

# WISEME SEGMENTATION BY LDA HYSTERESIS

Stanisław Badura<sup>1</sup>, Mariusz Leszczyński<sup>1</sup>, Marco Marcon<sup>2</sup>, Jürgen Rurainsky<sup>3</sup>, Władysław Skarbek<sup>1</sup>

<sup>1</sup>Warsaw University of Technology, Faculty of Electronics and Information Technology,

<sup>2</sup>Politecnico di Milano, Dipartimento di Elettronica e Informazione

<sup>3</sup>Fraunhofer Institute for Telecommunications, Heinrich-Hertz-Institute,  
Image Processing Department

## ABSTRACT

In this paper a viseme segmentation method based on Linear Discriminant Analysis (LDA) is presented in order to separate facial skin from lips. The algorithm works on color pictures of human faces using a novel rbS color space. Final effect is achieved by applying a hysteresis concept of two thresholds to LDA features what enables a flexible viseme segmentation.

## 1. INTRODUCTION

This research has been conducted to support analysis of visemes i.e. shape and appearance of human lips, teeth and tongue during speech. That is a part of being developed talking head application. Besides our dedicated solution, lip segmentation is an important and often researched problem in the field of computer vision. Viseme segmentation is one of the first step in a way to our final goal, so it is required the step to be fast enough with satisfactory precision of lips localization. Therefore we have decided to find an alternative solution of the problem, which gives us better results than the existing known for us methods [1][2].

In this paper we present our viseme segmentation method based on LDA algorithm[3]. The main problem in the viseme segmentation is to find a boundary between skin and lip areas. General properties of LDA result in a good separation of previously trained classes of some objects. Let's assume that the skin and lip colors create two different classes, so a scalar feature LDA extractor is supposed to separate them correctly.

## 2. IMAGE PREPROCESSING

Our main research goal is focused on analysis of face reaction during speech, especially on the mouth region. For our viseme segmentation method we use only color pictures of face, restricting to the lower part of facial

picture taken from a video sequence. The example of such picture is shown in in the Fig. 1.



Figure 1 Example of source image.

During the preprocessing stage every picture before further processing is filtered to reduce some noise and distortion. We have used Gaussian 3x3 smoothing filter with a mask as follow:

1	2	1
2	4	2
1	2	1

## 3. COLOUR SPACE

All source images are acquired in RGB colour space, the color space appropriate for a single picture [4]. However in case when a video sequence is processed, light condition of the environment can change. Therefore we propose a change of the color space to a normalized (rbS) version of RGB color space. It partially eliminates the

effect of lighting conditions variability. We have used the following transform of color space:

$$r=R/(R+G+B);$$

$$b=B/(R+G+B);$$

$$S= 1- \min(R,G,B)/\max(R,G,B);$$

$$r=r*S; b=b*S;$$

#### 4. LDA ALGORITHM

After the color transform we have optimized LDA algorithm to our case. There are two class of data (lip colour and skin colour) represented in three dimensional vector space  $([r,b,S])$ .

For the training data, manually defined sets of points in face pictures were created. There were checked from few hundreds to few thousands of pixels for each class. (cf. Fig. 2).

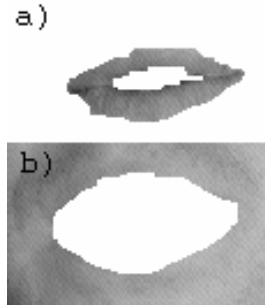


Figure 2 Manually selected training data: a) lips class; b) skin class.

For that collected data the following steps are performed:

1) Class mean shifting of the training sequence:

$$X = [x_1, \dots, x_L],$$

2) Grand mean shifting for class means:

$$M = [m_1, \dots, m_C],$$

3) Singular Value Approximation for  $X$  with subspace dimension

$$[U_q \Sigma_q] = sva(X, q); A_q = U_q \Sigma_q^{-1};$$

4) Whitening of columns in  $M$ :

$$M = A_q^t M;$$

5) Singular Value Approximation for  $M$  with subspace dimension

$$V_r = sva(M, q); W = A_q V_r;$$

6) Return  $W$ ;

In our case the only possible value of  $r = C - 1 = 1$ . The LDA feature  $y = W^t x$  for the pixel  $x$  is classified by the distance  $d_i$  to LDA features  $y_i = W^t x_i$ , representing the lips or skin colour class  $i = 1, 2$ :

$$d_i = \|y - y_i\|^2;$$

$$i_{opt} = \arg \min_{1 \leq i \leq 2} d_i;$$

Experiments have helped to tune the parameter  $q$  to  $q=2$ .

#### 5. HYSTERESIS

Above classifier discriminates only two classes with very similar color properties, so it can still gives wrong classification. The condition  $\min(d_i)$  can indicates some pixels as skin because the distance  $d_i$  to skin class is smaller then distance to lips class. But the distances can be nearly equal and addition the distance to lips can be smaller then distance of some pixels classified as lips class. Therefore we have applied a hysteresis for the classifier based only on the distance  $d_i$  to lip class.

The hysteresis has two thresholds:

- lower  $\gamma_l$
- upper  $\gamma_u$

The distance  $d_i < \gamma_l$  means that given pixel belongs to lips class.

The distance  $d_i > \gamma_u$  means that given pixel doesn't belong to lips class.

The distance  $\gamma_l < d_i < \gamma_u$  results in the class selection which has been assigned for the last visited pixel.

#### 6. EXPERIMENTS

The viseme segmentation is performed by classifying each pixel on the basis of the given color picture. A part of the picture with the majority of pixels classified as lips represent a viseme area.

Thresholds  $\gamma_l$  and  $\gamma_u$  were fixe on the basis of training data. In the intermediate stage the classifier produces a grayscale image which corresponds to LDA distances to class means. Total black pixels are meaning absolutely

lips pixels, white pixels are meaning non lips pixels and intermediate values mean pixels must be further analyzed.

The following pictures include example lips pixels localization for different parameters of hysteresis.

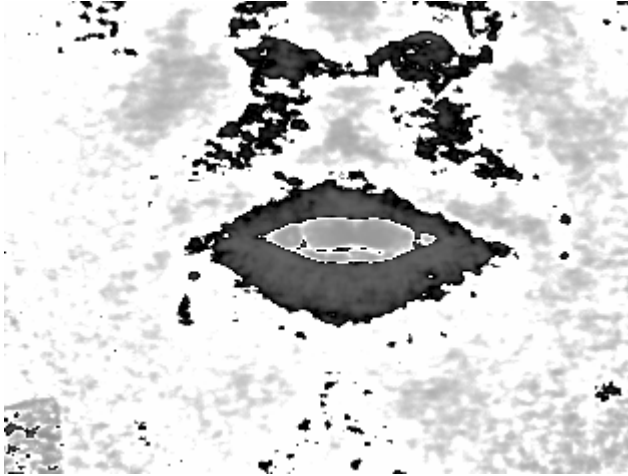


Figure 3 Lips pixels segmentation for hysteresis parameters  $\gamma_l=0.1$  and  $\gamma_u=2.0$ .



Figure 4 Lips pixels segmentation for hysteresis parameters  $\gamma_l=0.5$  and  $\gamma_u=1.5$ .



Figure 5 Lips pixels segmentation for hysteresis parameters  $\gamma_l=0.8$  and  $\gamma_u=1.2$ .

### 6.1. Connected components

As we can see, simple lips pixels segmentation gives lots of small noisy blobs. To remove them we have applied a connected components algorithm [5]. The algorithm allows to localize all blobs and their size measured by the amount of the black pixels. The lips are supposed to be the biggest connected components, so all smaller blobs are removed. The result of such operation is illustrated in the Fig. 6.



Figure 6 Lips pixels segmentation for hysteresis parameters  $\gamma_l=0.5$ ,  $\gamma_u=1.2$  and applied connected component algorithm.

## 6.2. Viseme Segmentation

The last step of the whole processing is the selection of the viseme area. As we have mentioned in the introduction this research has been conducted to support the viseme recognition. Among the analyzed methods of viseme recognition [6], the best one is based on the segmented visemes in the form of a rectangle which encloses the lips, the teeth and the tongue.

Performing connected components algorithm gives us an image with clear selected area of lips. In that case is easy to find the required viseme rectangle (see Fig. 7).



Figure 7 Example of segmented viseme.

## 7. CONCLUSION

The proposed method for the viseme segmentation using LDA hysteresis gives us an improved tool for lips segmentation. Presented results demonstrate the high quality of segmentation. In addition we have obtained a satisfactory time complexity. Namely, the single picture of resolution 320x240 pixels is processed in 20ms on PC Pentium IV 3.2GHz. The time includes all pre and post processing steps. Further time reduction can be obtained using ROI tracking in the video sequence. Further segmentation improvement can be achieved if the number of color classes is extended by adding teeth, tongue and facial hair as subclasses for non lips colors..

## 8. ACKNOWLEDGEMENT

The work presented was developed within VISNET, a European Network of Excellence (<http://www.visnet-noe.org>), funded under the European Commission IST FP6 programme.

## 8. REFERENCES

[1] L. Lucchese and S.K. Mitra, "Color Image Segmentation: A State-of-the-Art Survey," Image Processing, Vision, and Pattern Recognition, Proc. of the Indian National Science Academy (INSA-A), New Delhi, India, 2001.

[2] L. Lucchese and S. K. Mitra, "An algorithm for fast segmentation of color images, " Proc. IEEE 10-th Tyrrhenian Workshop on Digital Communication , Ischia, Italy, September 1998.

[3] Bober M., Kucharski K., and Skarbek W., Face Recognition by Fisher and Scatter Linear Discriminant Analysis, in Computer Analysis of Images and Patterns, eds. Petkov N., Westenberg M., Springer LNCS 2756, 638:645, 2003.

[4] Moritz Störring, "Computer Vision and Human Skin Colour", Ph.D. Dissertation, Computer Vision & Media Technology Laboratory, Faculty of Engineering and Science, Aalborg University 2004.

[5] Ali Rahimi, "Fast Connected Components on Images, " MIT Media Lab, Jul 22, 2001.  
<http://web.media.mit.edu/~rahimi/connected/>

[6] M.Leszczynski, W.Skarbek, S.Badura , Fast Viseme Recognition for Talking Head Application, ICIAR 2005.