A&QT-R 2002 (THETA 13)
2002 IEEE-TTTC International Conference on Automation, Quality and Testing, Robotics
May 23-25, 2002, Cluj-Napoca, Romania

A NEW FUZZY C-MEANS BASED SEGMENTATION STRATEGY. APPLICATIONS TO LIP REGION IDENTIFICATION

Mihaela Gordan[*], Constantine Kotropoulos[**], Apostolos Georgakis[**], Ioannis Pitas[**]

[*]*Basis of Electronics Department, Technical University of Cluj-Napoca*
mihag@bel.utcluj.ro
Constantin Daicoviciu Street, No. 15, Cluj-Napoca, RO-3400, Romania
[**] *Artificial Intelligence and Information Analysis Laboratory, Department of Informatics, Aristotle University of Thessaloniki, Box 451, GR-54006 Thessaloniki, Greece*
{costas,apostolos, pitas}@zeus.csd.auth.gr

ABSTRACT

The problem of lip contour detection is critical in the lipreading systems based on contour processing. The typical contour detection strategy based on image segmentation in homogeneous regions fails in the case when the mouth images available for lipreading are low-contrast gray level images. Most of the solutions adopted require manual marking of some contour points. Here we propose a new solution from the image segmentation class suitable to lip contour extraction, using a modified version of the fuzzy c-means image segmentation algorithm. The novelty of the solution proposed consists in the types of features used for segmentation, which include not only luminance information (as in the standard use of fuzzy c-means), but also spatial information about the pixels in the image. After a simple filtering of the outliers, the contours of the resulting segmented objects are extracted. The experimental results obtained are superior to the ones obtained by standard or other versions of geometrically constrained fuzzy c-means, or by gradient-based edge detection strategies, without the need of manual marking of any contour points. Thus we consider the strategy proposed promising for automatic lip contour extraction applications.

KEYWORDS: fuzzy c-means, spatially constrained image segmentation, lipreading, lip contour detection

1. INTRODUCTION

A relatively large class of lipreading algorithms are based on lip contour analysis. Examples of such algorithms can be found in [1-3]. In these cases, lip contour extraction is needed as the first step. By lip contour extraction, we usually refer to the process of lip contour detection in the first frame of an audio-visual image sequence. Obtaining the lip contour in subsequent frames is usually referred as lip tracking. While for lip contour tracking there are well-developed techniques and algorithms to perform this task automatically, in the case of lip contour extraction in the first frame the things are different. This is a much more difficult task than tracking, due to the lack of a good a-priori information in respect to the mouth position in the image, the mouth size, the approximate shape of the mouth, mouth opening etc. So, while in lip contour tracking we have a good initial estimate of the mouth contour from the previous frame, this

initial estimate is not always available for the first frame, but it has to be produced by some means.

Different authors tried different procedures to solve the extraction of a good lip contour in the initial frame. Of course, the goal would be to solve this task automatically; approaches like region-based image segmentation and edge detection have been proposed. These methods work quite well in profile images and also in frontal images where the speaker wears lipstick or reflective markers. However, in the frontal images without any marking of the lips, the above-mentioned techniques unfortunately fail; and these images are the most used for speechreading. The problem of automatic extraction of the lip contour becomes even harder in the gray-level images, where the chromatic information differentiating between lips and skin is no longer present. Usually these images have a low contrast, so region-based segmentation and edge detection algorithms fail to provide good results [1,2]. In these cases, the solution adopted is based on marking manually more or less points on the lip contour and lip contour detection from interpolation or geometric modeling, or even on drawing manually the entire lip contour.

One of the most successful image segmentation algorithms into homogeneous regions is fuzzy c-means algorithm. There are a lot of visual applications reporting the use of fuzzy c-means, e.g. in medical image analysis, soil structure analysis, satellite imagery. [4-6]. Unfortunately the experiments demonstrated that it is still not suitable for the segmentation of mouth images into lip and skin regions. The problem comes from the fact that, in its standard use, the resulting regions are not spatially continuous, due to the fact that only the gray level uniformity is checked. Therefore, we propose an enhancement of the use of fuzzy c-means, to ensure spatially continuous regions after segmentation. Previous approaches on using fuzzy c-means with geometric constraints were reported in [7-9]. These approaches were based on using an extra-step to update the fuzzy partition, step in which geometric properties of the pixels in different sized neighborhoods (typically 3×3) are considered. However, the features used in the fuzzy c-means algorithm were the same as in the standard approach: just the gray levels (color components' intensities) of the pixels.

The approach proposed in this paper is based on the following principle: the features are modified to include also information about the spatial position of the pixels, not only their gray level values. For the time being we use the easiest way to include the spatial information: each pixel is represented by its luminance, its x and y coordinates. For the application addressed, on small size images, this technique allows us to obtain very good segmentation results, superior to the standard use of fuzzy c-means and to other versions of spatially constrained fuzzy c-means [7,8], without the need of any manually marked contour points.

2. THE PROPOSED FUZZY C-MEANS BASED IMAGE SEGMENTATION STRATEGY

The need for a new segmentation strategy of the mouth image into two homogeneous regions, namely the lip region and the skin region, comes from the observation that, in the particular application of standard fuzzy c-means algorithm for gray level mouth images, the resulting objects are not compact, meaning, the separation of the pixels into lip and skin regions is not accurately done. In other words, simply the gray levels of the pixels are not enough to differentiate between the lips and the skin, also due to the low contrast of the images; many outliers are present in both classes, so they are difficult to be filtered out. An example is given in Figure 1 (b).

Therefore, we propose the addition of new features in the data to be classified/clustered, aiming at preserving the topology of the neighboring pixels, considering the fact that both lips and skin areas are spatially continuous regions. In other words, the *neighboring pixels* with similar luminance should be kept together, and the distance between classes should take into account both the space distance and the gray level distance. To do this, the most straightforward approach is to consider each data point represented by its spatial position and its gray level. Therefore we will have a three-dimensional feature space, in which each pixel of the mouth image $\mathbf{p}_i$ will be represented by its *x* and *y* coordinates and its luminance *l*:

$$\mathbf{p}_i = \begin{pmatrix} x_i & y_i & l_i \end{pmatrix}^{\mathrm{T}} \tag{1}$$

where: $x_i \in [0...W-1]$; *W*- the image width; $y_i \in [0...H-1]$; *H* – the image height; $l_i \in [0...L_{Max}-1]$; L_{Max} – the maximum luminance level, e.g. 256. All the three components of the feature vector have integer values.Thus, for the $W \times H$ sized image, the dataset to be partitioned is $\mathbf{P}=\{\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{\mathrm{WH}}\}$. This dataset will form the universe of discourse, to be partitioned in *C* classes. In our particular application, although the classes are skin and lips, sometimes is more advantageous to partition the skin into two regions (to ensure their convexity) which leads to a number of *C*=3 classes instead of *C*=2. The partition matrix containing the membership degrees to the *C* classes (*C*=2 or *C*=3), $\mathbf{U}=[u_{ij}]$ of size $C \times W \cdot H$, and the set of class centers, $V=\{\mathbf{v}_1, \ldots, \mathbf{v}_{\mathrm{C}}\}$, result finally as output of the fuzzy c-means algorithm. Here, each $\mathbf{v}_i$ is a vector of three components $\mathbf{v}_i = \begin{pmatrix} v_{xi} & v_{yi} & v_{li} \end{pmatrix}^{\mathrm{T}}$ comprising an *x* coordinate, *y* coordinate and a luminance value. The partition matrix should satisfy the same constraints as in standard fuzzy c-means. With the use of the Euclidian distance, the cost function to be minimized becomes:

$$J_m(\mathbf{U},\mathbf{V}) = \sum_{j=1}^{WH} \sum_{i=1}^{C} u_{ij}^{\ m} \left(\mathbf{p}_j - \mathbf{v}_i\right)^{\mathrm{T}} \left(\mathbf{p}_j - \mathbf{v}_i\right) \tag{2}$$

As in the standard use of fuzzy c-means, we set the weighting coefficient *m* to *m*=2. Then, the modified fuzzy c-means algorithm runs iteratively in the following steps:

Step 1. Set *C* to 2 or 3 (depending on the illumination variance of the mouth image).

Step 2. Set the convergence error ε=0.001%.

Step 3. Set *m*=2.

Step 4. Initialize randomly the partition matrix, $\mathbf{U}=\mathbf{U}^0$. Set *j*=0.

Step 5. If $\max_{i=1,\ldots,C;\,k=1,\ldots WH} \left(u_{ik}^{\ j} - u_{ik}^{\ j-1}\right) > \varepsilon$, go to Step 6. Otherwise, go to Step 7.

Step 6.

6.1. $j=j+1$;

6.2. Compute the class centers:

$$\mathbf{v}_i^j = \frac{\sum_{k=1}^{W \cdot H} \left(u_{ik}^{j-1}\right)^2 \cdot \mathbf{p}_k}{\sum_{k=1}^{W \cdot H} \left(u_{ik}^{j-1}\right)^2} \quad \text{for each } i=1,\ldots, C$$

6.3. Update the fuzzy partition:

$$u_{ik}^j = \left(\sum_{n=1}^{C} \left(\frac{\left(\mathbf{p}_k - \mathbf{v}_i\right)^{\mathrm{T}} \left(\mathbf{p}_k - \mathbf{v}_i\right)}{\left(\mathbf{p}_k - \mathbf{v}_n\right)^{\mathrm{T}} \left(\mathbf{p}_k - \mathbf{v}_n\right)} \right)^2 \right)^{-1} \quad \text{for all } i=1,\ldots, C \text{ and } k=1,\ldots, W \cdot H.$$

6.4. Go to Step 5.

Step 7. Set the final partition matrix $\mathbf{U}=\mathbf{U}^j$ and the final class centers vector $\mathbf{V}=\mathbf{V}^j$.

After the final segmentation results are obtained, we label all the data as belonging to the most plausible class. The class label is encoded by the gray level of the corresponding class center $\mathbf{v}_i$, i=1, …, C, namely:

for each k, k=1,…, $W \cdot H$, $\mathbf{p'}_k = \begin{pmatrix} x_k & y_k & v_{li} \end{pmatrix}^T$, such that $u_{ik} = \max_{n=1,\ldots,C} u_{nk}$

The last step of the region-based segmentation scheme proposed refers to the elimination of the outliers present inside each region. A pixel in the segmented image is defined as *outlier* if most of the pixels in a 3×3 neighborhood around it belong to another class than the pixel under investigation. In this case, the pixel is flagged as outlier, and in a second pass, it will be "moved" to the class to which most of its neighbors belong. In this way, the "noisy" class decisions will be eliminated, the result being a set of smooth continuous regions.

3. APPLICATION OF THE PROPOSED STRATEGY TO LIP CONTOUR EXTRACTION IN LOW CONTRAST GRAY LEVEL IMAGES

As stated in Introduction, the final goal of the proposed new strategy of fuzzy c-means based image segmentation is to improve the lip contour extraction. The critical problem appears in the case of low contrast gray-level mouth images. Here we differentiate two categories of images, which will require two different approaches:

(a) medium-low contrast gray level mouth images, with slightly variable illumination. In this case, we can process the entire mouth image at once, but due to the illumination variance inside the mouth image and to the non-convexity of the skin region, a segmentation in C = 3 classes might be needed. Namely, one class will represent the lips, and the other two classes will represent the skin areas, such that each skin area is mostly convex.

(b) very-low contrast gray level mouth images, with variable illumination. In this case, applying the proposed fuzzy c-means algorithm on the entire mouth image at once still fails to provide good segmentation results. A better approach is to divide the mouth image into four subimages , the upper-left, the upper-right, the lower-left and the lower-right ones. Thus, after applying the segmentation on each subimage and extracting the contour on each segmented subimage, we will obtain the piecewise lip contour, which can be finally joint by interpolation. An example of such a splitting of a mouth image into 4 subimages is given in Figure 1 (a).

For the time being only the lip contour extraction in mouth images with closed mouth is considered. For open mouth images, the number of classes will increase to include teeth region, tongue region etc. We must notice here that, in the case of splitting the mouth image into 4 subimages, only a segmentation in C = 2 classes is needed, because now the skin region in the subimage becomes convex.

The first three processing steps: dataset building, the modified fuzzy c-means algorithm and outlier filtering were described in detail in the previous section. The last step of the segmentation scheme is to ensure the final aim of the proposed region-based segmentation strategy: to find the lip contour. This becomes an easy task once we have available a good region-based segmentation. We must note that, in the case of mouth image segmentation in C = 3 classes, we will probably get in the gradient image an extra-boundary between the 2 skin regions, or inside the lip region, which does not represent a lip boundary, but these false boundaries can be easily recognized based on their shape or area enclosed or to the fact they are not closed boundaries as the lip contour must be. The quality of the extracted lip contour will be much better than in the case of classical edge detection schemes or simple fuzzy c-means region-based segmentation schemes.

4. EXPERIMENTAL RESULTS

In order to evaluate the performances of the new proposed region segmentation-based strategy in the lip contour detection application, we implemented software the above described system in C++. The estimation of the results is done by visual examination on a test set of different gray level mouth images, both of the segmented images and of the extracted lip contour, by superimposing it on the original mouth images.

The proposed segmentation strategy is compared to other variants of fuzzy c-means: standard fuzzy c-means; rule-base neighborhood enhancement (RB-NE) fuzzy c-means [7]. As test set, we selected two classes of mouth images:
(i) gray level mouth images with medium-low contrast: we manually selected rectangular regions comprising the mouth from the images "Lena", "Lisa" and "Girl";
(ii) gray level mouth images with very low contrast, from the audio-visual database Tulips1 [10]: closed mouth images of the subjects Anthony, Ben, Candace, George. These images were splitted into four subimages prior to segmentation.

The original images from class (a) and one example test image from class (b) (subject Anthony) from the test set are given in Figure 1 (a).
The segmentation results using standard fuzzy c-means, proposed fuzzy c-means and RB-NE fuzzy c-means are given in Figure 1 (b), (c) and respectively, (d). We can see visually the better performance of the proposed strategy in all the test cases. The only case where our algorithm, the same as the others, fails, is for the lower half of the mouth image depicted from Tulips1 database.

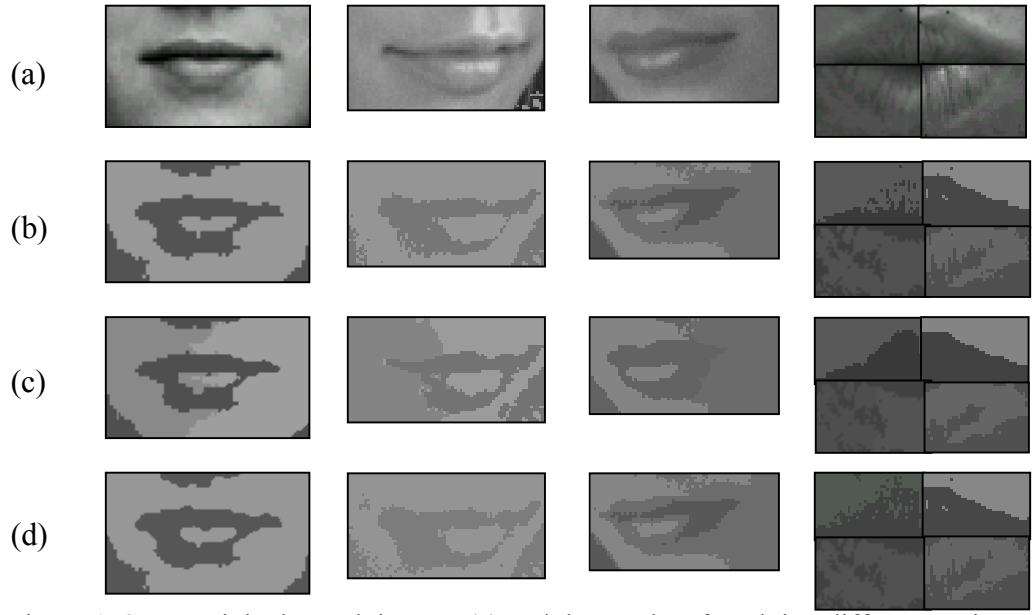

Figure 1. Some original mouth images (a) and the results of applying different variants of the fuzzy c-means algorithm for their segmentation: (b) standard fuzzy c-means; (c) the proposed enhancement of fuzzy c-means; (d) RB-NE fuzzy c-means. From left to right, the four original images represent: mouth areas of the images "Girl", "Lena", "Lisa"; the mouth image for subject "Anthony" from Tulips1, split in 4 subimages

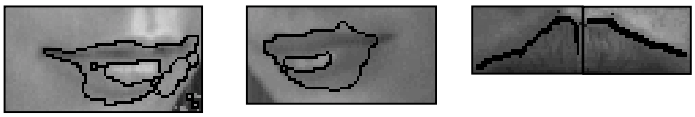

Figure 2. Examples of resulting region boundaries after the proposed modified fuzzy c-means segmentation; from left to right: for the mouth image of "Lena"; "Lisa"; "Anthony" (Tulips1).

Some resulting boundaries on the region-based segmented image, overlapped on the original mouth image/subimages, are demonstrated in Figure 2. Here we notice the good quality of the extracted lip contour. The false boundaries present in the images for the 3-class segmentation case can be easily filtered out, since they either are non-closed boundaries (so they cannot be lip boundaries), either the boundary is closed but is not the outermost one, so it cannot be the outer lip contour.

5. CONCLUSIONS

We proposed a new strategy for extracting object boundaries in low contrast gray level images, aiming to solve the problem of lip contour extraction in gray level low contrast mouth images in the first frame. Starting from a well known algorithm, fuzzy c-means, we modified its standard use by including in the feature vector spatial information about the pixels positions. The experimental results on a set of various low contrast gray level mouth images show better performance of our algorithm in terms of compactness of the segmented regions as compared to the standard fuzzy c-means or other versions of geometrically guided fuzzy c-means. The solution proposed makes possible a completely automated contour extraction, since no manually marked points on the lip contour are needed. In our future research we will examine the inclusion of some texture features in the feature vector and the use of more elaborated distance measures in the fuzzy c-means algorithm for a better shape variability of the clusters.

REFERENCES

[1] R. Kaucic, B. Dalton, A. Blake (1996), "Real-time lip tracking for audio-visual speech recognition applications", *Proc. European Conf. Computer Vision*, Cambridge, UK, pp. 376-387

[2] J. Luettin, N. A. Thacker, S. W. Beet (1996), "Active shape models for visual speech feature extraction", *Speechreading by Humans and Machine, NATO ASI Series, Series F: Computer and Systems Sciences*, Springer Verlag, Berlin, 150:383-390

[3] I. Matthews, T. Cootes, S. Cox, R. Harvey, J. A. Bangham (1998), "Lipreading using shape, shading and scale", *Proc. Auditory-Visual Speech Processing*, Sydney, Australia, pp. 73-78

[4] S. E. Crane, L. O. Hall (1999), "Learning to identify fuzzy regions in magnetic resonance images ", *18th Int. Conf. of NAFIPS*, pp. 352 -356

[5] P. Thitimajshima (2000), "A new modified fuzzy c-means algorithm for multispectral satellite images segmentation", *Proc. Geoscience and Remote Sensing Symposium*, Vol. 4, pp. 1684 -1686

[6] D. L. Pahm, J. L. Prince (1999), "Adaptive fuzzy segmentation of magnetic resonance images", *IEEE Transactions on Medical Imaging*, Vol. 18, no. 9, pp. 737-752

[7] Y. A. Tolias, S. M. Panas (1998), "On applying spatial constraints in fuzzy image clustering using a fuzzy rule-based system", *IEEE Signal Processing Letters*, Vol. 5, no. 10, pp.245-247

[8] T. D. Pham (2001), "Image segmentation using probabilistic fuzzy c-means clustering", *Proc. Int. Conf. On Image Processing*, Thessaloniki, Greece, Vol. 1, pp. 722-725

[9] J.C. Noordam, W.H.A.M. van den Broek (2000), "Geometrically Guided Fuzzy C-Means Clustering for Multivariate Image Segmentation", *Proc. Int. Conf. on Pattern Recognition*, pp. 462-465

[10] J. R. Movellan (1995), "Visual Speech Recognition with Stochastic Networks", *Advances in Neural Information Processing Systems, (G. Tesauro, D. Toruetzky, and T. Leen, Eds.)*, MIT Press, Cambridge, MA, Vol 7