Hypovigilence Analysis: Open or Closed Eye or Mouth ?
Blinking or Yawning Frequency ?

A.Benoit, A.Caplier

LIS-INPG
46, avenue Felix Viallet
38031, Grenoble, France
benoit@lis.inpg.fr, caplier@lis.inpg.fr

ABSTRACT

This paper proposes a frequency method to estimate the state open or closed of eye and mouth and to detect associated motion events such as blinking and yawning. The context of that work is the detection of hypovigilence state of a user such as a driver, a pilot ... In [1] we proposed a method for motion detection and estimation which is based on the processing achieved by the human visual system. The motion analysis algorithm the filtering step occurring at the retina level and the analysis done at the visual cortex level. This method is used to estimate the motion of eye and mouth : blinking are related to fast vertical motion of the eyelid and yawning is related to large vertical mouth opening. The detection of the open or closed state of the feature is based on the analysis of the total energy of the image at the output of the retina filter: this energy is higher for open features. The absolute level of energy associated to a specific state being different from a person to another and for different illumination conditions, the energy level associated to each state open or closed is adaptive and is updated each time a motion event (blinking or yawning) is detected.

No constraint about motion is required. The system is working in real time and under all type of lighting conditions since the retina filtering is able to cope with illumination variations. This allows to estimate blinking and yawning frequencies which are clues of hypovigilance.

I. Introduction

The aim of the presented work is the development of a real time algorithm for hypovigilence analysis. The degree of vigilance of a user can be related to the state open or closed of his eyes and mouth and to the frequency of his blinkings and yawnings.

Work about eye blinks detection is generally based on temporal image derivative (for motion detection) followed by image binarization analysis [2]. Also, feature point tracking on eyes and mouth is used to detect open / closed state and motion [3]. All these methods are based on spatial analysis of the eye/mouth region, they are sensitive to image noise and generally require a sufficient number of pixels to be accurate. Moreover, these methods often require morphological operations to avoid false blink detections generated by global head motion. Other methods can be used such as one based on «second order change» [4] but they always need binarization and thresholding, the choice of the threshold being of critical influence on the results.

Work on mouth shape detection is generally based on lips segmentation: work with lip models such as [5] use color and edge information but these methods are sensitive to lighting and contrast conditions. Other methods such as parametric curves [6] has been studied. Recently, statistical model approaches such as active shape and appearance models for example [7, 8] have been proposed and give accurate results for lips segmentations. Nevertheless all these methods cannot give information on the mouth state. In the case of mouth motion detection, lips segmentation or feature point tracking [9] can be used but these methods require much processing power and yield to a mouth shape estimation rather than yawnings detection.

In this paper, we use the spectral analysis method described in [1] that will allow the detection of eye and mouth states and blink/yawning with the same method. It involves a spatio-temporal filter modelling the human retina and dedicated to the detection of motion stimulus. It is used to estimate the motion of eye and mouth: blinking are related to fast vertical motion of the eyelid and yawning is related to large vertical mouth opening. The detection of the open or closed state of the feature is based on the analysis of the total energy at the output of the retina filter : this energy is higher for open features. In section 2 the general principle of the motion estimation method is explained and the properties of the motion estimator are given (see [1] for more details). . . Section 3 describes the proposed method to detect eye and mouth motions events (blinks and yawnings) and section 4 describes how to detect the open or closed feature state which is associated to an adaptive updating of the related level of energy of the image spectrum. Section 5 presents some results.

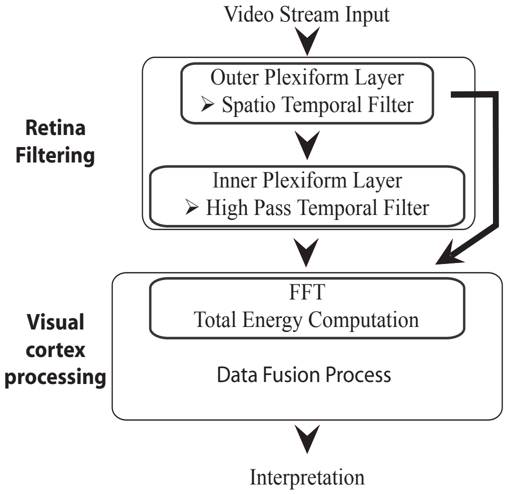


figure 1 : general algorithm

II. Motion estimator based on the human visual system

Figure 1 gives a general overview of the algorithm. It is made of two step : the retina filtering and the visual cortex processing.

II.1 Retina Filtering

The processing at the retina level consists in an efficient spatiotemporal filtering made of two stages [10] :

- at the Outer Plexiform Layer (OPL), all the treatments are modelled by a non separable spatio-temporal filter (see Figure 2), its transfer function is :

$$G_B(z,ft)=\frac{1}{1+\beta_c+\alpha_c[-z^{-1}+2-z]+j2\pi f_t\tau_c}.\frac{\beta_h+\alpha_h[-z^{-1}+2-z]+j2\pi f_t\tau_h}{1+\beta_h+\alpha_h[-z^{-1}+2-z]+j2\pi f_t\tau_h}$$

where

$$\alpha_i=\frac{r_i}{R_i},\beta_i=\frac{r_i}{r_{fi}},\tau_i=r_i.C_i$$

r_i, R_i are resistances and C_i are capacities that create the spatio-temporal effect. It models the filter generated by the synaptic network of the photoreceptors and horizontal cells of the retina.

This filter has a band pass spatial effect in low temporal frequencies which is responsible for contours enhancement. It has a wide band pass temporal effect for low spatial frequencies which smooths illumination variations. It has a low pass effect for high temporal frequencies and a low pass tendency for high spatial frequencies that minimizes spatio temporal noise.

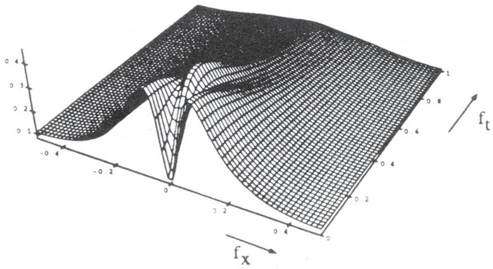
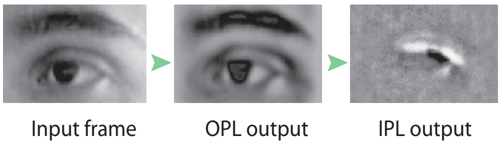


figure 2 : retina B transfer function [10]

- at the Inner Plexiform Level (IPL), the process is dedicated to the detection of moving stimulus. This process is modelled by a temporal derivation operator [11]. As a consequence, this filter enhances moving contours and removes static ones. The amplitude of the contours response at the output of the IPL depends on the contours orientation w.r.t. the motion direction (the optimal case are contours perpendicular to the motion direction) and it depends on the motion amplitude.

Figure 3 illustrates the effect of the retina filtering on an eye motion sequence in which the eye is closing. The OPL filter enhances all contours, attenuates the low spatial frequencies, and minimizes spatio temporal noise (note that the lateral illumination variation is cancelled). The IPL filter attenuates static contours and enhances only moving ones (especially the contour of the eyelid which is perpendicular to the motion direction). As a result, the spectrum of the IPL output only reports energy corresponding to the contours involved in the movement.

An other advantage of the retina filter compared with a cascade of classic band pass filters is that process can be achieved in real time [1].



On the Outputs, gray pixels correspond to null response, black and white to negative and positive responses

figure 3 : retina filtering

II.2 FFT in log polar domain and spectrum analysis

It was demonstrated in [1] that the FFT in the log polar domain of the IPL filter output allows an easy estimation of the motion direction. We use this property to extract vertical motions: the detection of motion direction is based on the analysis of the energy of the retina filtered image. We distinguish:

➢The global energy of the spectrum which is representative of the amplitude of the motion. In particular, this energy is minimum or null when no motion occurs.

➢The analysis of the same energy in the log-polar domain which exhibits maximum values for the contours perpendicular to the motion. Indeed the moving contours perpendicular to the motion direction are enhanced. This allows to estimate the motion direction. Eye blinking and mouth yawning are related to vertical motions. In order to extract such vertical motions, we compute a log polar spectrum composed of 15 orientations in order to detect motion orientation with 12° precision. Figure 4 presents an example of eye motion analysis: the same eye with two different motions is presented. In the first case the eye is closing (eyelid vertical motion) and in the second case the iris is rotating (i.e. focusing elsewhere). On figure 4, the cumulated energy per orientation curves are drawn. We can see that eye blinks (i.e. eye closing) report a maximum of energy at the 90° direction (vertical motion) which is not the case when the iris is horizontally moving.

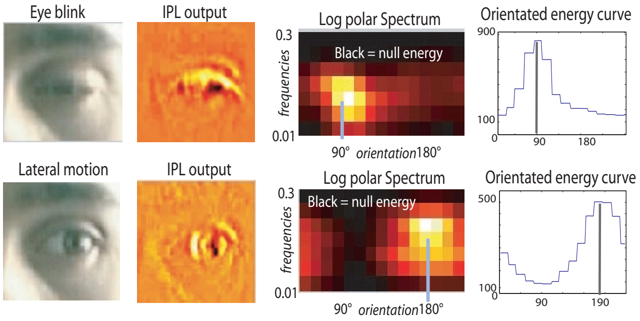


figure 4 : Log polar spectrum and oriented energy localisation in the case of an eye blink and eye lateral focusing motion

Note that the precision of the estimated orientation axis is influenced by the angle resolution of the log polar transformation and by the characteristics of the observed object. There is a higher precision if contours oriented perpendicular to the motion direction exist. This is the case with eye blink and mouth yawning.

III. Detection of blinking or yawning.

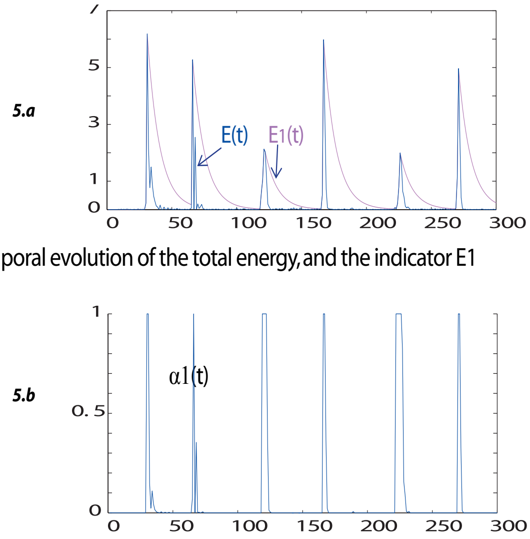
Here, we suppose that we are able to build a bounding box around each eye and around the mouth. The automatic extraction of such bounding boxes is under the scope of this paper. The detection of blinking and yawning is a difficult problem because the associated motions are non rigid and they can be of very different amplitude.

We focus on the total energy E(t), at time t, of the spectrum at the IPL output. Figure 5-a shows the temporal evolution of the total energy E(t) for a video sequence in which several eye blinkings are occurring. On that curve, each maximum is associated to a motion event and each minimum to a motion stop. But this example shows that the energy drops have not the same amplitude This represents a difficulty for the development of a motion detector able to detect all the eye blinks. We propose to use an adaptive motion detector.

III.1 Noise level estimation

Motion being related to the presence of energy at the IPL output, we have first to cope with residual noise, in order to avoid false motion detection. The spectrum noise level E_{noise} is computed at the beginning of the video sequence: it is the mean of the residual noise level of the *n* first frames (currently *n*=20) in which no motion occurs (the noise level estimation is computed for the region of interest only i.e. eye region or mouth region).

We consider that the current energy E(t) at the output of the IPL is related to a motion event of eye or mouth if $E(t)>3*E_{noise}$. This criteria is not restrictive because even for slow motions, related energies are more than 10 times the noise level even in noisy conditions. For example, in figure 5 each motion maximum reports an energy above 10 times the noise level (Enoise=0.05).



Temporal evolution of the total energy, and the indicator E1

Temporal evolution of the movement reliability marker α1

figure 5 : spectrum total energy temporal evolution during eye blinks

III.2 Motion level indicator

In order to detect motion events with precision, it is necessary to compare the current motion with the previous ones and the comparison must be adaptive to avoid false detection. Motion events are related to high level of energy at the output of the IPL but the absolute value of this high level is unknown and depends on lightning, contrast, contours sharpness... Since it is not possible to defined a unique threshold on the energy value, we propose to define a motion reliability marker called $\alpha_1(t)$.

First, an adaptive indicator $E_1(t)$ is introduced, it can be seen as the output of an electric analog/continuous current converter applied to the total energy time evolution. $E_1(t)$ reaches each maximum energy value and decreases temporally with an 1/t curve tendency (capacity effect). Figure 5-a illustrates this effect. When a maximum of energy is present, the indicator $E_1(t)$ reaches this maximum and decreases slowly as an electric analog/continuous current converter with low pass filter does.

This indicator $E_1(t)$ is used to estimate the reliability of each current energy level (associated to the current motion event) w.r.t. the last energy levels (associated to the previous motion events). Currently, the temporal integration is done on 0.5 second before the current event. We define the motion level indicator as :

$$\alpha_1(t) = E(t)^2/E_1(t)^2$$

$\alpha_1(t)=1$ when the current energy is high compared to the last energy values and $\alpha_1(t)=0$ when the current energy level is lower than the last energy values i.e. the last motions amplitude. $\alpha_1(t)$ indicates the reliability level of the amplitude of the current motion compared to the very last motion events. On figure 5-b, the graph shows the temporal evolution of the motion reliability marker α_1. It is minimum when no motion occurs, maximum when motion increases and decreases when motion slows down. The main advantage of this motion level indicator is that $\alpha 1(t)$ values are only in the range [0; 1] so that thresholding is easy. A threshold level of 0.2 allows to detect all significant motion events associated to low or fast motions. The risk of false detections introduced by the noise level is minimized while only considering the values of the total energy spectrum which are above 3*Enoise.

This motion reliability marker is well suited for eye blinks and mouth yawnings detection. Indeed, when such motion events are occurring, the indicator $\alpha 1(t)$ is above the threshold 0.2. Tests shows that with this marker, it is possible to detect more than 98% of such movements even in poor lighting conditions. False detections are only caused by the residual spatiotemporal noise. This occurs when the lightning is so low that contours extraction is impossible. Then, very fast motions such as eye blinks are detected even if the video acquisition system is too slow in regard of the eyelid motion (18 frames per second are sufficient). This detection could not be acheived with common optical flow algorithm [12] because of their hypothesis of low motion between frames.

IV. Eye and mouth state detection

IV.1 Closed or open state: High or Low output OPL filter energy

The goal is to detect the open or closed state of eyes or mouth. The states to be estimated can be considered as binary states. Taking advantage of the prefiltering, we propose an algorithm which is focusing on the total spectrum energy at the OPL filter output: the spectrum of the OPL output reports a level of energy which is proportional to the ”quantity” of contours present in the scene. As a consequence, the spectrum energy in case of an open eye is always higher than those obtained in case of a closed eye. This is also the case for an open mouth. Figure 6 presents an example of the output of the OPL filter with the associated level of energy for each different case. In the remaining of the paper, the energy level corresponding to an open feature is called HighEnergyLevel and the energy level corresponding to a closed feature is called LowEnergyLevel.

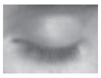
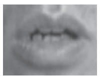
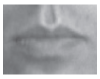
	OPL Output	Total Energy	Binary state
Eye open		4.7	High energy level
Eye closed		2.8	Low energy level
Mouth opened		1.9	High energy level
Mouth closed		1.1	Low energy level

figure 6 : facial features states and their related energy levels

Note that intermediate positions such as eyes half open, mouth nearly open ... present an intermediate energy level.

Closed feature is always related to a very low energy. As far as open feature is concerned, the associated energy is higher but the absolute value depends on the opening degree of the feature, and, the value of the energy for each state is different from one person to another. As a consequence, it is necessary to learn and to temporally update the energy value associated to each state. These updates can be done at each eye blinking or mouth yawning i.e. when motion events occurs.

The advantage of such a method is that we focus on energy quantity rather than facial features states (i.e. position of the eyelip for example). Then, if the bounding boxe around the eye or mouth area is not correctly adjusted, detection errors will be minimum even if the region of interest (eye/mouth area) is hiden up to 50%. Indeed, the detected part of the eye or mouth will give the same energy evolution tendancy as the full detected feature in the case of optimal bounding box adjustment. This require an adaptative algorithm able to cope with this kind of detection error cases.

IV.2 Adaptive computation and updating of HighEnergyLevel and LowEnergyLevel

Our idea is to combine the updating of the different levels of energy associated to each open or closed state with the detection of the motion events (blinking or yawning) occurring on the considered facial feature. When such an event is occurring, it is associated to a change in the state of the facial feature. For example, an eye blink is related to the successive transitions open/closed/open. When a movement event occurs, the OPL energy level in the frame before the motion event is computed and compared to the OPL energy level in the frame occurring just after the motion event. Both reference energy levels (HighEnergyLevel and LowEnergyLevel) are then updated.

In order to have a robust updating of both energy levels, the updating process is triggered when the event indicator $\alpha 1(t)$ is close to 1 and when the direction of the detected motion event is vertical: blinking and yawning both involve vertical motion. So we can cope with iris motion or mouth motion occurring when someone is speaking. The data fusion algorithm works as follow :

```
% HighEnergyLevel and LowEnergyLevel updating method
If a motion event associated to vertical direction is detected (α1(t)>0.2 and vertical motion)
Then    compute HighEnergyLevel and LowEnergyLevel
end
% Face feature state detection method
EnergyMean = (HighEnergyLevel + LowEnergyLevel)/2
if E(t) > EnergyMean
then    feature state = 'open'
else    feature state = 'closed';
end
```

V. Results

V.1 Eye state and blinking detection

The presented algorithm is applied to a video sequence of eye blinks.

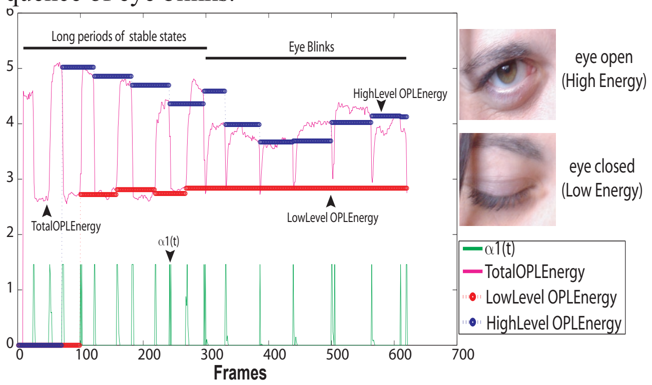


figure 7 : Temporal evolution of the OPL energy on an eye blink sequence

On figure 7, motions blinks are detected: they are related to $\alpha 1(t)>0.2$ with vertical motion direction and they corresponds to eye openings or closings. A short initialization step is required in order to adjust the HighEnergyLevel and LowEnergyLevel to the acquisition conditions. This is done from the first frame to frame 100 in which the eye stays successively in open and closed state more than 0.5 second. During this period, the algorithm computes and updates both levels with the energy of the states encountered before and after a blinking event. During non voluntary eye blinks period (from frame 300 to frame 600) the LowEnergyLevel is not updated because of the shortness of the blinks. The algorithm uses the previous low energy level computed during long period closed eye, so that detection remains reliable. This updating method can cope with very short blinks (shorter than the acquisition camera frame rate) and can avoid a non suitable updating of the LowEnergyLevel. Both energy levels are correctly computed and since we consider the eye as closed if the current OPL energy E(t) is lower than EnergyMean, the algorithm achieves 100% success of eye state/blinks detection from frame 100 to the end of the sequence.

The blink frequency is linked to the frequency of the maximum values of $\alpha_1(t)$. An eye is closed each time the current energy is close to the LowEnergyLevel so that the duration of an eye closing can be evaluated. Blinking frequency and eye closing duration are information about the state of human vigilance.

Several tests show that the LowEnergyLevel value does not change more than 10% from frame to frame. On the contrary, the HighEnergyLevel value evolves as the eye is open differently (from fully open up to frame 300 and less open after frame 300 in the case of the sequence of figure 7). This confirms the necessity of an adaptive algorithm.

Note that other motions such as translation and small rotations of eye due to focusing direction changes are not disturbing because these perturbations are related to motion which are smaller than the blinking motion and the energy level at the OPL output is proportional to the motion amplitude. Contours with small motion do not create sufficient energy changes compared to eye blinking motion. Moreover, focusing direction changes only translate the pupilly, but the quantity of contours does not change as much as in the case of eye blinks, and the motion direction of the pupilly should be vertical to generate false detection.

V.2 Mouth state and yawning detection

The same method is applied to mouth yawning detection. Figure 8 shows the results on a sequence in which the mouth exaggerates its open and closed state from frame 1 to frame 300, is closed from frame 301 to frame 500, and opens / closes normally after frame 500 because of natural speech.

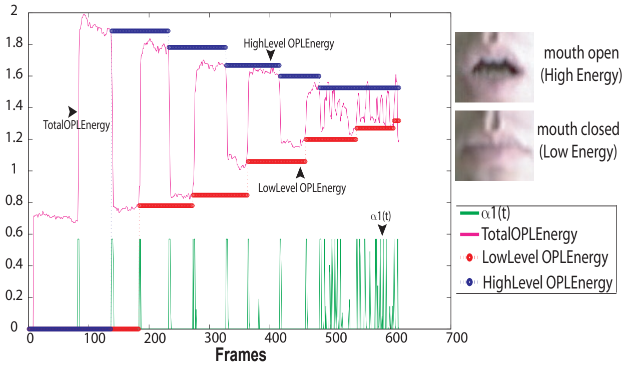


figure 8 : Temporal evolution of the OPL energy on a mouth motion sequence

We can see that the algorithm self adjusts its parameters HighEnergyLevel and LowEnergyLevel before frame 200, this is the initialization period where each mouth state is performed more than 0.5 second by the user in order to correctly initialize these parameters. Then the algorithm updates them with respect to the evolution of the OPL output spectral energy. The LowEnergyLevel corresponds to closed mouth because the closed lips generate a lower quantity of contours. The HighEnergyLevel corresponds to open mouth which let appear tooth and/or internal mouth details or a black area that generate high energy contours with the lips frontier.

Note that during the stable open/closed mouth periods, the HighEnergyLevel and LowEnergyLevel values are adjusted and when the speech periods happen (from frame 500 to the end), these levels are no more or few updated. This allows the correct detection of the mouth state even in case of fast mouth shape variation that occurs during speaking.

VI. PERFORMANCES AND APPLICATION

The performances of this facial feature state and motion event detector have been evaluated in various test condition : it detects states and movements events up to 99% success in standard office lighting conditions with the focused object occupying from 60% to 100% of the captured frame (currently 100*100 pixels). In low light conditions or noisy captured frames (Gaussian white noise of variance 0.04), the algorithm is able to detect the motion events and states with 80% success. Moreover, even if the algorithm is 'lost' at a moment, since it is adaptive, it automatically corrects its energy levels and works fine when the sequence returns to normal conditions.

The algorithm works in real time, reaching up to 80 frames per second on a standard PC desktop Pentium 4 running at 3.0Ghz on which a webcam is installed. The algorithm automatically adjusts its parameters during the analysis. This proposed approach is inspired from the capacities of the human visual system which is adaptive and is able to cope with various illumination and motion conditions.

VII. Conclusion

A real time method for facial feature state and motion events detection has been proposed, it works with eye and mouth in the same way. The algorithm inspired from the biological model of the human visual system shows its efficiency in terms of motion detection and analysis : the use of the retina filter prepares the data and yields to a spectrum easy to analyze.

The proposed algorithm proves its efficiency to estimate the open or closed state of eye and mouth and the frequency of blinking and yawning. This is well suited for the analysis of a user vigilance. The performances of the algorithm on video sequences of a car driver are under study.

VIII. References

[1] A. Benoit, A. Caplier. "Motion estimator inspired from biological model for head motion interpretation " WIAMIS 2005, Montreux, Switzerland, April 2005

[2] J. Coutaz, F. Berard, and J. L. Crowley. "Coordination of perceptual processes for computer mediated communication". In Proc. of 2nd Intl Conf. Automatic Face and Gesture Rec., pages 106--111, 1996.

[3] P. Smith, M. Shah, N. da Vitoria Lobo, "Determining Driver Visual Attention with One Camera", Accepted for IEEE Transactions on Intelligent Transportation Systems, 2004.

[4] D. Gorodnichy, "Towards Automatic Retrieval of Blink-Based Lexicon for Persons Suffered from Brain-Stem Injury Using Video Cameras," Proceedings of the First IEEE Computer Vision and Pattern Recognition (CVPR) Workshop on Face Processing in Video. Washington, District of Columbia, USA. June 28, 2004. NRC 47138.

[5] P. Delmas, N.Eveno, and M. Lievin, "Towards Robust Lip Tracking", International Conference on Pattern Recognition (ICPR'02),Québec City, Canada, August 2002

[6] N.Eveno, A. Caplier, and P-Y Coulon, "Jumping Snakes and Parametric Model for Lip Segmentation", International Conference on Image Processing, Barcelona, Spain, September 2003

[7] T. F. Cootes. "Statistical models of appearance for computer vision", Online technical report available from http://www.isbe.man.ac.uk/bim/refs.html, 2001.

[8] P. Gacon, P.-Y. Coulon, G. Bailly. "Statistical Active Model for Mouth Components Segmentation", 2005 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'05), Philadelphia, USA, 2005.

[9] Y. Tian, T. Kanade, J.F. Cohn "Robust Lip Tracking by Combining Shape, Color and Motion" Proc. of the 4th Asian Conference on Computer Vision (ACCV'00), January, 2000

[10] W.H.A. Beaudot, "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

[11] J. Ritcher&S.Ullman. "A model for temporal organization of X- and Y-type receptive fields in the primate retina". *Biological Cybernetics*, 43:127-145,1982.

[12] Barron J.L., Fleet D.J. and Beauchemin S.S., "Performance of Optical Flow Techniques", *International Journal of Computer Vision,* Vol. 12, No. 1, pp. 43-77, 1994.