Automatic Contour Extraction of Facial Organs for Frontal Facial Images with Various Facial Expressions*

Hiroshi KOBAYASHI**, Seiji SUZUKI***,
Hisanori TAKAHASHI†, Akira TANGE††
and Kohki KIKUCHI†††

This study deals with a method to realize automatic contour extraction of facial features such as eyebrows, eyes and mouth for the time-wise frontal face with various facial expressions. Because Snakes which is one of the most famous methods used to extract contours, has several disadvantages, we propose a new method to overcome these issues. We define the elastic contour model in order to hold the contour shape and then determine the elastic energy acquired by the amount of modification of the elastic contour model. Also we utilize the image energy obtained by brightness differences of the control points on the elastic contour model. Applying the dynamic programming method, we determine the contour position where the total value of the elastic energy and the image energy becomes minimum. Employing 1/30 s time-wise facial frontal images changing from neutral to one of six typical facial expressions obtained from 20 subjects, we have estimated our method and find it enables high accuracy automatic contour extraction of facial features.

Key Words: Image Processing, Recognition, Biomechanics, Human Engineering, Human Interface

1. Introduction

The importance of the face and facial expressions for communication makes machine understanding of them a crucial issue in the field of computer vision. Consequently, recently, many research works have been performed in an attempt to realize automatic extraction of facial features such as eyes, eyebrows

and mouth, and automatic recognition of facial expressions⁽¹⁾⁻⁽¹⁵⁾. With regards to automatic extraction of facial features, other than the report by some of the current authors (Ref.(7)), there are no research works which realize automatic contour extraction of facial features without any makeup. Most of the developed methods need human assistance mainly to make the initial templates for facial features and/or to acquire the initial positions of facial features. In the case of Ref.(7), although we realized automatic contour extraction of facial features and the face, the method can be applied only for the neutral face and, if the shape of facial features changes greatly, it might be impossible to extract contours.

Recognition of facial expressions depends on what kind of information is obtained from facial image processing. Because of the difficulty in automatic contour extraction of facial features, several methods are proposed. In Refs.(11), (12) and (15), the use of optical flow data in square-areas of facial features is reported. In the case of Refs.(11) and (12), the initial square-areas of facial features are obtained

* Received 25th June, 2001. Japanese original: Trans. Jpn. Soc. Mech. Eng., Vol. 66, No. 648, C (2000), pp. 2778-2783 (Received 8th December, 1999)

** Dept. of Mechanical Engineering, Science University of Tokyo, 1-3 Kagurazaka, Shinjuku, Tokyo 162-8601, Japan. E-mail: hiroshi@kobalab.com

*** Mitsubishi Heavy Industries, Ltd., 2-5-1 Marunouchi, Chiyoda-ku, Tokyo 100-0005, Japan

† Graduate School Department of Mechanical Engineering, Science University of Tokyo, 1-3 Kagurazaka, Shinjuku, Tokyo 162-8601, Japan

†† SONY Corporation, 6-7-35 Kitashinagawa, Shinagawa, Tokyo 141-0001, Japan

††† Department of Precision Engineering, Chiba Institute of Technology, 2-17-1 Tsudanuma, Narashino, Chiba 275-0016, Japan

manually and then these areas are tracked by optical flow computation. By using the modification value of these areas, correct recognition rate of 88% with respect to 105 facial images of six typical facial expressions for 30 subjects are obtained. Essa and Pentland^{(8),(9)} use the physical model of facial muscle and facial skin, and by applying optical flow computation, the physical model is modified in order to fit in facial images with various facial expressions using the FEM method. The motion energy is defined by the modification level of the physical model and is applied to recognize facial expressions by comparison with the values in the database. They achieved the recognition rate of 98% with respect to four facial expressions of eight subjects though, fitting to the face model is undertaken manually.

We have been using supervised neural networks for recognizing facial expressions. We now describe our previous works briefly. In terms of static facial images, we realized the machine recognition of six typical facial expressions⁽¹⁾ (surprise, fear, disgust, anger, happiness and sadness)⁽¹⁶⁾, the strength recognition of six typical facial expressions⁽²⁾ and the recognition of mixed facial expressions⁽³⁾. In case of dynamic facial image which display changes in facial expressions through time, we clarified the difference between static and dynamic recognition of facial expressions in our human visual perception, and then studied the performance and characteristics of machine recognition of dynamic facial expression by applying a recurrent neural network^{(4),(5)}. Although, in these cases, we have achieved the correct recognition rate of about 90% for six typical facial expressions for 30 subjects using supervised neural networks, due to the difficulty in automatic extraction of facial features such as the contour of eyes, eyebrows and mouth, we utilized the x - y coordination of 30 facial characteristic points (FCPs) manually inputted.

To overcome this issue, we employ the brightness distribution around the eyes, eyebrows and mouth. In the first place, using the brightness distribution along a vertical line crossing the pupil and eyebrow, and by calculating cross-correlation, we extract the center positions of right and left pupils. As the facial information for recognizing facial expressions, we employ the brightness distribution along the 13 vertical lines crossing the eyes, eyebrows and mouth obtained based on the positions of right and left pupils. We acquire the facial information for six typical facial expressions from 30 subjects, facial information for 10 and 15 randomly chosen subjects is used for training the neural network. The recognition test is done by using the facial information not used for the neural network learning. We achieved a high correct recognition rate

of about 85% in terms of six typical facial expressions in real time (55 ms)⁽⁶⁾.

Thus, we manage to accomplish a high correct recognition rate of facial expressions in real-time using many subjects. However, in order to recognize subtle movement, we need more precise information in terms of the contours of facial features. Therefore we decided to develop the method to extract contours of facial features automatically and succeeded in realizing it for the neutral face⁽⁷⁾. Following this result, in this paper, we show the method to accomplish automatic contour extraction of facial features not only for the neutral face but for any kind of facial expressions.

In our method, we employ the frontal face in which we can see right and left eyes and apply the image of the neutral face as the initial facial image. Using our previous method⁽⁷⁾, the contour of facial features is extracted from the neutral face. We then track it with respect to the changes in facial images with time by applying the dynamic programming method⁽¹⁸⁾ based on the elastic contour model⁽¹⁷⁾ and the image energy.

In section 2, we briefly describe the method used to extract the contour of facial features automatically for the neutral face (see Ref.(7) for details), and then explain how to accomplish automatic extraction of contours for a facial images with various facial expressions in section 3. Note that in this study, the resolution of the facial image is 256×240 pixels.

2. Automatic Contour Extraction of Facial Organs for the Neutral Face

In our method, we first obtain the center position of pupils using the cross-correlation method (section 2.1) and by using the relationship in terms of position of right-left pupils, we determine areas including eyes, eyebrows and mouth respectively (section 2.2). We then acquire contours of each facial organ by analyzing these areas (section 2.3).

2.1 Automatic extraction of pupils

In order to extract the pupils, we utilize the brightness distribution along a vertical line crossing the pupil and eyebrow (Fig.1), since the brightness distribution along the vertical line may have a dark-bright-dark configuration corresponding to the pupil, eyelid and eyebrow and may not change when we rotate the face or head. We therefore can expect to extract this brightness distribution pattern from the image easily and stably, and then acquire the position of each pupil. "Base data" (shown at the right side of Fig.1) is acquired by calculating the average of brightness distributions along the vertical line crossing the pupil of 10 subjects. Using the cross-correla-

Fig. 1 Brightness distribution along a vertical line crossing the pupil and eyebrow

tion method, we extract two positions which have first and second maximum cross-correlation values in the image compared with "base data" as right and left pupils. Table 1 shows the average of extracted pupil position error for a 10 second experiment in terms of four movements, i.e. right-left rotation, up-down rotation, right-left inclination and parallel displacement for three subjects. We find that high accuracy extraction is realized.

2.2 Area for each facial organ

We determine the facial organ areas, each of which contains both eyebrows (left and right), both eyes (left and right), the nose, and mouth, as shown in Fig. 2. In Fig. 2, X denotes the distance between both pupils and the values express the relative ones to X when we adjust X to one. These facial organ areas are examined as to whether they include facial features respectively of all of six typical facial expressions obtained from 30 subjects. These areas therefore are proved to contain with 100% certainty the human facial features.

2.3 Automatic contour extraction of facial features for the neutral face

We have developed image processing algorithms for automatic contour extraction of each facial organ respectively⁽⁷⁾. These algorithms fully depend on the shape and/or feature of each facial organ. Following the flowchart shown in Figs. 3, 4, and 5, we briefly explain our method for contour extraction of each facial organ (see Ref.(7) in detail).

Figure 3 shows the algorithmic flowchart for automatic extraction of the eyebrow shape, where the major procedures are as follows: Reinforcement of the horizontal edge (EB-1), positioning the eyebrow using template-matching method (EB-2), lower and upper edges of the eyebrow are approximated by quadratic curve fitting (EB-3 and EB-4), inner and outer parts of the closed contour of the eyebrow are binarized in terms of black and white (EB-5), and the maximum dark area is found by the labeling technique (EB-6).

Table 1 Accuracy of automatic pupil position extraction for 3 subjects

subject (diameter of the pupil)	right-left rotation	up-down rotation	right-left inclination	unit : mm	
				parallel displacement	
A (9mm)	2.529	2.911	3.050	2.692	
B (8.5mm)	2.714	1.907	1.682	1.861	
C (7.5mm)	2.847	1.705	2.701	2.684

Fig. 2 Facial organ areas containing each of facial features respectively

Fig. 3 Flow chart for extracting the eyebrow contour

Fig. 4 Flow chart for extracting the eye contour

Figure 4 shows the algorithmic flow chart for automatic extraction of the eye shape, where the major procedures are as follows: Transform the image data of the eye area into radial coordinates centered at each pupil position (E-1), find the eye contour line by using correlation technique (E-2), approximate the eye contour line by quadratic curve (E-3), and inversely transform the approximate eye contour line to determine the eye contour shape (E-4).

Figure 5 shows the algorithmic flow chart for automatic extraction of the mouth shape, where the major procedures are as follows: Edge reinforcement of the image in 8 directions (at 45 degrees differences) (M-1), template-matching to find approximate position of the mouth (M-2), binarization inside the template region (M-3), find the gravity center of the dark area for correction of the mouth position (M-4), transform the image data of the mouth into radial coordinates with respect to the gravity center (M-5), correlation calculation for finding the brightness-changing line (M-6), determine the both edge points of the mouth (M-7), determine the lip-closing line (M-8), and upper and lower lip line are approximated by quadratic curves (M-9).

3. Automatic Contour Extraction of Facial Features for the Facial Images with Various Facial Expressions

3.1 Issues of snakes

One of the most famous methods to extract facial contours might be "Snakes"⁽¹⁰⁾. It can extract facial contours of the heavy makeup face. This is because Snakes basically needs the binarized edge information, i.e. if edge is not clear and/or complicated, Snakes can not extract and track the contour. Also since there is no mechanism to keep the shape of contour and only edge information is used for computation, Snakes can not extract contour in case of parallel movement.

Fig. 5 Flow chart for extracting the mouth contour

3.2 Elastic contour model

Ueda et al. solved these issues⁽¹⁷⁾ by using the elastic contour model which represents the shape of contour line by control points and springs connecting them as shown in Fig. 6 so that the elastic contour model could keep the contour shape. In this method, the elastic energy (E_{elastic}) which represents deformation level of the elastic contour model and the image energy (E_{image}) which is similar to Snakes consisted by the edge information are defined. And contour line is determined where $E_{\text{elastic}} + E_{\text{image}}$ becomes minimum value. This method used the binary value (1 or 0) for the edge information in order to calculate E_{image} . To get the binarized value is very important factor for this method though, how to select the threshold value in order to get the binarized value for the edge information is not clear and its value fully depends on environment such as lighting condition. We then decide to use the analogue value directory obtained from the image for E_{image} . Let's explain this method in detail below.

E_{elastic} is defined as follows;

$$E_{\text{elastic}}(\mathbf{i}) = \alpha \{ \mathbf{r}(\mathbf{i}) - \mathbf{r}_0(\mathbf{i}) \}^2 + \beta \{ \theta(\mathbf{i}) - \theta_0(\mathbf{i}) \}^2 \quad (1)$$

where α and β are weighting factors which we define by heuristic approach and in this study we apply 600 and 500 respectively. $\mathbf{i}$ shows the number of a control

Fig. 6 Elastic contour model

point shown in Fig. 6 which is used for connecting springs. $r(i)$ represents the distance between control point i and $i+1$, and $\theta(i)$ denotes the angle between control point i and $i+1$. $r_0(i)$ and $\theta_0(i)$ show the average value of $r(i)$ and $\theta(i)$ manually obtained from 20 subjects in terms of neutral face. By using these average values, contour shape can be kept even if it is changed.

E_{image} is defined by following equation;

$$E_{\text{image}}(i) = \gamma E_{\text{bri}}(i) + \delta E_{\text{dif}}(i) \quad (2)$$

where E_{bri} shows the brightness value for each pixel on the control point and on the straight line between control point i and $i+1$. E_{dif} represents the average value of the brightness differences between one-upper pixel and one-lower pixel in terms of the straight line between control point i and $i+1$. γ and δ are weighting factors which we define by heuristic approach and in this study we employ 90 and 40 respectively.

We select each control point so that Eq. (3), i.e.,

$$E_{\text{total}} = \sum_i (E_{\text{elastic}}(i) + E_{\text{image}}(i)) \quad (3)$$

becomes minimum value by using dynamic programming method⁽¹⁸⁾, i.e. to move each control point in a sequential order in order to obtain the minimum value. As we see from values in terms of $\alpha(=600)$, $\beta(=500)$, $\gamma(=90)$, and $\delta(=40)$, E_{elastic} is more effective than E_{image} in Eq. (3). If not, from our experience, we can not expect to obtain smooth contour line precisely since E_{elastic} is employed for keeping the average smooth contour shape.

In this study, one of the important factors is the number of control points for expressing contour of each facial organ. Referring to our antecedent research works in terms of recognition of facial expressions by neural network where we utilize 30 feature points on contour of facial features⁽¹⁾⁻⁽⁶⁾, we employ 4 control points for each eyebrow, 8 ones for each eye and 8 ones for mouth as shown in Fig. 7. Thus there are 32 control points in total. No. 20 and no. 24 are employed in addition to 30 feature points for expressing eyebrow shape more clearly.

Fig. 7 Control points

3.3 Automatic contour extraction of facial features for facial images with various facial expressions

The most important factors to obtain the contour line of facial features precisely by using dynamic programming method is the initial position of the elastic contour model. If it were away from the position of real contour, it would be difficult to extract the contour of facial features. We therefore have decided to use video rate (1/30 s) time sequential image. Namely, only for the beginning, since subject normally expresses the neutral face, we employ the automatic extraction method for the neutral face mentioned in chapter 2 and extract the contour of facial features. To extract the contour of the next image, i.e. 1/30 s time-elapsd facial image, we apply the contour position of facial features obtained from the neutral face as the initial position of the elastic contour model. Thus we renew the initial position of the elastic contour model for the next facial image by using the contour position of facial features obtained from previous face image. Repeating this renewal manner, the contour of facial features is tracked.

3.4 Estimation of automatic contour extraction of facial features

By using time sequential facial image from neutral to six typical facial expressions of 20 Japanese, we estimate the accuracy of our method. Figure 8 shows an example of automatic contour extraction. Since duration for expressing facial expressions is arbitrarily, we select facial images for 5 seconds (150 images $(=30 \times 5)$ per one typical facial expression) so that in the end, one of six typical facial expressions is displayed. Thus tracking of facial features is undertaken for 5 seconds for each of six typical facial

Fig. 8 Examples of automatic contour extraction

expressions with respect to 20 people. In order to estimate the accuracy of our method, we employ the final result, i.e. the contour position of each facial organ obtained from the last facial image (one of six typical facial expressions) are utilized, and compare it with one manually inputted by 5 subjects.

Control points have constraint in terms of horizontal movement (x direction). That is to say, control points with $\bigcirc$ in Fig. 7 is used for determining x -coordinates of control points without $\bigcirc$. For example, control points 6, 7, and 8 are located between control points 1 and 2 at even intervals in terms of x direction. This constraint comes from the prerequisite of elastic contour model and then in case of manual input of control points by human beings, we employ the same constraint. Namely, following positions of control point 1 and 2 manually inputted, three vertical lines are shown up between them at even interval and then control point 3, 4, 5, 6, 7, and 8 are selected from on their lines respectively.

The total image number used for estimating our method is 120 (images for six typical facial expressions times 20 Japanese subjects). It is quite difficult to select unique control points even by human beings and then control points manually inputted have different values between subjects. We therefore utilize the average of 5 subjects. Figure 9 shows the standard deviation of position differences between control points manually inputted by 5 subjects with respect to each typical facial expression and each facial organ. The reason why we analyze the standard deviation in terms of each typical facial expression and each facial organ is that we think deformation of each facial organ is effected by each type of facial expression. From Fig. 9, we find that regardless of differences in facial expressions, standard deviation for eye is 1.1 pixels, one for eyebrow is 1.3 pixel, and one for mouth is 1.7 pixel. This result might show the fact that mouth moves more than eyebrow, and eyebrow moves more than eye. Also in case of mouth, the

Fig. 9 Standard deviation of control points extracted by humans

border between lip and face skin is less clear than ones for eye and eyebrow. As we see, because the position of control points selected are difference between subjects, we employ the average of 5 subjects in terms of position of each control point and use it for estimating automatic extraction result. In case of control points with $\bigcirc$ in Fig. 7, x and y coordinates are taken into account, and in case of control points without $\bigcirc$, only x coordinate is used for estimation. We compare these values with average ones by humans and acquire the absolute error between them. Figure 10 shows the average and standard deviation of the absolute error in terms of each typical facial expression and each facial organ.

We find from this figure that deformation and movement of each facial organ in terms of each typical facial expression are not similar though, the average of absolute error for eye and eyebrow is less than 2 pixels and one for mouth is between 2 and 3 pixels which is bigger than eye and eyebrow. In addition, the standard deviation for mouth is bigger than one for eyes and eyebrows. We think that this is because the movement of mouth is bigger than that of eyes and eyebrows and the border between lip and skin is not

Fig. 10 Accuracy estimation of automatic contour extraction

clear compare to one for eyes and eyebrows as we mentioned before. However, because the accuracy of automatic contour extraction does not depends on differences of facial expressions, we can say that automatic contour extraction of facial features is undertaken precisely despite of deformation differences of each facial organ. And also, since, as we find in Fig. 9, the standard deviation of contour extraction by human beings is about 1.5 pixel and the absolute error shown in Fig. 10 is the same value, we can say that the accuracy of automatic contour extraction is within the same range by human beings. Moreover, because 5 pixels corresponds to 1 mm, we can conclude that the automatic contour extraction is realized very precisely.

4. Conclusions

In this paper, we present the method in order to realize automatic contour extraction of facial features such as eyebrows, eyes and mouth for frontal facial images with various facial expressions. Because Snakes which is one of the most famous method to extract contour, has several issues such that only edge information is used, it can not track the contour in case of parallel movement and can not extract the contour for facial images with various facial expressions, we propose a new method to overcome these issues. We define the elastic contour model in order to keep the contour shape and then determine the elastic energy acquired by the amount of modification of the elastic contour model. Also we define the image energy obtained by brightness differences of the control points on the elastic contour model instead of using binary edge information which is used by Snakes. Applying the dynamic programming method, we select the contour position where the total value of the elastic energy and the image energy comes to minimum. By using 1/30 s time-wise facial image, we realize automatic contour tracking and

extraction for facial images with various facial expressions. We estimate the ability of our method by using six typical facial expressions obtained from 20 Japanese subjects and find that automatic contour extraction is undertaken very precisely.

For recognition of facial expressions, we imagine that texture information is also very important as well as contour information for each facial organ. To clarify what kind of information is efficient for recognizing facial expressions and then to pursue the important factors for displaying facial expression is our future research work.

References

- (1) Kobayashi, H. and Hara, F., The Recognition of Basic Facial Expressions by Neural Network, International Joint Conference on Neural Network, (1991), pp. 460-466.
- (2) Kobayashi, H. and Hara, F., Recognition of Six Basic Facial Expressions and Their Strength by Neural Network, IEEE International Workshop on Robot and Human Communication, (1992), pp. 381-386.
- (3) Kobayashi, H. and Hara, F., Recognition of Mixed Facial Expressions by Neural Network, IEEE International Workshop on Robot and Human Communication, (1992), pp. 381-386.
- (4) Kobayashi, H. et al., A Basic Study of Dynamic Recognition of Human Facial Expressions, Proceedings of IEEE International Workshop on Robot and Human Communication, (1993), pp. 271-275.
- (5) Kobayashi, H. and Hara, F., Dynamic Recognition of Basic Facial Expressions by Discrete-time Recurrent Neural Network, Proceedings of International Joint Conference on Neural Network, (1993), pp. 155-158.
- (6) Kobayashi, H., Tange, A. and Hara, F., Real-time Recognition of 6 Basic Facial Expressions, Proceedings of IEEE International Workshop on Robot and Human Communication, (1995), pp. 179-186.
- (7) Hara, F., Tanaka, K., Kobayashi, H. and Tange, A., Automatic Feature Extraction of Facial Features and Contour, Proceedings of IEEE International Workshop on Robot and Human Communication, (1997), pp. 386-391.
- (8) Essa, I.A. and Pentland, A., Facial Expression Recognition Using Visually Extracted Facial Action Parameters, Proceedings of International Workshop on Automatic Face- and Gesture-Recognition, (1995), pp. 35-40.
- (9) Essa, I.A. and Pentland, A., Facial Expression Recognition Using a Dynamic Model and Motion Energy, Proceedings of International Conference on Computer Vision '95, (1995).
- (10) Kass, A., Witkin, A. and Terzopoulos, D., Snakes: Active Contour Models, Int. J. Computer Vision,

- Vol. 1, No. 3 (1988), pp. 321-331.
- (11) Yacoob, Y. and Davis, L., Computing Spatio-Temporal Representation of Human Faces, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 18, No. 6 (1994), pp. 636-642.
 - (12) Black, M.J. and Yacoob, Y., Tracking and Recognizing Rigid and Non-Rigid Facial Motions Using Local Parametric Models of Image Motion, *Proceedings of International Conference on Computer Vision*, (1995), pp. 374-381.
 - (13) Moses, Y., et al., Determining Facial Expressions in Real Time, *Proceedings of International Conference on Computer Vision*, (1995), pp. 296-301.
 - (14) Mase, K., Recognition of Facial Expressions by Optical Flow, *IEICE Transactions, Special Issue on Computer Vision and Its Applications*, E74, No. 10 (1994).
 - (15) Prochazka, Z., Ito, T. and Okamoto, T., A Method for Automatic Extraction of Motion Based Expression Features, (in Japanese), *IEICE D-II*, Vol. J81-D-II No. 6 (1998), pp. 1150-1159.
 - (16) Ekman, P. and Friesen, W.V., *Unmasking the Face*, (1975), Prentice-Hall, Inc.
 - (17) Ueda, N., Mase, K. and Suenaga, Y., A Contour Tracing Method Using Elastic Contour Model and Eenergy Minimization Approach, (in Japanese), *IEICE D-II*, Vol. J75-D-II, No. 1 (1992), pp. 111-120.
 - (18) Amini, A.A., Weymouth, T.E. and Jain, R.C., Using Dynamic Programming for Solving Variational Problems in Vision, *IEEE Trans. Pattern Analysis & Machine Intelligence*, PAMI-12, 9 (1990), pp. 855-867.
 - (19) Grossberg, S., Adaptive Pattern Classification and Universal Recording, II: Feedback, Expectation, Olfaction, and Illusions., *Biological Cybernetics*, Vol. 23 (1976), pp. 187-202.
-