Lip Segmentation and Tracking Based on Chan-Vese Model

Aris Nasuha[1,2]
[1]Electronics Engineering Education Department
Universitas Negeri Yogyakarta
Yogyakarta, Indonesia
aris12@mhs.ee.its.ac.id

Tri Arief Sardjono[2], Mauridhi Hery Purnomo[2]
[2]Electrical Engineering Department
Institut Teknologi Sepuluh Nopember
Surabaya, Indonesia

***Abstract* — Lip reading has wide spread application, e.g. audio-visual Automatic Speech Recognition (AV-ASR), silent speech interface and person identification. Lip segmentation is one of important step in lip reading, because it provides basic information to be processed in subsequent steps. Lip tracking is a process of locating lip to associate lip in consecutive video frames. Chan-Vese model is a region-based segmentation algorithm, which also can be used as tracking method. This algorithm can detect boundary of object which not defined by gradient, while classical active contour can't be applied. This method also can detect object by any initial curve in the image, not necessarily surround the object. This paper investigates about lip segmentation and tracking based on Chan-Vese model, preceded by the color segmentation.**

Keywords—lip reading, segmentation, tracking, Chan-Vese

I. INTRODUCTION

Recently, lip reading becomes more popular research, because it has wide spread applications. The automatic speech recognition (ASR) system is widely used, for example in robotics, personal computer, and cellphone. But, in the noisy circumstances, the ASR performance will drop. By lip reading system, the problem will be overcome. Lip reading system also can be used as one of silent speech interface for laryngectomee (person who has no larynx) [1]. In specific security system, lip reading is used for person identification [2].

Lip reading system consist several steps, and lip segmentation is one of important step, because it provides basic information to be processed in next step. Several lip segmentation methods based on gray-scale image [3-5], other methods are based on color image. Some methods use color image directly [6, 7], the other method use color conversion to other color space to enlarge the color different of lip and the background, for example YCbCr [8], NTSC [9], and CIE-L*a*b* [10-12].

Another important step in lip reading is lip tracking. In this step, the key points of lip are tracked for each frame. There are several lip tracking methods, for example pattern matching snake [13] and watershed [4]. Chan-Vese model is a region-based segmentation algorithm, which can be used as segmentation method and also for tracking. The advantage of this algorithm can detect object in image with contours without gradient or weak gradient, which often presence in face image. In the case of image contour without gradient, classical active contour like Snake, can't be applied. This paper investigates about lip segmentation and tracking based on Chan-Vese model.

II. COLOR TRANSFORMATION AND IMAGE SEGMENTATION

*A. CIE-L*a*b* Color Transformation*

Although most of the image in our database have been stored in RGB color space format, we choose the color space CIE-L*a*b*. Because of the distance between any two points in color space is proportional to the perceived color difference, so we required a uniform space color. CIE-L*a*b* is one of the color space that satisfy uniform space color. Details of the transformation from RGB to CIE-L*a*b* are:

$$\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} 0.490 & 0.310 & 0.200 \\ 0.177 & 0.813 & 0.011 \\ 0.00 & 0.010 & 0.990 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} \tag{1}$$

$$L^* = \begin{cases} 116(Y')^{\frac{1}{3}} - 16 & \text{if } Y' > 0.008856 \\ 903.3Y' & \text{otherwise} \end{cases} \tag{2}$$

$$a^* = 500\left(K_1^{\frac{1}{3}} - K_2^{\frac{1}{3}}\right) \tag{3}$$

$$b^* = 200\left(K_2^{\frac{1}{3}} - K_3^{\frac{1}{3}}\right) \tag{4}$$

$$K_i = \begin{cases} \Phi_i & \text{if } \Phi_i > 0.008856 \\ 7.787\Phi_i + \frac{16}{116} & \text{otherwise} \end{cases} \tag{5}$$

For i = 1, 2, 3

$$\Phi_1 = X' = \frac{X}{X_0}, \Phi_2 = Y' = \frac{Y}{Y_0}, \Phi_3 = Z' = \frac{Z}{Z_0} \tag{6}$$

where X_0, Y_0 and Z_0 are the values of X, Y, Z, for the reference white, respectively. The reference white in the transformation is defined as {R = G = B = 255}

978-1-4799-0425-9/13/$31.00 ©2013 IEEE

B. Chan-Vese Algorithm Overview

Chan-Vese model is a segmentation method based on segmentation method by Mumford-Shah and level set method by Osher and Sethian. The detail of this model can be read in [14]. This algorithm is not based on edge function to stop evolution curve at desired edge. Chan-Vese algorithm can detect boundary of object which not defined by gradient, while classical active contour can't be applied. This method also can detect object by any initial curve in the image, not necessarily surround the object.

The objective of Chan-Vese model is to partition an input scalar image u_0: $\Omega \rightarrow \mathrm{R}$ defined over a d-dimensional image domain $\Omega \subset \mathrm{R}^d$ into two possibly disconnected regions Ω_1 (foreground) and Ω_2 (background) of low intra-region variance and separated by a smooth closed contour C ($\Omega = \Omega_1 \cup \Omega_2 \cup C$). Chan-Vese algorithm can be modeled in equation:

$$E_{CV}(C, c_1, c_2) = \mu|C| + \lambda_1 \int_{\Omega_1} (u_0(X) - c_1)^2\, dx + \lambda_2 \int_{\Omega_2} (u_0(X) - c_2)^2\, dx \quad (7)$$

where c_1 and c_2 represent the average intensity level inside Ω_1 and Ω_2, respectively, and μ, λ_1, and λ_2 are user-defined parameters. The optimal segmentation (C, c_1, c_2) corresponds to a global minimum of (7).

Chan-Vese model has been extended by using a multi-phase level set framework scheme, for piecewise constant (PC) [15] and piecewise smooth (PS) [16] optimal approximations. These models effectively solve the boundary blur or digital object segmentation and can detect interior contours in the image. However, their main drawback is the increased complexity of computing.

A lot of segmentation problems have been solved by Chan-Vese algorithm, for example in medical image segmentation: heart [17], CT bone [18], tooth [19], brain MRI [20]. This algorithm is also applied for segmenting nighttime vehicle license characters [21], and aerial photographs [22].

Chan-Vese algorithm can be applied in object tracking [23] and fluorescent cell tracking [24]. This algorithm can be described as sequential segmentation, where the final contour from one image frame is used as the initial contour for the next. The main strength of the algorithm is its ability to handle the topology changes that result from deformations in the object being tracked.

III. IMPLEMENTATION

In order to test our proposed method, we use image from FEI Face Database [25]. All images are colorful and taken against a white homogenous background. In this database, 20 subjects in frontal face image are randomly chosen. Each subject has two expressions, one with a neutral or non-smiling expression and the other with a smiling facial expression. Original image size is 360x260 pixels, then manually cropped to get only mouth region. The final image size is 56x114 pixels.

A. Lip Segmentation

Color image clustering is used in lip segmentation. There are many clustering algorithms, and k-means are often used in color image segmentation. To get a good clustering result, inputs of clustering should be chosen precisely. In order to enlarge the color different of lip and the background, input images are converted to CIE-L*a*b* color space. This color space consists of a luminosity layer L*, chromaticity-layer a* indicating where color falls along the red-green axis, and chromaticity-layer b* indicating where the color falls along the blue-yellow axis. Since the color information exists in the a* b* space, the clustering inputs are pixels with a* and b* values.

As a comparison, other color space, i.e. using Cb and Cr component of YCbCr color space, and also I and Q component of NTSC, is applied. The result of lip segmentation is used as input for Chan-Vese algorithm. The block diagram of implementation is sketched in Fig.1.

B. Lip Tracking

To simulate lip tracking, the first frame is taken from the neutral or non-smiling expression image, the second from the smiling expression and the third from the non-smiling expression. The result of Chan-Vese segmentation from first frame is used as initial curve for the second, and so forth. The pseudo-code of simple lip tracking algorithm using Chan-Vese model is sketched in Fig.2.

IV. EXPERIMENTAL RESULT

Original Chan-Vese algorithm is an image segmentation method which input is gray-scale image. Therefore, it can be used directly in lip image, with no color segmentation before. Input image is only converted to gray-scale image. But, the result provides that lip segmentation is failed. Examples of these images are seen in Fig.3.

Our proposed method is implemented in lip image of 20 subjects. To measure the accuracy of the segmentation, our proposed method is compared with manual segmentation. Of comparison, the accuracy is calculated by following equation:

$$\text{accuracy} = \frac{TP+TN}{TP+FP+FN+TN} = 1 - \frac{FP+FN}{TP+FP+FN+TN} \quad (8)$$

where TP is number of true positive pixel, TN is number of true negative pixel, FN is number of false negative pixel, and FP is number of false positive pixel.

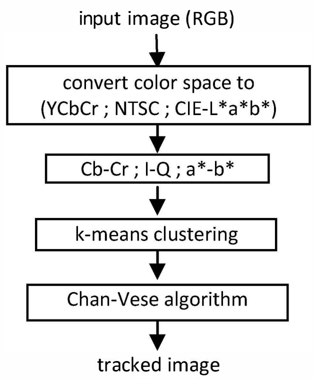

Fig.1. Block diagram of lip segmentation and tracking algorithm

```
contour 0 ← initial contour
For j = 1 to n_frames
{
   contour j ← Chan-Vese (contour j-1, image j)
   draw contour of frame-j
}
```

Fig.2. Pseudo-code of simple lip tracking

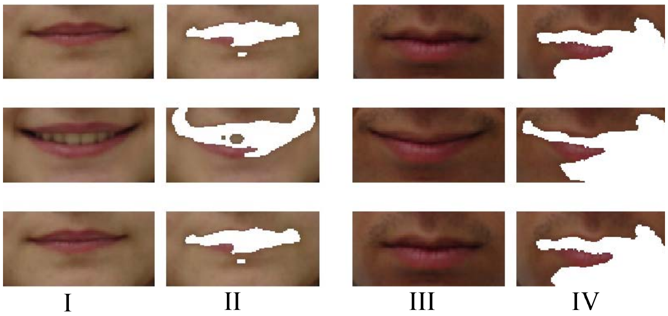

Fig.3. Lip tracking without prior color image segmentation, (I) original image with the presence of teeth, (II) lip tracking of (I), (III) original image without the presence of teeth, (IV) lip tracking of (III)

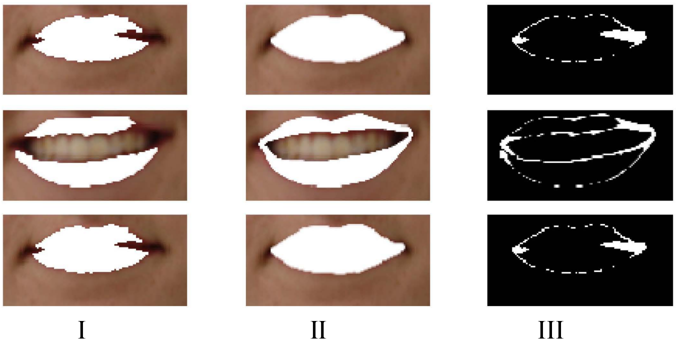

Fig.4. Example of lip segmentation and tracking using k-means where the inputs are Cb and Cr components in YCbCr color space, (I) automatic segmentation, (II) manual segmentation, (III) False Positive and False Negative

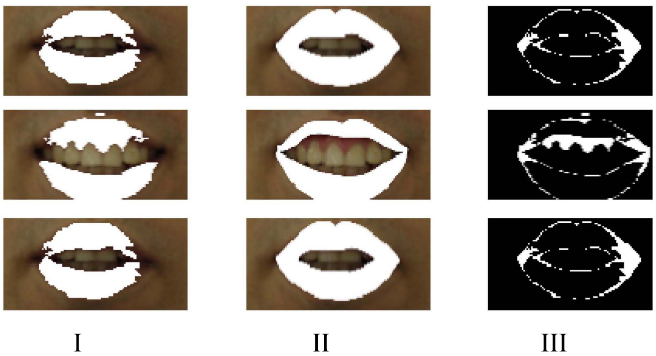

Fig.5. Example of lip segmentation and tracking using k-means where the inputs are I and Q components in NTSC color space, (I) automatic segmentation, (II) manual segmentation, (III) False Positive and False Negative

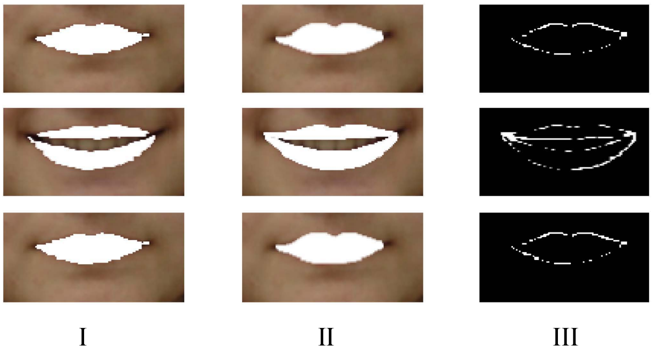

Fig.6. Example of lip segmentation and tracking using our proposed method (k-means where the inputs are a* and b* components in CIE-L*a*b* color space), (I) automatic segmentation, (II) manual segmentation, (III) False Positive and False Negative

The accuracy of segmentation of 60 images, (20 subjects, each of 3 images), for three kinds color image preprocessing, presented in Fig.7. The summary of lip segmentation accuracy is shown in Table 1.

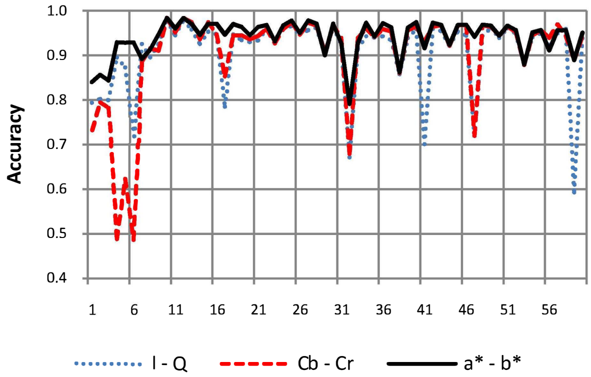

Fig.7. Curve of comparison of lip segmentation accuracy, for three kinds color image processing

Table 1. Summary of lip segmentation accuracy

	Accuracy		
	I - Q	Cb - Cr	a* - b*
The lowest	0.592888	0.486059	0.790570
The highest	0.978383	0.983083	0.983396
Mean	0.907092	0.906386	0.940666
Standard deviation	0.081548	0.107631	0.039814

It can be seen that lip segmentation and tracking using our proposed method is better than the two other methods.

V. Conclusion

This paper presented a new method for lip segmentation and tracking based on Chan-Vese model, preceded by color segmentation using k-means and a*-b* component of CIE-L*a*b* color space. The result shows that our proposed method achieves 94% mean accuracy, and is better than the two other methods, i.e. using I-Q and Cb-Cr components.

As future work, we intend to explore the implementation of our proposed method in real-time application.

REFERENCES

[1] B. Denby, T. Schultz, K. Honda, T. Hueber, J.M. Gilbert, and J.S. Brumberg, "Silent speech interfaces," Speech Communication 52 (2010) 270–287

[2] U. Saeed, "Person identification using behavioral features from lip motion," *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*, vol., no., pp.155,160, 21-25 March 2011

[3] Meng Li and Yiu-Ming Cheung, "Automatic Segmentation of Color Lip Images Based on Morphological Filter,"AMT'10 Proceedings of the 6th international conference on Active media technology Pages 296-305 Springer-Verlag Berlin, Heidelberg, 2010

[4] Siew Wen Chin, Kah Phooi Seng, Li-Minn Ang, and King Hann Lim, "New Lips Detection and Tracking System", Proceedings of the International Multi Conference of Engineers and Computer Scientists 2009 Vol I IMECS 2009, March 18 - 20, 2009

[5] Meng Li and Yiu-ming Cheung. "Automatic lip localization under face illumination with shadow consideration," *Signal Processing*, 89, 12 (December 2009), pp. 2425-2434.

[6] H. Talea, and K. Yaghmaie, "Automatic combined lip segmentation in color images," *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, vol., no., pp.109,112, 27-29 May 2011

[7] Young-Un Kim, Sun-Kyung Kang, and Sung-Tae Jung, "Design and implementation of a lip reading system in smart phone environment," *Information Reuse & Integration, 2009. IRI '09. IEEE International Conference on*, vol., no., pp.101,104, 10-12 Aug. 2009

[8] Xiaoping Wang, Yufeng Hao, Degang Fu, and Chunwei Yuan, "ROI processing for visual features extraction in lip-reading," *Neural Networks and Signal Processing, 2008 International Conference on*, vol., no., pp.178,181, 7-11 June 2008

[9] S.L. Wang, W.H. Lau, and S.H. Leung, "Automatic Lip contour extraction from color image", *Pattern Recognition*, 37, 2004, pp.2375-2387.

[10] E. Skodras, and N. Fakotakis, "An unconstrained method for lip detection in color images," *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*, vol., no., pp.1013,1016, 22-27 May 2011

[11] S. L. Wang, W. H. Lau, S. H. Leung, and A. W. C. Liew, "Lip segmentation with the presence of beards," *Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04). IEEE International Conference on*, vol.3, no., pp.iii,529-32 vol.3, 17-21 May 2004

[12] M.M. Hosseini, and S. Ghofrani, "Automatic Lip Extraction Based on Wavelet Transform," *Intelligent Systems, 2009. GCIS '09. WRI Global Congress on*, vol.4, no., pp.393,396, 19-21 May 2009

[13] Mark Barnard, Eun-Jung Holden, and Robyn Owens, "Lip tracking using pattern matching snakes," *ACCV2002: The 5th Asian Conference on Computer Vision*, 23–25 January 2002

[14] T.F. Chan, and L.A. Vese, "Active contours without edges," *Image Processing, IEEE Transactions on*, vol.10, no.2, pp.266,277, Feb 2001

[15] L.A. Vese, T.F. Chan, "A multiphase level set framework for image segmentation using the Mumford-Shah model", International Journal of Computer Vision, 50 (2002) 271-293.

[16] T.M. Le, and L.A. Vese, "Additive and multiplicative piecewise smooth segmentation models in functional minimization approach", Contemporary Mathematics, 445 (2007) 207-223.

[17] O. Rousseau, and Y. Bourgault, "Heart Segmentation With An Iterative Chan-Vese Algorithm", University of Ottawa, Ontario , 2009

[18] P. T. H. Truc, S. Lee, and T. Kim; "A Density Distance Augmented Chan-Vese Active Contour for CT Bone Segmentation", *Engineering in Medicine and Biology Society, 2008. EMBS 2008. 30th Annual International Conference of the IEEE*, vol., no., pp.482,485, 20-25 Aug. 2008

[19] S. Shah, A. Abaza, A. Ross and H. Ammar, "Automatic Tooth Segmentation Using Active Contour without Edges", *Biometric Consortium Conference, 2006 Biometrics Symposium: Special Session on Research at the*, vol., no., pp.1,6, Sept. 19 2006-Aug. 21 2006

[20] S. Gao, and Y. Yan, "Brain MR Image Segmentation via a Multiphase Level Set Approach" Journal of Information & Computational Science 9: 16 (2012) 4705–4711

[21] S. Gao, and L. Wang, "Local Chan-Vese Model for Segmenting Nighttime Vehicle License Characters", Journal of Information and Computing Science Vol. 6, No. 2, 2011, pp. 123-128

[22] P. Ahmadi, S. Sadri, R. Amirfattahi, and N. Gheissari, "Automatic Aerial Image Segmentation based on a Modified Chan-Vese Algorithm", *Image and Signal Processing (CISP), 2012 5th International Congress on*, vol., no., pp.643,647, 16-18 Oct. 2012

[23] M. Moelich and T. Chan. "Tracking Objects with the Chan-Vese Algorithm". UCLA CAM Reports, 03-14, 2003.

[24] M. Maska, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano, "Fast tracking of fluorescent cells based on the Chan-Vese model," *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*, vol., no., pp.1316,1319, 2-5 May 2012

[25] fei.edu.br/~cet/facedatabase.html