Lip Contour Extraction Based on Support Vector Machine

Xiaosheng Pan, Jiangping Kong, Alan Wee-Chung Liew

Peking University, Chinese University of Hong Kong, Griffith University

Tianchi03@163.com

Abstract: Lip contour extraction was a useful technique for obtaining a mouth shape in an image, and was one of the most important techniques for human-machine interface applications, such as lip reading and speech recognition. In this paper, a new method to extract lip contour from video was propose based on the fact that the lip color and skin-color was varied in the different color spaces. We first extracted frames from digital video first; then we classified face into lip area and non-lip area by the Support Vector Machine. At last we obtained some parameters from the lip area to reconstruct the lip contour. The experiment results proved that the proposed method was accurate and robust.

Keywords: Lip contour, Segment, Color space, Support Vector Machine

1. Introduction

Continuous lip information from a speaker is useful in speech recognition system especially in a noisy environment. In color image, we usually use lip and skin color information to separate lip and non-lip area. Many image segmentation techniques have been proposed [1][2][3][7]. For color image segmentation, histogram-based and clustering-based methods have been widely used. In [2], [3]Cheng et al. proposed a histogram segmentation technique which involves performing a fuzzy partition on a two-dimensional(2-D) histogram based on the maximum fuzzy entropy principle. In [7], alan et al. proposed a segmentation method which use fuzzy clustering and also take account for spatial information.

There are two main problems in lip segment. First, the lip color is similar with tongue color, so inner contour of the lip is difficult to separate from the tongue. Second, the outer contour of lip is often blurred due to the movement of the lip. The algorithm we proposed can successfully resolve the second problem but not the first one.

This paper is organized in four sections. Section two describes the approach proposed here. Section three explains the experimentation setup for testing the approach and gives the results on classification. Section four outlines major conclusions as well as gives directions for future work.

2. The method

2.1. Color Space

As we know, there are many kinds of color spaces that have been provided in existing research work, such as RGB, HSV, CMY, and Chromatic color space[4]. The original lip images are in the RGB color format. It is well known that the value [R G B] of the color image in the color space RGB not only represent chrominance but also luminance; they have high correlation. In most cases, mapping the original image into an adaptive color space can let the feature extraction task easier to be taken. After the comparison of the color spaces listed above, the chromatic color space is chosen as the color space applied in this paper. Three approximately color spaces are the 1976 CIELAB color space (L, a, b), the 1976 CIELUV color space (L, u, v) and YIQ color space (Y, I, Q). Every color component of lip image can be calculate by equation from[4]. Every color component of lip image is list in Figure.1. Because the component L in CIELAB and CIELUV has same value, so we only chose one of them. It is also found that the lip contour of the component v in color space CIELUV and the component b in color space CIELAB can be directly observed dissimilar with the original one, in particular in the area of lip corners.
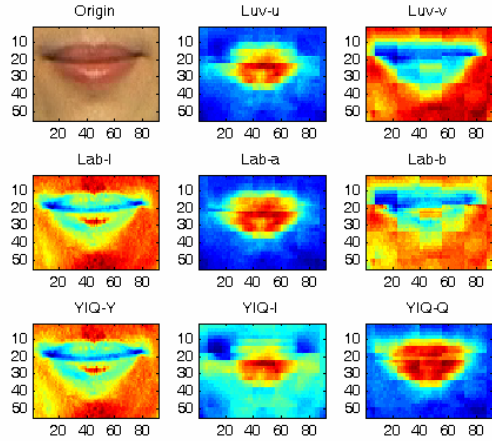
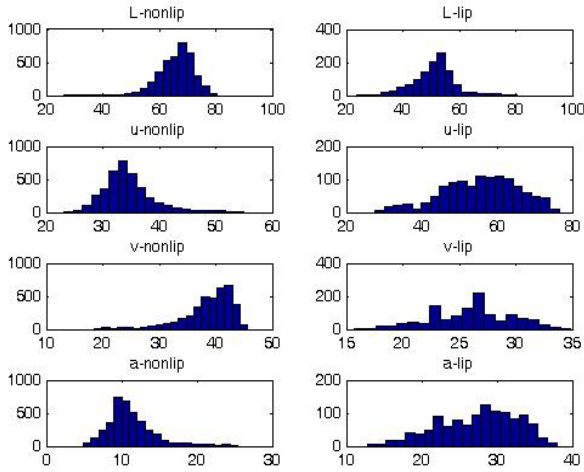**Figure.1** Different color component of lip image
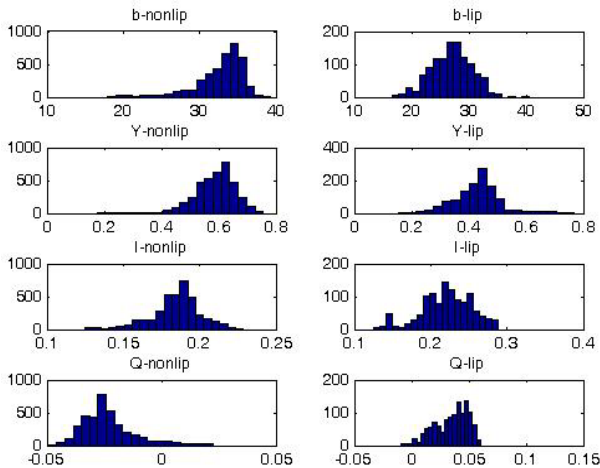

Figure.2 the histogram of color component

Figure.3 the histogram of color component

2.2. Histogram of lip and non-lip area

We know histogram represents the color distribution of lip and non-lip area. Figure.2 and

Figure.3 show the lip area and non-lip area of different color components. The histograms in the left column represent the non-lip area pixel distribution, whereas that in the right column indicated the lip area pixel distribution. If the overlap between the pixel distraction histogram of the lip area and that of the non-lip area exceeds a certain magnitude, then it is very difficult to distinguish lip and non-lip areas. The histograms of feature v have some overlaps between lip and non-lip area. Thus it is not useful for lip contour extraction, so is that of feature b. Then feature vector defined as below:

$$F = \{L, u, a, Y, I, Q\}$$

(1)

2.3. Support Vector Machine

Support Vector Machine (SVM) has recently been successfully applied to a number of applications ranging from face identification to text categorization. The approach is systematic and motivated by statistical learning theory[5]. SVM are based on the structural risk minimization principle, closely related to regularization theory. Here we focus on SVM for two-class classification. The mathematic expression of SVM is:

$$\min_w \phi(w, b) = \frac{1}{2}(w \cdot w)$$

（2）

subject　to the constraints:

$$y_i \mid (w \cdot x_i + b) \mid \geq 1, i = 1, ..., n$$

（3）

And the learning task can be reduced to minimization of primal Lagrangian:

$$L = \frac{1}{2}(w \cdot w) - \sum_{i=1}^{n} a_i(y_i(w \cdot x + b) - 1)$$

(4)

Where a_i are Lagrangian multipliers, hence $a_i \geq 0$.

Taking the derivatives with respect to b and w and resubstituting back the primal gives the Wolfe dual Lagrangian:

$$W(a) = \sum_{i=1}^{n} a_i - \frac{1}{2} \sum_{i=1}^{n} a_i a_j y_i y_j (x_i \cdot x_j)$$

(5)

which must be maximized with respect to a_i subject to the constraint:

$$a_i \geq 0 \qquad\qquad \sum_{i=1}^{n} a_i y_i = 0$$

(6)

The decision function is then given by:

$$f(x) = \text{sgn}((w \cdot x) + b)$$
$$= \text{sgn}(\sum_{i=1}^{n} a_i^* y_i (x_i \cdot x) + b^*)$$

(7)

An SVM is mainly characterized by its kernel function. In experiment, we use linear kernel function with parameter C=1.

3. Implementation and result

The lip video is provided by the Linguistics Lab of Department of Chinese Language and Literature of Peking University.

The lip contour extraction procedure is as following:

Step1. Get frames from the video.

Step2. Mark lip contour by handwork[6], separate the face into two component, lip and non-lip. And obtain the axes of every pixel.

Step3. Map the origin lip images into other color space and extract the feature vector F.

Step4. Select appropriate pixels as training data[8], then use SVM classify the lip images.

Step5. Extract parameters from lip area and use lip model provided from [6] to obtain a lip contour with smooth edge.

The result of the lip segment from the continuous frame is shown in Figure.4. The left column is four continuous frames extract from video. The middle column is the result of we classified the face into lip area and non-lip area by the Support Vector Machine. We are not very satisfied with the result. Because the lip edges we get are not smooth enough, and it is found that the non-lip area in the lip corners with dark color is misidentified as lip area. It is also the case for the concave non-lip area immediately below the under lip. Because the color in those position is more similar to lip area.

Alan et al. provide a lip model[6], the equations describing the lip shape Figure.5 are given by:

$$y_1 = \frac{-h_1}{(w - x_{off})^2}(|x - sy_1| - x_{off})^2 + h_1$$

(8)

$$y_2 = h_2\left(\left(\frac{x - sy_2}{w}\right)^2\right)^{1+\delta}$$

(9)

As its origin is at $(0,0)$, $x \in [-w, w]$. The parameter shows the skewness of lip shape. The skewing is attained with the ways of using the transformation matrix $\begin{bmatrix} 1 & s \\ 0 & 1 \end{bmatrix}$. The exponent δ describes the deviation of y_2 from a quadratic curve and it allows lip with different degree of curvature to be captured accurately.

Then we extract parameters from middle column and use equations (8) and (9) to get the lip contour. The result of lip contour is list in the right column of Figure.4.
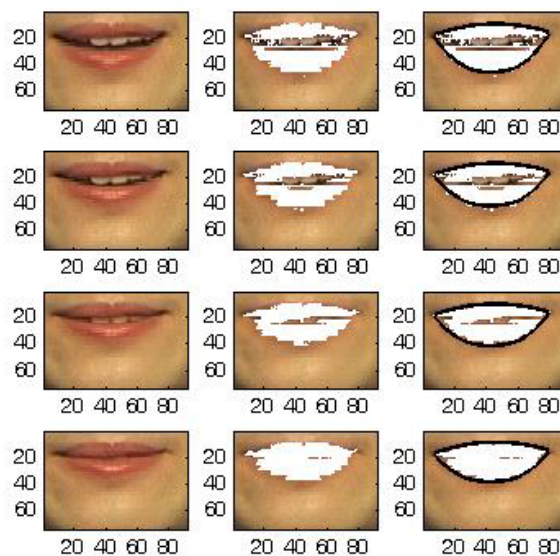
Figure.4 The left column is the lip image sequence of speaking. The middle column is corresponding the results of classify face into lip and non-lip area. The

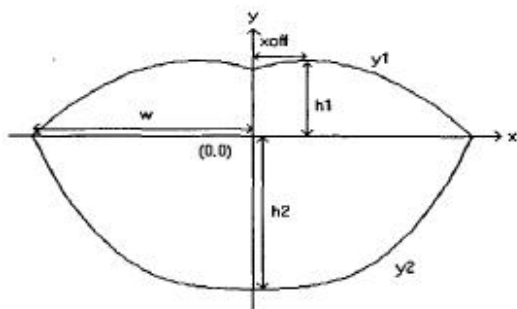right column is the lip contour optimized by lip model.

Figure.5 Lip model

4. Discussion

There still something can be done to improve the accuracy of my lip segment program. For example inner lip contour is still can not be extracted for the tongue color is similar to lip color, and get inner lip contour is our next stage target. And make some preprocessing to origin image maybe can eliminate the shadow at the underlip. Then the result of lip segment will be improved probably.

When using Support Vector Machine , we need to choose a kernel function, and decide the parameter values to kernel function. There is still no theory to prove which kernel function is best, so we need to do more experiment to find out which kernel function is the most suitable for our algorithm.

5. Reference

[1]Lewis TW, Powers D M, "Lip Feature Extraction Using Red Exclusion", In Proc. Selected papers from Pan-Sydney Workshop on Visual Information Processing, 2000, pp.61-67.

[2]. H. D. Cheng, Y. H. Chen, and X. H. Jiang, "Thresholding using two dimensional histogram and fuzzy entropy principle," IEEE Trans. Image Processing, vol. 9, pp. 732–735, Apr. 2000.

[3] H. D. Cheng, J. R. Chen, and J. Li, "Threshold selection based on fuzzy C-partition entropy approach," Pattern Recogn., vol. 31, no. 7, pp. 857–870, 1998

[4]Adrian Ford and Alan Roberts, "Colour Space Conversions", August 11, 1998 (coloureq.pdf)

[5] V. N. Vapnik. Statistical Learning Theory. Wiley, 1998.

[6] A.W.C. Liew, S.H. Leung, W.H. Lau, "Lip contour. extraction using a deformable model," In Proc. IEEE ICIP, Vol. 2, pp. 255-258, 2000.

[7]Liew A, Leung S H,Lau W H, et al. Segmentation of Color Lip Images by Spatial Fuzzy Clustering[J].IEEE Trans. on Fuzzy Systems,2003,11(4):542-549.

[8]边肇琪 张学工等编著 《模式识别》第二版 清华大学出版社 2002 年 10 月