1, P_2 is the mean image intensity level in region 2, and P_m is the mean intensity level of the entire region. The separability is contained within the range $0 < \eta \leq 1.0$. We have set P_i to the brightness level, but some other characteristic image value can be used, such as hue, color, or texture. The separability for the entire flat region cannot be defined, since the denominator’s dispersion becomes 0. In reality, the separability is set to 0 when the total dispersion within the region σ_T is less than a standard threshold σ_L (which is set to a value of 0 or more, for example, 10).

The basic approach to this separability filter is similar to that of the blob detector in Ref. 16, but in contrast to the blob detector, the calculation formulas are much simpler, and there is no need to use an extra parameter which must be determined experimentally. In addition, the denominator σ_T^2 in Eq. (1) can be simplified to $\sigma_T^2 = \sum_{i=1}^N P_i^2 - (n_1 + n_2)\overline{P_m}^2$. If a square table reference is prepared in advance to be used instead of computing P_i^2 , then the compu-

tation will mostly consist of summing operations, which is fast and convenient in terms of hardware.

Moreover, Ref. 16 mentions only an application to extraction of round pupils. As seen from Fig. 3, the separability filter in addition to the round feature points also outputs local maximum points for the edge points (in the figure the circle center is a local maximum point, while the radius corresponds to the mask radius). Figure 4 demonstrates results of application of the separability filter to a facial image. It can be seen that in addition to pupils, nostrils, and mouth edges, local maximum points are present for the eyebrow edges, lip edges, top of the eye edges, eye corners, and so forth. In this paper, we have used these characteristics to also extract mouth edges.

2.3. Pattern matching

Pattern matching is used to select the correct feature points from the feature point candidates. In Ref. 16 the correct pupil candidates among those detected by the blob detector were identified by fitting the entire frame model using energy minimization. However, the edge presence was considered as a prerequisite, and this technique is by no means robust to illumination deviations and face direction variations. Other drawbacks include the high cost of convergence computations and the dependence of detection on the initial parameters. Conversely, we have used pattern matching based on the subspace method [10] to identify correct feature points. In pattern matching, it is not necessary to extract edges. It is also robust to noise, since full pattern information is used. In addition, using the subspace technique results in a higher tolerance for pattern deformation. We shall briefly describe the subspace method. The image data are treated as $n \times n$ dimensional vectors. The subspaces which correspond to the feature point vector (pattern) in the $n \times n$ dimensional feature space can be expressed as linear subspaces spanned by multiple eigenvectors. The projection angle θ of an input vector (pattern)

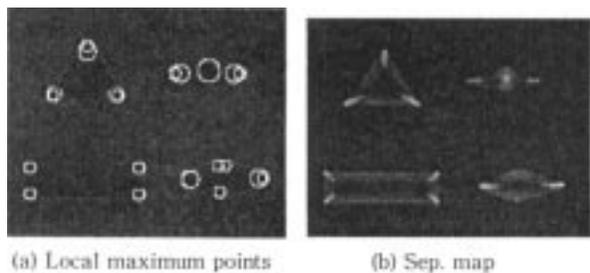


Fig. 3. Output examples: (a) local maximum points, (b) separability distribution (output by 255 times).

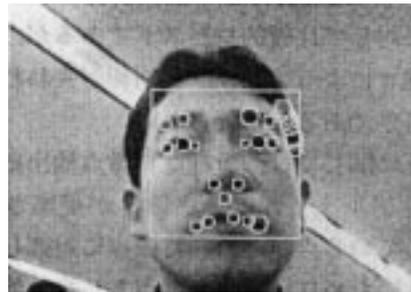


Fig. 4. Local maximum points of separability.

projected onto the subspace represents the extent to which the input vector is analogous to the feature point vector. For identification, the values of θ between the input vector and each feature point's subspace are obtained, and the input vector is that for the feature point which falls into the subspace with the highest θ . In particular, $\cos\theta$ of the projection component can be obtained from the following equation:

$$\cos^2 \theta = \frac{\sum_{i=1}^L (Vect, Eigenvect_i)^2}{|Vect|^2} \quad (4)$$

where L is the number of eigenvectors which form a subspace and $(Vect, Eigenvect_i)$ is the internal product of the input vector and the i -th eigenvector.

In order to obtain the eigenvectors forming the subspace for each feature point (dictionary image d1), one collects learning patterns of the left and right pupils, nostrils, and left and right mouth edges from different people under a variety of illumination and face direction conditions. The learning patterns are cropped as square images of a certain size based on the radius of the separability filter at the local maximum points of separability. Some learning patterns cropped in the vicinity of pupils and mouth edges are shown in Figs. 5 and 6. Principal component analysis of these learning patterns was used to obtain the eigenvectors (dictionary images d1). However, using only d1 is not enough to give high-reliability pattern matching. For instance, in the case of pupils, a high similarity is also obtained for the eye corners, top of the eye edges, brow edges, and other feature points, as seen in Fig. 7. This results



Fig. 5. Examples of left pupil image.



Fig. 6. Examples of left mouth edge image.

from the fact that the subspace method is based on the similarity, and does not consider relations to other categories. Therefore, we have generated a “negative dictionary” (referred to hereafter as nd1) by identifying and collecting from the feature point candidates erroneous patterns of eye corners, top of the eye edges, brow edges, and so forth characterized by an above-threshold similarity with the dictionary d1. The final similarity is obtained by subtracting the negative similarity from the previously obtained similarity. In this way, the extraction rate was improved by about 3%.

2.4. Extraction location identification

Assuming a perspective projection model, and abandoning the three-dimensional model, the two-dimensional coordinates of the facial feature points viewed from a random direction can be expressed as a linear combination of two-dimensional coordinates of the corresponding feature points in four images viewed from different directions, in accordance with Eqs. (5) and (6) [18]. Using this feature, one can verify the two-dimensional coordinates (\bar{x}_i, \bar{y}_i) of a newly detected feature point by using the two-dimensional coordinates (x_{if}, y_{if}) of the correct feature point i in the four frames ($f = 1-4$) already extracted from the dynamic image sequence. This relationship holds for the feature points i of nostrils, pupils, and the like. Using the least-squares method, one can obtain the linear combinatorial coefficients a_i, b_i ($i = 1-4$) from the new coordinates (\bar{x}_i, \bar{y}_i) and the two-dimensional coordinates of the lowest four feature points i for the four-frame images. The next step is computing the approximate coordinates (x'_i, y'_i) from the obtained combinatorial coefficients, and calculating the error Res with respect to (\bar{x}_i, \bar{y}_i) :



Fig. 7. Examples of images from nonpupil region: patterns similar to a pupil image among the feature point candidates detected by the separability filter.

$$\bar{x}_i = a_1x_{i1} + a_2x_{i2} + a_3x_{i3} + a_4x_{i4} \quad (5)$$

$$\bar{y}_i = b_1y_{i2} + b_2y_{i2} + b_3y_{i3} + b_4y_{i4} \quad (6)$$

$$Res = \sum_{i=1}^4 \sqrt{(\bar{x}_i - x'_i)^2 + (\bar{y}_i - y'_i)^2} \quad (7)$$

where a case in which the error Res is greater than some threshold value indicates that an error has occurred at the extraction location.

3. Processing Flow

In our method, it is possible to independently extract pupils, nostrils, and mouth edges; however, we shall explain the processing flow of Fig. 8 assuming that all feature points are observed at once.

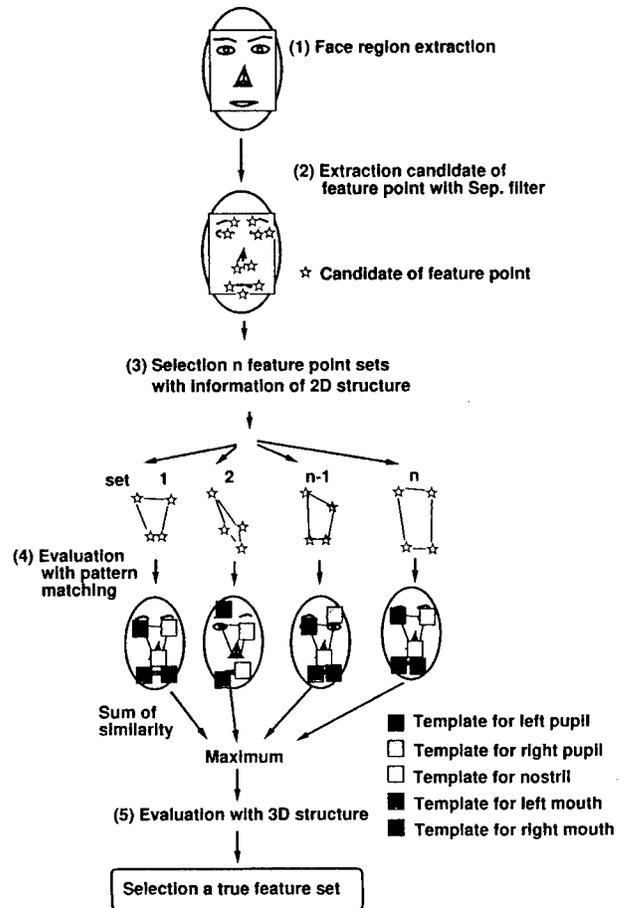


Fig. 8. Processing flow of feature point extraction.

(1) Facial region extraction by subspace method

For the facial region extraction, the subspace method is used [2]. The face dictionary patterns of 30×30 pixel size containing nostrils and pupils (referred to hereafter as d2) are moved over the entire image to obtain similarity values, and the local maximum points of the similarity are extracted as facial regions. This is performed by applying three-step scale variation to the input images.

(2) Extraction of feature point candidates

To extract the mouth edges, the facial feature point candidates are extracted by applying the separability filter to the lower section of the extracted facial region magnified by 30%. The application range of the separability filter is restricted to the dark regions by applying the p-tile method (with the scale set to 25%) to the entire facial region. By applying a variable separability filter radius r of 3 to 6 for all images in the restricted regions, a maximum output value is obtained, which is used as the output value for this image, and the mask radius r is set as the feature point radius. On smoothing by Gaussian processing ($\sigma = 1.0$), the local maximum points are obtained and considered as feature point candidates.

(3) Narrowing down by space location of feature points

To reduce the computing cost of pattern matching, the feature point sets for pupils, nostrils, and mouth edges are narrowed down by applying roughly defined restrictions to the entire facial region. Based on the facial region width L_0 , lengths c_1L_0 , c_2L_0 , c_3L_0 between the pupils, nostrils, and mouth edges are restricted; c_1 , c_2 , c_3 are empirically determined coefficients.

(4) Verification based on pattern information

The total similarity for the feature points is obtained by performing pattern matching for the feature point set. That feature point set which has the highest similarity total is considered the correct set.

(5) Sorting of learning images

Images with an error Res as obtained from Eq. (7) greater than some threshold value are excluded from the sample images.

4. Evaluation Results and Discussion

4.1. Evaluation system

The evaluation system shown in Fig. 9 was comprised of an SGI Indy workstation (CPU R5000, 180 MHz) and a color CCD camera (Toshiba IK-C40). The CCD camera

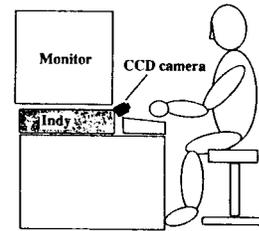


Fig. 9. Testing system (for acquisition of data 1, 3, 4).

equipped with a lens with a focal length of 6.5 mm was mounted at the bottom of the monitor. The camera input images were digitized to fit into a size of 320×240 pixels, using an image capturing board in the Indy, and a series of processing operations, including facial region extraction, facial feature point extraction, and facial image cropping, were performed by a software program.

Four types of images taken of a single person were prepared as evaluation images. Data 1 were obtained from static images of the face photographed according to computer commands while facing in various left–right and up–down directions relative to the display (17 directions in all). $17 \times 50 \times 4 = 3400$ images were collected by taking two pictures each of 50 persons under various illumination conditions in a lodging place and in the laboratory during different times of the day. Data 2 were 124 static images collected with changed camera location and illumination conditions as compared to those for data 1. Data 3 are dynamic images taken by changing the facial direction while acquiring data 1. Data 4 are dynamic images of the face taken during computer operation in the natural state.

4.2. Dictionary generation

Dictionary image d2 for face region extraction was generated using the facial region images cropped from 1700 images of data 1 (first). Dictionaries d1 and nd1 for extracting pupils, nostrils, and mouth edges were also generated based on 1700 images of data 1 by cutting out approximate images of the feature points with the aid of the separability filter (15×15 pixels). Examples in Fig. 10 show 10 eigenimages taken sequentially from the higher eigenvalue direction for the right pupil (a) and a region not containing the pupil (b). The number of eigenvectors used in each dictionary was set empirically to 10.

4.3. Extraction results for static images (data 1, 2)

Evaluation was performed for 1700 images of nonlearning data from data 1. Facial region extraction

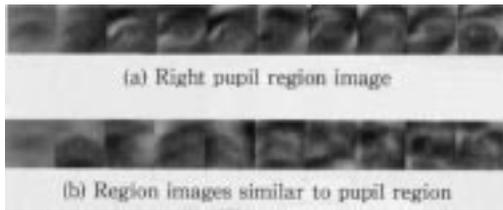


Fig. 10. Dictionary images for pupil extraction.

was considered successful in all cases when the extracted images included pupils and nostrils. Among 1700 cases, extraction failed in 8 cases, including profile pictures, smiling face, face with tilted head, and similar postures.

For the 1692 cases of successful facial region extraction, extraction of pupils, nostrils, and mouth edges was performed. Figure 11(a) shows some extraction results. Pupils, nostrils, and mouth edges in the figure are marked with “+.” Extraction was considered successful if the visually extracted feature points were located within 1 pixel of the correct feature points. The pupil extraction failed in 3 cases, yielding an extraction rate of 99.8%; the extraction rate for the nostrils was 100.0%; the mouth edge extraction failed in 14 cases, giving an extraction rate of 99.1%.

The reason for pupil extraction failure was the presence of highlight regions within the pupil region caused by reflection from eyeglasses. This problem is difficult to solve using only partial images, and in addition it requires using an entire face model. As for the mouth edge extraction, it failed, first, because of a higher brightness due to the lip surface reflection, so that the mouth edges were not included to the separability filter application region as defined by the p-tile method. Another reason was that, due to a substantially deformed pattern near the mouth edges because of smiling, the degree of similarity with the dictionary decreased. For the mouth edges, extraction might be improved by modifying the scale in the p-tile method or by enlarging the learning patterns.

However, facial croppings are possible if, among the six points pertaining to the pupils, nostrils, and mouth edges, any four points of two sets are extracted. From this point of view we had only one failed case of facial cropping, which can be explained by the fact that the camera position was restricted to the bottom of the display. This sufficiently confirms the efficiency of the proposed method.

Figure 11 demonstrates partial extraction results for data 2. Facial region extraction partial failed in 7 of 124 cases. Pupil extraction failed in 2 cases, and mouth edge extraction in 3 cases. The failure reasons are exactly the same as for data 1. Good results were obtained despite greatly varied illumination conditions and camera positions, such as back

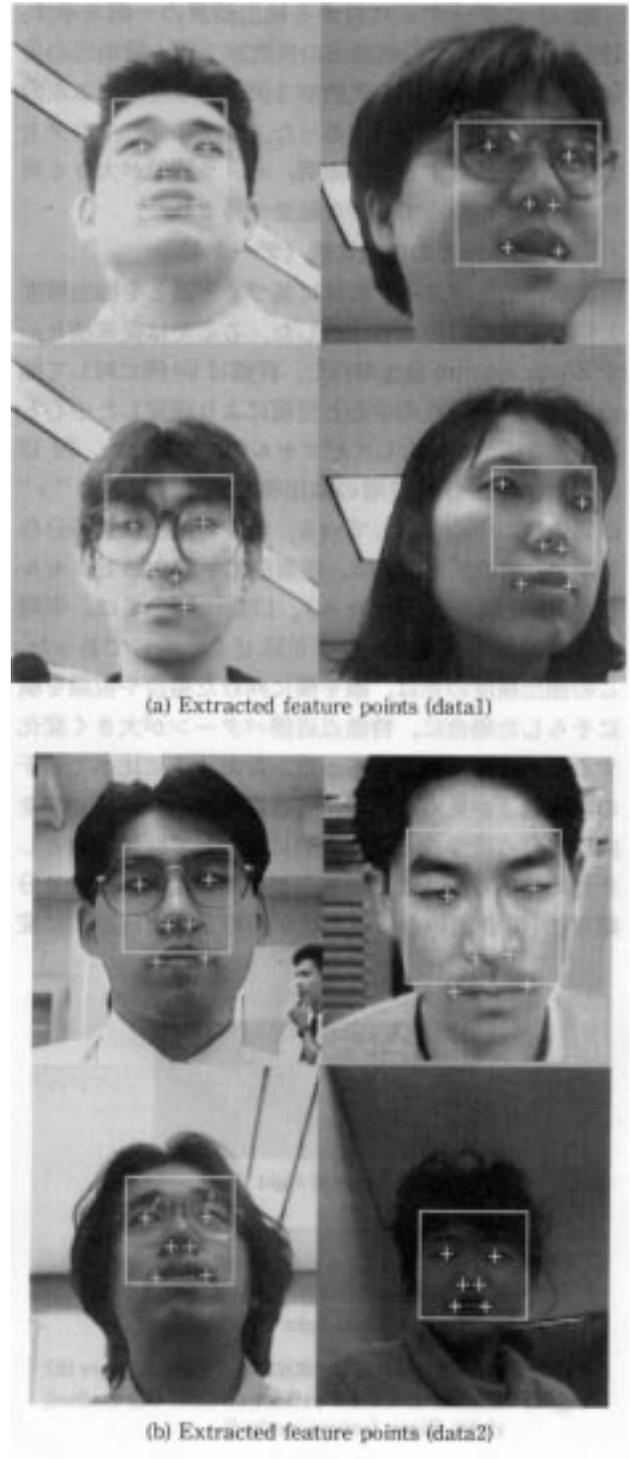


Fig. 11. Results of feature point extraction.

lighting, camera installed on the upper side of the monitor, and so forth.

4.4. Comparison of performance with conventional methods (data 1)

The proposed method and a method based on pattern information were compared in terms of both extraction accuracy and processing speed. The conventional method was the eigenfeature method. For evaluation, we examined magnified images for 60 cases to measure the errors (in pixels) between the center of each extracted feature point and the visually determined center. Figure 12 demonstrates extraction results for the right pupil and right mouth edge. The feature point centers are marked with “+.” The pupil extraction error is an average of 0.15 pixel for the proposed method (0.77 pixel for the conventional method). As for the mouth edge, it was 0.16 pixel on average (0.33 pixel for the conventional method). These extraction accuracy differences were especially conspicuous in the case of profile views and side glances, due to substantial changes of the pattern near the feature points. Accuracy was to some extent reduced as compared to that of section 4.3, because we had intentionally selected images of faces turned sideways, as evaluation objects for comparison. It is evident that the combinatorial separability filter [6] proposed by the authors allows extraction with a high position accuracy even with the pupil shape deformed from a circle to an ellipse.

As for the processing speed, it was 0.1 s with the proposed method and 1.7 s with the eigenfeature method for the same search region, constituting a substantial computing cost reduction.

4.5. Extraction results for dynamic images (data 3, 4)

We performed tests for extraction of pupils and nostrils based on dynamic image data 3. Considering that evaluation of dynamic image data is associated with a

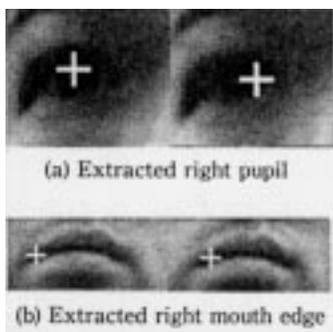


Fig. 12. Examples of feature extraction results (left: proposed method; right: eigenfeature method).

Table 1. Comparative results showing extraction errors: (a) Right pupil and (b) right mouth edge (60 samples)

Error (pixel)	0	1	2	3	4	5	6
(a) Our method	76%	21%	-	3%	-	-	-
(a) Eigenfeature	31%	26%	24%	12%	-	3%	2%
(b) Our method	88%	6%	6%	-	-	-	-
(b) Eigenfeature	64%	23%	11%	-	-	-	-

greatly increased number of object frames, visual evaluation is difficult. Therefore, we estimated whether the feature points have been extracted correctly by examining variations of the error Res obtained from Eq. (7). Figure 13 shows a partial transition diagram of Res : under common circumstances the extraction yields small errors, but when the pupil could not be extracted because of blinking, or owing to a similar problem which leads to an erroneous extraction of a brow, there was a large error, as at point A. Based on this diagram, the frames with an error Res greater than some threshold value (set empirically to 80) were visually estimated on accumulating processing results. Among 9880 frames processed with a speed of about 10 frames/s, facial region extraction was successful for 9786 frames (extraction rate 99.0%), and among these 9786 frames, all four feature points were successfully extracted for 9675 frames, an extraction rate of 98.8%.

In addition, we determined how much time passed from the face cropping until the completion of the dictionary generation, for data 4. The mean processing time measured for five runs from learning pattern cropping (15×15 pixels) until the completion of dictionary generation was 45 s. Figure 14 illustrates the extraction results obtained at 0.25-s intervals. This processing time demonstrated that dictionary generation and face recognition can be imple-

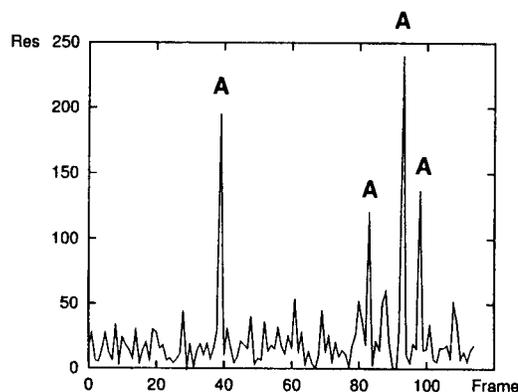


Fig. 13. Variation of error Res .

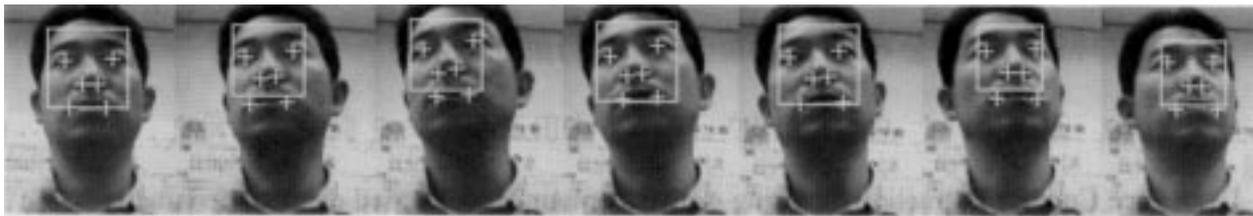


Fig. 14. Results of feature extraction from sequence images (frames 1 to 7).

mented in real time. However, the collected learning patterns contain some deviations because their objects are nonintentional movements. How an accurate dictionary can be generated from fewer learning patterns is a problem yet to be solved. It might be approached by using methods based on generating facial images in randomly selected viewing directions from learning facial images obtained in multiple viewing directions.

5. Conclusions

We have proposed a method for fast and accurate extraction of facial regions from images and extraction of facial feature points such as pupils, nostrils, and mouth edges. In the proposed method, high position accuracy and low computing cost are realized by using the subspace method and shape information obtained with the aid of a separability filter. In the case of static images obtained by a camera positioned at the bottom of the computer display, the rate of extraction at least four feature points at which it was possible to obtain croppings of normalized facial images was as high as 98.8%. Application to recognition of persons from dynamic images is planned for future research.

REFERENCES

1. S. Akamatsu, T. Sasaki, H. Fukamachi, and Y. Suenaga. Robust face identification by pattern matching based on KL expansion of the Fourier spectrum. *Trans. I.E.I.C.E. (D-II)*, **J76**, No. 7, pp. 1363–1373 (July 1993) (in Japanese).
2. Y. Komatsu and Y. Ariki. Orientation invariant face extraction and recognition based on subspace method. *Tech. Rep. I.E.I.C.E.*, **PRU95-292** (1996) (in Japanese).
3. Y. Murase and S.K. Nayar. Learning recognition of 3D object from appearance: Parametric eigenspace representation. *Trans. I.E.I.C.E. (D-II)*, **J77**, No. 11, pp. 2179–2187 (Nov. 1994). (in Japanese)
4. T. Sasaki, S. Akamatsu, and Y. Suenaga. Face image normalization based on color information. *Tech. Rep. I.E.I.C.E.*, **IE91-2**, pp. 9–15 (1991). (in Japanese)
5. K. Fukui. Contour extraction method based on separability of image features. *Trans. I.E.I.C.E. (D-II)*, **J80**, No. 6, pp. 1406–1414 (June 1997). (in Japanese)
6. O. Yamaguchi and K. Fukui. Face image analysis based on separability feature: Eye and pupil detection. *Proc. Semiannual Meeting of Information Processing Society of Japan*, **2**, pp. 187–188 (1996). (in Japanese)
7. Y. Mukaigawa, Y. Nakamura, and Y. Ohta. Synthesis of arbitrarily oriented facial views from two images. *Trans. Inf. Process. Soc. Japan*, **37**, No. 4, pp. 635–644 (April 1996). (in Japanese)
8. X. Song, C.-W. Lee, G. Xu, and S. Tsuji. Extraction of facial organ features using partial feature template and global constraints. *Trans. I.E.I.C.E. (D-II)*, **J77**, No. 8, pp. 1601–1609 (Aug. 1994). (in Japanese)
9. S. Sakamoto, Y. Miyao, and J. Tajima. Extracting feature points on human eye photographs. *Trans. I.E.I.C.E. (D-II)*, **J76**, No. 8, pp. 1796–1804 (Aug. 1993). (in Japanese)
10. E. Oja. *Subspace Methods of Pattern Recognition*. Research Studies Press (1983).
11. M. Turk and A.P. Pentland. Face recognition using eigenfaces. *Proc. CVPR 11*, pp. 453–458 (1993).
12. A. Pentland, B. Moghaddam, and T. Starner. View-based and modular eigenspaces for face recognition. *CVPR '94*, pp. 84–91.
13. R. Brunelli and T. Poggio. Face recognition: Features versus templates. *IEEE PAMI*, **15**, No. 10, pp. 1042–1052 (1993).
14. A. Gee and R. Cipolla. Adaptive visual tracking by temporal consensus. *ACCV '95*, Singapore, I, pp. 44–48.

15. K. Fukui. Edge extraction method based on separability of image features. *IEICE Trans Inf. Syst.*, **E78-D**, No. 12 (1995).
16. A.L. Yuille, P.W. Hallinan, and D.S. Cohen. Feature extraction from faces using deformable templates. *IJCV*, **8**, No. 2, pp. 99–111 (1992).
17. H.P. Graf, T. Chen, E. Petajan, and E. Cosatto. Locating faces and facial parts. *Int. Workshop on Automatic Face- and Gesture-Recognition*, pp. 41–46 (1995).
18. S. Ullman and R. Basri. Recognition by linear combinations of models. *IEEE Trans. PAMI*, **13**, No. 10, pp. 992–1006 (1991).

AUTHORS (from left to right)



Kazuhiro Fukui. Education: University of Kyushu, B.S., 1986, M.S., 1988. Industry positions: Toshiba Corporation, 1988 (Kansai Research Center). Research interests: computer vision, human interface.

Osamu Yamaguchi. Education: Okayama University, B.S., 1992, M.S., 1994. Industry positions: Toshiba Corporation, 1988 (Kansai Research Center). Research interests: image recognition, human interface.