Chapter III
Lip Modelling and Segmentation

A. Caplier
GIPSA-lab/DIS, France

S. Stillittano
GIPSA-lab/DIS, France

C. Bouvier
GIPSA-lab/DIS, France

P. Y. Coulon
GIPSA-lab/DIS, France

ABSTRACT

Lip segmentation is the first step of any audio-visual speech reading system. The accuracy of this segmentation has a major influence on the performances of the global system. But this is a very difficult task. First of all, lip shape can undergo strong deformations during a speech sequence. As many other image processing algorithms, the segmentation task is also influenced by the illumination conditions and by the orientation of the object to be segmented. In this chapter, we present an overview about lip modeling and lip segmentation (region-based and contour-based methods). We limit our study to the problem of lip segmentation in frontal faces. Section I gives an overview about the chrominance information that is used for lip segmentation and a comparison between different chrominance cues is proposed. Section II presents region-based approaches and training steps. Section III focuses on contour-based approaches and parametric lip models. Section IV inventories methods for lip segmentation accuracy evaluation. Some specific applications are briefly presented in section V.
WHICH IS THE BEST COLOR SPACE?

The main issue is to determine the most appropriate color space to make the difference between skin pixels and lip pixels. The problem of skin or non skin pixels characterization has been extensively studied in the context of face detection. But the point here is to be able to distinguish skin pixels from lip pixels (all these pixels belonging to a face). Concerning region based lip segmentation algorithms, the optimal color space is the one in which skin pixels and lip pixels are represented by two compact and distinct groups of pixels (low intra-class variances and high inter class variance). Regarding contour based lip segmentation algorithms, the optimal color space is a space in which the gradient information between lip and skin pixels is accentuated.

We propose a summary of the different color spaces that have been used for lip and skin pixels separation. We first focus on current color spaces such as RGB, HSV, and YCbCr which appear not completely efficient. Then we describe some specific color information that has been specially introduced in order to increase the difference between skin and lip pixels. We try to demonstrate the discriminative power of each considered color space by comparing the repartition of some skin pixel samples and some lip pixel samples. A specific database of skin and lip pixel samples has been built. All these samples have been manually extracted from a database of 150 frames representing 20 subjects acquired with the same camera and with the same conditions of illumination. For the purpose of comparison the dynamic of all the chromatic measures has been normalized to the range [0, 1] and the intra-class and inter-class variances have been computed.

In a second step, we present a set of specific luminance and/or chrominance cues that have been developed in order to accentuate lip boundaries gradient.

Current Color Spaces

**RGB Color Space**

We propose to study the RGB color space relevance for lip segmentation using our database of skin and lip pixel samples. In Figure 1, we give the histograms of the lip and skin pixel samples for the 3 color components R, G and B and the pixels repartition in the subspaces (R,G), (G,B) and (R,B). The intra-class and inter-class variances are also given in Table 1.

It is obvious that the color distributions of skin and lip pixel samples overlap each other for each R, G and B color component. We can also see the overlapping of the pixel repartition in the RGB subspaces. On Table 1, the inter-class variances are particularly low which also displays a strong overlapping.

We show in Figure 2 an example of the RGB components for a given mouth image. Visually it is very difficult to interpret the image. On the input image the lips seem to have more red color than the skin, but on the R component the values corresponding to the skin seem higher than those for the lips. Despite the fact that the RGB color space is very often used for displaying images and for computer graphics applications, it is not easy to interpret how the chromatic and luminance information is mixed in the RGB components. As we can see R, G, B channels cannot be used directly for lip segmentation because of the strong correlation between light and color information.