Chapter XIV

Fuzzy Clustering-Based Approaches in Automatic Lip Segmentation from Color Images

Shilin Wang, Shanghai Jiaotong University, China
Wing Hong Lau, City University of Hong Kong, China
Alan Wee-Chung Liew, Chinese University of Hong Kong, China
Shu Hung Leung, City University of Hong Kong, China

ABSTRACT

Recently, lip image analysis has received much attention because the visual information extracted has been shown to provide significant improvement for speech recognition and speaker authentication, especially in noisy environments. Lip image segmentation plays an important role in lip image analysis. This chapter will describe different lip image segmentation techniques, with emphasis on segmenting color lip images. In addition to providing a review of different approaches, we will describe in detail the state-of-the-art classification-based techniques recently proposed by our group for color lip segmentation: “Spatial fuzzy c-mean clustering” (SFCM) and “fuzzy c-means with shape function” (FCMS). These methods integrate the color information along with different kinds of spatial information into a fuzzy clustering structure and demonstrate superiority in segmenting color lip images with natural low contrast in comparison with many traditional image segmentation techniques.
INTRODUCTION

It is well known that the dynamics of the lip shape contain much information related to both the speaker and the utterance. The increasing interest in extracting such lip gestures from lip image sequence stimulates a variety of applications for which the extracted visual information can improve the performance of the respective systems. One major type of application is audio-visual speech recognition (Kaynak et al., 2004; Nakamura, 2002). The famous McGurk effect (McGurk & MacDonald, 1976) demonstrates the bimodal nature of speech understanding, and also shows a promising way to improve the performance of automatic speech recognition (ASR) systems by integrating visual information. Speaker identity authentication is another type of applications exploiting the visual information extracted from lip images. Recent research shows that lip dynamics is an important cue for personal identity verification (Kanak et al., 2003; Mok et al., 2004) and it can also enhance the performance of voice-based and face-based methods (Wark et al., 2000; Li et al., 2003; Yemez et al., 2003).

Facing the growing demands of extracting visual information from lip movements, various techniques have been proposed in the past decades. These techniques can be divided into two major categories: image-based, that is, model-less approaches (Bregler & Konig, 1994) and model-based approaches (Liew et al., 2002; Wang et al., 2004). Information retention is the key advantage of the image-based approaches, however, feature selection and extraction and linguistic information extraction are nontrivial tasks. They are also of high dimensionality in general. On the other hand, the lip features extracted by a model-based approach are usually related to the physical lip characteristics and of much lower dimensionality; they are also invariant to translation, rotation, scaling and illumination. Hence, the model-based approaches have been widely studied in recent years and various lip models have been proposed.

The most critical problem for model-based approaches is how to extract the lip region from the entire lip image, which is called the lip region segmentation problem. The accuracy and robustness of the lip segmentation process is of key importance for subsequent processing. Recently, segmenting color lip images has become more popular than segmenting gray-scale images due to the availability of low-cost hardware and increasing computing power. Color provides additional information that is not available in gray-scale images and thus the robustness of a lip segmentation algorithm can be enhanced. However, the large variations exhibited in different images caused by different speakers, utterances, lighting conditions or makeup create difficulties for this task. Low chromatic and luminance contrast between the lip and facial skin for an unadorned face is another major problem for lip region detection.

In this chapter, we describe how to automatically segment the lip region from color lip images. The chapter is organized as follows. We first provide a brief review of various kinds of lip segmentation techniques reported in recent literature. Major merits and drawbacks of these methods are also explicated. Then we describe in detail state-of-the-art fuzzy clustering-based techniques recently proposed by our group for color lip segmentation: The “spatial fuzzy c-mean clustering” (SFCM) method (Liew et al., 2003) and the “fuzzy c-means with shape function” (FCMS) method (Leung et al., 2004). From the experimental results, we show that our methods are able to provide accurate and robust lip region segmentation result even for lip images with undesirable color contrast.