Lip Shape Extraction for Word Recognition by Using Hardware Active Contour Model

Yuusuke SASAKI, Takao KAWAMURA and Kazunori SUGAHARA
Faculty of Engineering, Tottori University
Koyama-Minami, Tottori 680-8552, JAPAN
{kawamura,sugahara}@ike.tottori-u.ac.jp

Abstract

Even in noisy environments such as in machine factories or in crushes, we are supposed to control equipments by using human voices. Although word recognition techniques have been successfully developed, noises in environments causes serious recognition problems. In such cases, lip shape movements are expected to be useful as supporting data to improve the performance of recognition. In this paper, a method to extract outer and inner lip shape movements from input face images by using Active Contour Model is discussed. The proposed method is implemented as hardware circuits in a FPGA chip to process NTSC video signals in real-time. Experimental results of lip shape extraction are shown to confirm effectiveness of the proposed method.

1 Introduction

Word recognition techniques have been developed, so that even unspecified voices uttered by unspecified person can be recognized in high recognition rate. However, most of these techniques are applicable in the calm places. In noisy environments, noise causes serious recognition problems and as a result the recognition rate becomes low. But word recognition is practically expected to be utilized in noisy environments. Considering these points, supplemental use of visual information of lip shape movements is expected to improve the performance. In order to use lip shape movements for word recognition, a fast and accurate extraction technique of lip shapes from face images is required. In this paper, the Active Contour Model is realized as hardware circuits in one FPGA chip and is used to extract outer and inner lip shape movements. Input face image signals are compliant with NTSC video specifications and each frame is 640 x 480 [pixels] size and 24bit color images.

2 Binary transformation according to HLS components

The Active Contour Model is operated on binary images. The binary transformation according to HLS (Hue, Luminance, Saturation) values is introduced here. The relation among three parameters of HLS are represented as shown in Fig.1. The colors of lips are contained in the region surrounded by solid lines in Fig.1. Six parameters such as h_u, h_l, l_u, l_l, s_u, s_l are used to adjust color. The resultant binary images obtained from three vowel-speaking images are shown in Fig.2. The light source has no influence on these three images, whereas the conditions of skin can cause noises as seen around nose and cheek in Fig.2.

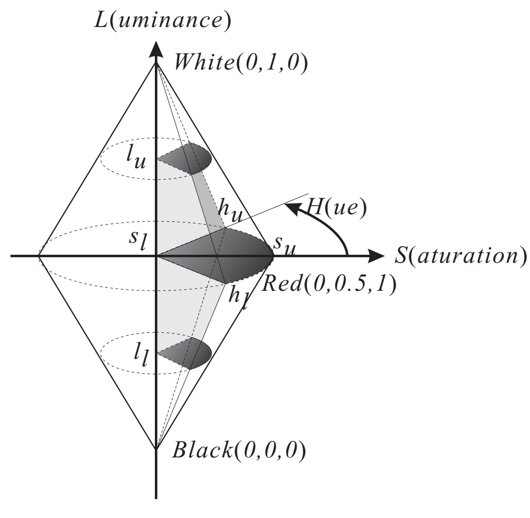


Figure 1: HLS cone model.

(a) 'a'

(b) 'i'

(c) 'o'

Figure 2: Binary transformed images according to HLS parameters.

3 Active Contour Models

3.1 Active Contour Models (snakes)

Active Contour Models are energy minimizing closed curves guided by external constraint forces, and they are influenced by image potentials that pull them toward certain features such as lines or edges [1].

3.2 The Sampled Active Contour Model

Hashimoto et. al. developed the idea of snakes and proposed the Sampled Active Contour Models (S-ACM) [2]. The original S-ACM is made up of a closed curve with some contour points. Three forces, i.e. the pressure F_P, the attractions $F_a's$ and the repulsion F_r, work on every contour points. According to the resultant of these forces, each contour point updates its position and finally extracts an area which has certain property. Considering that lips have convex shapes, pressures are counted out in this paper.

The outline of the S-ACM used in this paper is shown in Fig.3. The closed curve with some contour points updates its position according to the attractions worked on each contour point. The two attractions F_{a1}, F_{a2} work on every contour point and their magnitudes are proportional to the distance between the target point and adjacent contour points. The attractions F_a is resultant of F_{a1} and F_{a2}.

In order to improve the performance of exact area extraction, the additional force, i.e. the vibration factor F_v, is introduced here [3]. The magnitude of F_v is constant, and F_v works perpendicularly to F_a. The direction of the vibration factor reverses each turn of convergence. Due to the vibration factor, every contour point moves zigzag while it converges. As a result of that, the performance is superior to that of the ordinal S-ACM in the ability of escaping from noises. New position of the contour point is decided by the resultant forces of F_a and F_v.

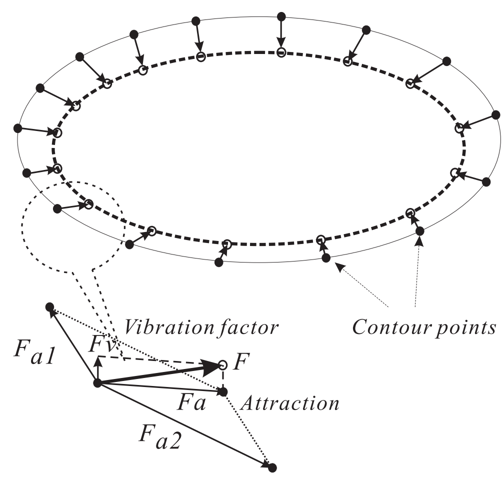


Figure 3: Attraction and vibration factor.

When a contour point hits the edge of an input image, the repulsion F_r works to cancel the vertical component of $F = F_a + F_v$ and the contour point stays at the edge as shown in Fig.4. When it comes to the model like this, the contour point caught in the noise behaves as follows.

- Slip along the noise and escape from it.
- Penetrate the noise when the vertical component of the F exceeds the threshold value Break-edge which is selected in advance.

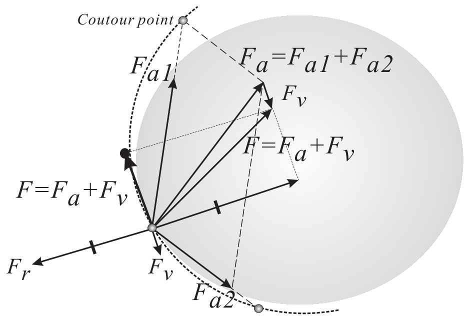


Figure 4: Repulsion.

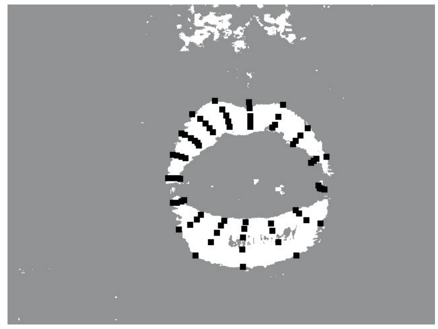


Figure 6: Extraction example of inner shape of a lip.

4 Extraction of inner and outer shape of lips

Applying the S-ACM to the binary images obtained by the method mentioned in Sec.2, outer shapes of lips can be obtained [4]. One example of obtained outer lip shape is shown in Fig.5. In this figure, initial positions of the contour points are set around the given image.

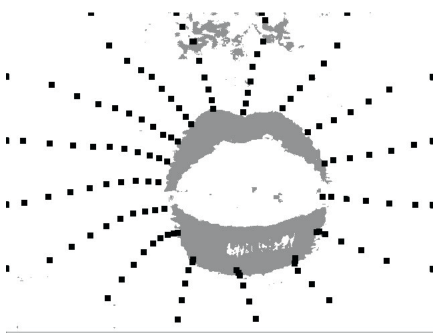


Figure 5: Extraction example of outer shape of a lip.

Inner lip shape is able to be obtained by applying the S-ACM to inverse of binary facial images as shown in Fig.6. Here, the initial positions of the contour points are set to the positions obtained by above outer lip shape extraction processes. An example of obtained inner lip shape is shown in Fig.6.

5 Hardware realization of S-ACM and experimental results

5.1 Hardware realization of S-ACM

The S-ACM is realized as hardware circuits in a FPGA and is equipped in the real-time lip shape extraction system. The system is constructed with the hardware S-ACM, image memories and NTSC video signal decoder. The appearance of the system is shown in Fig.7.

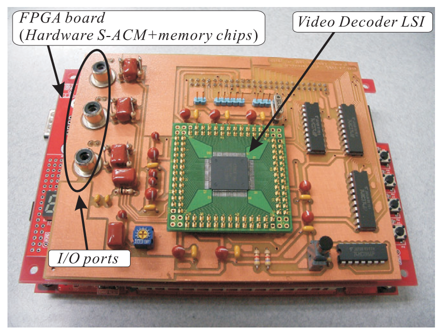


Figure 7: Realtime lip shape extraction system.

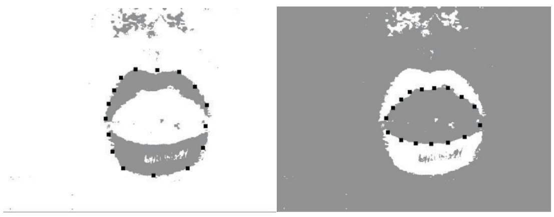
5.2 Experimentation and Results

For the experimentation of outer and inner lip shape extraction, several numbers of facial images are taken with a NTSC video camera. The input signal has 640 x 480 size and full color images. The lip shapes extraction results by using the proposed system are shown in Fig.8.

In Fig.9, a failure example is shown. In this figure, because the binarized nose area is fairly wide, some contour points are caught in the area and the edge is not extracted successfully.

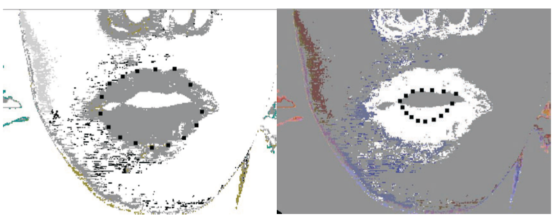
6 Conclusion

In this paper, the S-ACM with the vibration factor is proposed for lip shape extractions. The proposed S-ACM is realized as hardware circuits and is equipped in the real-time lip shape extraction system. The performance of the proposed system is confirmed by the experimentation. By using the proposed system, the performance of the speech

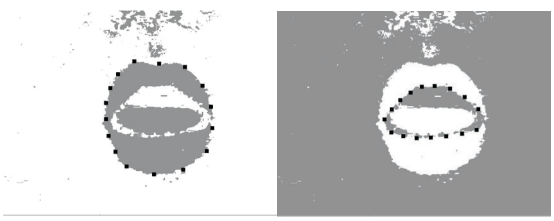


(a) “*a*”

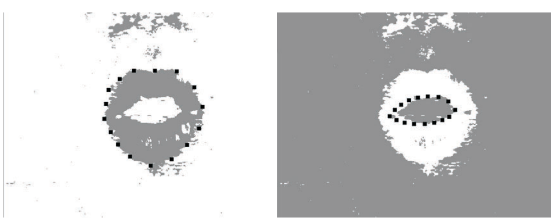
(b) “*i*”



(c) “*u*”



(d) “*e*”



(e) “*o*”

Figure 8: Outer and inner lip extraction examples by using the proposed real-time system.

Figure 9: Failure example of outer lip shape extraction.

recognition can be improved even in the noisy environments.

References

[1] M. Kass, A. Witkin and D. Terzopoulos, “Snakes: Active Contour Models,” International Journal of Computer Vision, pp.321-331, 1998.

[2] M.Hashimoto, H.Kinoshita and Y.Sakai, “An Object Extraction Method Using Sampled Active Contour Model,” IEICE Trans. D-II, Vol.J77-D-II, No.11, pp.2171-2178, 1994.

[3] K.Sugahara, T.Shinchi and R.Konish, “Active Contour Model with Vibration Factor,” IEICE Trans. D-II Vol.J80-D-II, No.12, pp.3232-3235, 1997.

[4] T. Shinchi, Kazunori Sugahara and Ryosuke Konishi, “Vowel Recognition According to Lip Shapes by Using Neural Network,” Proceedings of 1998 IEEE International Joint Conerence on Neural Networks, pp.1772-1777 (1998). Anchorage Alaska