Lip Extraction for Lipreading and Speaker Authentication

Shilin Wang
Shanghai Jiaotong University, China

Alan Wee-Chung Liew
Griffith University, Australia

INTRODUCTION

In recent years, there is a growing interest in using visual information for automatic lipreading (Kaynak, Zhi, Cheok, Sengupta, Jian, & Chung, 2004) and visual speaker authentication (Mok, Lau, Leung, Wang, & Yan, 2004). It has been shown that visual cues, such as lip shape and lip movement, would greatly improve the performance of these systems. Various techniques have been proposed in the past decades to extract speech/speaker relevant information from lip image sequences. One approach is to extract the lip contour from lip image sequences. This generally involves lip region segmentation and lip contour modeling (Liew, Leung, & Lau, 2002; Wang, Lau, Leung, & ALiew, 2004), and the performance of the **visual speech recognition** and **visual speaker authentication** systems depends much on the accuracy and efficiency of these two procedures.

Lip region segmentation aims to label the pixels in the lip image into lip and non-lip. The accuracy and robustness of the lip segmentation process is of vital importance for subsequent lip extraction. However, large variations caused by different speakers, lighting condition, or make-ups make the task difficult. The low color contrast between lip and facial skin, and the presence of facial hair, further complicate the problem. Given a correctly segmented lip region, the lip extraction process then involves fitting a lip model to the lip region. A good lip model should be compact, that is, with a small number of parameters, and should adequately represent most valid lip shapes while rejecting most invalid shapes. As most lip extraction techniques involve iterative model fitting, the efficiency of the optimization process is another important issue.

BACKGROUND

Accurate and robust lip region segmentation is of key importance for subsequent lip extraction. Techniques developed for lip segmentation are generally based on color space analysis, edge detection, Markov random field, or fuzzy clustering.

The color space analysis approach (Eveno, Caplier, & Coulon, 2001) identifies the lip pixels solely by their color information. However, color space-based methods are sensitive to poor color contrast and noise, and would give large segmentation error if the color distributions of lip and background regions overlap. The edge detection approach (Caplier, 2001) relies on the luminance or color edge information to detect the lip boundary. It works well when the speakers use lipstick or reflective markers. However, it would have difficulty dealing with unadorned lips. Markov random field (MRF) technique has also been used in lip region segmentation (Lievin & Luthon, 1999). MRF exploits local neighborhood information to enhance the robustness of the segmentation. However, MRF-based segmentation usually produces erroneous patches outside and inside the mouth region due to the presence of pixels with the wrong color distribution class.

Fuzzy clustering is another powerful tool for image segmentation. **Fuzzy clustering** attempts to assign a probability value to each pixel in order to minimize the fuzzy entropy. Since it is an unsupervised learning method, fuzzy clustering is capable of handling lip and skin color variation caused by make-up. Recently, we have proposed several novel fuzzy-clustering-based segmentation techniques that take the local (Liew, Leung, & Lau, 2000, 2003) and **global spatial information** (Leung, Wang, & Lau 2004; Wang, Lau, Liew, & Leung, 2007) into account to improve the segmentation performance. In our approaches, **spatial information** is seamlessly incorporated into the cost function and the optimization process.

Many techniques have been proposed for **lip modeling and extraction,** and they differ from each other in the following aspects:

- **The lip model used:** Active contour models (Snakes) (Eveno, Caplier, & Coulon, 2003; Lievin, Delmas, Coulon, Luthon, & Fristot, 1999), deformable templates (Hennecke, Prasad, & Stork 1994; Liew et al., 2002), active shape models (ASM) (Cootes, Hill, Taylor, & Graham, 1994; Luettin, Thacker, & Beet, 1996), and active appearance models (AAM) (Cootes, Edwards, & Taylor, 2001; Matthews, Cootes, Bangham, Cox, & Harvey 2002) are some of the widely used lip models.

- **The cost function used:** Edge-based, intensity-based, and region-based cost functions are some of the typical approaches for evaluating the model fitness.
- **The optimization procedure:** Since iterative technique is often required to search for the best parameters, the convergence speed and the stability are the two important issues in optimization. The choice of cost function would affect the optimization scheme.

In our recent work (Wang, Lau, & SLeung, 2004), we have proposed a new lip modeling and extraction algorithm that fits a 16-point lip model by minimizing a region-based **cost function** using a point-driven **optimization scheme**.

FUZZY-CLUSTERING-BASED LIP IMAGE SEGMENTATION

Lip region segmentation is a difficult problem. Figure 1 shows a lip image and its corresponding color distribution for the lip and non-lip pixels in the CIE-1976 CIELAB color space, where * and + represent the lip and background (or non-lip) pixels, respectively. The hue image, with the hue definition given in Zhang and Mersereau (2000), and the edge map are also shown. We see that lip and non-lip pixels overlap severely in the color space, and cannot be easily

Figure 1. (a) Original lip image; (b) Color distribution in CIELAB color space; Color distribution projection on (c) L-a plane, (d) L-b plane, (e) b-a plane; (f) Hue map of the lip image; (g) Edge map based on hue information; (h) Edge map based on intensity information