# Open or Closed Mouth State Detection: Static Supervised Classification Based on Log-polar Signature

Christian Bouvier, Alexandre Benoit, Alice Caplier, Pierre-Yves Coulon

▶ **To cite this version:**

Christian Bouvier, Alexandre Benoit, Alice Caplier, Pierre-Yves Coulon. Open or Closed Mouth State Detection: Static Supervised Classification Based on Log-polar Signature. ACIVS 2008 - International Conference on Advanced Concepts for Intelligent Vision Systems, Oct 2008, Juan-Les-Pins, France. pp.1093-1102. hal-00372148

**HAL Id: hal-00372148**
**https://hal.science/hal-00372148**

Submitted on 31 Mar 2009

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Open or Closed Mouth State Detection: Static Supervised Classification Based on Log-polar Signature

Christian Bouvier[1], Alexandre Benoit[1], Alice Caplier[1], Pierre-Yves Coulon[1]

[1] GIPSA_lab, INPG, CNRS, UJF, U.Stendhal
46 av. F. Viallet, 38031 Grenoble, France
{christian.bouvier, benoit, caplier, Pierre-Yves.Coulon}@lis.inpg.fr

**Abstract.** The detection of the state open or closed of mouth is an important information in many applications such as hypo-vigilance analysis, face features segmentation or emotions recognition. In this work we propose a supervised classification method for mouth state detection based on retina filtering and cortex analysis inspired by the human visual system. The first stage of the method is the learning of reference signatures (Log Polar Spectrums) from some open and closed mouth images manually classified. The signatures are constructed by computing the amplitude log-polar spectrum of the retina filtered images. Principal Components Analysis (*PCA*) is then performed using the Log Polar Spectrum as feature vectors to reduce the number of dimension by keeping 95 % of the total variance. Finally a binary *SVM* classifier is trained using the projections the principal components given by the *PCA* in order to classify the mouth.

**Keywords:** Face Analysis, Open or closed mouth state detection, classification, log-polar signature.

## 1 Introduction

Previous works have been done on human face components detection and segmentation such as mouth, lips, eyes… Here we are interested in determining the state, open or closed, of the mouth. Knowledge about the state of the mouth is very important for applications such as hypo-vigilance analysis, emotions recognition and facial features detection.

Mouth analysis methods are mainly focused on segmentation. Many techniques have been developed and we can, mainly, classify those techniques in 2 families, contour based approaches and region based approaches. Most of these methods deal implicitly with the mouth state as the inner contours of the mouth is extracted.

In [1] a region based approach is used to segment the lips. Markov random fields combining color and movement information are used to segment the area of the mouth and then an active contour is defined on the mask to extract the outer and inner contours of the mouth. This method can give accurate results but the problem with Markov random fields is the initialization of the color distributions for the relaxation

process. Moreover, the final mask can lead to impossible results because the shape of the mouth is not constraint.

Statistical methods have been developed to extract facial features and particularly the mouth [2, 3]. The model composed of a limited number of key points for the outer and inner contours is directly optimized, but the nonlinearity of the model, especially for the interior of the mouth, imposes to have a very good initialization. In [4] a supervised method has been developed to explicitly classify mouth shape. An Active Shape Model (*ASM*) using key points as initialization points is optimized for contours extraction. The best parameters found for the *ASM* are then used for mouth state classification using Support Vector Machine (*SVM*).

More recently, Benoit et al. [5] used motion information and a frequency approach inspired by a human visual system (*HVS*) modeling in order to find the state of the mouth and more generally to characterize the state of hypo-vigilance of a human subject.

Our goal in this work is to achieve the best static detection of the mouth open/closed state and to compare our classifier to standard classification methods. We make the hypothesis that the mouth has been roughly detected in a preliminary processing [5, 6]. We work with mouth image center on the mouth excluding other face features. Section 2 shortly describes the retina and cortex models used to compute Log Polar Spectrum corresponding to the signature of the images. Though the retina and cortex models [7, 8] are inspired by the *HVS* they are used has processing step in our classifier. Section 3 describes the state detection algorithm. It includes the training of log polar spectrum models used for classification and the classification procedure to detect the states of the mouth. In section 4 we present experimental results.

## 2 Retina and Visual Primary Cortex Modeling

Figure 1 gives an overview of the processing steps. First the ROI of the input is processed by the retina filter in order to enhance the contours of the picture. The filtered image is then sent to the Cortex V1 analysis stage in order to compute the log-polar spectrum of the image which is then used at a classification stage to determine the state of the feature.

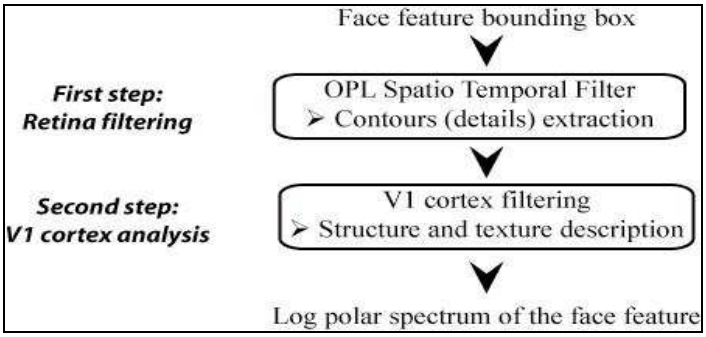

**Fig. 1.** Overview of the processing steps

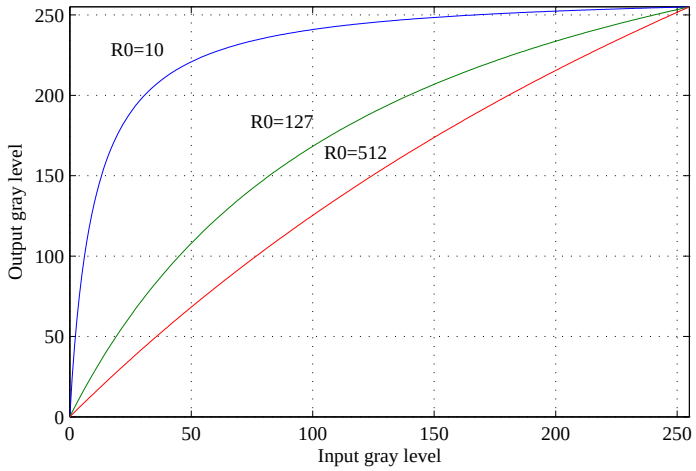


**Fig. 2.** Photoreceptors gain adjustment using $R_0(p)$

### 2.1 Retina Filtering

The retina filter is inspired by the *HVS* and consists basically in 3 different steps of filtering [7, 8]. The retina is composed of 3 layers of specialized cells, the photoreceptor layer, the Outer Plexiform layer (*OPL*) and the Inner Plexiform layer (*IPL*). The goal of the photoreceptors is to convert the light stimulus into electric potential and have the ability to adapt their dynamic to the local luminance [7]. This property of adaptive compression called the photoreceptor compression can be modeled by the Michaelis-Menten equation [7] adapted to 8 bits luminance pictures (1).

$$\begin{cases} r(p) = \dfrac{R(p)}{R(p) + R_0(p)} \cdot (255 + R_0(p)) \\ R_0(p) = \dfrac{V_0}{256} \cdot L(p) + (255 - V_0) \end{cases} \quad (1)$$

$p$ corresponds to the spatial position, $r(p)$ to the corrected luminance, $R(p)$ to the input luminance and $R_0(p)$ adjust the photoreceptors gain depending on the local luminance $L(p)$. $V_0$ has been experimentally set to 230.

Fig. 2 shows the non linear gain adaptation for different values of $R_0$(p). Fig. 3 shows the effect of the photoreceptor compression on a mouth image. We can see that the contrast of the dark areas is greatly improved by the adaptive compression.

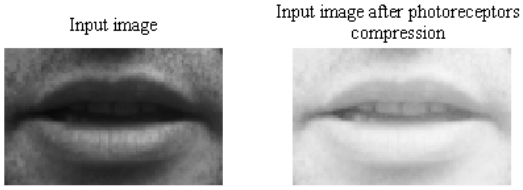


**Fig. 3.** Photoreceptors compression

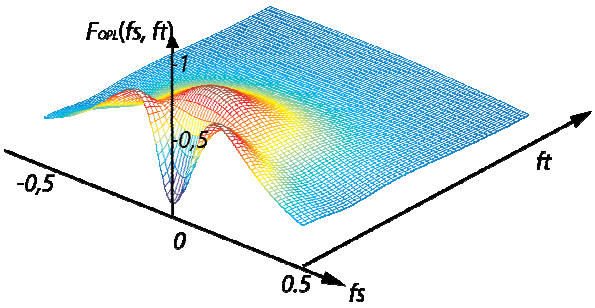


**Fig. 4.** *OPL* transfer function.

The Outer Plexiform Layer (*OPL*) and the Inner Plexiform Layer (*IPL*) of the retina process the visual information with specific filtering. Two information channels are then extracted from those filtering steps, the Parvocellular (Parvo) channel dedicated to detail analysis (static contours enhancement, see Fig. 5) and the Magnocellular (Magno) channel dedicated to motion analysis (moving contours enhancement). In the present work we are interested in finding mouth state on static images only. As a consequence, we will only use the Parvo channel which can be extracted at the *OPL* level [7]. The *OPL* is modeled by the $G_{OPL}$ transfer function (2). In the case of static image analysis it corresponds to $ft$=0 (see Fig. 4).

$$
\begin{aligned}
&G_{OPL} = G_c(fs, ft)\left[1 - G_h(fs, ft)\right] \\
&With: G_i(fs, ft) = \frac{1}{1 + \beta_i + 2\alpha_i\left(1 - \cos(2\pi fs)\right) + j2\pi\tau_i ft} \\
&With: \alpha_i = r_i / R_i,\ \beta_i = r_i / r_{m,i},\ \tau_i = r_i C_i
\end{aligned}
\qquad (2)
$$

$fs$ spatial frequency, $ft$ temporal frequency, $r_i$, $R_i$ are resistors and $C_i$ are capacities that model the retina synaptic network of photoreceptors and horizontal cells.

This filter is a non separable spatio-temporal filter which has a band pass effect in low temporal frequencies (Fig. 4) which induces static contours enhancement. It also

induces a spectral whitening effect. Fig. 5 illustrates the effects of the retina filter on a mouth image.

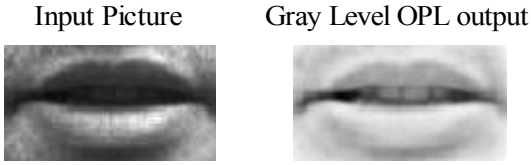


**Fig. 5.** Example of retina filter output

### 2.2 Primary visual cortex modeling and spectrum analysis

The Parvo channel is sent to the cortex V1 model [9]. In the V1 area, neurons are organized in layers and information is preferably transmitted in neurons columns dedicated to specific orientations. In this cortex area, the visual information is decomposed in frequency bands and orientations [10]. The model used in this work consists in the image spectrum processing proposed in [9]: the *FFT* of the Parvo channel signal is first computed and is then sampled using a set of specific filters. In [9] the author propose Gabor in log polar filters (*Glop*) that sample the *FFT* spectrum at specific orientations and frequency bands (see eq. (3)) which have the property to be symmetric in log scale. This property is important if we consider zoom effect because this would yield to a simple energy translation along the frequency axis rather than a more complex spectrum transformation.

$$G_{i,k}(f,\theta) = \frac{1}{\sigma\sqrt{2\pi}}\left(\frac{f_k}{f}\right)^2 \exp\left(-\frac{\ln\left(\frac{f}{f_k}\right)^2}{2\sigma^2}\right)\cos\left(\frac{1+\cos(\theta-\theta_i)}{2}\right)^2 \quad (3)$$

These filters are centered on normalized frequency $f_k$ at the orientation $\theta_i$, with the scale parameter $\sigma$ with $0 < \theta_i \leq \pi$ and $0.1 \leq f_k \leq 0.3$. Those filters are also normalized so that the integral of a filter is always equal to 1. In this work 15 orientations and 15 frequency bands are used which is close to the biological model. Then by computing the output energy of each filter, we obtain a sampled amplitude spectrum of the signal coming from the *OPL* in log polar domain (Fig. 6-c). This spectrum contains information about the structure and the texture of the input image which corresponds to specific energies by frequency bands and orientations. Given the sampled energy spectrum, the spectrum in *dB*, $LPS_{dB}$, is computed. The goal here is to enhance the secondary orientations of the spectrum. Fig. 6-c) shows normalized linear log polar spectrum and Fig. 6-d) shows the log polar spectrum in *dB*. We can see one main orientation in the horizontal direction (180°) on the linear spectrum whereas on the spectrum in *dB*, the secondary orientations are enhanced and give more information on the global energy distribution of the input image.

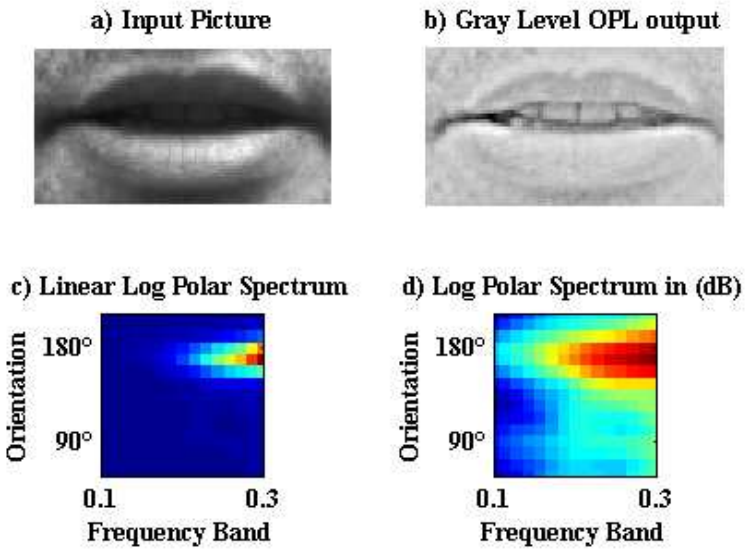


**Fig. 6.** Log Polar Spectrum of a mouth image, a) Input Picture, b) Gray Level OPL output, c) *LPS* (Linear Log Polar Spectrum), d) $LPS_{dB}$ (Log Polar Spectrum in dB).

# 3 Mouth State Detection

## 3.1 Problem Statement

Our goal is to detect if the mouth is open or closed by considering the luminance information of static images. Using the models described in section 2, the idea is to use the image description given by the Log Polar Spectrum as an image signature in order to determine the state of the input feature.

In [5], Benoit et al. use the temporal evolution of the log polar spectrum total energy to compute adaptive thresholds mouth state detection. The hypothesis is that the energy of an open mouth is higher than the energy of a closed one because of the presence of more contours on open mouth pictures. The adaptive thresholds are used to compute the dynamic state of the mouth that is opening or closing based on the variation of the global energy. The spatial energy distribution is not taken into account.

In [9] the authors developed a method for natural image classification using Log Polar Spectrum. The images are classified in cluster such as city, beach and mountain. The mean log polar spectrum is computed for each cluster using spectra normalized by frequency band. Then images are classified by using criteria such as Euclidean distances, Minkowski distances between the input normalized log-polar spectrum and

the cluster mean spectra. In that case, the energy is not taken into account because the models and the classification are based on models normalized by frequency band.

In the present work we choose a supervised classification method based on the log polar spectrum database of mouth state and a Support Vector Machine (*SVM*) classifier. The log polar spectrums are very suitable for that kind of classifier because the number of orientation and frequency band are set and fixed for all images and independent of the input image resolution.

### 3.2 Learning database

The mouth database is composed of 900 mouth images from 18 different subjects. Each image only contains the mouth region of interest (a bounding box around the mouth). The image has been manually classified as "open" or "closed". The closed mouth set is composed of 230 images. Only images of completely closed mouth have been used. The open mouth set is composed of 670 images of different shapes with different levels of mouth opening. Mouth images have been considered as open when teeth, tongue or interior of the mouth is visible. See Figure 7 for training images examples.

**Fig. 7.** Examples of feature images used to compute the log-polar signature. The first row gives examples of open mouth images; the second row shows examples of closed mouth.

### 3.3 SVM classifier

Given our database of mouth images manually classified, the log polar spectrums for the entire database are computed. The spectrums columns are then concatenated to form vectors of 225 values. A Principal component analysis is then run to reduce the number of dimensions. We choose to keep 95% of the total variance. This leads to keep only 6 principal components. The log polar spectrums are then projected on these principal components and a binary *SVM* classifier is trained using the projection parameters and the associated state. On Fig. 8 we give the projections of the log polar spectrums on the 2 first principal components computed by *PCA* using the entire database.

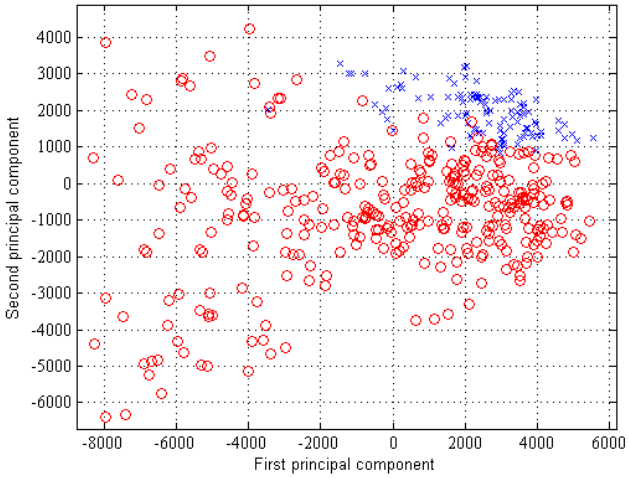


**Fig. 8.** Log polar spectrum projections on the 2 first principal components given by *PCA* when the retina filter is used as pre-processing. Blue cross correspond to the closed mouth and red circles correspond to the open mouth.

## 4 Experimental Results and Conclusion

The performances of the algorithm have been tested on the training database of 900 pictures from 18 different subjects.

We computes the mean classification rate for all the subjects in our database in the case where their log polar spectrum are include for the *PCA* computation and the *SVM* training and in the case where there are removed from the training step (leave one out test). All the percentages given correspond to percentage of good classification. The results are given in Table 1 for the proposed algorithm.

**Table 1.** Experimental results for the proposed algorithm

| Correct classification (%) | Mouth |
|---|---|
| With the subject log polar spectrums in the *SVM* training | 98.4 % |
| The subject log polar spectrums are removed from the *SVM* training | 95.3 % |

In order to show the efficiency the photoreceptor compression and of the retina filter we have tested our algorithm in the case where the photoreceptor compression and the retina filter are replaced by a simple illumination correction (zero mean and unit variance). The results are given in Table 2.

**Table 2.** Experimental results using "zero mean unite variance" illumination correction

| Correct classification (%) | Mouth |
|---|---|
| With the subject log polar spectrums in the *SVM* training | 86 % |
| The subject log polar spectrums are removed from the *SVM* training | 81 % |

The results show clearly a performance improvement when the photoreceptor compression and the retina filter are applied. We also tested the effect of the cortex V1 model by comparing the performance of the algorithm to a multiresolution Local Binary Patterns (*LBP*) based algorithm with 3 different scales [11]. The classification is also done by training a binary *SVM* classifier on the *LBP* histogram for all the images in the database. In the first case the photoreceptor compression and the retina filter are applied to the input image. The results are given in Table 3. Then we tested the *LBP* based classification with "zero mean unite variance" illumination correction (Table 4).

**Table 3.** Experimental results using Photoreceptor compression and retina filter with *LBP* as texture feature.

| Correct classification (%) | Mouth |
|---|---|
| With the subject log polar spectrums in the *SVM* training | 78 % |
| The subject log polar spectrums are removed from the *SVM* training | 75 % |

**Table 4.** Experimental results using "zero mean unite variance" illumination correction and *LBP* as texture feature.

| Correct classification (%) | Mouth |
|---|---|
| With the subject log polar spectrums in the *SVM* training | 77 % |
| The subject log polar spectrums are removed from the *SVM* training | 74 % |

We can see that the log polar spectrum gives stronger performance for mouth classification than the *LBP* based algorithm.

Finally we tested the model trained on our database on the AR database [12] for validation. The database is composed of face images from 126 subjects with different facial expressions, different illumination conditions and occlusions. We extracted 473 images of closed mouth and 537 of open mouth from 126 subjects. The classification rate is 97.4 %.

We can see that for all our tests the classification rate is above 95% even for the case of completely unknown images. We can also see the pertinence of the retina filtering and cortex V1 model in order to classify mouth by state. Currently the retina

filter and the cortex V1 model algorithms are implanted using Matlab and the complete process can be achieved at a rate of 1 image per second.

We presented a supervised method for mouth state detection. The algorithm based on the analysis of a bio-inspired signature gives good results and proves the relevance of the approach. The retina and the cortex V1 models yield to a compact spectrum, easy and fast to analyze. The next step will be estimate the opening degree of the mouth. The performance for the opening degree estimation is not satisfying for now. The recognition of some particular shapes for the mouth such as smiling mouth or wide open mouth is also an objective.

**References**


1. Liévin, M., Luthon, F.: Nonlinear Color Space and Spatiotemporal MRF for Hierarchical Segmentation of Face Features in Video. In : IEEE Transactions on Image Processing, Vol 13, No. 1, pp. 63 -- 71 (2004)
2. Cootes, T. F.: Statistical Models of Appearance for Computer Vision. Technical report, free to download on http://www.isbe.man.ac.uk/bim/refs.html, (2004)
3. Gacon, P., Coulon, P.-Y., Bailly G.: Non-Linear Active Model for Mouth Inner and Outer Contours Detection. In: 2005 European Signal Processing Conference (EUSIPCO'05), Antalya, Turkey (2005)
4. Yuen, P. C., Lai, J. H., Huang, Q. Y.: Mouth State Estimation in Mobile Computing Environment. In: Proc. Sixth International conference on Automatic Face and Gesture Recognition, Seoul, Korea (2004)
5. Benoit, A., Caplier, A.: Hypo-vigilance Analysis: Open or Closed Eye or Mouth? Blinking or Yawning Frequency ? In: IEEE AVSS, Como, Italy (2005)
6. Bouvier, C., Coulon, P-Y., Maldague, X.: Unsupervised Lips Segmentation Based On Roi Optimization and Parametric Model. In IEEE International Conference on Image Processing, San Antonio (2007)
7. Beaudot, W. H. A.: The neural information processing in the vertebrate retina: A melting pot of ideas for artificial vision. PhD Thesis in Computer Science, INPG (France) December 1994
8. Hérault, J., Durette, B.: Modeling Visual Perception for Image Processing. In: F. Sandoval et al. (Eds.): IWANN 2007, LNCS 4507, 662–675, Springer-Verlag Berlin Heidelberg (2007)
9. Guyader, N., Chauvin, A., Massot, C., Héraul, J., Marendaz, C.: A biological model of low-level vision suitable for image analysis and cognitive visual perception. In: Perception vol.35., ECVP (2006)
10. Webster, M. A., De Valois R. L.: Relationship between spatial-frequency and orientation tuning of striate-cortex cells. In Journal of Optical Society of America A, vol. 2, no. 7, pp. 1124--1132 (1985)
11. Ojala, T., Pietikäinen, M., Mäenpää, T.: Multiresolution Gray-Scale and Rotation Invariant Texture Classification with Local Binary Patterns. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol 24, No. 7, pp. 971 – 987 (2002)
12. Martinez, A. M., Benavente, R. : The AR Face Database. CVC Technical Report #24 (1998)