Robust Lip Contour Extraction using Separability of Multi-Dimensional Distributions

Tomokazu Wakasugi,　Masahide Nishiura　and　Kazuhiro Fukui

Corporate Research and Development Center, Toshiba Corporation
1, KomukaiToshiba-cho, Saiwai-ku, Kawasaki 212-8582 Japan
e-mail: tomokazu.wakasugi@toshiba.co.jp

Abstract

We present a lip contour extraction method using separability of color intensity distributions. Usually it is difficult to robustly extract the outer lip contour mainly because of the following two problems. First, the outer lip contour is often blurred. Secondly, the contrast between the skin and the lip region is often reduced by transformation from the color intensity to the gray scale intensity. To overcome these two problems we propose an edge detection method in which edge strength is defined as separability of two color intensity distributions. We apply the proposed method to lip contour extraction using an active contour model. We present several experimental results demonstrating the effectiveness of the proposed method.

1. Introduction

Lip contour extraction is a useful technique for obtaining a shape of a mouth in an image, and is one of the most important techniques for human-machine interface applications such as lip reading and facial expression analysis. However, there are two main problems with regard to robust lip contour extraction in practical situations. Firstly, the outer contour of the lip, particularly of the lower lip, is often blurred. Secondly, the contrast between the skin and the lip region is often reduced by transformation from the color intensity to the gray scale intensity.

To extract the lip contour we use an active contour model. Most previous lip contour extraction methods use an active contour model which extracts a curve with high edge strength. For this process it is important to robustly detect edges on the lip contour. However, most of the previous edge detection methods using gradient of gray scale intensity [1][2] are insufficient to robustly detect edges on the outer lip contour because of the above two problems. A method using separability of two gray scale intensity distributions within each of two local regions [3] is one of the most effective methods for detecting such edges since the method is effective for detecting blurred edges. This is mainly because in the method an edge is defined not as a point where the gray scale intensity changes steeply but as
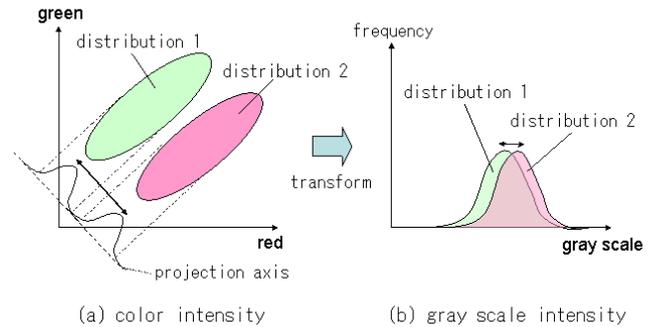

Figure 1: Reducing separability by transforming (a) 2-dimensional color intensity distributions to (b) 1-dimensional gray scale intensity distributions.

a boundary of two adjoining regions which are well separated.

However, this method is ineffective in coping with the second problem since it uses the gray scale intensity as an image feature. Transformation from the color intensity to the gray scale intensity often changes sufficiently-separated three-dimensional distributions to insufficiently-separated one-dimensional distributions (Fig.1). In such cases, an edge which originally has a high edge strength may not be detected. Alternatively the hue can be used to distinguish the lip region from the face [4][5] and has proven to be more effective for detecting the lip region than the gray scale intensity. But, in the case of most of the methods using the hue, it has difficulties coping with various human lips. In addition, the hue is often insufficient to enhance the contrast between the lower lip region and the skin region. Consequently, we believe that separability of color intensity distributions will prove effective for detecting edges on the outer lip contour.

As described in detail in section 2, we must determine a *projection axis* to calculate separability. A N-dimensional distribution (N : a natural number) of variables is projected onto the *projection axis* and consequently becomes a one-

dimensional distribution. In the case of one-dimensional variables, the *projection axis* corresponds to the axis of the variable. On the other hand, in the case of multi-dimensional variables, the *projection axis* (Fig.1(a)) is not determined. Therefore, we introduce a method for determining the *projection axis* in the multi-dimensional feature space by Fisher discriminant analysis so that separability is maximized. Kaucic et al. proposed a similar method to enhance the contrast between the skin and the lip region [6]. Their method is based on projection of the color intensity onto an axis, determined by Fisher discriminant analysis, to maximize separation of two distributions composed of color intensities within the skin and the lip region respectively. This method has proven to be more effective for identifying the lip than transformation to the hue image. However, the optimum Fisher axis is thought to be different for each person and each part of the lip. Therefore, the proposed method determines the optimum Fisher axis at each local region in each image.

In this paper, we apply the proposed edge detection method to extract the outer lip contour using an active contour model. The experimental results show that the proposed method is more effective for extracting the blurred outer lip contour than are the conventional methods.

2. Edge Detection using Separability

2.1. Separability of One-Dimensional Distributions

Firstly, we explain the definition of separability of one-dimensional distributions. We use a separability mask composed of two regions (Fig.2). Separability η is calculated as

$$\eta = \frac{\sigma_b{}^2}{\sigma_T{}^2} \tag{1}$$

$$\sigma_T{}^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \overline{x}_m)^2 \tag{2}$$

$$\sigma_B{}^2 = \frac{1}{N} \{ n_1(\overline{x}_1 - \overline{x}_m)^2 + n_2(\overline{x}_2 - \overline{x}_m)^2 \}, \tag{3}$$

where σ_T^2 is the variance of all image intensities in region 1 and region 2. N is the number of all pixels in the two regions. x_i is the intensity at pixel i and $\overline{x}_m$ is the average intensity in the two regions. σ_B^2 is a between-class variance of region 1 and region 2. $\overline{x}_1$ and $\overline{x}_2$ are average intensities in region 1 and region 2, respectively. n_1 and n_2 are the numbers of pixels in region 1 and region 2, respectively. Separability η is a normalized value which has a range from 0 to 1.0. We use this separability as edge strength.

Next we explain our method for detecting edges using separability on the supposition that the method is applied to contour extraction using an active contour model [8]. We
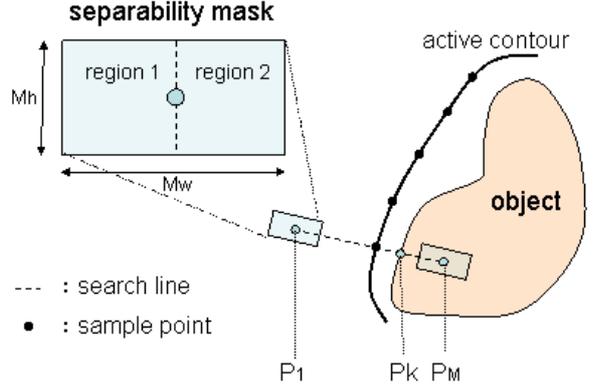

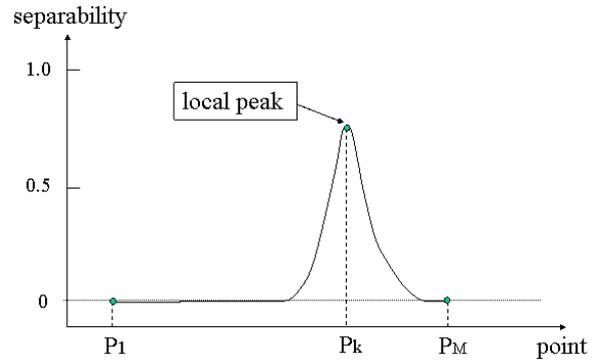Figure 2: Edge detection using separability.

Figure 3: Separability on the search line.

suppose an active contour near the object and sample points on the active contour (Fig.2). Also we suppose a search line set so as to be vertical to the active contour at a sample point. Separability masks are located in a direction parallel to the search line, at each point $P_m(m = 1, 2, ..., M)$ on the search line, where M is the number of masks. Separability η_m is evaluated at corresponding P_m. A point P_k which gives a local peak value is defined to be an edge location (Fig.3).

2.2. Separability of Multi-Dimensional Distributions

It is known that the gray scale intensity is inadequate for identifying the outer lip contour [7]. This is mainly since transformation from the color intensity to the gray scale intensity often reduces the contrast between the skin and the lip region. Therefore, the method using separability of gray scale intensity described in the previous subsection is disadvantaged in that it isn't able to make full use of the color

information. Hence we propose an edge detection method, *separability of multi-dimensional distributions*. The proposed method is able to calculate separability of two distributions composed of multi-dimensional variables such as the color intensity. Also, since the definition of an edge in the proposed method is similar to that in the method using separability of one-dimensional distributions, the proposed method naturally has robustness against noise and effectiveness in detecting blurred edges. *Separability of multi-dimensional distributions* η_M is calculated as the largest eigenvalue λ_{max} of the matrix $\hat{A}$ written as

$$\hat{A} = \hat{S_T}^{-1} \hat{S_B} \tag{4}$$

$$\hat{S_T} = \sum_{i=1}^{N} (\mathbf{x}_i - \overline{\mathbf{x}}^{(m)})(\mathbf{x}_i - \overline{\mathbf{x}}^{(m)})^T \tag{5}$$

$$\hat{S_B} = \frac{n_1 n_2}{N} (\overline{\mathbf{x}}^{(1)} - \overline{\mathbf{x}}^{(2)})(\overline{\mathbf{x}}^{(1)} - \overline{\mathbf{x}}^{(2)})^{\mathbf{T}} \tag{6}$$

$$\mathbf{x}_i = (x_{i1}, x_{i2}, ..., x_{iD}), \tag{7}$$

where $\mathbf{x}_i$ is an image intensity vector, and D is the number of dimensions of the variable. $\hat{S_T}$ is a scatter matrix of all image intensity vectors $\mathbf{x_i}$ in both region 1 and region 2 (Fig.2). N is the number of all pixels in both regions. $\overline{\mathbf{x}}$ is the average vector of color intensity vectors in both regions. $\hat{S_B}$ is the between-class scatter matrix of region 1 and region 2. n_1 and n_2 are the numbers of pixels in region 1 and region 2, respectively. $\overline{\mathbf{x}}^{(1)}$ and $\overline{\mathbf{x}}^{(2)}$ are the average vectors of color intensity vectors in region 1 and region 2, respectively. η_M is a normalized value which has a range from 0 to 1. In the proposed method we define edge strength as this separability η_M.

Usually calculation of eigenvalues has a high computational cost and is sometimes unstable. In addition the problem is potencially more serious since matrix $\hat{A}$ is not symmetric. Nevertheless, in the proposed method we can quickly and stably calculate η_M as

$$\eta_M = \frac{\left| \hat{S_B} \mathbf{u} \right|}{\left| \hat{S_T} \mathbf{u} \right|} = \frac{\left| \hat{S_B} \hat{S_T}^{-1} (\overline{\mathbf{x}}^{(1)} - \overline{\mathbf{x}}^{(2)}) \right|}{|\overline{\mathbf{x}}^{(1)} - \overline{\mathbf{x}}^{(2)}|} \tag{8}$$

$$\mathbf{u} = \hat{S_T}^{-1} (\overline{\mathbf{x}}^{(1)} - \overline{\mathbf{x}}^{(2)}), \tag{9}$$

where $\mathbf{u}$ is an eigenvector of the matrix $\hat{A}$ corresponding to an eigenvalue λ_{max}. The eigenvector $\mathbf{u}$ is the Fisher axis, that is, the *projection axis* determined so that separability is maximized. Calculation of $\hat{S_T}^{-1}$ has a much lower cost than calculation of an eigenvalue of matrix $\hat{A}$.
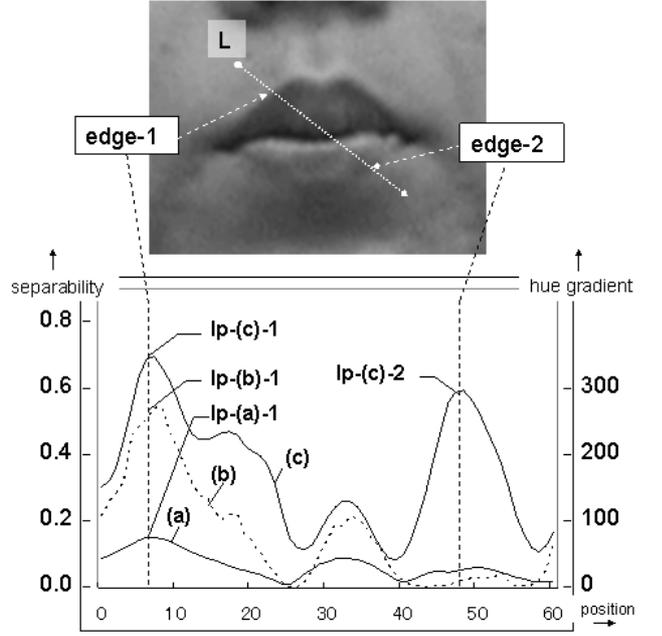

Figure 4: Edge strength on the search line. (a) hue gradient (b) gray scale separability (c) color separability

2.3. Basic Performance of Separability of Multi-Dimensional Distributions

In order to evaluate the effectiveness of the proposed method, we experiment with detecting edges on the outer lip contour. We assign a search line L on the mouth (Fig.4). Three edge detection methods are performed along the search line: (a) hue gradient, (b) gray scale separability, and (c) color separability (the proposed method). The image size is 100×75 pixels, and separability mask sizes M_w and M_h are 16 and 8 pixels, respectively. Each result is plotted in Figure.4.

All local peak points lp-(a)-1, lp-(b)-1, and lp-(c)-1 correspond to edge-1 on the outer contour of the upper lip. This means that all three methods are able to detect edge-1. However, only one local peak point lp-(c)-2 of the proposed method corresponds to edge-2 on the outer contour of the lower lip, whereas methods (a) and (b) are unable to detect edge-2. With respect to method (a), this result can be understood since transformation to the hue image is inadequate for enhancing the contrast between the lip region and the skin region for this sample image. With respect to method (b), the result can be explained as follows. Although there was originally an obvious separation of two color intensity distributions corresponding to the skin and the lip region respectively, the separation has vanished through the trans-
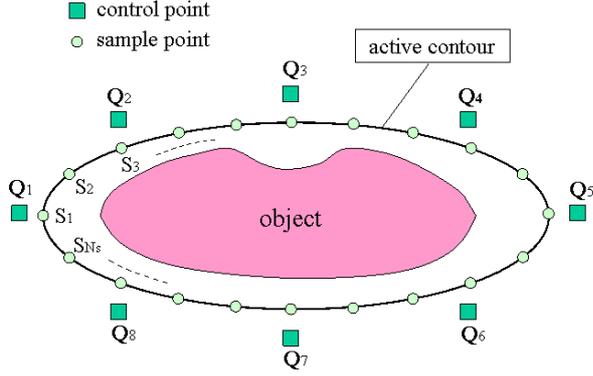
3

Figure 5: Formation of an active contour.

Figure 6: Convergent process of an active contour.

formation to the gray scale intensity. This problem often occurs in detection of edges on the outer contour of the lower lip. Thus, the proposed method appears to be the most effective for extracting the outer contour of the lower lip among the three choices.

3. Lip Contour Extraction

As a method for lip contour extraction, we use an active contour model [8]. This method is based on minimization of an energy function which is defined so that it decreases as an active contour gets close to the contour of an object. We form the active contour with the B-spline curve $\mathbf{L(t)}$ with eight control points $\mathbf{Q}_i$ $(i = 1, 2, ..., 8)$ (Fig.5) and B-spline functions $N_i(t)$ as

$$\mathbf{L}(t) = \sum_{i=1}^{8} N_i(t)\,\mathbf{Q}_i, \qquad (10)$$

where i indicates the number of control points, and t is a parameter to indicate the position on the B-spline curve which has a range from 0 to 8 [9]. The energy function E is defined as

$$E = E_{int} + E_{image}, \qquad (11)$$

where E_{int} is the internal energy, which represents smoothness of the active contour and E_{image} is the image energy, defined to be minimum when the active contour is correctly on the contour. In this paper we introduce a simplification in that we don't take account of the internal energy since the B-spline representation maintains smoothness via hard constraints implicit in the representation [9]. Therefore, the energy function E is composed of only the image energy E_{image}. We define the image energy E_{image} as the average of separabilities calculated at all sample points. Consequently, the energy function E is written as
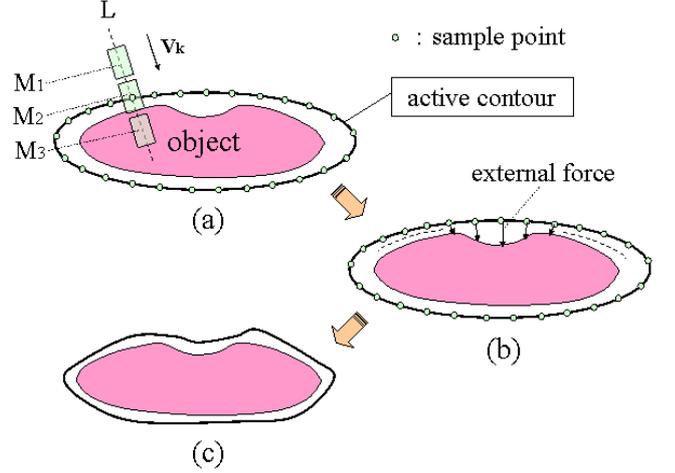
$$E = E_{image} = -\frac{1}{N_s} \sum_{k=1}^{N_s} \eta(s_k), \qquad (12)$$

where s_k denotes the kth sample point on the B-spline curve (Fig.5), and N_s is the number of sample points.

The active contour is controlled using separability as follows:

- N_s sample points are placed on the active contour.

- At each sample point s_k, the following three processes are executed.

    - Set three masks M_c $(c = 1, 2, 3)$ on the search line L (Fig.6 (a)).
    - Calculate the $\eta^{(c)}$ at each of the three masks.
    - Determine the *external force* $\mathbf{f}_k$ (Fig.6 (b)).

- Move the active contour according to all the *external forces* $\mathbf{f}_k$ $(k = 1, 2, ..., N_s)$ (Fig.6 (c)).

- Repeat the above three processes until convergence.

The search line L is determined so as to be vertical to the tangential line of the active contour at the sample point s_k. *External force* $\mathbf{f}_k$ is a vector which determines the direction and the amount of the movement of the active contour near sample point s_k. It is written as

$$\mathbf{f}_k = \begin{cases} \alpha\,\mathbf{v_k} & : \quad \eta^{(3)} > \eta^{(1)},\,\eta^{(2)} \\ \mathbf{0} & : \quad \eta^{(2)} > \eta^{(1)},\,\eta^{(3)} \\ -\alpha\,\mathbf{v_k} & : \quad \eta^{(1)} > \eta^{(2)},\,\eta^{(3)}, \end{cases}$$

4

where α is a constant and $\mathbf{v_k}$ is a vector whose length is 1.0 in a direction parallel to the search line L and toward the inside of the active contour. The active contour is moved according to all the *external forces* $\mathbf{f}_k$ via the principle of virtual work [9]. By repeating these processes the energy function (12) is minimized and the active contour approches the correct lip contour. This repetition continues until the number of iterations exceeds a threshold T_{th}.

4. Experimental Results

We applied the proposed edge detection method to lip contour extraction and evaluated its effectiveness compared to gray scale separability, a conventional method. We have collected 200 sample color images of 10 persons $I_n(n = 1, 2, ..., 10)$, of size 320×240 pixels. We tested lip contour extraction with all the sample images using the conventional method and the proposed method. In two lip contour extractions of the conventional method and the proposed method which are tested on the same image, initial locations of two active contours are the same and are randomly determined near the mouth. The separability mask sizes M_w and M_h are 16 and 8 pixels, respectively. Figure 8 shows 10 examples of results of lip contour extraction.

For the conventional method, the contours of the upper lip are correctly extracted except (I_6) and (I_9), whereas the contours of the lower one are almost all incorrectly extracted. This result is due to the fact that the contour of the lower lip is usually less distinct than that of the upper one. In examples of the conventional method except (I_1), (I_2) and (I_{10}), the edges on the outer contour of the lower lip are partially undetected, and the other parts such as the inner contour of the lower lip are partially extracted by mistake. On the other hand, the proposed method has robustly extracted not only the outer contour of the upper lip but that of the lower one for all examples except (I_9). This difference between the two results is mainly due to the fact that the proposed method can detect the edges on the outer contour of the lower lip which the conventional method cannot. In such cases, separability of the skin and the lip region, which is originally high, vanishes through the transformation from the color intensity to the gray scale intensity.

To compare the effectiveness of the conventional and the proposed methods, we calculated the error ratios of the two methods for each person I_n. We defined the error ratio regarding each person as N_{error}/N_{image}, where N_{image} is the number of sample images of each person, and N_{error} is the number of failed sample images of each person. In order to decide whether a result of contour extraction is correct, we have introduced a factor F_c as
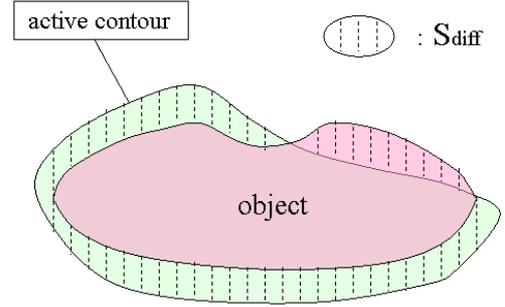
$$F_c = \frac{S_{\text{diff}}}{L_c^2}, \qquad (13)$$

Figure 7: An illustration of the aera S_{diff} which represents the area between the correct lip contour and an active contour.

where S_{diff} is the area between the correct lip contour and an active contour (Fig.7), L_c is a length of the correct lip contour. The factor F_c represents a degree of incorrectness of the result of contour extraction. If F_c is larger than a threshold value, the result of contour extraction is incorrect. The correct contour is manually traced in advance. Also, in this paper, the threshold value is determined as 0.04. This threshold value of 0.04 means that a permissible amount of an average difference between a correct lip contour and an active contour is 4 pixels when L_c is 100 pixels. We show the error ratios for each person (Table 1).

The error ratio of the proposed method is smaller than that of the conventional method with respect to all persons. This result shows that the proposed method can robustly extract a lip contour despite individual. Almost all the cases in which the proposed method cannot correctly extract the outer lip contour occurred when the initial location of the active contour is closer to other parts which have a high edge intensity, such as the inner lip contour, than the outer lip contour.

Table 1: Error ratios (%) of lip contour extractions
(a) conventional method   (b) proposed method

	I_1	I_2	I_3	I_4	I_5	I_6	I_7	I_8	I_9	I_{10}	total
(a)	20	25	40	55	30	45	60	50	60	40	42.5
(b)	0	10	20	20	15	25	30	20	35	10	18.5

5. Summary and Conclusions

In this paper we have proposed a method for edge detection using *separability of multi-dimensional distributions*. We have defined edge strength as the separability of two multi-dimensional distributions maximized by Fisher discriminant analysis. We have used the color

intensity as a variable composing multi-dimensional distributions, and applied the proposed method to detect edges on the outer lip contour. Experimental results have demonstrated that the proposed method can detect edges whose separation is such that it vanishes following transformation from the color intensity to the gray scale intensity.

Subjects for future work include solving the following problem. The proposed method sometimes detects more edges than the conventional method, including some which are not on the contour of the target object. This is mainly since the proposed method more sensitively detects the separation of two regions than the conventional method does. Therefore, compared with the results of the conventional method, those of the proposed method have a tendency to be more dependent on the initial location of the active contour.

References

[1] D. Marr and E. Hildreth, "Theory of Edge Detection," Proc. Roy. Soc. London, Vol.B-207, pp.187-217, (1980).

[2] J. Canny, "A Computational Approach to Edge Detection," IEEE Trans. Pattern Anal. Machine Intell., Vol.PAMI-8, no.6 pp.679-698, (1986).

[3] K. Fukui, "Edge Extraction Method Based on Separability of Image Features", IEICE Trans. Inf.&Syst., Vol.E78-D, no.12 pp.1533-1538, Dec (1995).

[4] M. Lievin, P.Delmas, P.Y.Coulon, F.Luthon and V.Fristot, "Automatic Lip Tracking: Bayesian Segmentation and Active Contours in a Cooperative Scheme," IEEE Int. Conf. on Multimedia Computing and Systems, Vol.1, pp.691-696, (1999).

[5] T. Coianiz, L. Torresani and B. Caprile, "2D Deformable Models for Visual Speech Analysis," In Stork and Hennecke, pp.391-398 (1996).

[6] R. Kaucic and A. Blake, "Accurate, Real-Time, Unadorned Lip Tracking," 6th Int. Conf. on Computer Vision, pp.370-375, (1998).

[7] A. Yuille, "Feature Extraction from Faces Using Deformable Templates," Int. Journal of Computer Vision, 8(2):99-111, (1992).

[8] M. Kass, A. Witkin, and D. Terzopouls, "Snakes: Active Contour Models," in Proc. 1st Int. Conf. on Computer Vision, p259-268, (1987).

[9] R. Cipolla and A. Blake, "The Dynamic Analysis of Apparent Contours," in Proc. 3rd Int. Conf. on Computer Vision, p616-623, (1990).
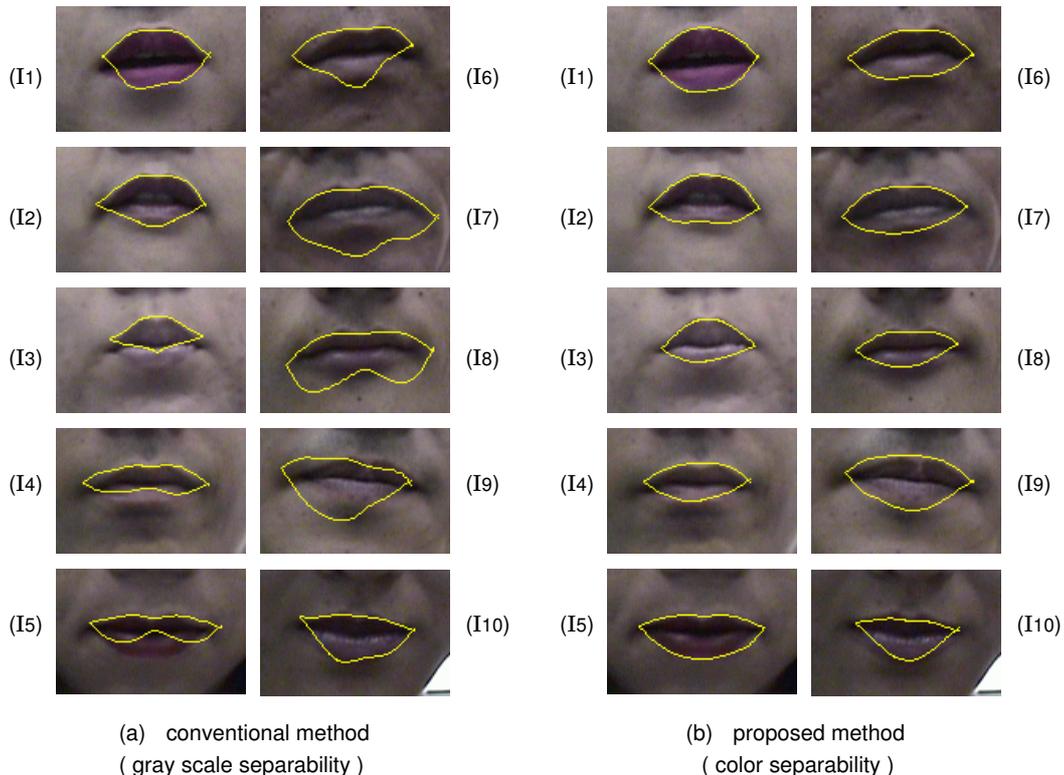
(a)   conventional method
(gray scale separability)

(b)   proposed method
(color separability)

Figure 8: Examples of the results of lip contour extraction using separability. I_n denotes each of 10 persons.