LIP SEGMENTATION WITH THE PRESENCE OF BEARDS

S.L.Wang[†], W.H.Lau[†], S.H.Leung[*] and A. W. C. Liew[†]

[†]Department of Computer Engineering and Information Technology
*Department of Electronic Engineering
City University of Hong Kong, 83 Tat Chee Avenue, Hong Kong

ABSTRACT

Lip image analysis has attracted much interest in recent years because some important speech information is contained in the shape and movement of the lip. To extract such information from the lip images, accurate and robust lip region segmentation is of vital importance. However, most of the current lip segmentation methods fail to provide accurate results if the person has beards. In this paper, we propose a "one object, multiple background" clustering method to solve the problem. Since the non-lip region becomes inhomogeneous in the presence of beards, multiple background clusters can produce better fitting to a rather complex background region than one single cluster. Spatial information in terms of the physical distance towards the lip center is incorporated to enhance the differentiation between the lip and background region. Experimental results demonstrate that our algorithm provides accurate lip segmentation results for the images with beards.

1. INTRODUCTION

The performance of automatic speech recognition (ASR) system decreases drastically due to the presence of the noise. It has been shown that incorporating visual information extracted from lip images analysis can greatly enhance the accuracy of speech recognition systems solely based on acoustic signal [1,2]. Lip region segmentation becomes very important since it provides lip pixel information for the subsequent lip modeling process.

Many researchers have proposed various algorithms for robust lip region segmentation. Some methods are based on color space analysis, i.e., segmenting lip region directly from the color space [3] or transforming the color representation to other color space to enlarge the color difference between the lip and the background [4]. These algorithms can segment the lip region in a very efficient manner. However, it is difficult to find the optimum value of the threshold setting for various lip images. Markov Random Field (MRF) techniques have been used in some lip segmentation algorithms [5,6] to improve the performance by exploiting the local spatial information. The hue edges are also considered to provide useful information for lip segmentation [6]. These MRF based methods perform well in dealing with "pepper" noise. Fuzzy clustering is another kind of widely used image segmentation technique. The classical fuzzy c-means (FCM) algorithm attempts to assign a probability value for each pixel to minimize a fuzzy entropy measure [7]. As the clustering method is an unsupervised learning method where neither prior assumption about the underlying feature distribution nor training is needed, it is able to handle various lip and skin color due to different human race or make-up.

However, for lip images with various kinds of beards, most of the methods mentioned above cannot segment the lip region accurately. In this paper, we proposed a new fuzzy clustering algorithm that provides accurate lip region segmentation even for the lip images with beards. Detail description of the objective problem is shown in Section 2. In Section 3, the theory and implementation of the proposed algorithm are presented. Experimental results are given in Section 4 and Section 5 draws the conclusion.

2. OBJECTIVE PROBLEM DESCRIPTION

Segmenting lip images with beards of different color in the skin region remains an open question for most of the previous lip segmentation techniques [5]. In order to demonstrate the difficulty of segmenting these images, an image with beards is shown in Fig.1 (a) along with the color distribution of the lip pixels (in red) and non-lip pixels (in blue) in the CIELAB color space. To facilitate the investigation of the usefulness of edge information in segmentation, the edge maps of the hue [6] and luminance of the original image are also given in Fig. 1.

Three important aspects of the lip image are observed from Fig. 1. Firstly, the pixels of lip region overlap with the background pixels in the color space (Fig1 (b)-(e)). Thus, methods solely depend on the color information will experience unsolvable troubles in partitioning these overlapping pixels. Exploiting local spatial information using MRF may improve the performance to some extent. However, if the number of pixels of similar color is large enough, holes inside the lip and patches outside are often observed. Secondly, as the color distribution of the background region is quite complex, the traditional one background class is not enough to model the background region sufficiently. Thirdly, there are many erroneous edges in the edge maps as shown in Figures 1(f) and 1(g) and they

The work described in this paper is fully supported by a research grant (CityU 1215/01E) from the RGC of the HKSAR, China.

can hardly be used as an indicator of the boundary information between the lip and the background.

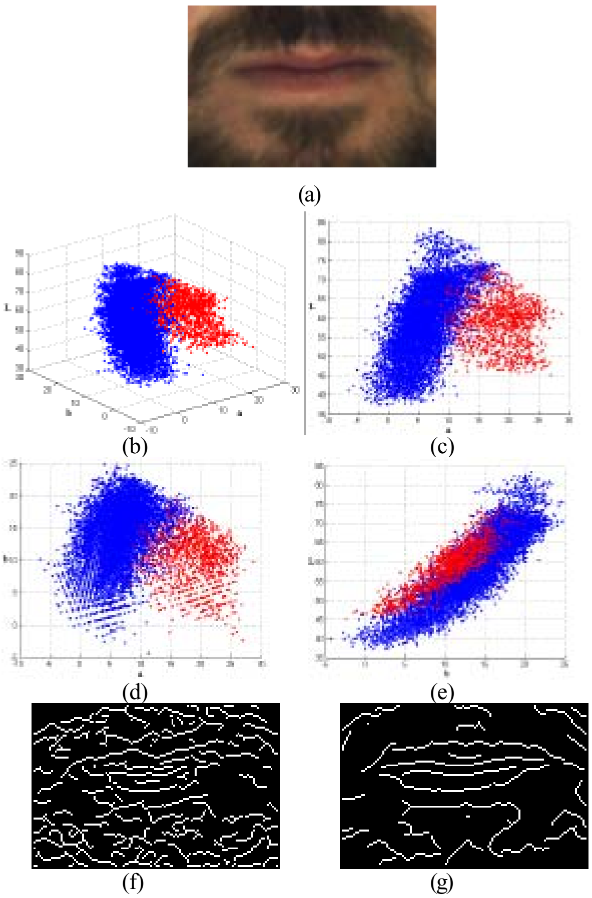

Fig.1 (a) original lip image, (b) color distribution of (a) in CIELAB color space, (c)-(e) color distribution projection on the *L-a*, *b-a*, *L-b* plane, respectively, (f) the edge map of the hue image, (g) the edge map of the luminance image.

3. THE PROPOSED FUZZY CLUSTERING ALGORITHM

3.1 General description

In order to overcome the difficulties mentioned above, we propose a "one object, multiple background" clustering method for lip region segmentation. In our algorithm, one cluster is set for representing the lip region and several clusters for the background. The physical distance towards the lip center is included in the dissimilarity measure as the spatial information. We can formulate the spatial dissimilarity measure of the lip and non-lip respectively and thus the differentiation between the lip region and the background is enhanced. The purpose of using " multiple background" is to sufficiently model the non-lip region. Since the FCM clustering method uses Euclidean distance as its dissimilarity measure, the clusters extracted by FCM are of spheroidal shapes. Thus, one cluster is not sufficient to model the non-lip region due to the large color variation in the background. Increasing the number of clusters improves the fitness of the background clusters catering for the wide color distribution.

3.2 Objective function formulation

Let $\boldsymbol{X}=\{\boldsymbol{x}_{1,1}, \ldots, \boldsymbol{x}_{r,s}, \ldots, \boldsymbol{x}_{N,M}\}$ be the set of feature vectors associated with an image I of size N by M, where $\boldsymbol{x}_{r,s} \in \mathrm{R}^q$ is a q-dimensional color vector for a pixel located at (r,s). Let $d_{i,r,s}$ represents the Euclidean distance between the feature vector $\boldsymbol{x}_{rs}$ and the centroid $\boldsymbol{v}_i$ of the i^{th} cluster (i=0 for the lip cluster and $i \neq 0$ for the background cluster).

A spatial distance measure is designed in an elliptic form since the shape of the outer lip contour is resembled to an ellipse. Let $\boldsymbol{p} = \{x_c, y_c, w, h, \theta\}$ denotes the parameter set that describes the elliptic distance, where (x_c, y_c) is the center of mass of the ellipse, w and h are respectively the semi-major axis and semi-minor axis, and θ is the inclination angle about (x_c,y_c). With the lip cluster specified, the dissimilarity measure combining both the color and spatial information, denoted by $D_{i,r,s}$, is formulated as follows:

$$\begin{aligned} D_{i,r,s} &= d_{i,r,s}^2 + \alpha \times dist(r,s)^{\rho_L} \quad & i = 0 \\ D_{i,r,s} &= d_{i,r,s}^2 + \alpha \times dist(r,s)^{-\rho_B} \quad & i \neq 0 \end{aligned} \tag{1}$$

where

$$dist(r,s) = \frac{\left((r-x_c)\cos\theta + (s-y_c)\sin\theta\right)^2}{w^2} + \frac{\left((s-y_c)\cos\theta - (r-x_c)\sin\theta\right)^2}{h^2}.$$

The weighting parameter α in (1) is for adjusting the weight of the physical distance against the color feature. The exponents ρ_L and ρ_B determine the shape of the spatial dissimilarity function in lip and background clusters. Generally they are assigned with positive values so that the spatial dissimilarity provides a correct guidance for differentiating the lip and background region.

The objective function of our algorithm is given by,

$$J_m(\boldsymbol{U},\boldsymbol{V},\boldsymbol{p}) = \sum_{r=1}^{N}\sum_{s=1}^{M}\sum_{i=0}^{C-1} u_{i,r,s}^m D_{i,r,s} = J_{mc} + \alpha \times (J_{mL} + J_{mB}) \tag{2}$$

subject to
$$\sum_{i=0}^{C-1} u_{i,r,s} = 1, \quad \forall (r,s) \in I \tag{3}$$

where the $N \times M \times C$ matrix $\boldsymbol{U} \in M_{fc}$ is a fuzzy c-partition of $\boldsymbol{X}$, $\boldsymbol{V}=\{\boldsymbol{v}_0, \boldsymbol{v}_1, \ldots \boldsymbol{v}_{C-1}\} \in R^{Cq}$ with $\boldsymbol{v_i} \in R^q$ is the set of fuzzy cluster centroids, $m \in (1,8)$ defines the fuzziness of the clustering and the value $u_{i,r,s}$ gives the membership of the (r,s)-th pixel in the fuzzy cluster C_i. J_{mc} in (2) is the same as the objective function for fuzzy c-means, which is responsible for the color part. J_{mL} and J_{mB} represent the spatial cost function of lip and background clusters, respectively.

3.3 Parameter updating method

Since the optimum solution for $\min\{J_m(\boldsymbol{U},\boldsymbol{V},\boldsymbol{p})\}$ is the stationary point of the objective function J_m, Picard iteration is used to solve for the optimum point $(\boldsymbol{U}^*,\boldsymbol{V}^*,\boldsymbol{p}^*)$. The derivation for the parameter updating formula in each iteration is outlined in the following.

By keeping $\boldsymbol{V}\in\boldsymbol{R}^{cq}$ and $\boldsymbol{p}\in\boldsymbol{R}^{5}$ unchanged and taking the partial derivative of J_m with respect to $\boldsymbol{U}$ subject to the constraints in (3), we have:

$$u^{+}_{i,r,s}=\left[\sum_{j=0}^{C-1}\left(D_{i,r,s}/D_{j,r,s}\right)^{1/(m-1)}\right]^{-1} \tag{4}$$

Next, we fix $\boldsymbol{U}\in\boldsymbol{M}_{fc}$ and $\boldsymbol{p}\in\boldsymbol{R}^{5}$ and take the derivative of J_m with respect to $\boldsymbol{V}$. Since both J_{mL} and J_{mB} remain constant when $\boldsymbol{U}\in\boldsymbol{M}_{fc}$ and $\boldsymbol{p}\in\boldsymbol{R}^{5}$ are fixed, the derivative is the same as that of FCM. Hence we have:

$$\boldsymbol{v}_i^{+}=\sum_{r=1}^{N}\sum_{s=1}^{M}u^{m}_{i,r,s}\boldsymbol{x}_{r,s}\Big/\sum_{r=1}^{N}\sum_{s=1}^{M}u^{m}_{i,r,s} \tag{5}$$

Similarly, the partial derivative of J_m with respect to $\boldsymbol{p}$ is given by:

$$\frac{\partial J_m(\boldsymbol{U},\boldsymbol{V},\boldsymbol{p})}{\partial \boldsymbol{p}}=\frac{\partial J_{mc}}{\partial \boldsymbol{p}}+\alpha\left(\frac{\partial J_{mL}}{\partial \boldsymbol{p}}+\frac{\partial J_{mB}}{\partial \boldsymbol{p}}\right) \tag{6}$$

It should be noted that the first term of the right hand side in (6) vanishes since the spatial parameter set $\boldsymbol{p}$ is independent to the color part of the objective function. Setting the derivative equal to zero, we have:

$$\sum_{r=1}^{N}\sum_{s=1}^{M}\left(\rho_L u^{m}_{0,r,s}dis(r,s)^{\rho_L-1}-\rho_B\sum_{i=1}^{C-1}u^{m}_{i,r,s}dis(r,s)^{-\rho_B-1}\right)\frac{\partial dis(r,s)}{\partial \boldsymbol{p}}=0 \tag{7}$$

Since directly computing $\boldsymbol{p}^{+}$ from (7) is quite complicated, instead the Conjugate Gradient (CG) method is adopted to solve $\boldsymbol{p}^{+}$ numerically for its fast convergence.

In each parameter updating process, the objective function J_m is a continuous function in $(\boldsymbol{U},\boldsymbol{V},\boldsymbol{p})$ with positive values and it is always decreasing. Hence, the iterative process described above must converge to a local minimum of J_m.

3.4 Implementation procedure

In our experiments, the RGB lip images are quantized to 8 bits per color. Since the RGB color space is not visually uniform, the lip image will be transformed to the approximately uniform CIELAB and CIELUV color space with details given in [8]. The color vector $\{L^*, a^*, b^*, u^*, v^*\}$ is used to represent the color information of pixels. Since the presence of teeth can disturb the membership distribution which in turn causes bias to the color centroids, the method described in [9] is used to mask the teeth pixels if teeth do appear in the image.

The clustering algorithm runs as follows:

1. Initialize the color centroids $\boldsymbol{V}$.
2. With the spatial dissimilarity set to zero, compute the initial membership distribution $\boldsymbol{U}$ given in (4).
3. Update $\boldsymbol{V}$ given in (5) and compute the spatial parameter set $\boldsymbol{p}$ using the CG method.
4. Calculate the dissimilarity measure in (1) and update $\boldsymbol{U}$ given in (4).
5. Repeat steps 3 and 4 for k = 1, 2, 3, … until $\left\|U^{(t+1)}-U^{t}\right\|_{\infty}<\varepsilon_T$ or k=$k_{\max}$, where ε_T is a small threshold and $k_{\max}$ is the maximum number of iterations.

After clustering, the membership values in each cluster are smoothed by a 3x3 Gaussian low-pass filter before segmentation takes place. The final lip-background segmentation is a hard classification process by assigning the pixel to the lip cluster if the corresponding membership value for the lip cluster is the highest among all the clusters.

4. EXPERIMENTAL RESULTS

110 lip images with beards and 90 lip images without beards have been taken from the " AR Face Database" [10] for testing the performance of our algorithm. In our experiment, the fuzziness m is set to 2 and the parameter set $\{\alpha,\rho_L,\rho_B\}$ in (1) are set to {10,5,3} empirically. These choices are shown to provide good balance between the color and spatial information. Since the estimation of an appropriate number of background cluster is rather complicated and time-consuming, the number of background clusters is then initially set to one, i.e. C = 1. After clustering, the standard deviation of the color distance to the centroid for all pixels in each cluster is calculated. The clustering process will be repeated with C=C+1 if any of these standard deviations is larger than a preset value σ (σ =7 in our experiment). To find the initial color centroids, FCM is applied to analyze an arbitrary image. The color centroid of the cluster situated in the center portion of the image is assigned as the lip cluster. These initial color centroids are used for all the 200 lip images.

An example image with beards is used to illustrate the segmentation results, as shown in Fig. 2, of the proposed algorithm. The results obtained from both the traditional FCM and the Zhang's method [6] are also given for comparison. It can clearly be seen from Fig. 2 that the segmentation result produced by our algorithm is much better than that of the others. It has been observed that the traditional FCM approach is unable to differentiate the color similarity for some of the lip pixels and non-lip pixels. As a result, patches close to or far away from the lip are resulted as shown in Fig. 2(b). The two-class setting of Zhang's method does not produce satisfactory segmentation result as shown in Fig. 2(c). Even with the aid of the hue edges, its performance is still unsatisfactory.

To investigate the number of clusters required for segmenting the lip region from the 110 images with beards, it is found that 12 images require 2 clusters, 92 images require 3 clusters and only 6 images requires 4 clusters. It is clear that majority of the images require 3 clusters for segmenting the lip region accurately. If the color of the beards is close to that of the skin or it is light color, 2 clusters are sufficient for the segmentation.

In order to objectively evaluate the performance of the proposed algorithm, we apply a quantitative technique described in the following. The boundary of the lip for the original image in Fig. 2 is manually drawn. The Segmentation Error (SE) is defined as [11]:

$$SE = P(O) \cdot P(B \mid O) + P(B) \cdot P(O \mid B) \qquad (7)$$

where $P(B|O)$ is the probability of classifying background as object, $P(O|B)$ is the probability of classifying object as background. $P(O)$ and $P(B)$ are the a priori probabilities of the object and the background of the image, respectively. The SE as well as the two misclassifying probability of the three algorithms are shown in Table 1. It is observed from the table that the SE of our algorithm is much smaller than that of the other two algorithms. More segmentation results for images with beards produced by our algorithm are given in Fig. 3. The SE for the 200 lip images selected from the "AR Face Database" is ranged from 0.5% to 5%. These results demonstrate that the segmented lip region obtained by our algorithm fits well to the actual lip.

	$P(B\|O)$	$P(O\|B)$	SE (%)
FCM	1.683	0.080	26.77
Zhang's	3.119	0	36.56
Our method	0.014	0.030	**2.81**

Table.1 Comparison of $P\{B|O\}$, $P\{O|B\}$ and SE among FCM, Zhang's and our method for the lip image shown in Fig.2.

5. CONCLUSIONS

A fuzzy clustering based algorithm for segmenting the lip region in the presence of beards is presented in this paper. Since the background region of a lip image is inhomogeneous, we formulate the lip region segmentation process as a "one object, multiple background" clustering problem. With the aid of the spatial information which is incorporated in the dissimilarity measure, our algorithm is able to differentiate pixels with similar color features but located in different regions. Experimental results show that our algorithm provides accurate lip segmentation results even for lip images with beards.

6. REFERENCES

[1] M. T., Chan, " HMM-based audio-visual speech recognition integrating geometric and appearance-based visual features", *Proc of IEEE 4th Workshop on Multimedia Signal Processing*, pp. 9-14, Cannes, Oct. 2001.

[2] E. D. Petajan, " Automatic Lipreading to enhance Speech Recognition" , *Proc of IEEE Conf on Computer Vision and Pattern Recognition*, pp. 40-47, 1985.

[3] T. Wark, S. Sridharan, V. Chandran, "An approach to statistical lip modelling for speaker identification via chromatic feature extraction," *Proc of 14th Int'l Conf on Pattern Recognition*, vol.1, pp.123-125, Brisbane, Aug. 1998.

[4] N. Eveno, A. Caplier, P. Y. Coulon, "New color transformation for lips segmentation," *Proc of IEEE 4th Workshop on Multimedia Signal Processing*, pp. 3-8, Cannes, Oct. 2001.

[5] M. Lievin, F. Luthon, " Lip features automatic extraction," *Proc of IEEE ICIP*, vol.3, pp. 168-172, Chicago, March 1999.

[6] X. Zhang, R. M. Mersereau, " Lip feature extraction towards an automatic speechreading system," *Proc of IEEE ICIP*, vol.3, pp.226-229, Vancouver, Sept. 2000.

[7] J. C. Bezdek, *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, New York, 1981.

[8] R. W. G.. Hunt, *Measuring Color*, 2nd Ed., Ellis Horwood Series in Applied Science and Industrial Technology, 1991.

[9] A. W. C. Liew, S. H. Leung and W. H. Lau, " Segmentation of color lip images by spatial fuzzy clustering," *IEEE Trans on Fuzzy Systems*, vol.11, pp.542-549, August 2003.

[10] A. M. Martinez, R. Benavente, "The AR face database" [Online] CVC Technical Report #24, June 1998.

[11] S. U. Lee, S. Y. Chung, R. H. Park, " A comparative performance study of several global thresholding techniques for segmentation," *Computer Vision, Graphics and Image Processing*, vol. 52, pp.171-190, 1990.

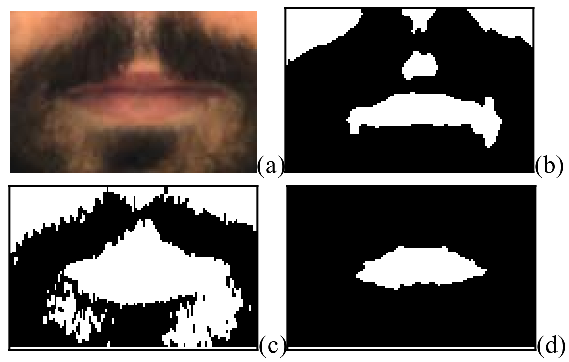

Fig.2 (a) The original lip image; segmentation results by: (b) FCM; (c) Zhang's; and (d) the proposed method.

Fig.3 More lip segmentation results obtained by the proposed algorithm (with the lip region shown in white color).