Lip Motion Capture and Its Application to 3-D Molding

Masashi Okubo
Okayama Prefectural University
111 Kuboki,Soja
Okayama 719-1197, Japan
okubo@cse.oka-pu.ac.jp

Tomio Watanabe
Okayama Prefectural University
111 Kuboki,Soja
Okayama 719-1197, Japan
watanabe@cse.oka-pu.ac.jp

Abstract

It is essential for machine lip reading to process not only static images but also moving images. This paper proposes a method, which is named 'Optical-Snakes', for lip motion extraction from moving images using the cooperation between the active contour models called Snakes and the optical flow. The effectiveness of the method is demonstrated by the exact extraction of a series of lip contours from time-varying images without any markers, and by the analysis of lip contours for machine lip reading.

Some real products which characterize a speaker's pronunciation are made from the lip motion data by a photoforming system. It is demonstrated that the real products are better than the virtual products on a computer display for the evaluation of shapes. The application of lip motion to molding would lead to a new method for the analysis of lip motion and the training of pronunciation.

1. Introduction

The extraction of lip motion is essential for machine lip reading. A lot of methods have been proposed in this field. Some of them utilize the optical flow to extract the lip motion[1]. However, this kind of methods can not utilize the outcomes of study on static images effectively. The active contour models called Snakes can highly extract object edges from static images[2]. One of the problems of this method is how to obtain the initial set of points. Another one is that Snakes are the method not for moving images but for static images fundamentally. To solve these problems at the same time, a method named Optical-Snakes is proposed to obtain the initial set of points in a target frame by the concurrent processing between the result of Snakes in the previous frame and the optical flow between the target frame and the previous one[3].

Some real 3D products are made from the lip motion data obtained by Optical-Snakes to evaluate the lip motion clearly. The lip motion data are transformed to 3D CAD data to make the real objects by a photoforming system. The possibility of new molding from human motion is demonstrated.

2. Lip Motion Capture

2.1. Active Contour Models (Snakes)

The active contour models which are defined by an energy function as follows , present the contour of an object by the set of points in the spline curve which surrounds the object smoothly.

$$E_{snake}(\boldsymbol{v}(s)) = \int_0^1 (E_{int}(\boldsymbol{v}(s)) + E_{ext}(\boldsymbol{v}(s)))ds \quad (1)$$

$$E_{int}(\boldsymbol{v}(s)) = (\alpha|\ \boldsymbol{v}_s(s)\ |^2 + \beta|\ \boldsymbol{v}_{ss}(s)\ |^2)/2 \quad (2)$$

where, $\boldsymbol{v}_s = d\boldsymbol{v}/ds$ $\boldsymbol{v}_{ss} = d^2\boldsymbol{v}/ds^2$

$$E_{ext}(\boldsymbol{v}(s)) = -(G_\sigma(\boldsymbol{v}(s)) * \nabla^2 I(\boldsymbol{v}(s)))^2 \quad (3)$$

where,

$$G_\sigma(\boldsymbol{v}(s)) = exp\left(-\frac{|\ \boldsymbol{v}(s)\ |^2}{2\pi\sigma^2}\right)$$

One of the problems of Snakes is how to get the initial set of points. The points should be near around the contour of an object, because the active contour models are easily influenced by small noise. Figure 1 shows the result of contour extraction by Snakes under condition that the initial points are near the object. This indicates the exact extraction. The result in Fig.1(a) was utilized as the initial set of points to obtain the lip contour in Fig.1(b). As shown in Fig.1(b), it is obvious that the lip motion can not be extracted correctly only by Snakes when the lip moved quickly.

0-8186-8344-9/98 $10.00

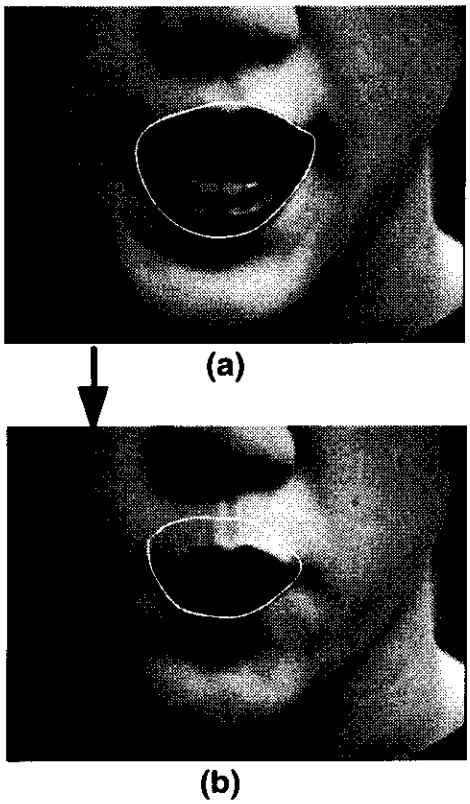

Figure 1. Lip contour extraction using only Snakes.

2.2. Optical flow

To estimate an appropriate initial set of points for each frame by Snakes, an optical flow between the target frame and the previous frame is utilized. The optical flow used in this study is obtained by a simple template matching method because of its speed and the restricted region of mouth movement. The matching method uses the spatial correlation function between the matrices on both frames as follows:

$$r = \frac{S_{xy}}{S_x S_y} \tag{4}$$

where,

$$\begin{aligned} S_x &= \sqrt{\sum (x_i - \bar{x})^2 / n} \\ S_y &= \sqrt{\sum (y_i - \bar{y})^2 / n} \\ S_{xy} &= \sum (x_i - \bar{x})(y_i - \bar{y}) / n \end{aligned}$$

The template size is 41 pixels(height) x 61 pixels(width) which centers the points of Snakes on the previous frame.

2.3. Optical-Snakes

The Optical-Snakes determine the initial set of points by the concurrent processing between the result of Snakes in the previous frame and the optical flow between the present frame and the previous one.

Figure 2 shows the procedure of lip contour extraction by using the Optical-Snakes from moving images obtained from a video camera.

First of all, the system operator draws the line which lies on an edge of lip contour using the mouse only for the first frame. Then, Snakes extract the lip contour on the first frame. Optical-Snakes determine the set of points as a initial contour on the second frame using both the result of Snakes on the first frame and the optical flow between the first frame and the second one. On and after third frame, Optical-Snakes iterate the procedure and obtains the set of points which indicates the lip contour frame by frame.

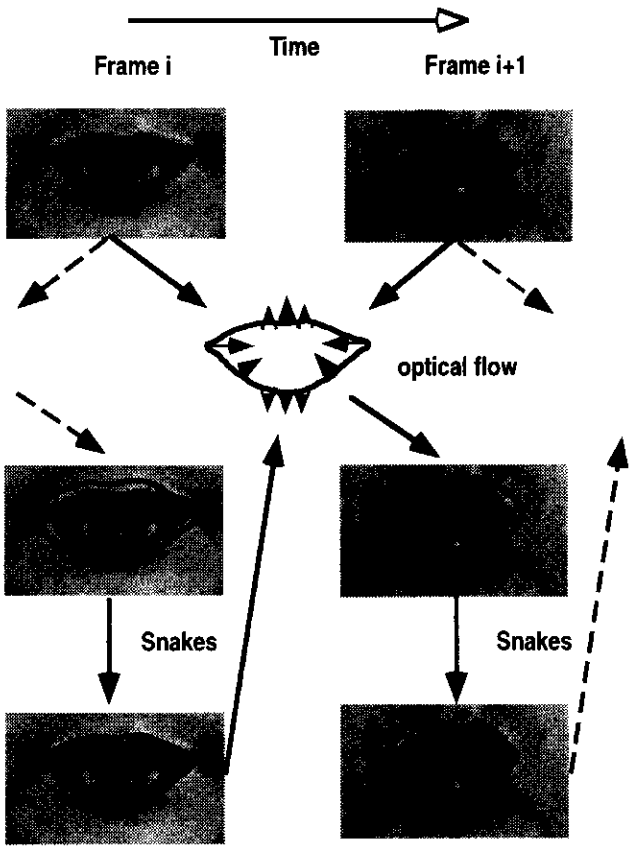

Figure 2. Procedure of Optical-Snakes.

2.4. Experiment

Figure 3 shows the result of Optical-Snakes applied to the images shown in Fig.1 where the values of param-

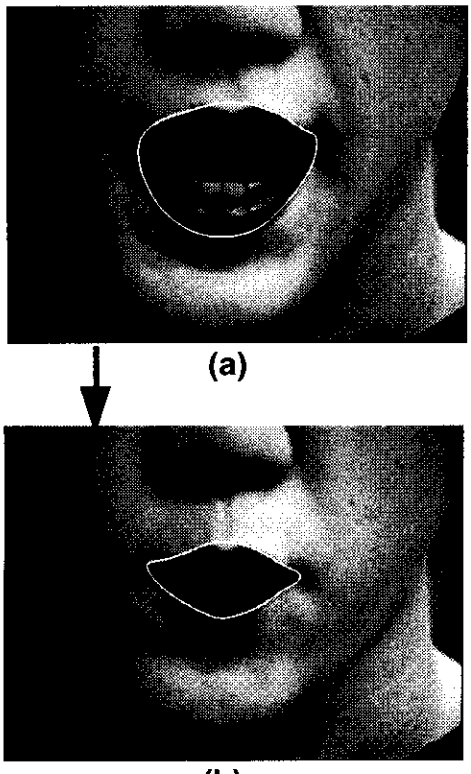

Figure 3. Lip contour extraction using Optical-Snakes.

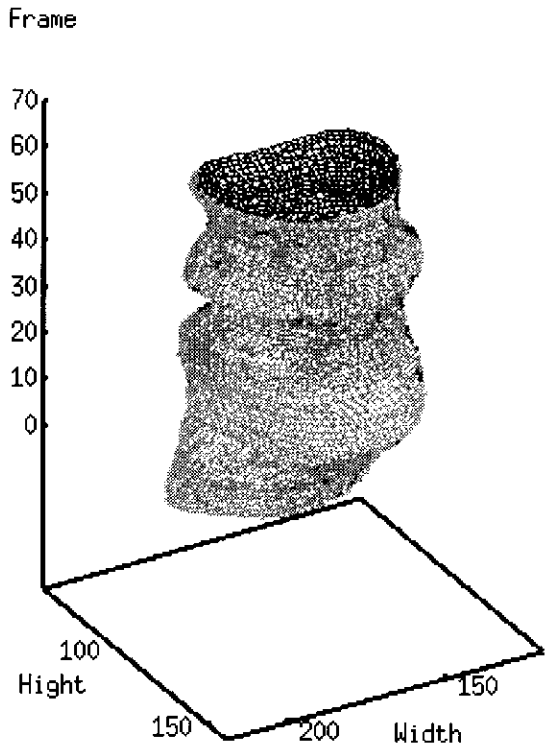

Figure 4. Example of lip motion extraction.

eters of the energy function in Formula(2) are the same as Fig.1. This result indicates that Optical-Snakes can extract the lip contour more easily.

Figure 4 shows an example of extracted lip motion data using the above system when a person pronounced Japanese five concatenated vowels (/a/, /i/, /u/, /e/, /o/) in 2.3 seconds (70 frames). The vertical axis indicates the frames, i.e., the time. In this figure, it is easy to find the frames corresponding to each vowel. Figure 5 shows the time changes of some parameters which are often applied to analyze the lip contour on a static image for machine lip reading [4][5][6]. All figures are made from the same data as Fig.4. As is obvious from theses figures, the characteristics of time changes of parameters enhance the discrimination of vowels.

3. Molding

In our previous paper, it is demonstrated that there is the high possibility of disagreement between the evaluation of shape in virtual space and that in real space [7]. From the point of view, the lip motion data are transformed to 3D CAD data and some photoforming products are made from them.

3.1. Molding Instruments

Figure 6 shows the system configuration to make the photoforming products using the lip motion data. The lip motion images are captured by EWS1 and then sliced into each frame.

The proposed Optical-Snakes extract the lip contour from each frame continuously. The 3D models are made from a set of lip contour data on the 3D CAD (Richo Design Base V5) on a graphic workstation (Silicon Graphics Indy) as shown in Fig.7. They are converted to IGES formatted data. Then the data are sent to EWS2 and the real objects are made from the data using the photoforming system (NTT DATA CMET SOUP400GH-SP) as shown in Fig.8.

3.2. Photoforming Products

Figure 9 shows an example of photoforming product. The size is 65 mm in mouth width, 50 mm in mouth height and 70 mm (1 mm per frame) in length. Figure 10 shows the cross sections of the 3D shape in Fig.9 to denote the representative mouth shape of each vowel. These products characterize the speaker's pronunciation where the speaker's under lip is distorted to left side when he pronounces the vowel /a/ as shown

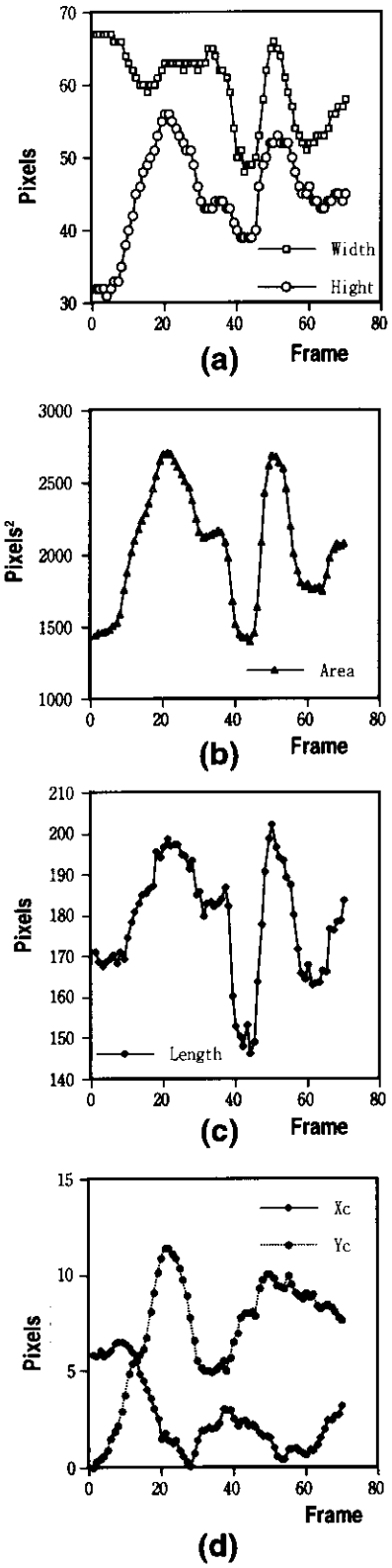

Figure 5. Time changes of parameters. (a) Lip width and height;(b) Lip area;(c) Lip circumference;(d) Center of gravity of lip area.

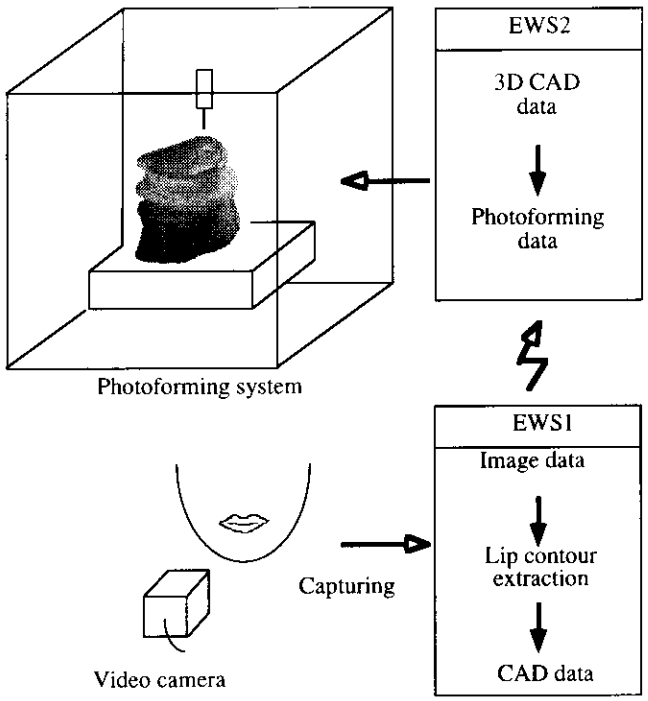

Figure 6. System configuration.

in Fig.10(c). Figure 11 shows two examples of photoforming products made from lip motion data. The left one shows 'momotaro' which is a popular hero's name in a famous fairy tale in Japan. And another one shows the miniaturized lip motion data when another speaker pronounced the vowels 'aiueo'. It is obvious at a glance that the photoforming products made from lip motion data must become the mementos characterizing the speaker as the speaker pronounces something from his or her heart. This would apply to new molding from human motion because we feel familiar with the natural shape and motion.

4. Conclusion

The new method for lip contour extraction is proposed. The method named Optical-Snakes is able to extract the lip motions from moving images exactly. Optical-Snakes are the effective concurrent processing between the optical flow and Snakes. Moreover, some photoforming products are made from the lip motion data. The possibility of new molding from human motion is demonstrated. The molding is helpful for the analysis of lip motion and for machine lip reading. The products must become the mementos characterizing the speaker.

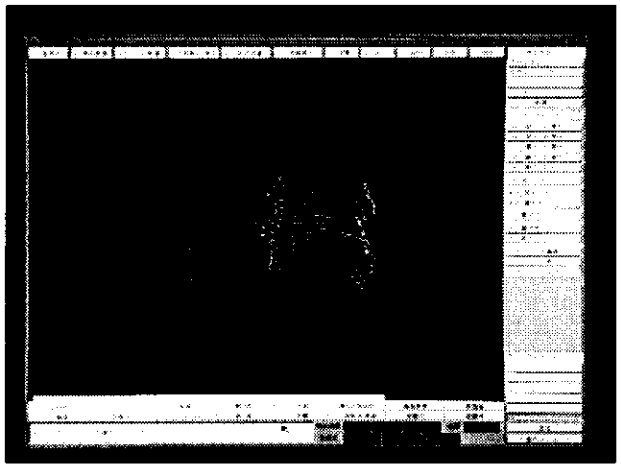

Figure 7. Lip motion data on 3D CAD.

Figure 8. Photoforming system.

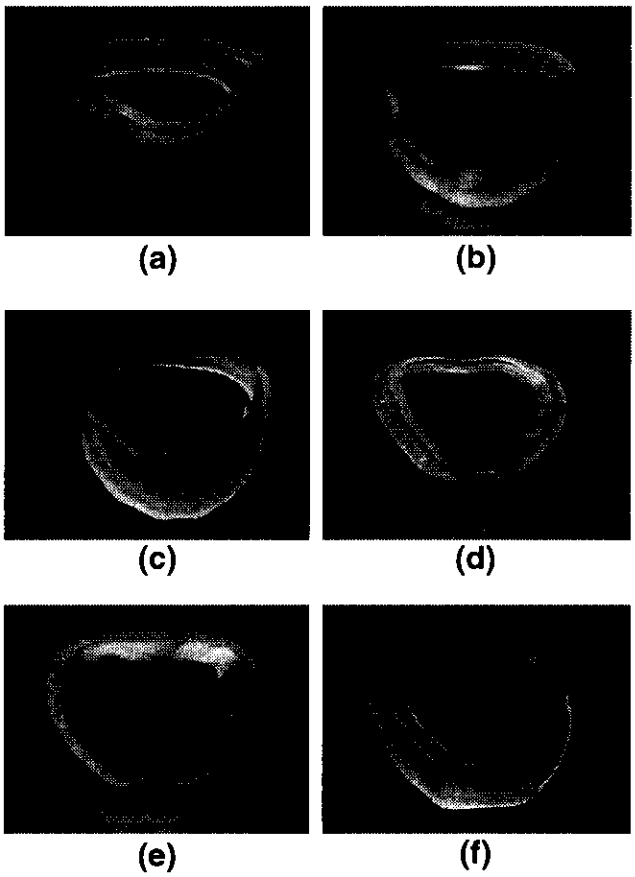

Figure 10. Cross sections of photoforming products, (a) closed mouth; (b) /a/; (c) /i/; (d) /u/; (e) /e/; (f) /o/.

Figure 9. Photoforming product from lip motion 'aiueo'.

Figure 11. Photoforming product from lip motion 'momotaro'.

References

[1] Kenji MASE and Alex PENTLAND. Automatic Lipreading by Optical-Flow Analysis, IEICE, Vol.J73-DII, No.6, 1990.

[2] Michael KASS, Andrew WITKIN and Demetri TERZOPOULOS. Snakes : Active Contour Models, International Journal of Computer Vision, 1988.

[3] G.AVID. Determining three-dimensional motion and structure from optical flow generated by several moving objects, IEEE Trans. Vol.PAMI-7, No.4, 1985.

[4] Tomio WATANABE. Vowels Recognition by Machine Lip Reading, JASME Vol.53, No.496(C), 1987 (in Japanese).

[5] Tomio WATANABE. Machine Lip Reading of Two Concatenated Vowels, JASME Vol.55, No.509(C), 1989 (in Japanese).

[6] Masami NAKANO and Tomio WATANABE. Machine Lip-Reading of Japanese Vowels Utilizing A Stereoscopic Vision System, JASME Vol.60, No.570(C), 1994 (in Japanese).

[7] Masashi OKUBO and Tomio WATANABE. Sensory Evaluation of Preference of 3D shape in Virtual and Real Environments, 6th IEEE Workshop on Robot and Human Communication, 1997.