A Cooperative Learning-Based Clustering Approach to Lip Segmentation Without Knowing Segment Number

Yiu-ming Cheung, *Senior Member, IEEE*, Meng Li, Qinmu Peng, and C. L. Philip Chen, *Fellow, IEEE*

*Abstract***— It is usually hard to predetermine the true number of segments in lip segmentation. This paper, therefore, presents a clustering-based approach to lip segmentation without knowing the true segment number. The objective function in the proposed approach is a variant of the partition entropy (PE) and features that the coincident cluster centroids in pattern space can be equivalently substituted by one centroid with the function value unchanged. It is shown that the minimum of the proposed objective function can be reached provided that: 1) the number of positions occupied by cluster centroids in pattern space is equal to the true number of clusters and 2) these positions are coincident with the optimal cluster centroids obtained under PE criterion. In implementation, we first randomly initialize the clusters provided that the number of clusters is greater than or equal to the ground truth. Then, an iterative algorithm is utilized to minimize the proposed objective function. For each iterative step, not only is the winner, i.e., the centroid with the maximum membership degree, updated to adapt to the corresponding input data, but also the other centroids are adjusted with a specific cooperation strength, so that they are each close to the winner. Subsequently, the initial overpartition will be gradually faded out with the redundant centroids superposed over the convergence of the algorithm. Based upon the proposed algorithm, we present a lip segmentation scheme. Empirical studies have shown its efficacy in comparison with the existing methods.**

*Index Terms***— Clustering, cooperative learning, lip segmentation, number of clusters.**

I. Introduction

SEGMENTING out person's lip from face image has received much attention in the past decades due to the wide range of possible attractive applications, such as lip-reading, audio-visual speech recognition in noisy environment, face detection, biometric person identification, lip synchronization, human expression recognition, and so forth [1]–[6]. In the past decades, a number of image segmentation methods based on different theories and methodologies have been proposed, e.g., see the surveys in [7]–[10]. However, due to the low chromatic and luminance contrast between lip region and skin, which make the segmentation task become challenging, few of them have been applied to lip segmentation successfully.

Wark *et al.* [11] and Zhang and Mersereau [12] utilized some basic image process techniques, such as threshold in the specific color channels of the input image, to obtain the lip region. Although these methods are conducive to implement and have low computation complexity, they are not essentially applicable for the practical cases with complexion difference or various illumination conditions. Pardàs and Sayrol [13], Delmas *et al.* [14], and Eveno *et al.* [15] utilized the gradient-based methods to extract the lip boundary, while the input image is viewed as a vector map. However, the accuracy of these methods is easily affected by false boundary edges caused by shadow, skin pigmentation, and so forth. Matthews *et al.* [16], Eveno [17], and Seyedarabi *et al.* [18] utilized the shape template model-based methods (e.g., snake, active shape model, and active appearance model) to obtain the lip region and achieved the promising results. Nevertheless, the final segmentation accuracy of such a method depends on the initial template position. Moreover, its performance is sensitive to the noisy boundaries brought from the segmentation process.

Recently, clustering-based approach has provided a promising way for lip segmentation. For example, fuzzy C-means (FCM) and K-means clustering-based methods have been employed to perform lip segmentation [19]–[21]. Moreover, the works described in [22]–[24] utilize the statistical models, e.g., Gaussian mixture model and FCM, to estimate the lip membership maps as well. Nevertheless, such methods would miscalculate the membership due to the similarity and overlap between the lip and nonlip pixels in color space. As a result, lip segmentation methods that solely depend on the edge or color information will not deliver the satisfactory performance [25]. Along this category, Liew *et al.* [26] have, therefore, proposed a clustering algorithm by taking spatial restriction into account, which con-

Manuscript received May 14, 2013; revised November 1, 2015; accepted November 2, 2015. Date of publication December 11, 2015; date of current version December 22, 2016. This work was supported in part by the Faculty Research Grant of Hong Kong Baptist University (HKBU) under Project FRG2/12-13/082, Project FRG1/14-15/041, and Project FRG2/14-15/075, in part by the Knowledge Transfer Office of HKBU under Grant MPCF-005-2014/2015, in part by the National Natural Science Foundation of China under Grant 61272366 and Grant 61572540, and in part by the Macau Science Technology Development Fund under Grant 019/2015/A. *(Corresponding author: Y.M. Cheung.)*

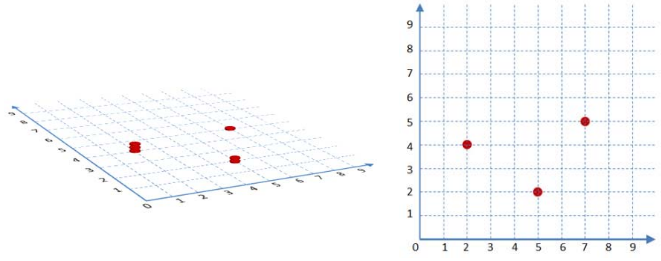
Y. M. Cheung is with the Department of Computer Science and Institute of Research and Continuing Education, Hong Kong Baptist University, Hong Kong, and also with the United International College, Beijing Normal University – Hong Kong Baptist University, Hong Kong (e-mail: ymc@comp.hkbu.edu.hk).

M. Li and Q. Peng are with the Department of Computer Science, Hong Kong Baptist University, Hong Kong (e-mail: mli@comp.hkbu.edu.hk; qmpeng@comp.hkbu.edu.hk).

C. L. P. Chen is with the Faculty of Science of Technology, University of Macau, Macau 999078, China (e-mail: philipchen@umac.mo).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2015.2501547



(a) (b)

Fig. 1. (a) and (b) Relationship between position and centroid, where the plane with grid represents the pattern space. As shown in (a), there are six centroids (denoted by circles) with the coordinates $(2, 4)$, $(2, 4)$, $(2, 4)$, $(5, 2)$, $(5, 2)$, and $(7, 5)$, but they only occupy three positions $(2, 4)$, $(5, 2)$, and $(7, 5)$, as shown in (b).

siders both of the distributions of data in feature space and the spatial interactions between neighboring pixels during the clustering process. In addition, Hara and Chellappa [27] utilized Bayesian information criterion as a measure to choose the suitable cluster number, but it heavily depends on the distribution estimation of the samples. Cheung [28] proposed to learn the number of clusters via maximizing a weighted likelihood.

Another clustering-based lip segmentation method proposed in [29] obtains the spatial continuity constraints by using a dissimilarity index that allows the spatial interactions between the image voxels. Similarly, Leung *et al.* [30] dealt with the lip segmentation using fuzzy clustering with spatial restriction as well. Although these methods have achieved the promising results, their accuracy highly depends on the predefined number of segments, whose selection is, however, often a nontrivial task in practice. As a variant of [30], Wang *et al.* [31] have proposed a multiclass, shape-guided clustering algorithm. This method determines the number of clusters by using the I-index, and employs a penalty term considering the spatial location information to differentiate the pixels that have similar color but are located in different regions. However, the number of clusters is determined by an individual local exhaustive search before the segmentation, i.e., there are redundant data traverses, whose computation is quite laborious. Moreover, similar to the I-index, some existing cluster validity measures, e.g., those in [32]–[37], have the same problem as well.

In this paper, we shall present a fuzzy clustering-based segmentation method, whose objective function is derived from the classical partition entropy (PE) and implemented using Havrda–Charvat's structural α-entropy. This objective function features that the coincident cluster centroids in pattern space (also called input space interchangeably) can be equivalently substituted by one centroid with the function value unchanged. It is shown that the minimum of the proposed objective function can be obtained provided that: 1) the number of positions occupied by the centroids in pattern space is the same as the true number of clusters, as shown in Fig. 1 and 2) these positions are coincident with the optimal cluster centroids obtained under the PE criterion. Thus, the optimal partition can be acquired by minimizing the proposed objective function regardless of whatever the preassigned number of clusters is as long as it is greater than or equal to the ground truth. From the practical viewpoint, it is generally feasible to estimate an upper bound of the number of clusters. In implementation, we, therefore, first assign some cluster centroids (i.e., the learnable data points in the input space toward the cluster centers), whose number is greater than or equal to the ground truth, and initialize them randomly. Subsequently, an iterative algorithm is utilized to minimize the proposed objective function. At each iterative step, not only is the winner, i.e., the centroid with the maximum membership degree, updated to adapt to the corresponding input data (also called observation hereinafter), but also the other centroids are adjusted with a specific cooperation strength, so that they are each closer to the winner. Subsequently, some neighboring centroids will be gradually merged into one, so that the overpartition caused by redundant centroids can be eventually faded out. That is, the clustering performance of the proposed algorithm is robust against the preassigned number of clusters. Based upon the proposed algorithm, a lip segmentation scheme is presented, which is robust against the visibility of mustache, teeth, and tongue. Experiments have shown the efficacy of the proposed approach.

The remainder of this paper is organized as follows. Section II overviews the minimum entropy-based fuzzy clustering method. Section III describes the proposed method in detail. Section IV presents the unsupervised lip segmentation scheme based upon the proposed method. Section V shows the experimental results. Finally, the conclusion is drawn in Section VI.

II. Minimum Entropy Method in Fuzzy Clustering

Clustering is the process of assigning data elements into classes or clusters, so that data in the same class are as similar as possible under a certain similarity measure. In general, the task of image segmentation can be formulated as a clustering problem, i.e., the image segments turn into data clusters, in which the specific property measured in feature space of each pixel can be viewed as the data to be divided.

Fuzzy clustering [38] is a class of algorithms for cluster analysis, in which data elements may belong to more than one cluster, and associated with each element is a set of membership levels. To discuss the image segmentation problem under this framework, we first suppose that the image of interest has s pixels. For the ith pixel, the feature vector utilized in clustering process is denoted by x_i. Then, we define m segments whose centroids are denoted by $c_1, c_2, \ldots, c_m$. The purpose of fuzzy clustering algorithm is to optimize the centroid collection, i.e., $C = \{c_1, c_2, \ldots, c_m\}$, and corresponding partition matrix

$$U = \begin{bmatrix} u_{11} & \cdots & u_{1m} \\ \vdots & \vdots & \vdots \\ u_{s1} & \cdots & u_{sm} \end{bmatrix} \tag{1}$$

with

$$\sum_{j=1}^{m} u_{ij} = 1, \quad (i = 1, 2, \ldots, s) \tag{2}$$

where $u_{ij} \in [0, 1]$ indicates the strength of the association between an input data x_i and cluster c_j.

One of the most popular fuzzy clustering algorithms is the FCM algorithm [39]. In this algorithm, the optimal cluster centroids and partition can be achieved by minimizing the following objective function:

$$J = \sum_{i=1}^{s}\sum_{j=1}^{m}(u_{ij})^p \|x_i - c_j\|^2 \tag{3}$$

where p is a weighting exponent which is a real number greater than or equal to 1.

Moreover, there are several variants of fuzzy clustering methods. From the viewpoint of information theory, the information entropy can be viewed as a measure of the uncertainty. Moreover, the uncertainty of belonging of each input data is reduced during the clustering procedure. Thus, the relationship between clustering and entropy is naturally close. Mathematically, Shannon's entropy [40] of a random variable x with the probability $p(x)$ is defined as

$$H(x) = -\sum_{x} p(x)\log p(x). \tag{4}$$

Based upon Shannon's entropy, Bezdek [39], [41] has proposed a fuzzy clustering criterion named PE to measure the fitness of a fuzzy partition which is shown as

$$H(U, m) = -\frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{m} u_{ij}\log u_{ij}. \tag{5}$$

Bezdek [42] indicates that the partition matrix and cluster number (U^*, m^*) are optimal as long as

$$(U^*, m^*) = \mathop{\arg\min}_{1<m\le m_{\max}}\left\{\mathop{\arg\min}_{U\in\Omega_m}\{H(U, m)\}\right\} \tag{6}$$

where $m_{\max}$ denotes the maximum value of the number of clusters, and Ω_m is the collection of partition matrices with the cluster number m.

Furthermore, Li *et al.* [43] have proposed another version of the minimum entropy criterion of fuzzy clustering, in which the membership degree of x_i in cluster c_j can be measured by the conditional probability. Thus, given s observations, denoted by $x_1, x_2, \ldots, x_s$, (5) can be rewritten as

$$H(C|X) = -\frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{m} p(c_j|x_i)\log p(c_j|x_i) \tag{7}$$

where $X = \{x_1, x_2, \ldots, x_s\}$.

Moreover, for the sake of analysis, Li *et al.* [43] utilized Havrda–Charvat's structural α-entropy [44]

$$H^\alpha(x) = (2^{1-\alpha} - 1)^{-1}\left[\sum_{x} p^\alpha(x) - 1\right] \tag{8}$$

as a substitution of Shannon's entropy, where $\alpha > 0$ and $\alpha \neq 1$. Evidently, different values of α can lead to different entropy measures.

In this paper, the following quadratic entropy with $\alpha = 2$ is selected:

$$H^2(x) = 1 - \sum_{x} p^2(x). \tag{9}$$

Thus, similar to (4) and (7), we can let

$$H(C|X) = 1 - \frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{m} p^2(c_j|x_i) \tag{10}$$

based on (9). To show the validity of this criterion, we formulate the probability of clustering error as

$$P_e = P(C \neq C^*) \tag{11}$$

where C^* denotes the optimal cluster centroid collection.

Based on Fano's inequality [45], we then have

$$H(P_e) + P_e\log(m-1) \geq H(C|X) \tag{12}$$

where $H(P_e)$ is the Shannon's entropy of P_e. As $H(P_e) \leq 1$ and $m \geq 2$, (12) can be further rewritten as

$$P_e \geq \frac{H(C|X) - 1}{\log(m-1)}. \tag{13}$$

Equation (13) indicates that C^* can be estimated with a low error probability only if $H(C|X)$ is small. This implies that minimum $H(C|X)$ could be an appropriate choice for fuzzy clustering [43].

III. Fuzzy Clustering Without Knowing True Number of Clusters

As stated in Section II, $H(C|X)$ is a classical criterion for fuzzy clustering, which, however, depends on the number of centroids. Although the optimal partition can be achieved under this criterion, the oversegmentation or undersegmentation almost always occurs if the number of centroids is not preassigned appropriately. In this section, we propose a variant of $H(C|X)$ which depends on the number of positions occupied by centroids instead of the number of centroids. Moreover, the proposed one inherits the property of $H(C|X)$. That is, when the number of positions occupied by centroids is equal to the true cluster number, the proposed objective function reaches the minimum value. In the following, we will present this method in detail.

A. Proposed Objective Function for Fuzzy Clustering

Given an observation data x_i and the centroid collection C, by adjusting the order of the elements in C, we can obtain

$$\tilde{C}_i = \{\tilde{c}_1^i, \tilde{c}_2^i, \ldots, \tilde{c}_m^i\} \tag{14}$$

satisfying $p(\tilde{c}_k^i|x_i, \bar{C}_k^i) \leq p(\tilde{c}_j^i|x_i, \bar{C}_j^i)$ if and only if $k < j$, where $j, k = 1, 2, \ldots, m$, and $\bar{C}_j^i = \tilde{C}_i - \{\tilde{c}_j^i\}$.

Similar to (10), we propose a new objective function

$$\begin{aligned}&\delta H(C \mid X)\\ &= 1 - \frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{m}\left[\frac{p(\tilde{c}_j^i|x_i, \bar{C}_j^i) - p(\tilde{c}_{j-1}^i|x_i, \bar{C}_{j-1}^i)}{p(\tilde{c}_m|x_i, \bar{C}_m^i)}\right]^2\end{aligned} \tag{15}$$

with $p(\tilde{c}_0^i|x_i, \bar{C}_0^i) = 0$. For simplicity, we will denote $p(\tilde{c}_j^i|x_i, \bar{C}_j^i)$ as $p(\tilde{c}_j^i|x_i)$ without ambiguity. Here, we utilize the difference of conditional probability between adjacent clusters that are two neighboring clusters in $\tilde{C}_i$ (e.g., $\tilde{c}_{j-1}^i$ and

$\tilde{c}^i_j$ in $\tilde{C}_i$), i.e., $p(\tilde{c}^i_j|x_i, \bar{C}^i_j) - p(\tilde{c}^i_{j-1}|x_i, \bar{C}^i_{j-1}) = p(\tilde{c}^i_j|x_i) - p(\tilde{c}^i_{j-1}|x_i)$, to measure the membership degree of x_i to the cluster with the centroid $\tilde{c}^i_j$. Such membership degree depends not only on the distance between x_i and $\tilde{c}^i_j$, but also on the distance between x_i and the other clusters. The learning toward maximizing such membership degree can make the similar centroids speedup approaching the same position, meanwhile forcing other centroids to move away from it.

We define $p(\tilde{c}^i_j|x_i)$ as

$$p\left(\tilde{c}^i_j|x_i\right) = \frac{1}{\sum_{k=1}^{m}\left(\frac{\left\|x_i - \tilde{c}^i_j\right\|}{\left\|x_i - \tilde{c}^i_k\right\|}\right)^2}. \tag{16}$$

Singularity in $p(\tilde{c}^i_j|x_i)$ occurs when one or more of the distances $\|x_i - \tilde{c}^i_k\|$ is equal to zero. In this case, $p(\tilde{c}^i_j|x_i)$ of (16) will be calculated in the sense of a limit. Similar to [46], we assign zeros to each nonsingular class and distribute memberships equally to the singular classes. Note that, (15) is responsible only for cluster center updating, and the membership in (16) is not necessarily optimized by (15).

Equation (15) can be further rewritten as

$$\delta H(C\,|\,X) = 1 - \frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{m}\left[\frac{\left\|x_i - \tilde{c}^i_m\right\|^2}{\left\|x_i - \tilde{c}^i_j\right\|^2} - \frac{\left\|x_i - \tilde{c}^i_m\right\|^2}{\left\|x_i - \tilde{c}^i_{j-1}\right\|^2}\right]^2. \tag{17}$$

The basic property of this objective function is shown as follows.

Theorem 1: Given a centroid collection $C = \{c_1, c_2, \ldots, c_m\}$, a new centroid collection obtained by adding an element c' into C is denoted by $C' = \{c_1, c_2, \ldots, c_m, c'\}$. We have $\delta H(C \mid X) = \delta H(C' \mid X)$ if there exists $c_j \in C$ ($j \in [1, m]$) satisfying $c_j = c'$.

Proof: For specific x_i, C and C' can be written as the ordered forms [see (14)], i.e., $\tilde{C}_i = \{\tilde{c}^i_1, \ldots, \tilde{c}^i_m\}$ and $\tilde{C}'_i = \{\tilde{c'}^i_1, \ldots, \tilde{c'}^i_{m+1}\}$, respectively.

In $\tilde{C}'_i$, we assume the corresponding element of c_j is $\tilde{c'}^i_k$ ($k \in [1, m]$). Since $c_j = c'$, the corresponding element of c' can be written by $\tilde{c'}^i_{k+1}$.

Thus, we have

$$\begin{aligned}\delta H&(C' \mid X)\\ &= 1 - \frac{1}{s}\sum_{i=1}^{s}\left\{\left[\frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_1\right\|^2}\right]^2\right.\\ &\quad + \cdots + \left[\frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_{k+1}\right\|^2} - \frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_k\right\|^2}\right]^2\\ &\quad \left. + \cdots + \left[\frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2} - \frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_m\right\|^2}\right]^2\right\}.\end{aligned} \tag{18}$$

As $\tilde{c'}^i_k = \tilde{c'}^i_{k+1}$, we have

$$\left[\frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_{k+1}\right\|^2} - \frac{\left\|x_i - \tilde{c'}^i_{m+1}\right\|^2}{\left\|x_i - \tilde{c'}^i_k\right\|^2}\right]^2 = 0. \tag{19}$$

Moreover, as $\tilde{C}'_i \backslash \tilde{C}_i = \tilde{c'}^i_{k+1}$, we have

$$\tilde{c'}^i_l = \begin{cases}\tilde{c}^i_l, & l = 1, 2, \ldots, j\\ \tilde{c}^i_{l-1}, & l = k+2, k+3, \ldots, m+1.\end{cases} \tag{20}$$

Thus, (18) can be written as

$$\begin{aligned}\delta H(C'\,|\,X) &= 1 - \frac{1}{s}\sum_{i=1}^{s}\sum_{j=1}^{m}\left[\frac{\left\|x_i - \tilde{c}^i_m\right\|^2}{\left\|x_i - \tilde{c}^i_j\right\|^2} - \frac{\left\|x_i - \tilde{c}^i_m\right\|^2}{\left\|x_i - \tilde{c}^i_{j-1}\right\|^2}\right]^2\\ &= \delta H(C \mid X).\end{aligned} \tag{21}$$

■

According to Theorem 1, the value of (15) depends on the number of the positions of centroids in pattern space but not the number m of centroids. For the sake of description, we define two functions named $PNum(C)$ and $ENum(C)$. The former one returns the number of positions of the centroids in C, and the latter one returns the number of centroids in C. Moreover, we employ $Pos(C)$ to obtain a collection composed by the positions of the centroids in C, where $Pos(C) = \{p_1, p_2, \ldots, p_{PNum(C)}\}$. Furthermore, we let

$$H(C \mid x_i) = 1 - \sum_{j=1}^{m} p^2(c_j \mid x_i) \tag{22}$$

and

$$\delta H(C \mid x_i) = 1 - \sum_{j=1}^{m}\left[\frac{p\left(\tilde{c}^i_j \mid x_i\right) - p\left(\tilde{c}^i_{j-1} \mid x_i\right)}{p\left(\tilde{c}^i_m \mid x_i\right)}\right]^2. \tag{23}$$

Then, we have the following lemma.

Lemma 1: Given an input x and a constant $\eta_m \in (0, (m-1)/m]$, the minimum of $H(C \mid x)$ is approximately equal to η_m subject to $\delta H(C \mid x) = \eta_m$, where $PNum(C) = ENum(C) = m$.

The detailed proof of Lemma 1 is given in Appendix I, and the experimental justification is shown in Appendix II.

Based on Lemma 1, we have the result as follows.

Theorem 2: Given an input x, there exist two centroid collections C_1 and C_2 satisfying $C_1 = Pos(C_2)$, such that $H(C_1 \mid x)$ reaches the minimum value approximately when $\delta H(C_2 \mid x)$ reaches the minimum.

Proof: We utilize the notation $\mathcal{C}$ to represent the solution space of $\delta H(C \mid x) = \vartheta$, where ϑ denotes the global minimum value of $\delta H(C \mid x)$.

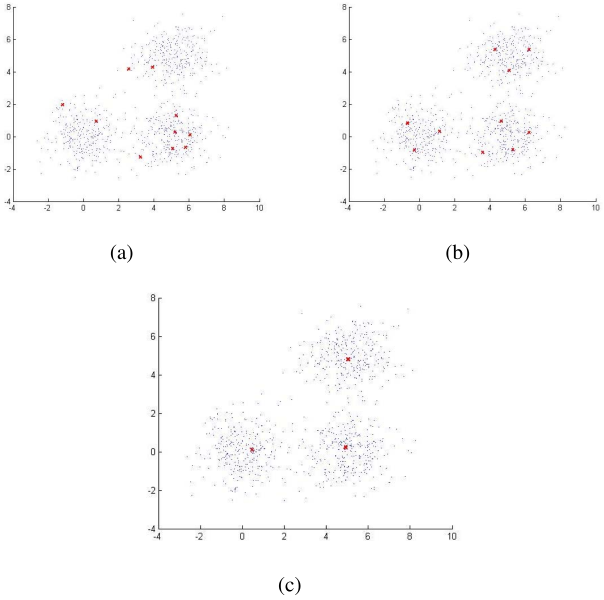
According to Lemma 1, the minimum value of $H(C \mid x)$ with $C \in \mathcal{C}$ is approximately equal to ϑ as well. The corresponding centroid collection is denoted by C_1 with $ENum(C_1) = PNum(C_1)$.

According to Theorem 1, there exists the centroid collection C_2 with $C_1 = Pos(C_2)$, such that $\delta H(C_2 \mid x) = \delta H(C_1 \mid x)$. ■

Recalling the property of PE presented in [42], i.e., see (6), the proposed objective function $\delta H(C \mid X)$ will reach the minimum value provided that:

1) The number of centroid positions in pattern space is equal to the true cluster number, i.e., $PNum(C) = ENum(C^*)$.
2) The positions are coincident with the optimal centroids under the PE criterion, i.e., $Pos(C) = C^*$, where C^* is the centroid collection obtained by minimizing (7) with $m = m^*$.

Please note that we construct (15) inspired by (10). Both (10) and (15) can obtain similar results if the number of clusters is appropriately determined. The main difference is that (10) only works well when choosing an appropriate cluster number. Under the circumstances, the cluster centroids will move to the appropriate position, and H in (10) could reach the minimum. By contrast, (15) can work well as long as the assigned number of clusters is greater than or equal to the true one. When the assigned number of cluster is greater than the true one, optimizing (15) can make the redundant cluster centroids superimposed, thus resulting in the number of positions occupied by the cluster centroids is exactly the true number of clusters.



(a)

(b)

(c)

Fig. 2. (a) Initial positions of centroids marked by "$*$." (b) Demonstrate that the FCM method cannot learn an appropriate cluster number as the number of centroids obtained by it is always equal to the preassigned one. (c) Indicate that the proposed approach can superpose redundant centroids and learn an appropriate number and positions of the centroids automatically.

B. Iterative Algorithm

This section presents an iterative algorithm to perform fuzzy clustering by minimizing the proposed objective function shown in (15). At each iterative step, based upon the idea of cooperative learning initially proposed in [47], this algorithm not only updates the winner centroid in terms of membership degree to adapt to the corresponding input data, but also the other centroids are adjusted with a specific cooperation strength, so that they are each close to the winner. Subsequently, the initial overpartition will be gradually faded out with the redundant centroids superposed over the convergence of the algorithm.

Specifically, we first preassign the segment number m, a value which is greater than or equal to the ground truth, and initialize the centroid collection C randomly. Then, the subsequent implementation is given as follows.

Step 1: Fixing C, we calculate $p(c_j \mid x_i)$ and obtain the collection $\tilde{C}_i$ by (16) for each input data x_i.

Step 2: For each x_i, we update C via

$$c_{j_i^w}^{(\text{new})} = c_{j_i^w}^{(\text{old})} - \eta_w \cdot \left. \frac{\partial \delta H(C|x_i)}{\partial c_{j_i^w}} \right|_{c_{j_i^w}^{(\text{old})}} \tag{24}$$

and

$$c_{j_i^r}^{(\text{new})} = c_{j_i^r}^{(\text{old})} - \eta_r \cdot \left. \frac{\partial \delta H(C|x_i)}{\partial c_{j_i^r}} \right|_{c_{j_i^r}^{(\text{old})}} \tag{25}$$

where $c_{j_i^w}$ is the winner centroid in terms of membership degree with $j_i^w = \arg\max_j (p(\tilde{c}_j^i \mid x_i) - p(\tilde{c}_{j-1}^i \mid x_i))$, $j_i^r = 1, 2, \ldots, m$ but $j_i^r \neq j_i^w$, and η_w and η_r are the positive learning rates. Actually, η_r is also the specific cooperation strength.

The above two steps are processed iteratively. When the number of positions occupied by centroids is equal to the true cluster number, the proposed objective function reaches the minimum value; for more details, refer to [47]. Fig. 2 is to compare the performance of FCM [i.e., use (24) only in Step 2] and the proposed method when redundant centroids exist.

IV. Lip Segmentation and Postprocessing

In this section, we apply the proposed method in Section III to the unsupervised lip segmentation problem. The task is to extract the lip boundary from a color image consisting of a part of face between nostril and chin. A sample of original image is shown in Fig. 3(a).

A. Pattern Space

In general, the image segmentation methods are based on color space rather than gray level because color image can provide more useful clue for segmentation. Furthermore, since hue, saturation, and value (HSV) color space is similar to the way human being perceives [48], we utilize a modified HSV color space as our pattern space.

In HSV color space, the S–H space is represented by polar coordinate system. The distance utilized in our method is Euclidean distance. Thus, a polar-Descartes coordinate transformation is required. We first transform the original image into the HSV color space, in which the HSV components for site i (i.e., pixel i) are denoted by H_i, S_i, and V_i, respectively.

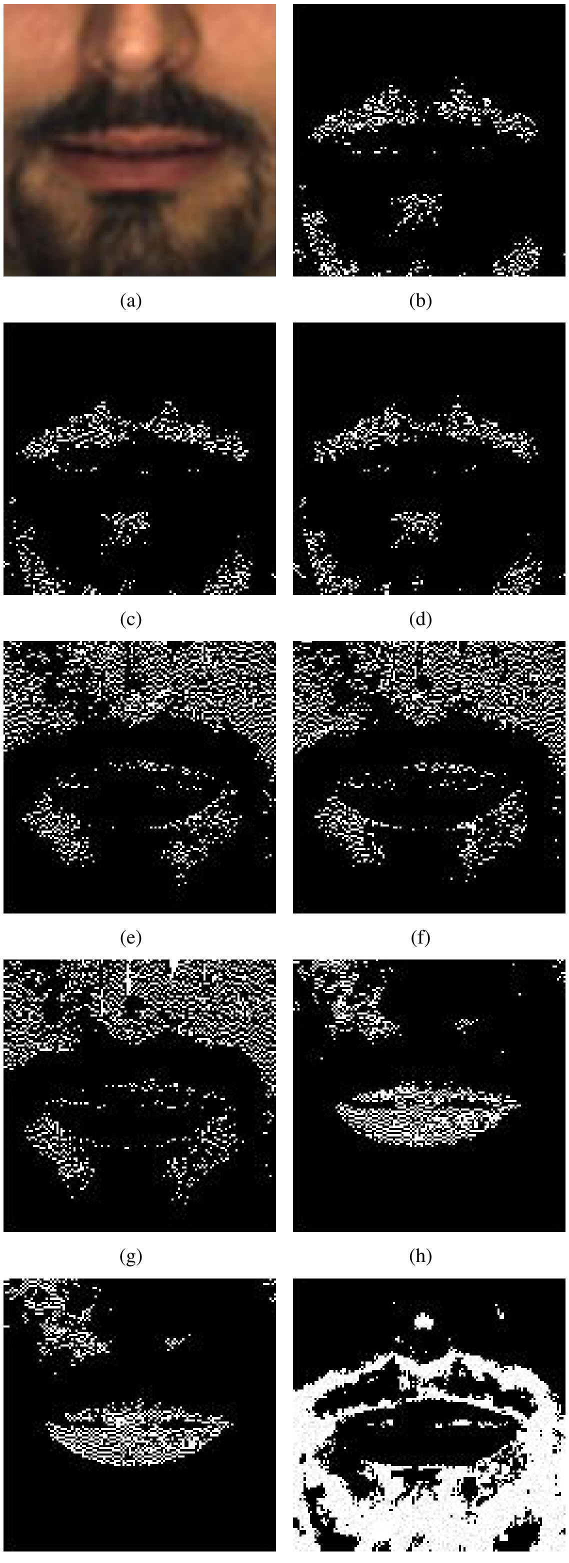


Fig. 3. Segmentation (clustering) results with $m = 9$. (a) Source image. (b)–(j) White pixels represent the pixels falling into clusters 1–9.

For each site, we then perform the following transform to get the pattern vector:

$$x_i = [H_i \cdot \cos(2\pi \cdot S_i), H_i \cdot \sin(2\pi \cdot S_i)]^T, \quad i \in S. \tag{26}$$

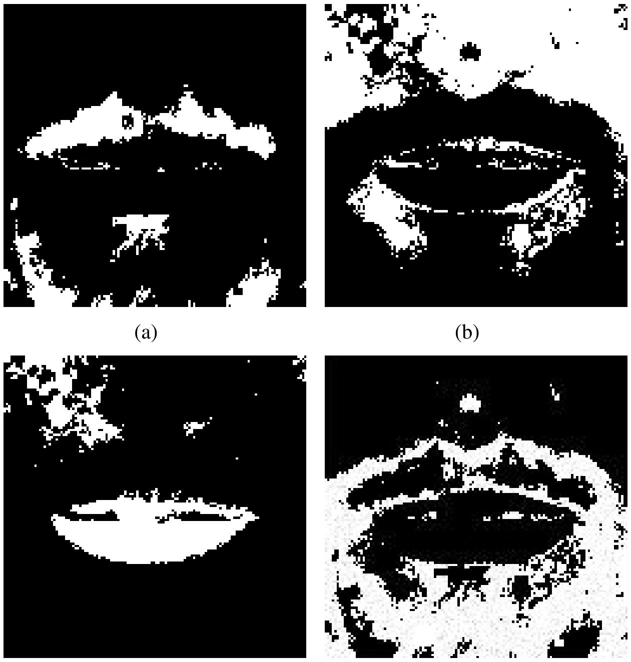


Fig. 4. Lip segmentation results shown in (a)–(d) after the redundant cluster centroids have been merged.

B. Segmentation and Binarization

Subsequently, the centroid collection C is calculated via the proposed method as introduced in Section III. We utilize the following equation to obtain the hard segmentation result:

$$S^{(j)} = \Big\{ i \mid j = \arg\max_{j} \big(p\big(\tilde{c}^{i}_{j} \mid x_i\big) - p\big(\tilde{c}^{i}_{j-1} \mid x_i\big)\big), \\ 1 \le i \le n, 1 \le j \le m \Big\} \tag{27}$$

where $S^{(j)}$ denotes the set of data falling into cluster j.

A sample of $S^{(j)}s$ with $m = 9$ is shown in Fig. 3(b)–(j). Obviously, the site sets: $\{S^{(1)}, S^{(2)}, S^{(3)}\}$, $\{S^{(4)}, S^{(5)}, S^{(6)}\}$, and $\{S^{(7)}, S^{(8)}\}$ are similar because the corresponding redundant centroids are coincident in pattern space.

Then, for any two centroids c_j and c_k in pattern space, if

$$\|c_j - c_k\| \le \varepsilon \tag{28}$$

where ε is a small threshold value, they can be replaced by a new centroid c_l

$$c_l = \frac{c_j + c_k}{2}. \tag{29}$$

Thus, in the example shown in Fig. 3, the number of centroids is reduced from 9 to 4. The new clustering result after the merger of centroids is shown in Fig. 4.

C. Lip Segment Selection

We utilize the method proposed in [24] to extract a patch of lip region. Then, the mean of $x_i s$ restricted by the patch is calculated and denoted by $\hat{\mu}$. It is regarded as an estimate of the mean of $x_i s$ that fall into the true lip region. To save space, interested readers may refer to [24] for more details about this method. In the following, we summarize the major steps of this method as follows.

Step 1: We transform the source lip image into 1976 CIELAB color space. a^* component for site i is mapped into the range of $[0, 255]$ via the histogram equalization and denoted by a_i^*. Meanwhile, we utilize the equation

$$U_i = \begin{cases} 256 \times \dfrac{G_i}{R_i}, & R_i > G_i \\ 255, & \text{otherwise} \end{cases} \tag{30}$$

proposed in [2] to calculate U component for each pixel, where R_i and G_i denote the red and green component, respectively, for site i in a source lip image.

Step 2: Let $B_i = a_i^* - U_i$, we establish a Gaussian model for the positive B_is with the mean μ_B and the standard deviation σ_B. The following equation is employed to binarize the source lip image:

$$\hat{B}_i = \begin{cases} 0, & B_i \le \mu_B - 2\sigma_B \\ 1, & \text{otherwise.} \end{cases} \tag{31}$$

Step 3: Considering the site set $\hat{S} = \{i \mid \hat{B}_i = 1, 1 \le i \le n\}$ as a lip patch, as shown in Fig. 5(a), $\hat{\mu}$ can be calculated by

$$\hat{\mu} = \frac{1}{s} \sum_{i \in \hat{S}} x_i \tag{32}$$

where s denotes the number of elements in the set $\hat{S}$.

Thus, the index of lip segment layer can be determined by

$$j^{\text{lip}} = \arg\min_{j} \|c_j - \hat{\mu}\|, \quad 1 \le j \le m. \tag{33}$$

The site set corresponding to lip segment is denoted by $S^{(j^{\text{lip}})}$.

D. Postprocessing

Suppose $S^{(j^{\text{lip}})}$ can be viewed as a binary image with c columns and r rows in pixel. For the sake of description, we map the index i into a 2-D coordinate $\{(p, q) \mid 1 < p \le c, 1 < q \le r\}$ by $i = (q - 1) \cdot r + p$. We hereby represent the binary image as

$$B(p, q) = \begin{cases} 1, & (q - 1) \cdot r + p \in S^{(j^{\text{lip}})} \\ 0, & \text{otherwise.} \end{cases} \tag{34}$$

Suppose the lip region is not connected to the border of image. The morphological reconstruction-based method in [49] is, therefore, employed to clear border connected noisy structures, as shown in Fig. 5(b). Furthermore, we utilize the morphological close operation with 5×5 structuring element and open operation with 3×3 structuring element, respectively. The result is denoted by B_m, as shown in Fig. 5(c). For the foreground elements in B_m, the corresponding positions

$$\{(p, q) \mid B_m(p, q) = 1, 1 < p \le c, 1 < q \le r\} \tag{35}$$

are recorded and compose a matrix P as follows:

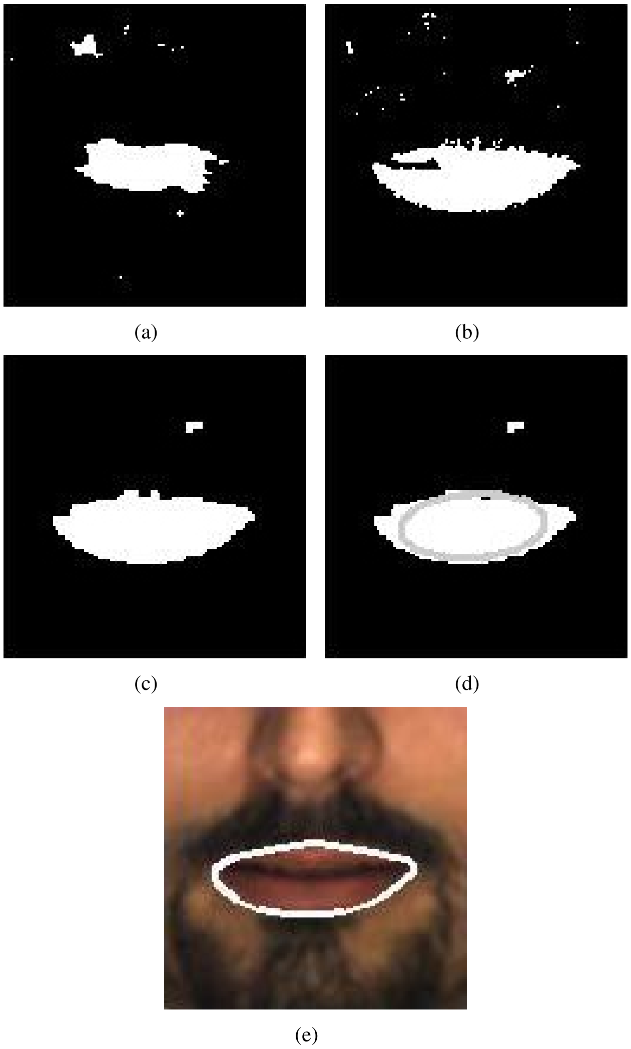

(a)
(b)
(c)
(d)
(e)

Fig. 5. (a) Lip patch which is used to estimate the mean of x_is falling into the true lip region. (b) Result of clearing border connected noisy structures. (c) Result of morphological filter (closing with 5×5 structuring element and opening with 3×3 structuring element). (d) Shape of gray ellipse is defined as the eigenvectors and eigenvalues of the covariance matrix of P. The continued objects on the outside of this ellipse are viewed as noises and masked out. (e) Final extraction result obtained via the quickhull algorithm.

$$P = \begin{bmatrix} q_1 & p_1 \\ q_2 & p_2 \\ \vdots & \vdots \\ q_r & p_r \end{bmatrix} \tag{36}$$

where r is the number of foreground elements in B_m.

Then, the eigenvectors and eigenvalues of the covariance matrix of P are calculated. We can further obtain an ellipse, whose position and inclination are defined as the eigenvectors with the length of major and minor axis defined as 1.5 times the square root of eigenvalues, respectively. The continued objects on the outside of this ellipse are masked out, as shown in Fig. 5(d).

Finally, given the prior knowledge of human mouth shape, the quickhull algorithm proposed in [50] is employed to draw

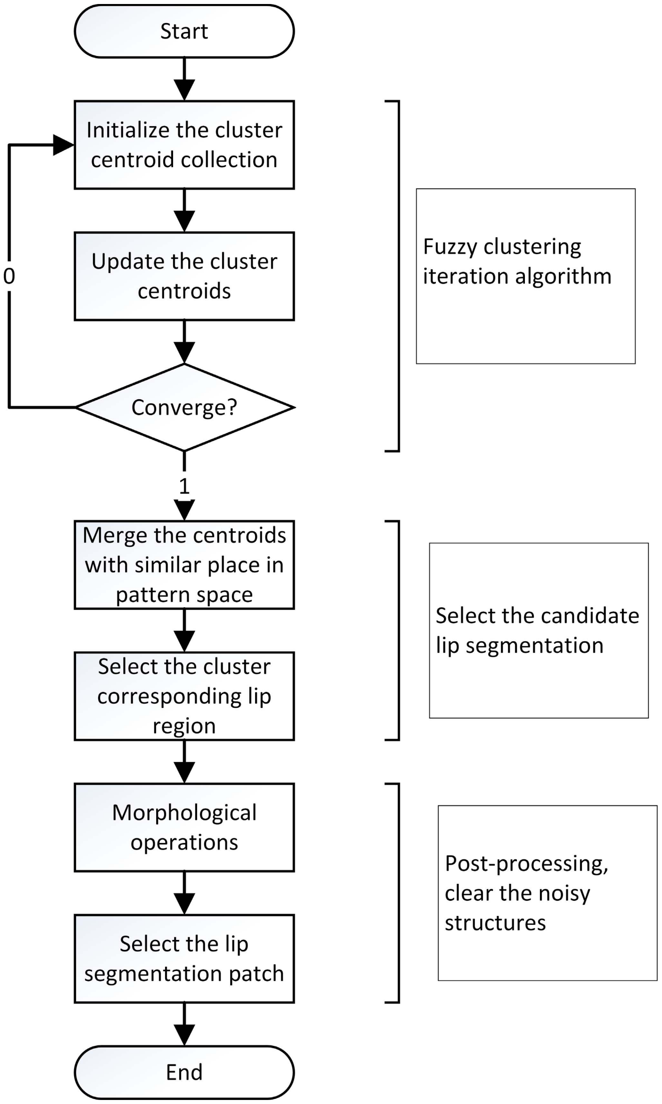

Start
Initialize the cluster centroid collection
Update the cluster centroids
Converge?
0
1
Fuzzy clustering iteration algorithm
Merge the centroids with similar place in pattern space
Select the cluster corresponding lip region
Select the candidate lip segmentation
Morphological operations
Select the lip segmentation patch
Post-processing, clear the noisy structures
End

Fig. 6. Procedure of the lip segmentation in the proposed approach.

the boundary of lip, as shown in Fig. 5(e). Fig. 6 shows the segmentation procedure in the proposed approach.

V. Experimental Results

A. Database and Initialization

To show the performance of the proposed approach, we utilized three databases.

1) Fisher's iris database [51] consisting of three classes, each of which has 50 instances with four attributes.
2) AR face database [52] with 126 persons, each of whom has 26 images.
3) CVL face database [53] with 114 persons, each of whom has seven images.

For each image in AR and CVL databases, the part of face between nostril and chin was clipped by a window of 128 × 128 pixels as the source of segmentation experiment.

Moreover, in the following experiments, we utilized:

$$c_j = x_{\text{rand}(j)}, \quad 1 \le j \le m \tag{37}$$

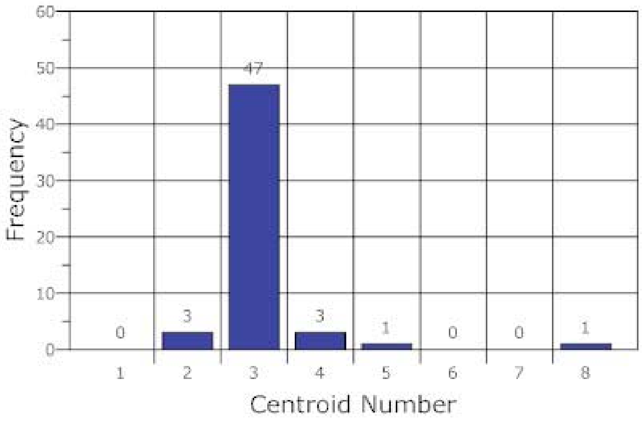


Fig. 7. Final centroid number after the clustering performed in Fisher's iris database by the proposed method with the different values of m and initial C.

to initialize the centroids, where rand(i) denotes a number randomly selected from the set of $\{1, 2, \ldots n\}$ at the jth selection, and n is the total number of samples. Furthermore, we let $\varepsilon = 0.5$, $\eta_w = 0.01$, and $\eta_r = 0.001$.

B. Experiment 1

In this experiment, the method described in Section III was employed to perform the fuzzy clustering in Fisher's iris database. This experiment was conducted with $m = 5, 6, \ldots, 15$, respectively. Moreover, for each specific m, the experiments were repeated five times with the different initial values of C. After the redundant centroids merged based on (28) and (29), the histogram of final centroid number is shown in Fig. 7. It can be seen that 47 out of 55 results kept three centroids, which implies that the true number of classes can be determined automatically by the proposed method. Under these 47 trials, in each of which three centroids were finally kept, we further evaluated the difference between the final centroids obtained by the proposed method and the classical FCM, respectively, using the following equation:

$$\text{error} = \sqrt{\frac{\sum_{j=1}^{3} \left\| \hat{c}_j - c_j^* \right\|^2}{3}} \tag{38}$$

where $\hat{c}_j$ is the final centroid obtained by the proposed method, c_j^* is the corresponding centroid obtained by the classical FCM method with the number of clusters set at the true number of classes, i.e., 3. The histogram of error is shown in Fig. 8. It can be seen that, with the same number of centroids, the values of these final centroids obtained from the proposed approach still have the moderate differences from those obtained by the classical FCM, although both of these two methods are clustering-based ones. In general, such a difference will lead to quite different segmentation results. We will further demonstrate this in Experiment 4.

C. Experiment 2

To demonstrate the accuracy and robustness of the proposed method, we separated the source images from AR database

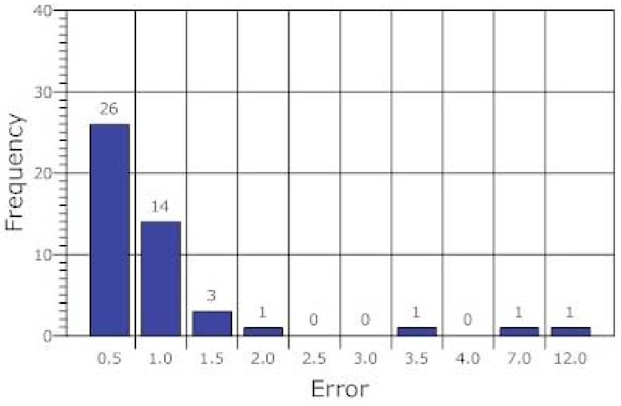


Fig. 8. Histogram of the error between the cluster centroids obtained by the proposed method, and the classical FCM with the number of clusters set at the true number of classes, i.e., 3.

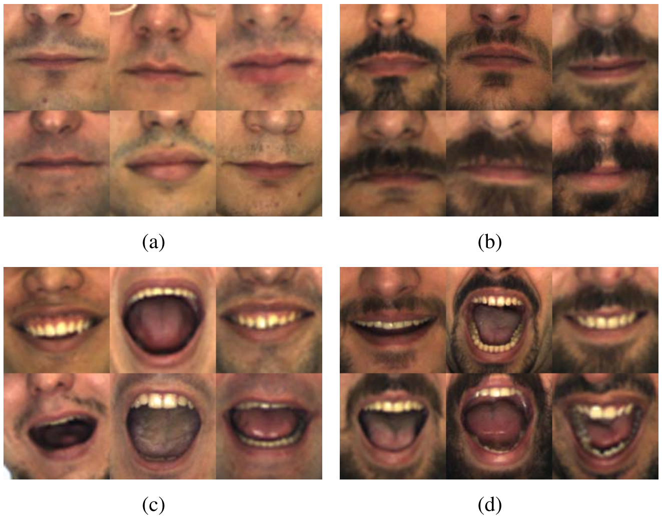


Fig. 9. (a)–(d) Some sample images, which belong to Groups 1–4, respectively, from AR database.

into the four groups based on the different appearances. The details are as follows.

1) *Group 1:* People have no evident mustache with the mouth closed, as shown in Fig. 9(a).
2) *Group 2:* People have evident mustache with the mouth closed, as shown in Fig. 9(b).
3) *Group 3:* People have no evident mustache with the mouth opened, as shown in Fig. 9(c).
4) *Group 4:* People have evident mustache with the mouth opened, as shown in Fig. 9(d).

Our experiment was conducted on each group, respectively. For each group, we randomly selected 20 images as the input, and manually segmented the lip to serve as the ground truth. Two measures defined in [26] were used to evaluate the performance of the algorithms. The first measure

$$\text{OL} = \frac{2(A_1 \cap A_2)}{A_1 + A_2} \times 100\% \tag{39}$$

determines the percentage of overlap between the segmented lip region A_1 and the ground truth A_2. The second measure

TABLE I

OVERLAP OF SEGMENTED LIPS WITH THE GROUND TRUTH

Group / m	1	2	3	4
5	91.5%	93.5%	85.4%	88.3%
6	92.3%	93.3%	87.6%	90.6%
7	90.7%	94.6%	85.0%	86.5%
8	94.5%	89.7%	88.2%	90.3%
9	91.0%	92.7%	90.3%	87.6%
10	91.3%	95.3%	91.5%	89.4%
11	93.0%	92.5%	88.4%	90.4%
12	90.7%	93.6%	87.9%	87.8%
13	92.8%	93.0%	88.1%	88.3%
14	92.3%	92.9%	94.2%	88.2%
15	90.4%	92.4%	88.3%	89.8%

TABLE II

SE OF SEGMENTED LIP

Group / m	1	2	3	4
5	9.2%	6.0%	17.3%	12.9%
6	8.3%	7.1%	14.3%	10.4%
7	10.0%	5.4%	17.5%	14.4%
8	6.4%	14.6%	11.0%	13.8%
9	9.5%	9.3%	9.0%	14.3%
10	10.4%	5.0%	10.2%	10.5%
11	8.6%	8.2%	14.1%	10.7%
12	11.5%	6.7%	14.3%	13.6%
13	9.1%	6.3%	13.7%	12.9%
14	7.4%	8.3%	5.8%	12.8%
15	11.3%	8.5%	11.5%	13.0%

is the segmentation error (SE) defined as

$$\text{SE} = \frac{\text{OLE} + \text{ILE}}{2 \times \text{TL}} \times 100\% \tag{40}$$

where OLE is the number of nonlip pixels classified as lip pixels (i.e., outer lip error), ILE is the number of lip pixels classified as nonlip ones (inner lip error), and TL denotes the number of lip pixels in the ground truth.

We repeated the experiments with $m = 5, 6, \ldots, 15$. Tables I and II list the average OL and SE on the different image groups and m. It can be seen that the segmentation performance of the proposed approach is robust against m in all cases we have tried so far. Furthermore, we also utilized the AR database to investigate the robustness of the proposed approach against the value selection of the parameters: ϵ, η_w, and η_r. From Table III, it can be seen that the performance of the proposed approach changes slightly over the moderate variation of these parameters. That is, its performance is robust against the selection of these parameters to a certain degree when performing the lip segmentation.

D. Experiment 3

To evaluate the performance of the proposed method under the different capture environments, we randomly selected 50 images from AR and CVL databases, respectively. These raw images were further clipped by 128×128 and reindexed. Moreover, for each image, we also randomly assigned m an integer from the set $\{5, 6, \ldots, 15\}$ to conduct the lip segmentation. For images from either AR or CVL database, the average

TABLE III

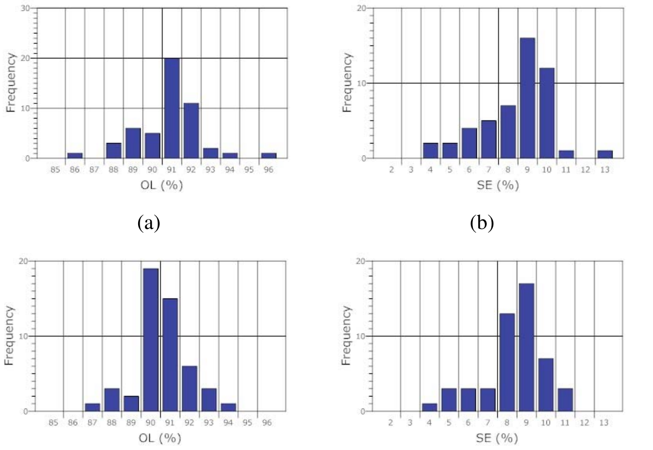
SE on AR Database by the Proposed Approach With the Different Settings of Parameters: ϵ, η_w, and η_r

$\eta_w = 0.01, \eta_r = 0.001$ ϵ	**OL**	**SE**	$\epsilon = 0.5, \eta_r = 0.001$ η_w	**OL**	**SE**	$\epsilon = 0.5, \eta_w = 0.01$ η_r	**OL**	**SE**
0.3	91.6%	8.4%	0.010	90.8%	8.4%	0.001	90.8%	8.4%
0.4	89.2%	9.2%	0.015	91.5%	8.1%	0.002	91.4%	8.0%
0.5	90.8%	8.4%	0.020	90.6%	9.7%	0.003	90.1%	9.2%
0.6	91.7%	7.7%	0.025	90.9%	9.0%	0.004	89.2%	10.9%
0.7	90.5%	8.3%	0.030	90.1%	8.9%	0.005	87.3%	11.5%

TABLE IV

Average Overlap and SE Obtained by Liew03, Leung04, Wang07, FCM, and the Proposed Method for the Images From AR and CVL Databases

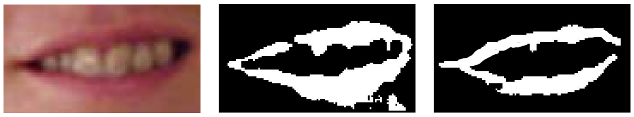
Measure **Database**	Liew04		Leung04		Wang07		FCM		Proposed	
	OL	**SE**	**OL**	**SE**	**OL**	**SE**	**OL**	**SE**	**OL**	**SE**
AR	83.50%	17.31%	89.92%	13.25%	87.35%	21.13%	65.25%	42.82%	***90.80%***	***8.40%***
CVL	88.03%	10.57%	93.33%	11.00%	***93.53%*** %	10.05%	75.70%	24.37%	91.90%	***8.40%***



(a) (b)

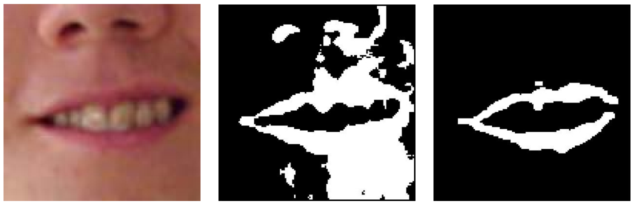
(c) (d)

Fig. 10. (a) and (b) Histograms of OL and SE of the selected images from AR database. (c) and (d) Histograms of OLs and SEs of the selected images from CVL database.



(a) (b) (c)

Fig. 11. (a) Sample of the input image for Wang07 in the experiment. (b) and (c) Corresponding segmentation results obtained by Wang07 and the proposed method without postprocessing, respectively.



(a) (b) (c)

Fig. 12. (a) Sample of the input image for the proposed method in the experiment. (b) and (c) Corresponding segmentation results obtained by Wang07 and the proposed method without postprocessing, respectively.

OL and SE were calculated. The two rightmost columns in Table IV list the average OL and SE values obtained by the proposed method. Fig. 10 shows the histograms of OL and SE for images from each database. Once again, it can be seen that the proposed approach is robust against the preassigned number of clusters.

E. Experiment 4

We demonstrated the performance of the proposed approach in comparison with four existing methods, i.e., Liew03 [26], Leung04 [30], Wang07 [31], and classical FCM. We chose the methods based on two rules: 1) they have been utilized for lip segmentation or extraction, and 2) they have similar clustering-based theoretical background.

We implemented these algorithms on the same images utilized in Experiment 3. The experimental result is shown in Table IV. The algorithm parameters in the existing methods are set according to the original paper. As for our work, the preassigned number of clusters is easily set as long as it is greater than or equal to the true one. That is, it was set to ten for the lip segmentation task.

It can be seen that the proposed approach outperforms Liew2003, Leung04, and FCM methods in most cases we have tried so far, and has a competitive advantage with the much smaller SE values in comparison with Liew2003 and Wang07. Furthermore, when we implemented Wang07 for comparative studies, we actually utilized the image clips employed in [31] (the size of input image is various so as to make the most parts occupied by lip region), as shown in Fig. 11, rather than 128 × 128 image clips as the input of Wang07. It is found that the performance of Wang07 somewhat depends on the image clips. For example, if we utilize 128 × 128 as the inputs of Wang07, the segmentation results become deteriorate, as shown in Fig. 12, where the image in Fig. 11(a) is the same as the one in Fig. 12(a), i.e., $4 - MVC - 007F$ in CVL. This implies that the proposed algorithm has more robust performance in comparison with Wang07. The results of Liew2003 and Leung04 are sensitive to the setting of cluster number. It can be seen that the lip segmentation given by these two methods becomes worse when the cluster number is not appropriately selected. In addition, Leung04 method utilized the elliptic shape to model the lip. Its result would degrade if the beard and teeth around the mouth disturb the

TABLE V
RUNNING TIME OF DIFFERENT METHODS

	Liew03	Leung04	Wang07	FCM	Proposed one
AR	3.8s	5.4s	8.7s	3.2s	8.9s
CVL	4.3s	5.2s	9.1s	3.7s	8.7s

clustering process. Compared with the proposed method, it can be seen that the SE in Liew2003 and Leung04 is much higher, and the OL is much lower. That is, the proposed method outperforms both of them.

The average running time of these methods running at a machine with an Intel(R) Core(TM) Two Duo CPU E7500 2.93-GHz CPU is shown in Table V. It can be seen that the proposed method does not show its superiority on the computational cost. In fact, the computation cost would not become the bottleneck of lip-reading system due to the development of high-speed CPU. The real bottleneck is the automatic selection of cluster number for the lip segmentation.

VI. CONCLUSION

This paper has proposed a cooperative learning-based clustering method for lip segmentation without knowing the true cluster number in advance. This method features that the overlapped (or close) cluster centroids in pattern space can be merged into one from the viewpoint of objective function value. Then, an iterative algorithm is utilized to minimize the proposed objective function by superposing the redundant centroids. At each iterative step, not only is the winner updated to adapt to an input data, but also the other centroids are adjusted with a specific cooperation strength, so that they are each close to the winner. As a result, the clustering performance is robust against the preassigned number of clusters. Based upon this method, a lip segmentation scheme has been presented. Experimental results have shown its efficacy in comparison with the existing counterparts.

APPENDIX I

We can obtain a mapping $h_m : \eta_m \to H_{\eta_m}^{\min}$ by solving the following optimization problem, where $H_{\eta_m}^{\min}$ is the minimum of $H(C \mid x_i)$ subject to $\delta H(C \mid x_i) = \eta_m$:

$$\begin{aligned}&\min:\ 1-\sum_{j=1}^{m}\left[p\left(\tilde{c}_j^i \mid x_i\right)\right]^2\\ &\text{s.t.}:\begin{cases}1-\sum_{j=1}^{m}\left[\dfrac{p\left(\tilde{c}_j^i \mid x_i\right)-p\left(\tilde{c}_{j-1}^i \mid x_i\right)}{p\left(\tilde{c}_m^i \mid x_i\right)}\right]^2=\eta_m\\ \sum_{j=1}^{m}p\left(\tilde{c}_j^i \mid x_i\right)=1\\ p\left(\tilde{c}_j^i \mid x_i\right)\geq 0\\ p\left(\tilde{c}_j^i \mid x_i\right)\leq p\left(\tilde{c}_k^i \mid x_i\right)\quad \text{if } j<k.\end{cases}\end{aligned}\tag{41}$$

Using the substitution

$$p\left(\tilde{c}_j^i \mid x_i\right)=\sum_{k=1}^{j}a_k^2,\quad a_k\in\mathbb{R}\tag{42}$$

the optimization problem can, therefore, be simplified as

$$\begin{aligned}&\min:\ 1-\sum_{j=1}^{m}\left(\sum_{k=1}^{j}a_k^2\right)^2\\ &\text{s.t.}:\begin{cases}\sum_{k=1}^{m}a_k^4+(\eta_m-1)\left(\sum_{k=1}^{m}a_k^2\right)^2=0\\ \sum_{k=1}^{m}(m+1-k)a_k^2=1.\end{cases}\end{aligned}\tag{43}$$

Subsequently, the corresponding Lagrange function is

$$\begin{aligned}&\Lambda_m(a_1,\ldots,a_m,\alpha,\beta)\\ &=1-\sum_{\tilde{j}=1}^{m}\left(\sum_{k=1}^{\tilde{j}}a_k^2\right)^2+\alpha\left[\sum_{k=1}^{m}a_k^4+(\eta_m-1)\left(\sum_{k=1}^{m}a_k^2\right)^2\right]\\ &\quad+\beta\left[\sum_{k=1}^{m}(m+1-k)a_k^2-1\right]\end{aligned}\tag{44}$$

where α and β are Lagrange multipliers.

Thus, the constrained extrema of (43) are the extreme points of (44), which can be obtained by solving the following equations:

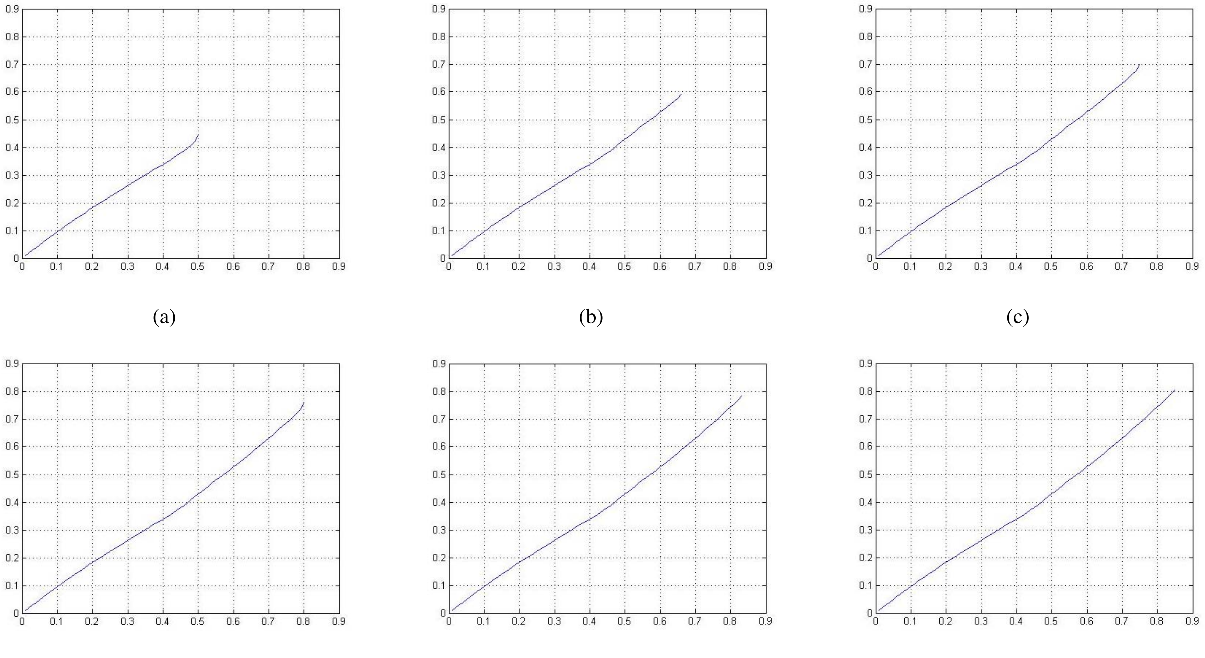
$$\nabla_{a_l,\alpha,\beta}\Lambda_m=0\quad(l=1,2,\ldots,m)\tag{45}$$

which can be further expressed as

$$\begin{cases}a_1\sum_{\tilde{j}=1}^{m}\left(\sum_{k=1}^{\tilde{j}}a_k^2\right)-\alpha a_1\left[a_1^2+(\eta_m-1)\sum_{k=1}^{m}\left(a_k^2\right)\right]=\dfrac{m\beta a_1}{2}\\ a_2\sum_{\tilde{j}=2}^{m}\left(\sum_{k=1}^{\tilde{j}}a_k^2\right)\\ \quad-\alpha a_2\left[a_2^2+(\eta_m-1)\sum_{k=1}^{m}\left(a_k^2\right)\right]=\dfrac{(m-1)\beta a_2}{2}\\ \qquad\cdots\\ a_l\sum_{\tilde{j}=l}^{m}\left(\sum_{k=1}^{\tilde{j}}a_k^2\right)\\ \quad-\alpha a_l\left[a_l^2+(\eta_m-1)\sum_{k=1}^{m}\left(a_k^2\right)\right]=\dfrac{(m-l+1)\beta a_l}{2}\\ \qquad\cdots\\ a_m\sum_{k=1}^{m}a_k^2-\alpha a_m\left[a_m^2+(\eta_m-1)\sum_{k=1}^{m}\left(a_k^2\right)\right]=\dfrac{\beta a_m}{2}\\ \sum_{k=1}^{m}a_k^4+(\eta_m-1)\left(\sum_{k=1}^{m}a_k^2\right)^2=0\\ \sum_{k=1}^{m}(m+1-k)a_k^2=1.\end{cases}\tag{46}$$

For any of the first m equations in (46), i.e., $\nabla_{a_l}\Lambda_m=0$, we fix $a_{\tilde{l}}$ $(\tilde{l}=1,\ldots,l-1,l+1,\ldots,m)$ and α. Therefore, a_l^2 can be represented as a linear function with respect to β:

$$\begin{aligned}&(m-l+1-\alpha\eta_m)a_l^2+\frac{l-m-1}{2}\beta\\ &\quad+\sum_{j=l}^{m}\left(\sum_{k=1,k\neq l}^{j}a_k^2\right)-\alpha(\eta_m-1)\sum_{k=1,k\neq l}^{m}\left(a_k^2\right)=0.\end{aligned}\tag{47}$$



(a) (b) (c)

(d) (e) (f)

Fig. 13. Functional relationship h_m between η_m and $H^{\min}_{\eta_m}$ with (a) $m = 2$, (b) $m = 3$, (c) $m = 4$, (d) $m = 5$, (e) $m = 6$, and (f) $m = 7$, respectively, where the horizontal axis represents the value of δH_m, and the vertical axis represents the value of $H^{\min}_{\eta_m}$.

Subsequently, we can eliminate a_l and obtain a quadratic polynomial of β via substituting (47) into $\nabla_\alpha \Lambda_m = 0$.

On the other hand, β can be calculated by solving (46). To be specific, adding the first m equations in (46), then using the last two equations in (46), we have

$$a_1^2 \sum_{\tilde{j}=1}^{m} \left(\sum_{k=1}^{\tilde{j}} a_k^2 \right) + a_2^2 \sum_{\tilde{j}=2}^{m} \left(\sum_{k=1}^{\tilde{j}} a_k^2 \right) + \cdots + a_m^2 \sum_{\tilde{j}=1}^{m} a_k^2 = \frac{\beta}{2}. \tag{48}$$

Finally, we can obtain

$$\beta = 2 \sum_{j=1}^{m} \left(\sum_{k=1}^{j} a_k^2 \right)^2 = 2\left(1 - H^{\text{sta}}_{\eta_m}\right) \tag{49}$$

where $H^{\text{sta}}_{\eta_m}$ can be calculated by the possible stationary point of (44). Substituting (49) into the quadratic polynomial determined by (47) and $\nabla_\alpha \Lambda_m = 0$, we can obtain a quadratic polynomial with respect to $H^{\text{sta}}_{\eta_m}$. That is, for (44), the number of stationary points is 0, 1, or 2. Based on the extreme value theorem, this number can be further fixed to 2, corresponding to global maximum and minimum, respectively.

Suppose the minimum of Lagrange function Λ_{m-1} is obtained at the point $(a_1, a_2, \ldots, a_{m-1}, \alpha, \beta)$. According to (46), Λ_m has the stationary point at $(0, a_1, a_2, \ldots, a_{m-1}, \alpha, \beta)$ as long as $H^{\text{sta}}_{\eta_m} = H^{\text{sta}}_{\eta_{m-1}}$. Let the Hessian matrix of Λ_{m-1} at $(a_1, a_2, \ldots, a_{m-1}, \alpha, \beta)$ be $\mathcal{H}_{m-1}$. Then, the Hessian matrix of Λ_m at $(0, a_1, a_2, \ldots, a_{m-1}, \alpha, \beta)$ can be represented recursively as

$$\mathcal{H}_m = \begin{bmatrix} 4\alpha a_1^2 + 2\beta & A \\ B & \mathcal{H}_{m-1} \end{bmatrix} \tag{50}$$

where $A = [0, 0, \ldots, 0]$ and $B = [0, 0, \ldots, 0]^T$.

As we know, the entropy value of a random variable will tend to zero as the variable becomes certainty. Thus, we suppose that $p(\tilde{c}_1^i \mid x_i) = a_1^2 \to 0$ and $p(\tilde{c}_m^i | x_i) = \sum_{j=1}^m a_j^2 \to 1$ when the constrained minimum in (43) is obtained. Under this situation, $\mathcal{H}_m$ is a positive definite matrix. Moreover, as stated above, since there is only one minimum stationary point in Λ_m as given a specific η_m, $(0, a_1, a_2, \ldots, a_{m-1}, \alpha, \beta)$ must be the global minimum of (44). Thus, h_m can be represented by the following recursion approximatively:

$$h_m(\eta_{m-1}) \approx h_{m-1}(\eta_{m-1}) \tag{51}$$

as shown in Fig. 13.

When $\eta_{m-1} \in (0, (k-1)/k]$ with $k = 2, 3, \ldots, m-1$, the curves of $h_m(\eta_{m-1})$ and $h_k(\eta_{m-1})$ are coincident (see Fig. 14). Then, (51) can be further formulated as

$$h_m(\eta_{m-1}) = \begin{cases} h_{m-1}(\eta_{m-1}), & \eta_{m-1} \in \left(0, \dfrac{m-2}{m-1}\right] \\ h_{m-2}(\eta_{m-1}), & \eta_{m-1} \in \left(0, \dfrac{m-3}{m-2}\right] \\ \cdots \\ h_2(\eta_{m-1}), & \eta_{m-1} \in \left(0, \dfrac{1}{2}\right] \\ 0, & \eta_{m-1} = 0. \end{cases} \tag{52}$$

Subsequently, substituting (49) and $\nabla_\alpha \Lambda_m = 0$ into $\nabla_{a_1} \Lambda_m = 0$, we can obtain

$$H^{\text{sat}}_{\eta_m} = \frac{\alpha \sum_{k=1}^m a_k^2}{m} \eta_m + \frac{\alpha a_1^2 - \alpha \sum_{k=1}^m a_k^2 + m - 1}{m}. \tag{53}$$

When the minimum of (53) is achieved, and $m \to +\infty$, we have

$$H^{\min}_{\eta_{+\infty}} = \eta_{+\infty}. \tag{54}$$

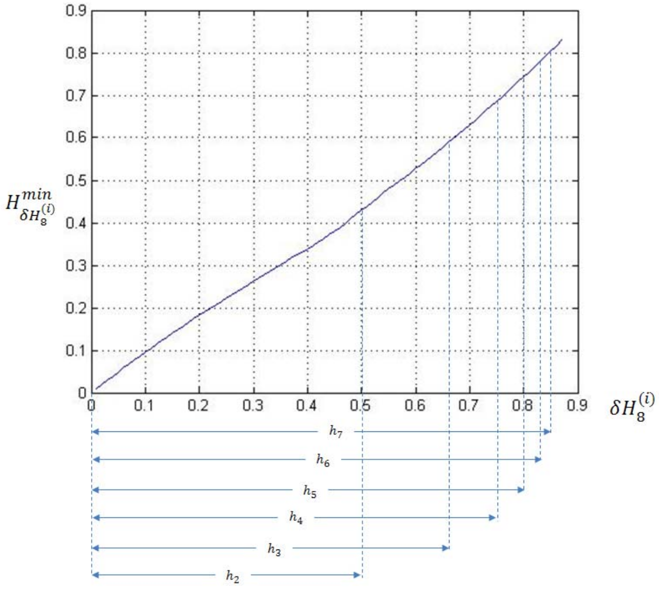


Fig. 14. Curve of $H^{\min}_{\eta_8} = h_8(\eta_8)$. When $\eta_8 \in (0, (k-1/k)]$ with $k = 1, 2, \ldots, 7$, the corresponding curve segments are coincident with $h_k(\eta_8)$.

TABLE VI

MSE BETWEEN THE NUMERICAL SIMULATION RESULT $\hat{H}^{\min}_{\eta_m}$ AND IDEAL VALUE $H^{\min}_{\eta_m} = \eta_m$ OVER m

m	2	3	4	5	6	7	8
MSE	0.0016	0.0025	0.0028	0.0028	0.0028	0.0028	0.0028

Based on (52), the relationship between η_m and $H^{\min}_{\eta_m}$ can be written approximatively as

$$H^{\min}_{\eta_m} \approx \eta_m. \tag{55}$$

APPENDIX II

We conduct an experiment to justify the validity of Lemma 1. First, we select an input x_i in pattern space randomly, and calculate the corresponding $H^{\min}_{\eta_m}$ for different $\eta_m \in (0, (m-1)/m]$ by interior point method. Then, we utilize the mean square error (MSE) to evaluate the bias between the numerical simulation result, denoted by $\hat{H}^{\min}_{\eta_m}$, and the desired value, i.e., $H^{\min}_{\eta_m} = \eta_m$. Moreover, this experiment is repeated with $m = 2, 3, \ldots, 8$. For each m, we select five different values of x_i, each of which is a sample generated by (26).

Table VI lists the average MSE over m. It can be seen that the error is tiny and tends to constant over m with the ignorable fluctuation. This implies that Lemma 1 is indeed valid empirically.

REFERENCES

[1] S. Ben-Yacoub, Y. Abdeljaoued, and E. Mayoraz, "Fusion of face and speech data for person identity verification," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1065–1074, Sep. 1999.

[2] M. Lievin and F. Luthon, "Nonlinear color space and spatiotemporal MRF for hierarchical segmentation of face features in video," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 63–71, Jan. 2004.

[3] H. E. Cetingul, Y. Yemez, E. Erzin, and A. M. Tekalp, "Discriminative analysis of lip motion features for speaker identification and speech-reading," *IEEE Trans. Image Process.*, vol. 15, no. 10, pp. 2879–2891, Oct. 2006.

[4] G. Chetty and M. Wagner, "Robust face-voice based speaker identity verification using multilevel fusion," *Image Vis. Comput.*, vol. 26, no. 9, pp. 1249–1260, 2008.

[5] M. Sorci, G. Antonini, J. Cruz, T. Robin, M. Bierlaire, and J.-P. Thiran, "Modelling human perception of static facial expressions," *Image Vis. Comput.*, vol. 28, no. 5, pp. 790–806, 2010.

[6] H. Ç. Akakın and B. Sankur, "Robust classification of face and head gestures in video," *Image Vis. Comput.*, vol. 29, no. 7, pp. 470–483, 2011.

[7] K. S. Fu and J. K. Mui, "A survey on image segmentation," *Pattern Recognit.*, vol. 13, no. 1, pp. 3–16, 1981.

[8] R. M. Haralick and L. G. Shapiro, "Image segmentation techniques," *Comput. Vis., Graph., Image Process.*, vol. 29, no. 1, pp. 100–132, 1985.

[9] N. R. Pal and S. K. Pal, "A review on image segmentation techniques," *Pattern Recognit.*, vol. 26, no. 9, pp. 1277–1294, 1993.

[10] Y. J. Zhang, "A survey on evaluation methods for image segmentation," *Pattern Recognit.*, vol. 29, no. 8, pp. 1335–1346, 1996.

[11] T. Wark, S. Sridharan, and V. Chandran, "An approach to statistical lip modelling for speaker identification via chromatic feature extraction," in *Proc. 14th Int. Conf. Pattern Recognit.*, Brisbane, QLD, Australia, Aug. 1998, pp. 123–125.

[12] X. Zhang and R. M. Mersereau, "Lip feature extraction towards an automatic speechreading system," in *Proc. Int. Conf. Image Process.*, Vancouver, BC, Canada, 2000, pp. 226–229.

[13] M. Pardàs and E. Sayrol, "Motion estimation based tracking of active contours," *Pattern Recognit. Lett.*, vol. 22, no. 13, pp. 1447–1456, 2001.

[14] P. Delmas, N. Eveno, and M. Liévin, "Towards robust lip tracking," in *Proc. IEEE Int. Conf. Pattern Recognit.*, Quebec City, Canada, 2002, pp. 528–531.

[15] N. Eveno, A. Caplier, and P.-Y. Coulon, "Accurate and quasi-automatic lip tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 5, pp. 706–715, May 2004.

[16] I. Matthews, T. F. Cootes, J. A. Bangham, S. Cox, and R. Harvey, "Extraction of visual features for lipreading," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 2, pp. 198–213, Feb. 2002.

[17] N. Eveno, A. Caplier, and P.-Y. Coulon, "Jumping snakes and parametric model for lip segmentation," in *Proc. IEEE Int. Conf. Image Process.*, Barcelona, Spain, Sep. 2003, pp. II-867–II-870.

[18] H. Seyedarabi, W. Lee, and A. Aghagolzadeh, "Automatic lip tracking and action units classification using two-step active contours and probabilistic neural networks," in *Proc. Can. Conf. Elect. Comput. Eng.*, Ottawa, ON, Canada, May 2006, pp. 2021–2024.

[19] B. Beaumesnil and F. Luthon, "Real time tracking for 3D realistic lip animation," in *Proc. 18th Int. Conf. Pattern Recognit.*, Hong Kong, 2006, pp. 219–222.

[20] R. Rohani, S. Alizadeh, F. Sobhanmanesh, and R. Boostani, "Lip segmentation in color images," in *Proc. IEEE Int. Conf. Innov. Inf. Technol.*, Al Ain, UAE, Dec. 2008, pp. 747–750.

[21] E. Skodras and N. Fakotakis, "An unconstrained method for lip detection in color images," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Prague, Czech Republic, May 2011, pp. 1013–1016.

[22] P. Gacon, P.-Y. Coulon, and G. Bailly, "Non-linear active model for mouth inner and outer contours detection," in *Proc. 13th Eur. Signal Process. Conf.*, Antalya, Turkey, Sep. 2005, pp. 1–4.

[23] C. Bouvier, P.-Y. Coulon, and X. Maldague, "Unsupervised lips segmentation based on ROI optimisation and parametric model," in *Proc. IEEE Int. Conf. Image Process.*, San Antonio, TX, USA, Sep./Oct. 2007, pp. IV-301–IV-304.

[24] M. Li and Y.-M. Cheung, "Automatic segmentation of color lip images based on morphological filter," in *Proc. 20th Int. Conf. Artif. Neural Netw.*, Thessaloniki, Greece, 2010, pp. 384–387.

[25] S. Wang, A. W.-C. Liew, W. H. Lau, and S. H. Leung, "Lip region segmentation with complex background," in *Visual Speech Recognition: Lip Segmentation and Mapping*, A. W.-C. Liew and S. Wang, Eds. Hershey, PA, USA: IGI Global, 2009.

[26] A. W.-C. Liew, S. H. Leung, and W. H. Lau, "Segmentation of color lip images by spatial fuzzy clustering," *IEEE Trans. Fuzzy Syst.*, vol. 11, no. 4, pp. 542–549, Aug. 2003.

[27] K. Hara and R. Chellappa, "Growing regression forests by classification: Applications to object pose estimation," in *Proc. 13th ECCV*, 2014, pp. 552–567.

[28] Y.-M. Cheung, "Maximum weighted likelihood via rival penalized EM for density mixture clustering with automatic model selection," *IEEE Trans. Knowl. Data Eng.*, vol. 17, no. 6, pp. 750–761, Jun. 2005.

[29] A. W.-C. Liew and H. Yan, "An adaptive spatial fuzzy clustering algorithm for 3-D MR image segmentation," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1063–1075, Sep. 2003.

[30] S.-H. Leung, S.-L. Wang, and W.-H. Lau, "Lip image segmentation using fuzzy clustering incorporating an elliptic shape function," *IEEE Trans. Image Process.*, vol. 13, no. 1, pp. 51–62, Jan. 2004.

[31] S.-L. Wang, W.-H. Lau, A. W.-C. Liew, and S.-H. Leung, "Robust lip region segmentation for lip images with complex background," *Pattern Recognit.*, vol. 40, no. 12, pp. 3481–3491, 2007.

[32] X. L. Xie and G. Beni, "A validity measure for fuzzy clustering," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 13, no. 8, pp. 841–847, Aug. 1991.

[33] M. R. Rezaee, B. P. F. Lelieveldt, and J. H. C. Reiber, "A new cluster validity index for the fuzzy c-mean," *Pattern Recognit. Lett.*, vol. 19, nos. 3–4, pp. 237–246, 1998.

[34] A.-O. Boudraa, "Dynamic estimation of number of clusters in data sets," *Electron. Lett.*, vol. 35, no. 19, pp. 1606–1607, 1999.

[35] D.-J. Kim, Y.-W. Park, and D.-J. Park, "A novel validity index for determination of the optimal number of clusters," *IEICE Trans. Inf. Syst.*, vol. E84-D, no. 2, pp. 281–285, 2001.

[36] D.-W. Kim, K. H. Lee, and D. Lee, "On cluster validity index for estimation of the optimal number of fuzzy clusters," *Pattern Recognit.*, vol. 37, no. 10, pp. 2009–2025, 2004.

[37] K.-L. Wu and M.-S. Yang, "A cluster validity index for fuzzy clustering," *Pattern Recognit. Lett.*, vol. 26, no. 9, pp. 1275–1291, 2005.

[38] E. H. Ruspini, "A new approach to clustering," *Inf. Control*, vol. 15, no. 1, pp. 22–32, 1969.

[39] J. C. Bezdek, "Pattern recognition with fuzzy objective function algorithms," in *Advanced Applications in Pattern Recognition*. New York, NY, USA: Plenum, 1981.

[40] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, no. 3, pp. 379–423, 1948.

[41] J. C. Bezdek, "Cluster validity with fuzzy sets," *J. Cybern.*, vol. 3, no. 3, pp. 58–73, 1974.

[42] J. C. Bezdek, "Mathematical models for systematics and taxonomy," in *Proc. 8th Int. Conf. Numer. Taxonomy*, San Francisco, CA, USA, 1975, pp. 143–166.

[43] X. R. Li, K. Zhang, and T. Jiang, "Minimum entropy clustering and applications to gene expression analysis," in *Proc. IEEE Comput. Syst. Bioinform. Conf.*, Aug. 2004, pp. 142–151.

[44] J. Havrda and F. Charvát, "Quantification method of classification processes. Concept of structural α-entropy," *Kybernetika*, vol. 3, no. 1, pp. 30–35, 1967.

[45] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: Wiley, 1991.

[46] N. R. Pal and J. C. Bezdek, "On cluster validity for the fuzzy c-means model," *IEEE Trans. Fuzzy Syst.*, vol. 3, no. 3, pp. 370–379, Aug. 1995.

[47] Y.-M. Cheung, "A competitive and cooperative learning approach to robust data clustering," in *Proc. IASTED Int. Conf. Neural Netw. Comput. Intell.*, Grindelwald, Switzerland, 2004, pp. 131–136.

[48] M. K. Agoston, *Computer Graphics and Geometric Modeling: Implementation and Algorithms*. London, U.K.: Springer-Verlag, 2005.

[49] P. Soille, *Morphological Image Analysis: Principles and Applications*. Berlin, Germany: Springer-Verlag, 1999.

[50] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The Quickhull algorithm for convex hulls," *ACM Trans. Math. Softw.*, vol. 22, no. 4, pp. 469–483, 1996.

[51] A. Frank and A. Asuncion. (2010). *UCI Machine Learning Repository*. [Online]. Available: http://archive.ics.uci.edu/ml

[52] A. Martínez and R. Benavente, "The AR face database," CVC, Barcelona, Spain, Tech. Rep. 24, Jun. 1998.

[53] F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovač, "Color-based face detection in the '15 seconds of fame' art installation," in *Proc. Conf. Comput. Vis./Comput. Graph. Collaboration Model-Based Imag., Rendering, Image Anal., Graph. Special Effects*, Versailles, France, 2003, pp. 38–47.

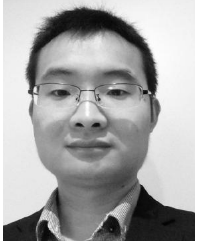
Yiu-ming Cheung (SM'06) received the Ph.D. degree from the Department of Computer Science and Engineering, Chinese University of Hong Kong, Hong Kong, in 2000.

He is currently a Full Professor with the Department of Computer Science, Hong Kong Baptist University, Hong Kong. His current research interests include machine learning, information security, signal processing, pattern recognition, data mining, and computer vision.

Dr. Cheung is a Senior Member of the Association for Computing Machinery. He is the Founding Chair of the Computational Intelligence Chapter of the IEEE Hong Kong Section. More details can be found at: http://www.comp.hkbu.edu.hk/∼ymc.

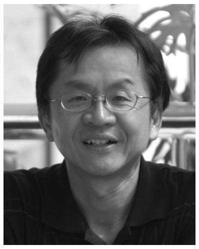
Meng Li received the B.E. degree from the Department of Automatic Test and Control, Harbin Institute of Technology, Harbin, China, in 2004, the M.E. degree from the Department of General and Fundamental Mechanics, Harbin Institute of Technology, in 2007, and the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, Hong Kong, in 2014.

His current research interests include human lip segmentation and Markov random field-based image processing.



Qinmu Peng received the B.E. degree from North China Electric Power University, Beijing, China, in 2008, the M.E. degree from the Huazhong University of Science and Technology, Wuhan, China, in 2011, and the Ph.D. degree from the Department of Computer Science, Hong Kong Baptist University, Hong Kong, in 2015.

His current research interests include image processing, pattern recognition, and machine learning methods in computer vision.



C. L. Philip Chen (S'88–M'88–SM'94–F'07) received the M.S. degree in electrical engineering from the University of Michigan, Ann Arbor, MI, USA, in 1985, and the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, USA, in 1988.

He was a Tenured Professor in the U.S. for 23 years, as a Department Head and Associate Dean in two different universities. He is currently the Dean of the Faculty of Science and Technology, University of Macau, Macau, China and a Chair Professor of the Department of Computer and Information Science. He is a Program Evaluator for the Accreditation Board of Engineering and Technology Education in USA in Computer Engineering, Electrical Engineering, and Software Engineering programs. His current research interests include systems, cybernetics, and computational intelligence.

Dr. Chen is a fellow of the American Association for the Advancement of Science. After being the IEEE SMC Society President from 2012 to 2013, he has been the Editor-in-Chief of the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS since 2014, and the Associate Editor of several IEEE TRANSACTIONS. He is also the Chair of TC 9.1 Economic and Business Systems of IFAC.