Computer Vision and Image Understanding 114 (2010) 758–773

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Using Human Visual System modeling for bio-inspired low level image processing

A. Benoit [a,*], A. Caplier [b], B. Durette [b], J. Herault [b]

[a] LISTIC – Polytech'Savoie, B.P. 80439, 74944 Annecy le Vieux Cedex, France
[b] Gipsa-lab, 961 rue de la Houille Blanche, Domaine Universitaire – B.P. 46, F – 38402 Saint Martin d'Hères Cedex, France

ARTICLE INFO

Article history:
Received 4 December 2008
Accepted 31 January 2010
Available online 4 March 2010

Keywords:
Retina model
V1 cortex model
Low level image processing
Real-time processing
Contour analysis
Motion analysis
Preprocessing
Texture and structure enhancement
Local adaptation

ABSTRACT

An efficient modeling of the processing occurring at retina level and in the V1 visual cortex has been proposed in [1,2]. The aim of the paper is to show the advantages of using such a modeling in order to develop efficient and fast bio-inspired modules for low level image processing.

At the retina level, a spatio-temporal filtering ensures accurate structuring of video data (noise and illumination variation removal, static and dynamic contour enhancement). In the V1 cortex, a frequency and orientation based analysis is performed.

The combined use of retina and V1 cortex modeling allows the development of low level image processing modules for contour enhancement, for moving contour extraction, for motion analysis and for motion event detection. Each module is described and its performances are evaluated.

The retina model has been integrated into a real-time C/C++ optimized program which is also presented in this paper with the derived computer vision tools.

1. Introduction

In this paper, we propose an image processing approach belonging to what we call "biological vision based approach". The basic idea is to copy the Human Visual System (HVS) by modeling some of its parts in order to develop low level image processing modules. Up to now, the most well-known parts of our visual system are the retina and the V1 cortex area which are the two parts on which we focus our work. The retina can be considered as a preprocessing step which conditions the visual data for facilitated high level analysis. The V1 cortex can be considered as a low level visual information describer. From these two "tools", we want to show how to design efficient low level image processing tools.

Biologically inspired methods for image processing are numerous and we choose to focus only on bio-inspired models dedicated to image processing in order to make the paper easier to read. For example, the Retinex filter proposed in [3,4] is a method that enhances a digital image in terms of dynamic range compression, color independence from the spectral distribution of the scene illumination, and color/lightness rendering as it is done in the retina and in the cortex. This algorithm is based on luminance analysis and its enhancement. It assumes that color perception is related to ratios of reflected light intensity in specific wavelength bands computed between adjacent areas. As a consequence, this algorithm is dedicated to color applications. Other models of the HVS are used, for example, for information coding [5]. These methods generally use high level information processing such as visual cortex modeling but do not take into account the low level processing that occurs at retina level.

Since our goal is to demonstrate the interest of using retina and V1 cortex modeling in order to proceed to low level image processing, the preliminary step of our work, which is to choose the most appropriate retina and cortex models, is described in the following. As discussed in [51], the definition of standard models is a rich research field and some approaches allow dedicated image processing implementations to be expected. As far as retina models are concerned, some have already been proposed with different degrees of precision. Mead and Mahowold [6] was a precursor for the modeling of the neurophysiological properties of vertebrate's retinas by considering analogies with electronic circuits. His model focuses on the link between the retinal architecture and its functionalities. Nevertheless, his work insists more on the spatial filtering properties of the retina than on temporal effects related to motion analysis. The modeling of the biological retina was also studied by Franceschini et al. [7] who worked on the retina architecture of the fly. He built robots working on the same model and showed their properties for target tracking or for flying in unstable wind conditions and for collision prevention. Spike based models

* Corresponding author at: Fax: +33 450 09 65 59.
E-mail addresses: alexandre.benoit@univ-savoie.fr (A. Benoit), alice.caplier@gipsa-lab.grenoble-inp.fr (A. Caplier), barthelemy.durette@gipsa-lab.grenoble-inp.fr (B. Durette), jeanny.herault@gipsa-lab.grenoble-inp.fr (J. Herault).
URLs: http://www.listic.univ-savoie.fr (A. Benoit), http://www.gipsa-lab.inpg.fr (B. Durette).

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.01.011

were also studied; an advanced model was presented with the SpikeNet toolbox [8]. It models the electrical impulse spikes exchanged by the neural cells at the retina ganglion cells and V1 cortex levels. It already demonstrates high speed computing properties for high level image analysis, but low level retina processing are not completely described. Other approaches are developed such as digital retinas. Some of them are methods dedicated to VLSI (very-large-scale integration) implementations [9,10]. These algorithms are efficient parallel methods generating binary or floating point output pictures but the models contain only parts of all the processes carried out in the retina.

The starting point of our work is an accurate model of the human retina. This model presents a global approach of the retina processing inspired from an analogy between electronic circuits and signal processing strategies of the biologic retina. It is based on Mead's work and has been improved in terms of spatial and temporal properties by Herault and Beaudot [1,2,11]. It describes the different computing carried out in the first cell layers of the retina (Outer and Inner Plexiform Layers). This model allows fine perception modeling. This emphasizes the different cell network properties of the retina and its implementation enables fast computing thanks to natural parallel processing properties.

Considering V1 cortex, several studies led to the creation of various models. Marcelja [12] showed that the cortical cells in the V1 cortex are sensitive to orientations and can be modeled with 1D Gabor filters. This work was extended to 2D by Daugman [13]. This modeling leads to a simple representation of scene information in the spectral domain. In this way, 2D Gabor filters are generally used in literature for texture classification [14], or saliency area research to extract relevant features in a scene [15]. Because of their properties in log scale (reliable zoom effects handling), we propose to use the modeling of the V1 cortex area described in [16] which uses log polar Gabor filters (GloP) instead of Gabor filters.

With the choice of Herault's model for retina modeling and the choice of Guyader's model for the V1 cortex modeling, we obtain a model for the parts of the visual system we are interested in. In order to situate the chosen global HVS model with regard to well known visual system models, we present the main orientations of these works and ours. The Itti and Koch model [17] focuses on the analysis of the visual scene in terms of scale and orientation description. This model exhibits the high level analysis achieved at the visual cortex level in order to compute saliency maps for visual attention modeling. These bottom-up orientation and scale description are indeed specific features of the V1 cortex area which we also propose to perform with the help of Guyader's model [16]. The work of Walter [18] also insists on the processing carried out at the cortex level and adds top-down interactions for visual attention simulation. Nevertheless, at the retina level, low level processing is not fully considered. Similar approaches have been proposed, for example by Daly's [19], the Irccyn Lab's model [20] and Gipsa Lab's model [63]. These models are suited for image and video quality evaluation and saliency area extractions. These accurate models insist more on high level cortex processing (even above V1 area) dedicated to image description than on the properties of low level processing done at the retina level. The Contrast Sensitivity Function (CSF) they use does not include some specific features of the retina such as local adaptation and temporal filtering. In comparison, our approach focuses more on the first low level retina processing and the V1 cortex in the aim of demonstrating the interest of the low level retina filtering properties. Future work will consist in fusing our model with aforementioned approaches in order to reach a higher step of complexity with a more accurate low level processing precision and to describe a wider area of image processing applications.

In order to show the potential of such human visual models for efficient low level image computing, we present, in this paper, a set of real-time image processing modules. A first set, based on a retina model, allows detail and motion information extraction. A second set based on the V1 cortex area and a motion event detector enable to describe the visual scene at a higher semantic level. Keep in mind that the motion analysis which will be exposed can be considered as being close to optical flow computation. Nevertheless, we insist more on the preprocessing aspect of the retina for motion energy extraction, its noise reduction and local motion information enhancement. Even if information about motion energy is offered, the extraction of the optical flow is the next step, as proposed in [21].

The paper is presented as follows: Section 2 gives a short description of the model proposed in [1,2] for retina and V1 cortex processes. Section 3 describes the four low level processing modules which we developed for contour enhancements, moving contour extraction, image orientation analysis and context aware motion event detection. Section 4 describes the developed real-time retina program which has been made available publicly and which justifies our global approach, consisting in using HVS modeling in order to build up efficient image processing algorithms. This section also summarizes the previous image processing algorithm which has already been achieved with the help of the presented models.

2. Human Visual System modeling

Fig. 1 gives a general overview of the parts of the HVS which are considered here and which have been modeled in [1,2,16]. In the retina, the spatial and temporal properties of the different cells layers are considered, from photoreceptors and the connected cell layers of the so called Outer Plexiform Layer (OPL) followed by the Inner Plexiform Layer (IPL). These processing steps are described in Section 2.1. The proposed model allows two information channels to be modeled. The former, Parvo, being related to details extraction while the latter, Magno, is dedicated to motion analysis. In the V1 cortex area (cf. Section 2.2), a frequency and orientation analysis in the log polar domain is carried out [16]. It is intended to process the information given by the retina model in order to perform low level visual scene properties description.

2.1. Retina modeling

Fig. 2 describes the different retina cells: photoreceptors, horizontal cells, bipolar cells, ganglion cells and amacrine cells. Photoreceptors are responsible for visual data acquisition and are also associated with a local logarithmic compression of the image lumi-

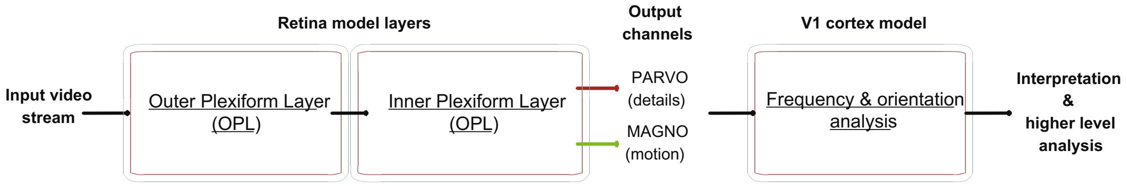

Retina model layers
Output channels
V1 cortex model
Input video stream
Outer Plexiform Layer (OPL)
Inner Plexiform Layer (OPL)
PARVO (details)
MAGNO (motion)
Frequency & orientation analysis
Interpretation & higher level analysis

Fig. 1. General algorithm.

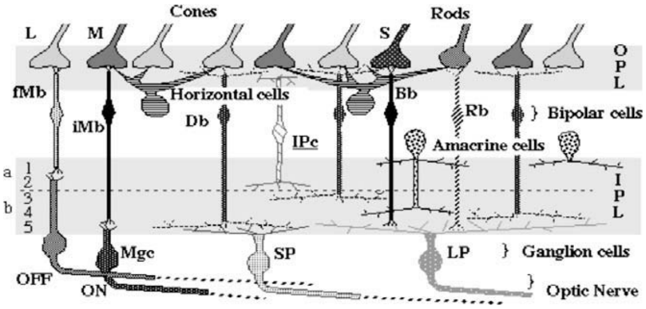


Fig. 2. Biologic architecture of the retina [2].

nance. The retina cells are connected to each other in order to form two cell layers: the Outer Plexiform Layer (OPL) and the Inner Plexiform Layer (IPL). Each layer is modeled with specific filters. Finally, at the IPL level which constitutes the retina output, different channels of information can be identified. We focus here on the most well known: the Parvocellular channel (Parvo) dedicated to detail extraction and the Magnocellular channel (Magno) dedicated to motion information extraction. Note that we consider here the parvocellular and Magnocellular processing on the whole visual scene. In the human retina, the parvo cellular channel is most present at the fovea level (central vision) and the Magnocellular channel is most important outside of the fovea (peripheral vision), because of the relative variations of specialized cells [25]. Considering both pieces of information in the same area of the image can be interesting for computer vision because detail and motion data become available as parallel information on the same area.

2.1.1. Photoreceptors and illumination variation removal

2.1.1.1. Model. Photoreceptors have the ability to adjust their sensitivity with respect to the luminance of their neighborhood [1,2]. This is modeled by the Michaelis–Menten [1,50] relation which is normalized for a luminance range of [0, V_{max}]

$$C(p) = \frac{R(p)}{R(p) + R_0(p)} \cdot V_{max} + R_0(p)) \tag{1}$$

$$R_0(p) = V_0 \cdot L(p) + V_{max}(1 - V_0) \tag{2}$$

In this relation, the adjusted luminance $C(p)$ of the photoreceptor p depends on the current luminance $R(p)$ and on the compression parameter R_0 (p) which is linearly linked to the local luminance $L(p)$ (cf. Eq. (2)) of the neighborhood of the photoreceptor p. This local luminance (p) is computed by applying a spatial low pass filter to the input image. This low pass filtering is actually achieved by the next cellular network: the horizontal cell which is presented in Section 2.1.2.1.

As shown in Eq. (2), R_0 (p) depends on the local luminance $L(p)$. Moreover, in order to increase flexibility and make the system more accurate, we add the contribution of a static compression parameter V_0 of value range [0; 1] which allows the local adaptation effect to be adjusted in order to enhance ease of use and precision. Its value is experimentally set to 0.90. A lower value reduces the local adaptation effect. This parameter allows a new degree of freedom for computer vision applications. The value range can be adjusted between 0.60 and 0.99 for optimal results with eight or more bits per pixel pictures. Note that V_{max} represents the maximum allowed pixel value in the image. In the case of standard eight bits images, its value is 255 but it can be very different in case of different coding (such as High Dynamic Range (HDR) images, such as openEXR images [37]). In this case, V_{max} should be set as the maximum pixel value of the image or the set of processed images.

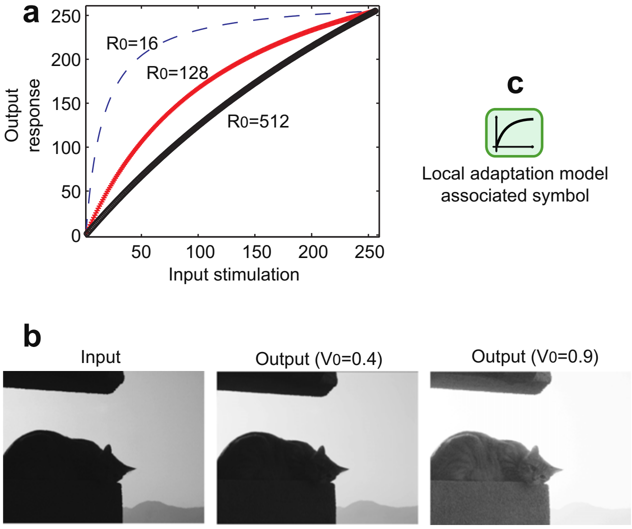


Fig. 3. Effect of the photoreceptors compression.

Fig. 3a shows the evolution of sensitivity with respect to parameter R_0 (p). Sensitivity is reinforced for low values of R_0 (p) and is kept linear for high values. As a result, this model enhances contrast visibility in dark areas while maintaining it in bright areas.

2.1.1.2. Properties. Fig. 3b illustrates the effect of such a compression on a back-lit picture with two different compression parameters V_0. The low-lit areas become brighter and more contrasted so that details appear. On the other hand, bright areas are not significantly modified. Note that in this example, a lower value of V_0 allows the high luminance value saturation to be limited.

The symbol depicted in Fig. 3c is associated with the logarithmic compression effect of the photoreceptors.

2.1.2. OPL: spatio-temporal filtering and contour enhancement

2.1.2.1. Model. The cellular interactions of the OPL layer can be modeled with a nonseparable spatio-temporal filter [1,2] whose transfer function for 1D signal is defined in Eq. (3) where the variable f_s is the spatial frequency and f_t is the temporal frequency.

$$F_{OPL}(f_s, f_t) = F_{ph}(f_s, f_t) \cdot [1 - F_h(f_s, f_t)]$$

where

$$F_{ph}(f_s, f_t) \frac{1}{1 + \beta_{ph} + 2\alpha_{ph} \cdot (1 - \cos(2\pi f_s)) + j2\pi\tau_{ph} f_t} \tag{3}$$

$$F_h(f_s, f_t) = \frac{1}{1 + \beta_h + 2\alpha_h \cdot (1 - \cos(2\pi f_s)) + j2\pi\tau_p f_t}$$

This filter can be considered as a difference between two low-pass spatio-temporal filters which model the photoreceptor network ph and the horizontal cell network h of the retina. The linked bipolar cells perform the final subtraction. As shown in [1], the output of the horizontal cell network (F_h) contains only the very low spatial frequency of the image. It is therefore used as the local luminance $L(p)$ which feeds back the luminance adaptation stage described in Section 2.1. Fig. 4a presents the global OPL scheme. In that figure, we represent the difference between F_{ph} and F_h by two operators BipON and BipOFF, respectively giving the positive and negative parts of the difference between the P_h and h images. This models the action of the bipolar cells which divide the OPL outputs in two channels, ON and OFF. As these outputs are complementary, they can be combined in order to visualize the global transfer function of the OPL (F_{OPL}) shown in Fig. 4b and corresponding to Eq. (3). This filter has a spatial band-pass effect in low temporal

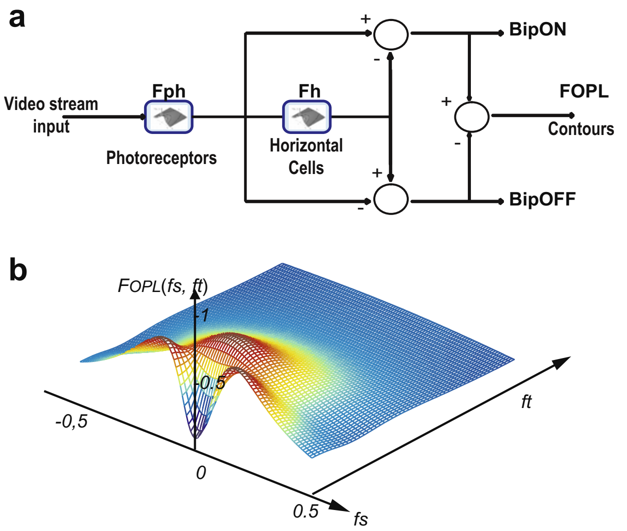


Fig. 4. OPL retina modeling and transfer function.

frequencies, a wide temporal band-pass effect for low spatial frequencies, a low-pass effect for high temporal frequencies and a low-pass effect for high spatial frequencies.

The F_{OPL} spatio-temporal filter involves several parameters: β_{ph} is gain of the filter F_{ph}, this parameter is generally set to 0 but it should be increased if the input dynamic range has to be increased. $h\beta_h$ is the gain of filter F_h. By setting this parameter to 0, only the contour information is extracted. Tweaking this parameter allows to adjust the gain at the null frequency, thus modifying the mean luminance of the image. τ_{ph} and τ_h are temporal filtering constants allowing the temporal noise to be minimized. α_{ph} and α_h are spatial filtering constants setting the spatial filtering capabilities: α_{ph} sets the high cut frequency and α_h sets the low cut frequency. An example of parameter setting is discussed in Section 4.

2.1.2.2. Properties. The OPL filter can remove spatio-temporal noise and enhance contours. These two properties are complementary because noise generates disturbing contours so that enhancing contours is often linked to noise enhancement. Fig. 5 illustrates the effect of the OPL filter. The two sequences were acquired with a commercial webcam (low quality of the CCD sensors and compression effect) in a situation of standard lighting conditions (see Fig. 5a) and in case of very low lighting conditions (small signal to noise ratio) (see Fig. 5c). In both cases, contours were enhanced and the signal to noise ratio has increased after OPL filtering. In addition, one of the most relevant effects is spectral whitening which compensates for the 1/*f* spectrum tendency of natural images [49] as shown in Fig. 5b. This acts as a decorrelation of the image input. Models and physiological reports have previously emphasized this decorrelation of the visual information processing in spatial and temporal domains. Some of the most important papers include [22–24].

As a result, the high spatial frequency contours are enhanced and the null frequency is lowered (i.e. the mean luminance is canceled here with $\beta_h = 0$). The structure and texture of the scene are then easily extracted.

2.1.3. IPL and Parvo channel: contours enhancement

2.1.3.1. Model. The ganglion cells of the Parvo channel (these ganglion cells are called "midget") receive the contour information coming from the BipON and BipOFF outputs of the OPL. On this information, they act as a local enhancer CgP which reinforces the contour data. This is modeled by a Michaelis–Menten law similar to the photoreceptors [26] (cf. Fig. 6a).

2.1.3.2. Properties. As the incoming information is about contours, the result is the enhancement of the contour contrast (cf. Fig. 6b). Here, the local adaptation law is exactly the same as that of photoreceptors. Nevertheless, the incoming data is different here. Indeed, at this point, only contour information is available with a reduced amount of luminance if $\beta_h > 0$ up to no luminance if $\beta_h = 0$ at the previous filtering step. As a consequence, the con-

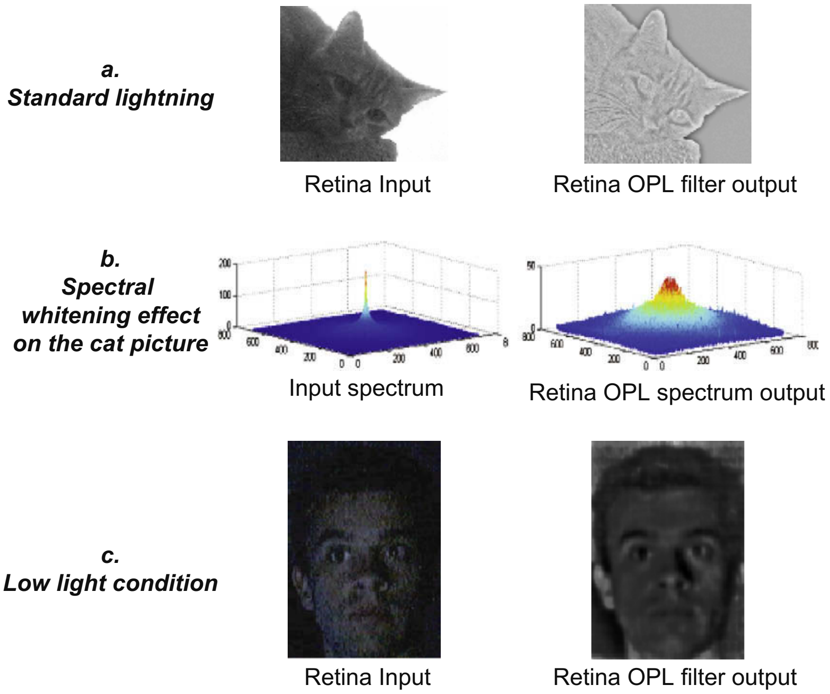


Fig. 5. Effect of the OPL filter (a) in standard lighting conditions and (b) in low light/noise conditions.

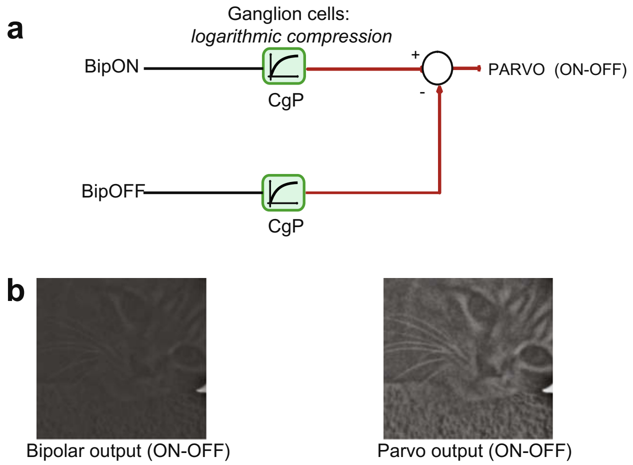


Fig. 6. (a) IPL modeling and (b) contours enhancement at the parvo channel of the IPL retina stage.

tour enhancement is less dependent on local luminance and depends more on contours. Moreover, as the information is divided into two channels (ON and OFF), each channel is enhanced independently in its own context and leads to image local contrast equalization.

2.1.3.3. Global Parvo channel properties. We propose to illustrate Parvo filtering including photoreceptors, OPL and IPL Parvo models in the picture presented in Fig. 7a. In this picture, small black arrows are set out on a background which varies linearly from deep white to deep black. The retina Parvo filtering applied to this image (cf. Fig. 7b) with parameter $\beta_h = 0$, allows all arrows to be extracted even those hidden in the black background while canceling the luminance information. In a second step, the picture was corrupted (cf. Fig. 7c) by high frequency spatio-temporal Gaussian noise ($\mu = 0$, $\sigma = 0.01$). Fig. 7d presents the effect of the retina Parvo filter on this noisy picture. Noise has reduced because of the low-pass temporal frequency effect and the band-pass effect for spatial frequencies (i.e. at null temporal frequencies). The global result is simultaneous noise reduction and contour enhancement.

A signal to noise analysis gives the results of Table 1, notation MSE meaning Mean Square Error. The signal to noise ratio is enhanced with the retina Parvo model: contours are preserved while noise and mean luminance are removed. Indeed, the difference between the image and its noisy counterpart after retina Parvo filtering is lower than the difference between the input image and the input noisy image, the gain being around 3.1 dB SNR. So the contour response is enhanced even if the input is subject to noise and background luminance variations. This also explains why computing the MSE and SNR between Parvo filter input and output would have been meaningless because the filter cancels mean luminance leading to a totally different image. As a comparison, the SNR obtained with the well-known Cany-Derich algorithm is 2.8 dB which is slightly lower than the retina processing because of a lower sensitivity in dark areas.

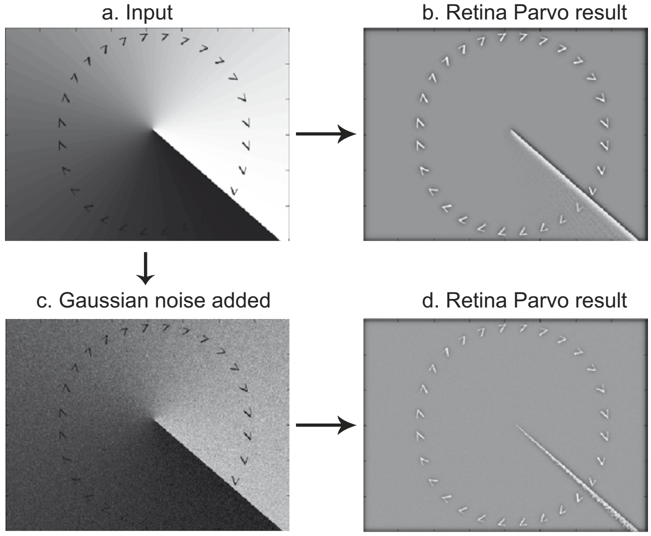


Fig. 7. Detail analysis with the retina Parvo filtering ($\beta_h = 0$).

Table 1
Noise filtering with retina Parvo filter.

Image comparison	MSE	SNR (dB)
Original input v.s. noise input	2e+004	1.3

This Parvo filter is even more interesting when working with High Dynamic Range images, i.e. natural images. This kind of image contains a much wider range of luminance values since the image coding is higher than the standard eight bits per color channel format; it is typically 32 bits. This allows real life like image storage. The captured images can contain very bright and very dark areas with a constant accuracy but all the luminance range cannot be seen on standard media such as standard displays or printers. Similarly, as discussed in [2], natural High Dynamic Range (HDR) images captured by the human eye cannot be coded by neurons but local luminance adaptation acts as a data compression tool and keeps all relevant information. Then, processing such images exhibits the real potential of the model. Fig. 8 shows a HDR image (“Memorial”) shown at different exposures in order to make all the details viewable on the paper. Then, filtering this High Dynamic Range image with the retina Parvo channel allows all the structural information in all the areas of the scene to be extracted whatever the luminance variations there are. Two goals can be reached: details extraction and luminance compression. Fig. 8b illustrates the first idea: the Parvo filter with a null horizontal cellular gain value (parameter $\beta_h = 0$) eliminates the mean luminance and extracts the structure and texture of the image. The second idea is shown in Fig. 8c where the Parvo filter, using a positive value of the gain (parameter $\beta_h = 1$), compresses the luminance range while preserving the global visual scene ambiance. In this example, because of the nonnull β_h value, the low frequencies are drastically attenuated but not canceled. Then, the local mean luminance variation amplitudes are reduced. As a consequence, the visual scene can be rendered on a lower dynamic range media such as the printed paper while preserving all the details of the scene in all its areas. This processing is referenced as a ‘tone mapping’ which is a topic that is discussed in a more dedicated paper [48].

2.1.4. IPL and Magno channel: motion dedicated filtering

2.1.4.1. Model. On the Magnocellular channel of the IPL, amacrine cells act as high pass temporal filters [27]. We model this effect by a first order filter

$$A(z) = b \cdot \frac{1 - z^{-1}}{1 - b \cdot z^{-1}} \quad \text{with } b = e^{-\Delta t/\tau_A} \qquad (4)$$

where $\Delta t = 1$ is the discrete time step, and τ_A is the time constant of the filter. This filter enhances areas where changes occur in space and time. Fig. 9a shows the IPL Magno model: the amacrine cells (A) are connected to the bipolar cells (BipON and BipOFF) and to the “parasol” ganglion cells. As on the Parvo channel, the ganglion cells perform local contrast compression (CgM), but they also act

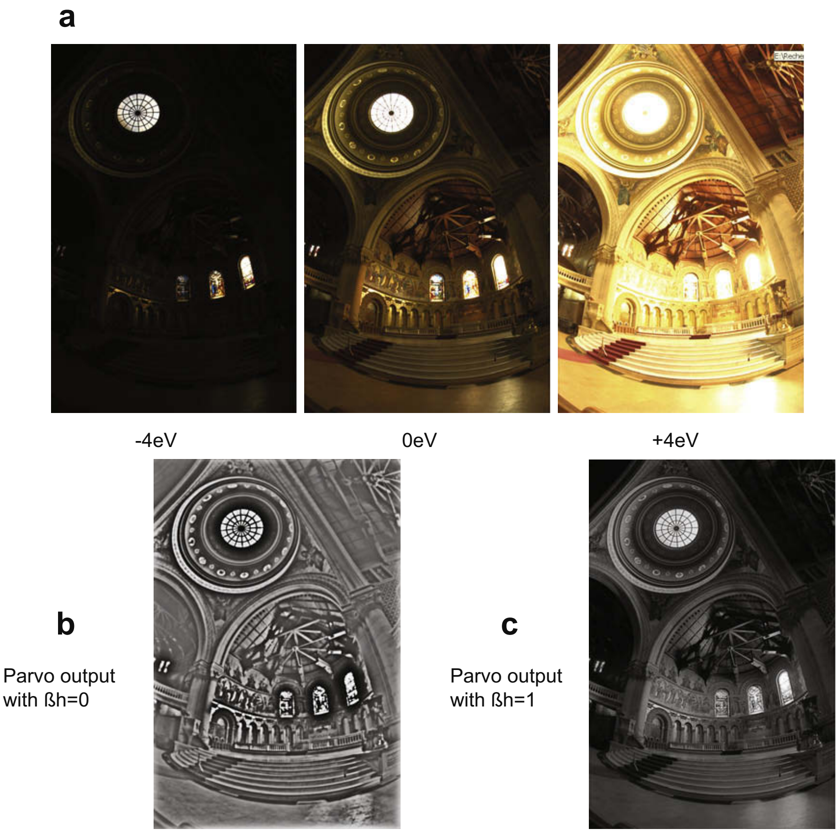


Fig. 8. (a) Thirty-two bits High Dynamic Range (HDR) image: "Memorial" viewed under different exposure in order to view all details in all luminance range on the paper (which is a low dynamic range media), (b) retina Parvo filtering eight bits output with $\beta_h = 0$: details extraction, (c) retina Parvo filtering eight bits output with $\beta_h = 1$: luminance compression: all details of the HDR image become viewable on a low dynamic range media and the luminance is preserved.

as a spatial low pass filter (*F*gM, a filter similar to the filters of the OPL model) thanks to their long range connections to their neighbors. The result is a high pass temporal filtering of the contour information (A filter) which is smoothed and enhanced (*FgM* filter and *C*gM compression). As a consequence, only low spatial frequency moving contours are extracted and enhanced (especially contours perpendicular to the motion direction). We focus here on the MagnoY output which represents the nonlinearity observed on cat Y-cells [28]. Note that it would be interesting to compare this model with the physiological measures obtained in [29].

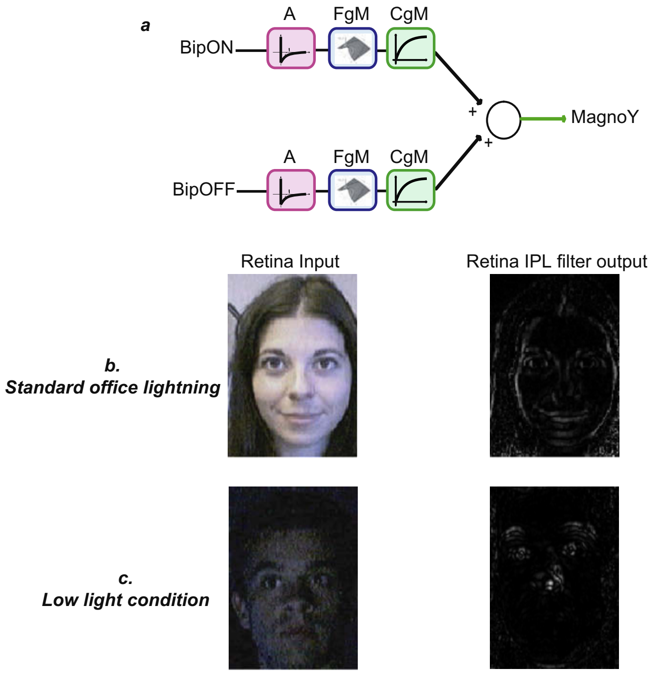


Fig. 9. Effect of the IPL MagnoY filter during a head motion sequence in standard and low light/noisy condition.

2.1.4.2. Properties. Fig. 9b and c give examples of IPL Magno outputs. Fig. 9b presents the case of a pan head motion. Fig. 9c presents the case of tilt head motion in a noisy and poorly illuminated scene. Moving contours perpendicular to the motion direction are accentuated while others are reduced or removed. The IPL Magno output amplitude is linearly dependent on the velocity (high response for fast moving areas and null response for static regions where no change occurs).

We propose to illustrate the effect of the retina Magno filtering including the photoreceptors, the OPL and the IPL Magno models on the synthetic video sequence presented in Fig. 10a. In this sequence, one object is translating while the other remains static on a static background. The retina Magno filtering applied to this image (see Fig. 10b) extracts only the moving object. Moreover, thanks to the temporal effect, the motion energy is high all over the moving objects surface. In the second step, the sequence was corrupted (see Fig. 10c) by high frequency spatio-temporal Gaussian noise ($\mu = 0$, $\sigma = 0.01$). Fig. 10d presents the effect of the retina Magno filter on this noisy sequence; the noise has been reduced. More precisely, the SNR between the retina Magno filtering outputs applied to the original and noisy sequences is 3.2 dB. As a comparison, the image difference which is the simplest motion extraction method gives 0.2 dB because noise is not minimized.

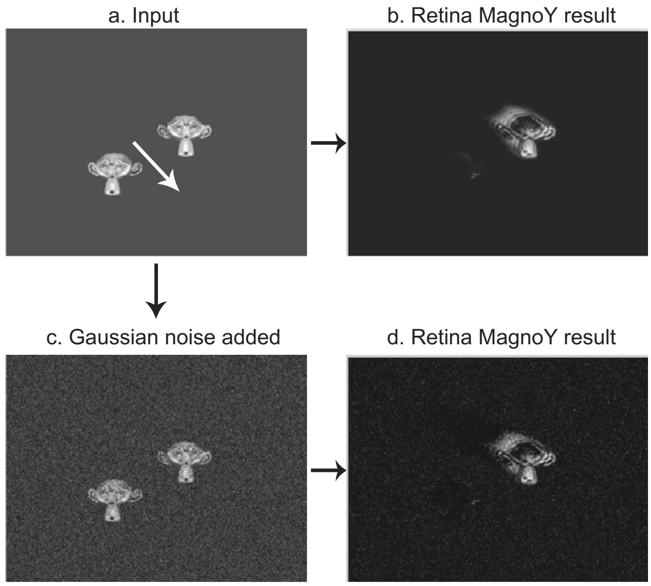


Fig. 10. Motion energy analysis with the retina MagnoY filtering.

Then, IPL Magno allows robust motion information extraction with the advantage of using the benefit of locally adapted contour extraction even in dark or noisy areas.

2.1.5. Overview of the retina model

Fig. 11a presents the global modeling of the retina with photoreceptors, OPL and IPL layers. Many retina cells are still unknown and not all the known cells have been modeled. Nevertheless, this “incomplete” model shows interesting properties for computer vision purposes. The combined effect of the photoreceptors and OPL stage makes up the basis of the system and extracts all the contours. Then, at the IPL level, two information channels exhibit information about details and motion. Fig. 11b represents the whole model. Table 2 summarizes the main advantages and drawbacks of this model: this model takes the main known advantages of the biological model and makes them available for image processing applications. Nevertheless, some work still remains in order to calibrate precisely its parameters in regard of the biological model to make it useful for bio-mimetic applications such as visual substitution. This possible application is currently under development [44].

One step further, a recent study exhibited the properties of other specific polyaxonal amacrine cells which are involved in specific motion background inhibition [65]. This article shows that differential motion detection actually begins early, at the retina level. The association of such properties with the proposed model would allow high level motion description to be improved for specific applications (background motion compensation, etc.).

2.2. Primary visual cortex modeling: FFT and log polar transform

2.2.1. Model

Signals filtered by the retina are received by the Lateral Geniculate Nucleus (LGN) and transmitted to the cortex area 17 called area V1 [30] (the LGN is here considered as an element that only transmits information from the retina to the cortex). It has been demonstrated that in the V1 area, the output of the retina is analyzed by orientation and frequency bands [31,53,54]. An interesting model for the processes occurring at the V1 cortex level is the combination of the FFT amplitude and log polar transformation as introduced by Schwartz [52]. The log polar transformation is generally performed with Gabor filters [14,17,20] which sample the Cartesian spectrum by orientation and frequency bands.

We propose to consider the model introduced in [16] which consists of the use of the GLOP filters (Log Polar Gabor Filters) defined by the following transfer function for a single GLOP filter:

$$G_{ik}(f,\theta) = \frac{1}{\sigma\sqrt{2\pi}}\left(\frac{f_k}{f}\right)^2 \exp\left(-\frac{\ln\left(\frac{f}{f_k}\right)^2}{2\sigma^2}\right)\left(\frac{1+\cos(\theta-\theta_i)}{2}\right)^{50} \qquad (5)$$

where the GLOP filter centered on frequency f_k in the θ_i orientation and with the scale parameter σ appears as a separable filter. Fig. 12

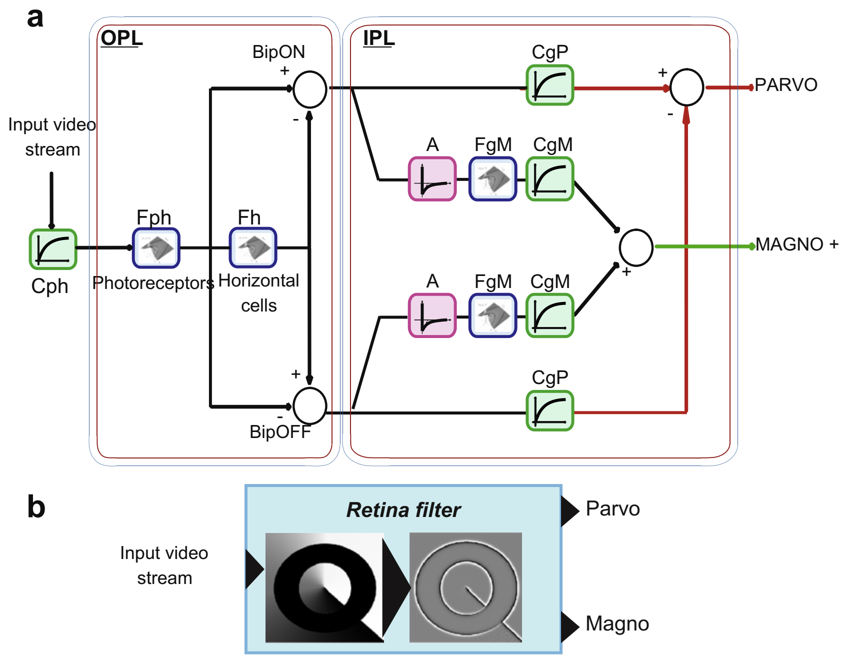


Fig. 11. Global architecture of the retina model.

Table 2
Advantages and drawbacks of the described retina model.

Advantages	Drawbacks
Cell level model: this ensues on a high biological plausibility	Parameters calibration has to been performed to correlate exactly with the biological model
Global approach: nonseparable spatio-temporal filtering, motion analysis and motion changes can also be detected. Also, the implementation naturally supports parallel processing (fast computing)	
All parameters are available and each one tweaks independently a particular property of the retina model. This allows the model to be easy to adapt to many computer vision problems (motion analysis, details extraction, luminance compression, etc.)	
Easy to upgrade/enhance following the same approach (electric circuits analogy)	

shows four amplitude normalized GLOP filters. These filters act as frequency and orientation analyzers which report out the energy related to each specific frequency and orientation band. Compared to standard Gabor filters, these GloP filters have the advantage of being symmetrical in frequency log scale which allows a better zoom effect analysis.

As a final remark, the more angle and frequency samples there are, the more precision we obtain, but at a higher level of complexity. As a compromise, we currently use a maximum of 15 angles by 15 oriented frequencies to obtain 12° angle resolution and fast computing time. Note that generally V1 cortex models use approximately seven frequency bands by seven orientations [16,20].

2.2.2. Properties

In our approach, we only consider the amplitude spectrum sampled by the set of GloP filters, i.e. we sum the energy of each filter in order to get a sampled spectrum of N orientations by M frequency bands. This simplified sampled spectrum allows an easy interpretation of the visual scene with good data dimension reduction and allows information about its main orientations to be obtained. As an example, Fig. 13 shows the log polar spectrum of a car in front of a uniform black background. Because of the high energy located on the vertical frequencies of the spectrum (90°), we can deduce that that the main orientations of the visual scene are horizontal.

In addition, the log polar spectrum has specific properties for rotations and zoom effects. In case of an object rotating around the axis of the camera (roll), the structural characteristics of the spectrum do not change. However the global spectrum is translated along the orientation axis. This effect is illustrated in Fig. 14: a synthetic square rotates in front of the camera. At frame 43, the square has vertical and horizontal diagonals resulting in a specific spectrum. At frame 52, a 45° rotation has occurred and the spectrum is translated by 45°.

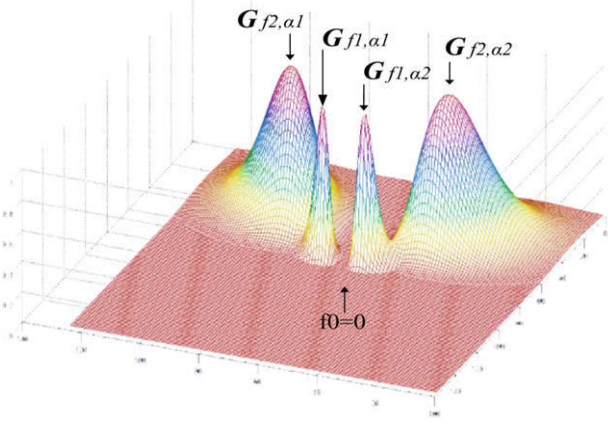


Fig. 12. Four GloP filters samples, placed on frequencies f_1 and f_2 with orientations α_1 and α_2, the null frequency f_0 is centered.

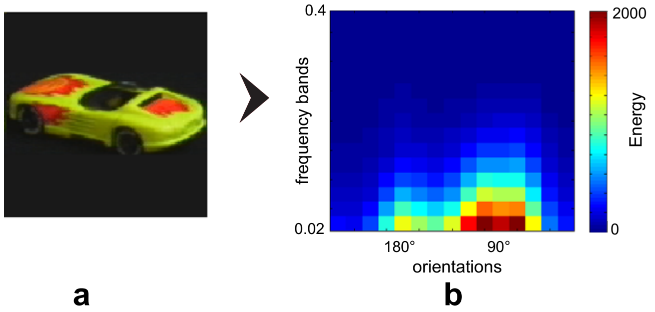


Fig. 13. Log polar spectrum sample of a car.

In the case of a zoom effect, the object captured at different distances from the video acquisition system (eye or camera) always presents the same contours. But these contours are more concentrated or relaxed depending on the viewing distance. Indeed, if a signal $i(x)$ with the corresponding spectrum $S(f)$ is zoomed with a a factor, the resulting signal $i(a \cdot x)$ has the following spectrum

$$\frac{1}{a^2} \cdot S\left(\frac{f}{a}\right) \quad (6)$$

This results in a translation of the spectrum along the log frequency axis

$$\frac{1}{a^2} \cdot S(\ln(f) - \ln(a)) \quad (7)$$

Fig. 15 illustrates this effect with rings captured at different distances (at frame 70, the rings are closer to the camera than at frame 81). On the spectrum, the energy has the same orientation distribution but the mean frequency is higher at frame 81 than at frame 70.

Here is the advantage using GLOP filters rather than standard Gabor filters. Because of the symmetrical shape of the GLOP filters

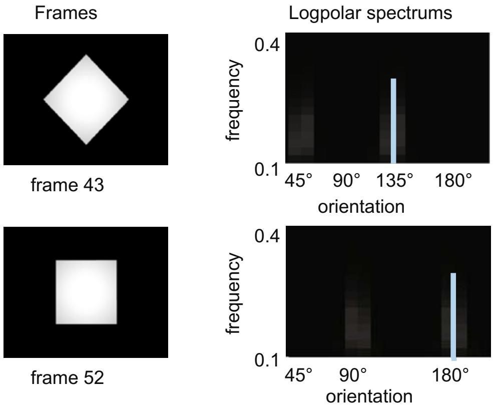


Fig. 14. Spectrum evolution in case of a rotating object around the camera axis.

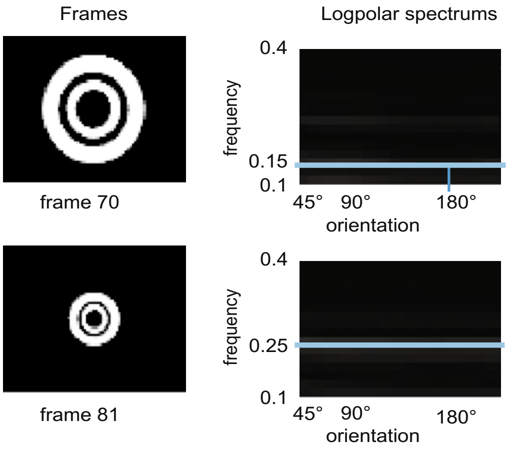


Fig. 15. Spectrum evolution of a zoomed object.

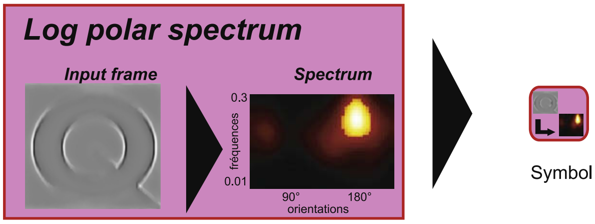


Fig. 16. Symbol of the spectrum analyzer.

on a logarithmic spatial frequency scale, the energy translation can be accurately estimated, which is not possible with Gabor filters as discussed in [16].

This log polar spectrum is used as a structure and texture analyzer. We assign it the symbol of Fig. 16. Its input is a picture, generally one of the retina outputs (Parvo or Magno) and the output is the corresponding sampled log polar spectrum. This analyzer can actually be considered as close to the SIFT algorithm [47]. Indeed, these algorithms and possible applications (object recognition, local features analysis) are similar; nevertheless, the main difference is the use of log polar filters which allow a better robustness against zoom effects. Also, the proposed V1 cortex model works in combination with the retina model which whitens the spectrum and allows a better description of the high frequencies.

3. Bio-inspired low level image processing modules

From a biological point of view, all the processing occurring in the visual cortex is usually related to high level processing. In that sense, the biological model we are considering combines low level analysis (retina processing) and high level analysis (V1 cortex processing). In parallel, from an image processing point of view, low level processing is related to pixel level process compared to high level process which is related to objects. In that sense, all the modules we are going to present are low level image processing modules.

The biological model exploits motion information coming from the retina at the V1 cortex level and computes high level motion analysis at the MT/V5 cortex level, in the occipito-temporal cortex [57, 61, 62]. In this area, specific cortical cells sensitive to speed gradient [55] and transient direction changes [56] have been pointed out. Such high level motion analysis models have been modeled in [58–61].

Here, we propose low level motion analysis tools, using a similar approach and remaining at the first processing steps of human vision. The combination of Magno retina channel and the V1 cortex model analyzer is considered (see Fig. 17) in order to perform motion event detection and direction analysis.

Also the proposed methods rely on a global analysis of the input video stream. Thus, it works in the case of static cameras. However, in the case of camera motion, retina Magno channel would give a high disturbing energy. This issue can be solved using a motion compensation algorithm as proposed in [63].

3.1. Motion analysis

In the proposed motion analysis tool (cf. Fig. 17), the log polar spectrum analyzer acts on the Magno output of the retina filter so that information about motion amplitude, motion direction and motion type can be extracted.

3.1.1. Motion amplitude analysis

The study of the evolution of the log polar spectrum energy of the Magno output of a video sequence v.s. motion amplitude shows that there is a linear relation between speed amplitude and the spectrum energy [32,33]. This is a known property of the Magno channel which is confirmed by the model. We generated 10 different synthetic head motion sequences (pan, tilt, translations in several directions, see captions in Fig. 19), each of them being recorded at 10 different speeds, from slow to fast motion with a linear speed factor. Then, for each sequence, we computed the total Magno energy during the motion and plotted it in Fig. 18. On this graph, each motion gives a linear energy evolution in regard to the motion speed. Considering very low speed, the energy evolution is no longer linear. Since only moving contours contribute to the spectrum energy, this energy is null when no motion occurs.

3.1.2. Motion direction estimation

The log polar spectrum reports the highest energy at the orientations linked to the contours perpendicular to the motion direction. In order to estimate the motion direction, we sum the energy of the log polar spectrum for each orientation [32]. This leads to a cumulated energy per orientation curve (see Fig. 19a–d). On this curve, the abscissa of the maximum amplitude corresponds to the orientation of the most energized moving contours which are perpendicular to the motion direction. More precisely, Fig. 19 gives frames of a synthetic moving head and the corresponding cumulated energy curves. Fig. 19a–c show that a single motion induces a single maximum on the cumulated oriented energy per orientation curve. The abscissa of this energy maximum corresponds to the orientation of the displacement. In the case of more complex motions (Fig. 18d), the curve reports two maximums corresponding to the two rotation axes involved (i.e. hori-

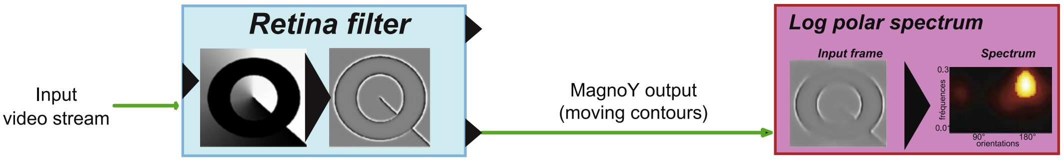


Fig. 17. Motion analyzer structure based on the retina processing followed by the log polar spectrum analysis.

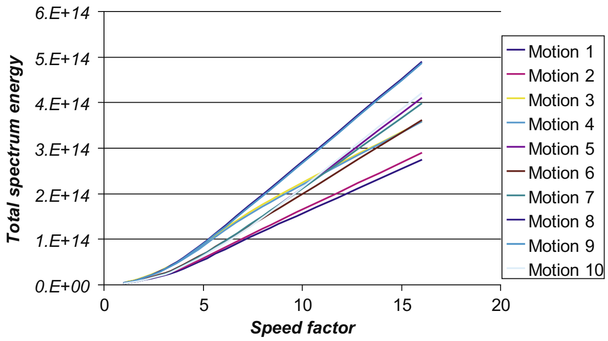


Fig. 18. Evolution of the total spectrum energy with the motion amplitude for different motion sequences.

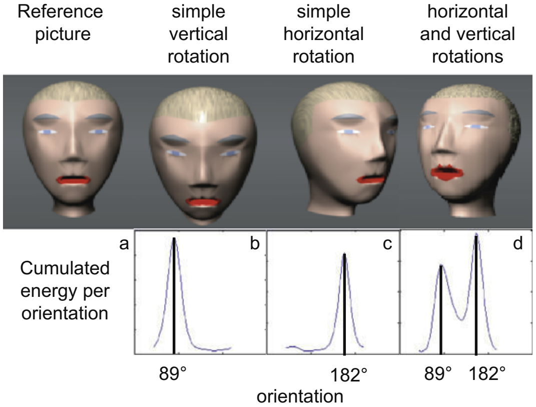


Fig. 19. Simple and mixed head motion estimation and their related cumulated energy per orientation curve.

zontal and vertical which are related to the two main orientations of the face). When achieving a complex rotation, these two orientations exhibit energy even if they are not exactly oriented along the motion direction. This is the well-known aperture problem. In our case, it becomes an advantage: in Fig. 19d, two maximums appear, one for each present rotation (tilt rotation related to 182° and pan rotation related to 89°) occurring at the same time. Moreover, this complex motion can be analyzed observing the amplitude variation of each maximum. In this case, tilt rotation is faster than pan rotation by considering normalized energies with regard to all the orientation statistics of the face.

As a final remark, the precision of the estimated motion orientation is linked to the angle resolution of the log polar transformation and to the frequency characteristics of the object observed. As a consequence, there is a higher precision if contours oriented perpendicular to the motion direction exist.

3.1.3. Motion type detection

Motion type (rotation, zoom or translation) is related to some simple transformations of the log polar spectrum of the retina MagnoY output. As explained in Section 2.2.2, zoom and roll motions induce global translations along the frequency and orientation axes respectively.

Pan and tilt rotations induce, in the log polar domain, localized frequency translations along the rotation axis. Indeed, when a pan or tilt rotation occurs, moving contours are compressed or dilated along the main rotation axis so that the associated spatial frequencies increase or decrease. Fig. 20 illustrates this effect for horizontal rotation of a ring textured object. The energy of the log polar spectrum is concentrated on vertical contours (i.e. horizontal frequencies) because of the rotation orientation. Between frames 11 and 23, the object does a 25° horizontal rotation and the maximum energy translates from $f_{11} = 0.16$ to $f_{23} = 0.22$ normalized frequencies. The white pixels of the log polar spectrum correspond to high energy values.

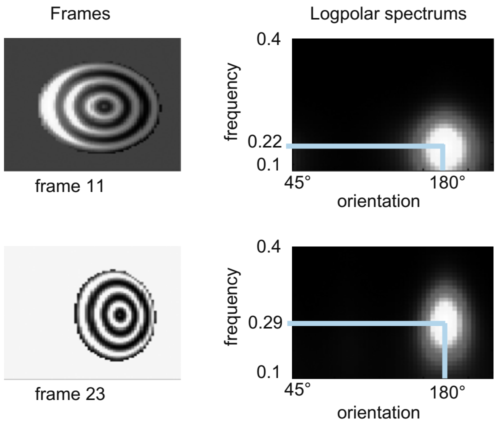


Fig. 20. Log polar retina MagnoY spectrum evolution of a 3D rotating object around the vertical axis. The perspective effect changes the retina image projection yielding to a specific change on its moving contours spectrum.

The last motion type is 2D translation, when an object translates in front of the camera. In such a case, there is no frequency change because contours are not modified. Fig. 21 illustrates this effect with the same object translating upwards. Only horizontal moving contours give a response on the spectrum and the spectrum does not change during the motion.

3.2. Context aware event detection

The goal is to detect as accurately as possible all the dynamic events with slow or fast motions. By considering the global energy of the IPL output or its global log polar spectrum energy, it is possible to deduce information about the motion temporal evolution on a video sequence. Indeed, the IPL output energy is maximum in case of motion and is minimum when no motion occurs, following a linear law. Minimum energy values are related to residual spatio-temporal noise of the acquired video sequence. When motion appears, acceleration induces a global spectrum energy increase and when motion stops, deceleration induces a global spectrum energy decrease.

Fig. 22 gives two examples of periodic motion events: Fig. 22a illustrates the evolution of the global spectrum energy in the case

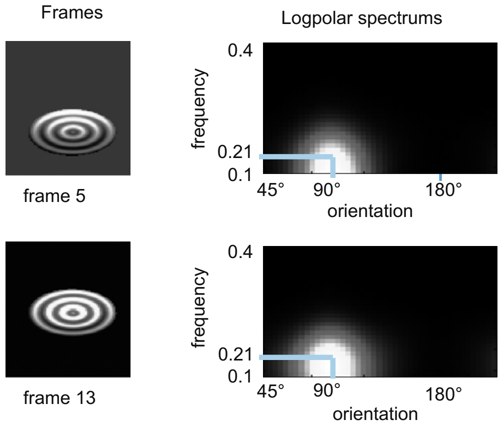


Fig. 21. Retina moving contours log polar spectrum evolution of a translating object in front of camera.

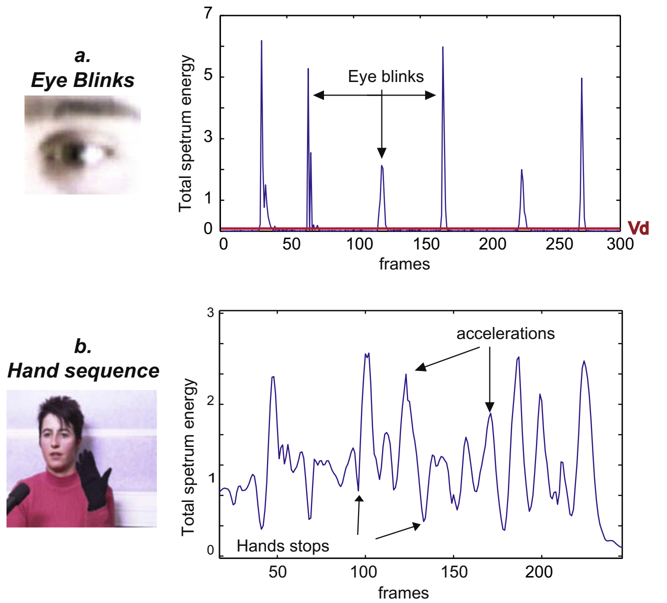


Fig. 22. Temporal evolution of the total spectrum energy in two cases: (a) eye blinks and (b) moving hand sequence.

of eye blinks and Fig. 22b illustrates the evolution of the global spectrum energy in the case of a moving hand gesture.

- On the energy curves, each maximum is related to maximum speed and each minima is linked to motion stops or motion transitions.
- The eye blinks example (Fig. 22a) shows that even if eye blinks are nearly the same from one blink to the other, the global spectrum energy is not the same for each blink because the period of the motion is too short with regard to the frame rate. The acquisition frame rate (30 images per second) is too low to catch the maximum speed and consequently the expected maximum energy value of the IPL spectrum is not reached.
- The moving hand example (Fig. 22b) shows that each energy minimum is related to a hand slow down, i.e. is related to a motion transition between two different specific hand gestures.

3.2.1. Noise level estimation

The mean noise level μ_{noise} and its standard deviation σ_{noise} associated to the frame acquisition system are estimated (assuming Gaussian noise). The noise level is computed as the mean of the residual noise level of the n first frames (currently $n = 40$) in which no motion occurs. The value of μ_{noise} can be updated during the sequence by using the frames with no motion. Indeed, it is quite easy to detect the frames with no motion since the Magno IPL energy is close to zero.

The current global spectrum energy $E(t)$ is supposed to highlight the presence of motion in the scene if:

$$E(t) > Vd, \quad \text{with } Vd = \mu_{noise} + 3 \cdot \sigma_{noise} \tag{8}$$

The threshold Vd can be considered as the minimal motion change that can be detected by the analyzer. From a biological point of view, such threshold brings out the minimal sensitivity of our visual system. Here, we focus on motion detection sensitivity but this threshold concept has already been taken into account for other tasks such as visual contrast sensitivity [19,20]. From a signal processing point of view, this allows false alarms to be limited by not considering small energy variations.

In practice, even for very slow motions (above 0.2 pixel displacement between 2 frames), the global spectrum energy $E(t)$ is higher than the considered threshold (see Fig. 22a where $E_{noise} = 0.15$ with $\sigma_{noise} = 0.01$).

3.2.2. Context aware motion level indicator

Our goal is to propose a temporal filtering acting on motion energy $E(t)$ in order to validate the presence of motion in the scene or not. This requires a reference for the decision to be taken, this reference being here a motion context level called $E_1(t)$. We finally propose the definition of a relative motion level indicator $\alpha(t)$ whose values exhibit the current motion strength with regard to the previous motions. The following describes the method chosen.

The reference indicator $E_1(t)$ introduced in [32], can be interpreted as the output of an electric analog/continuous current converter applied to the total IPL energy $E(t)$. $E_1(t)$ reaches each maximum energy value of $E(t)$ noted E_0 at time t_0 and decreases temporally according to an exponential curve law (capacity effect with temporal constant Δ):

$$E_1(t) = (E_0 - Vd) \cdot \exp[-(t - t_0)/\Delta] \tag{9}$$

Fig. 23a. illustrates this effect for the eye blink sequence presented in Fig. 22a and Fig. 24a shows the results on the hand sequence of Fig. 22b. The indicator $E_1(t)$ can be considered as a "motion context reference" because its value is the same as the current motion energy when a motion transition occurs and forgets the last energy peak other time as a memory effect.

Finally we build a relative motion indicator $\alpha(t)$ whose goal is to link the current energy level $E(t)$ to the motion energy context $E_1(t)$. $\alpha(t)$ allows the current energy level reliability w.r.t. the last motion events (currently Δ seconds between events) to be estimated. $\alpha(t)$ is defined as:

$$\alpha(t) = [E(t) - Vd]/E_1(t) \tag{10}$$

This can be considered as a motion level indicator in the current context. Its values remain between 0 and 1. $\alpha(t) = 1$ when the current energy $E(t)$ is high compared to the motion energy context represented by $E_1(t)$ (i.e. the last motion amplitudes) and $\alpha(t) = 0$ when the current energy level $E(t)$ is lower than the motion energy context. As a consequence, $\alpha(t)$ gives information about the reliability of the amplitude of the current motion compared to the previous motion events.

On Figs. 23b and 24b, the graphs show the temporal evolution of the indicator $\alpha(t)$. It is minimal when no motion occurs, maximal

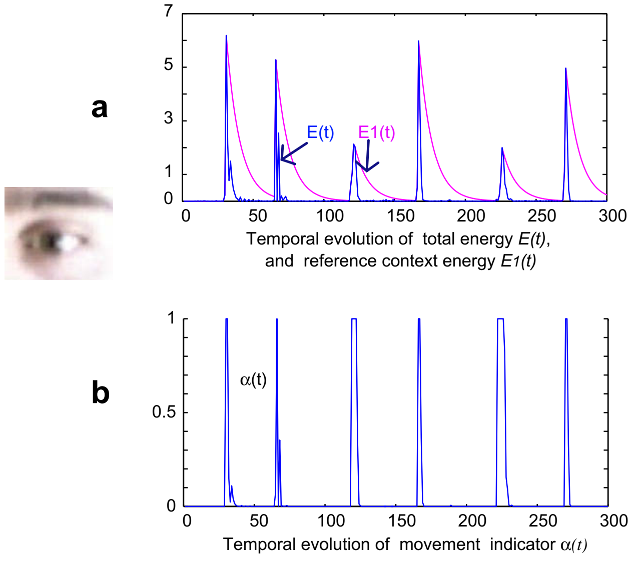


Fig. 23. Temporal evolution of spectrum energy $E(t)$, the temporal context energy reference $E_1(t)$ and the related motion level indicator $\alpha(t)$ during eye blinks.

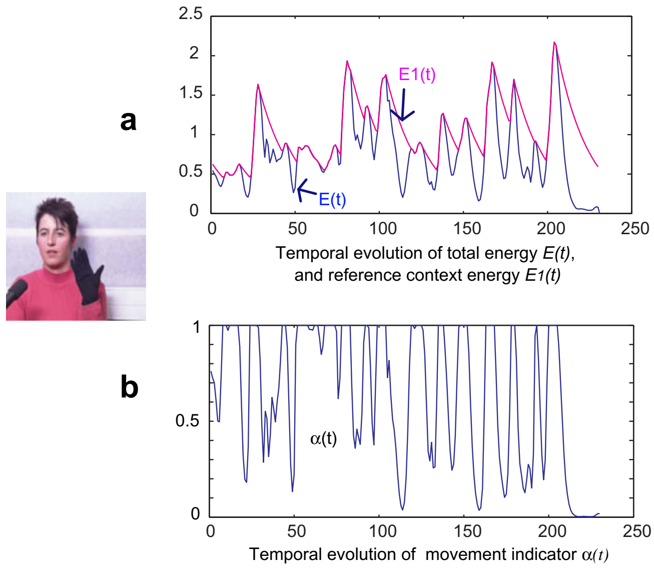


Fig. 24. Temporal evolution of spectrum energy $E(t)$, the temporal context energy reference $E_1(t)$ and the related motion level indicator $\alpha(t)$ for the hand motion sequence.

Table 3
Performances of the motion alert detector.

	Success rate (%)	False alarm rate (%)	Missing rate (%)
Standard lighting	97	2	1
Low light	96	2	2
Noise added	90	3	7

when motion increases and decreases when motion slows down. A threshold level $m\alpha$ can be used in order to take a decision on the presence of a motion in the scene: if $\alpha(t) > m\alpha$ then a motion is detected, none if the condition is not validated. A low value of $m\alpha$ (close to 0) makes the system sensitive to all motions and can also detect parasitic motions if the previous filters (here, the retina Magno filter) do not eliminate them. On the contrary, a value close to 1 makes the system sensitive to very high motions only or isolated motions (a motion preceded by a long no motion period). For example, a threshold level $m\alpha = 0.2$ allows all motion amplitudes (even low motions) to be detected. The risk of false detections introduced by the noise level is minimized while considering only total spectrum energy values higher than Vd (see 3.2.1).

We evaluate on Table 3 the performances of this indicator with a webcam device (Logitech Sphere), using 320 $*$ 240 image resolution. It was tested in standard office lighting conditions (300 lux ambiance), in low light conditions (10 lux ambiance) which force the used device sensor to get a lower signal to noise ratio and finally in noisy conditions (20% Gaussian noise added to the standard lighting videos). Fig. 25 shows some examples of the video sequences tested; it is made up of indoor and outdoor sequences in different lighting conditions. The ground truth is established by a human expert, it is made of 3500 motion alerts for a total 25 Hz video duration of 5 h.

We carried out all tests with the same set of parameters for the retina model: the spatial cut frequency of the photoreceptors stage (F_{ph} filter) α_{ph} is set to 1.0 pixel which allows high frequency noise minimization. The second filter F_h has its spatial cut frequency α_h set to 7.0 pixels to allow large object motion detection. The temporal constants τ_{ph} and τ_h are set to 1 frame which minimizes the high frequency temporal noise. Finally, the temporal constant τ_A of the high pass temporal filter at the IPL Magno level is set to five frames which allows mainly high frequency changes to be extracted. The system is able to detect motion events in very different conditions with a mean success rate higher than 90%. We can see that even if the retina parameters remain the same for the different capture conditions, the performances of the motion detector remain high. This exhibits, on the one hand, the adaptability of the retina model which allows the signals to be reinforced whatever the conditions are, and on the other hand, the reliability of the algorithm applied after the retina step. The lowest performances are obtained when the noise level is too high. Indeed, the mean noise energy level becomes close to the motion energy and this generates a higher missing rate of the motion event detector.

4. Human vision tools software

4.1. General presentation

Demonstration software is made available at [34]. It is possible to apply the retina filter, V1 cortex model and event detector to single pictures, image sequences, video files or live video captures with the help of a webcam. The demonstration also allows High Dynamic Range images to be compressed to Low Dynamic Range

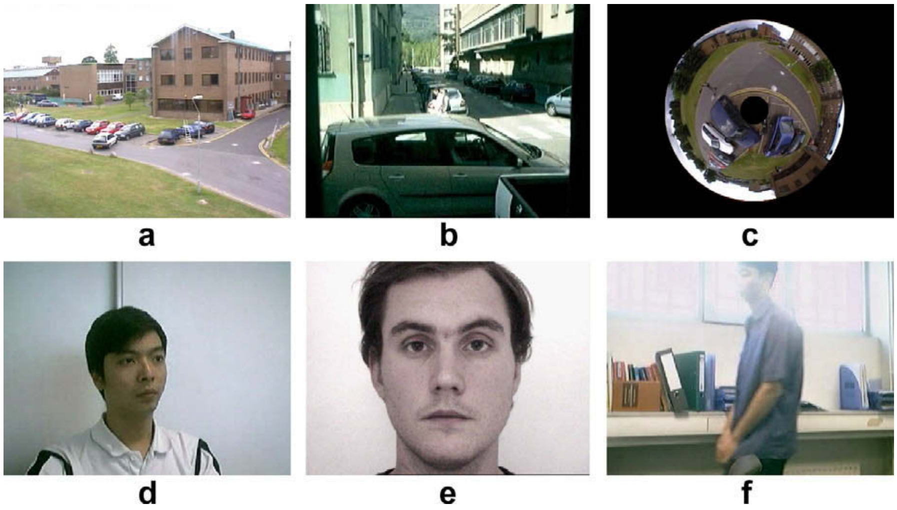


Fig. 25. Samples of event detector test database.

as proposed in Section 2.1 and [48]. In addition, color information can be processed using the demosaicing algorithm proposed by Chaix de Lavarène et al. [64]. Development libraries are also made available for academic experiments only.

The data acquisition step of this software is based on OpenCV [35] which provides an easy and portable image processing library. The toolbox is developed with C++ programming and the display allows the processing results to be seen in real-time with the help of the cross-platform SDL and OpenGL libraries [36].

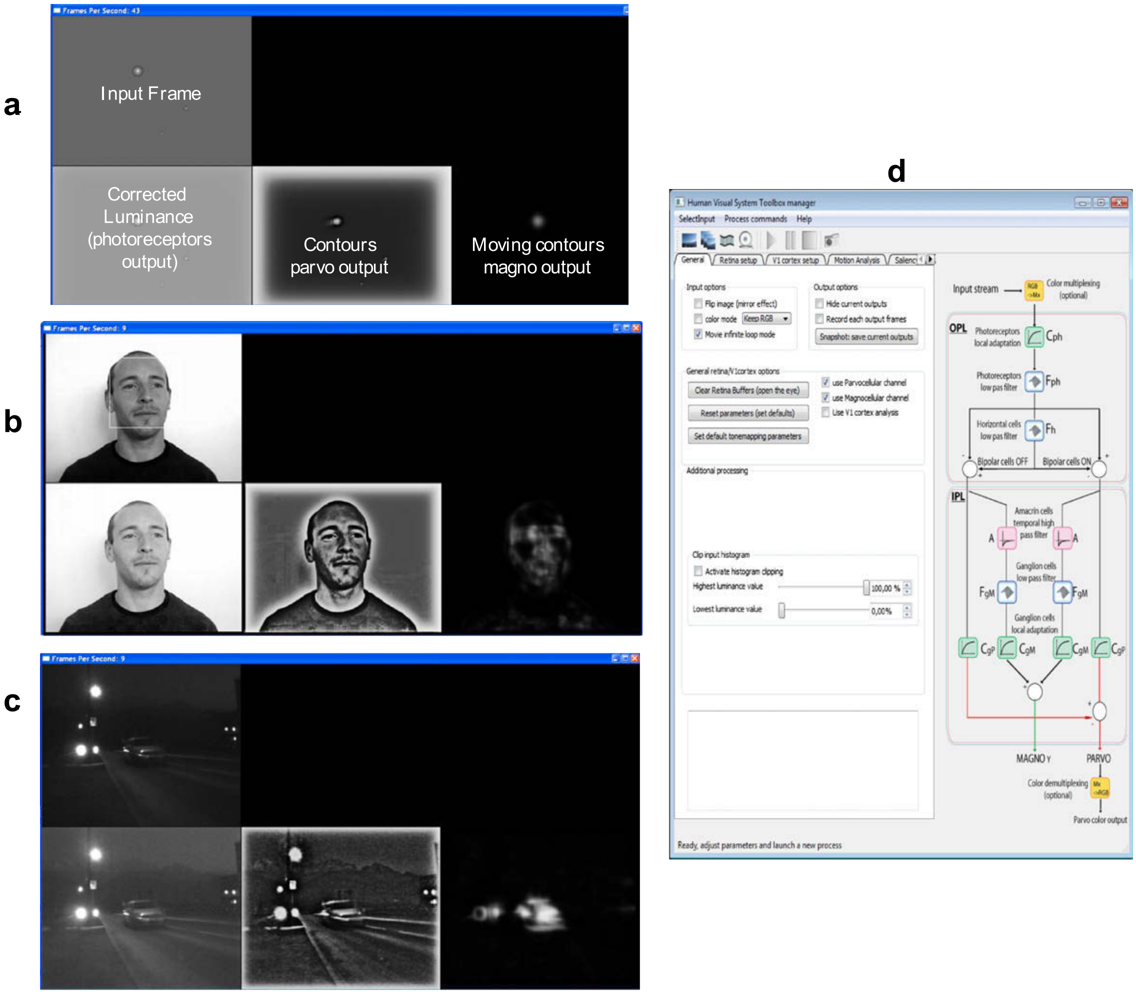
Considering retina processing, four different outputs are proposed (see Fig. 26a): the first one corresponds to the photoreceptor output. This output shows the local luminance adaptation with back-light correction and high frequency spatio-temporal noise filtering. This picture is close to the input but brighter and more contrasted in dark areas with an increased SNR. The second output corresponds to the retina Parvo channel. At this stage, the mean luminance energy is attenuated, spectrum is whitened and all contours are enhanced. The third output corresponds to the retina Magno channel which reports only energy on the moving low spatial frequency contours. The last output is displayed in the case of color processing, following [48] principle. Color image is multiplexed using Bayer color sampling before being processed by the retina. Color image is demultiplexed at the output of the Parvo retina channel. Depending on retina parameters setup, color image tone mapping can be performed (cf. Section 2.1.3 and [48]).

V1 cortex model is available and can be applied either to the input frames, retina Parvo or Magno channels and the resulting log polar sampled spectrum can be observed. Finally, event detection (cf. Section 3.2) is available and allows flexible initialization steps for situation-adapted event detection.

For a processing efficiency overview, we give in Table 4 the framerate obtained with the single threaded demonstration using different image input sizes, from $160 * 120$ pixels to $1024 * 768$ pixels. The test has been carried out on a laptop computer based on an Intel Core 2 2.5 GHz T9300 processor with Windows Vista operating system. Note that the reported values do not take into account image acquisition and display computational time in order to focus exclusively on the model. Also, since the code is regularly enhanced, better results can be expected on the same configuration while hardware configuration and external libraries versions can impact on performances (disk access, live frame grabbing efficiency, FFT processing speed, etc.).

In the table, the computational cost of each part of the retina model, V1 model and the global processing (see. Fig. 11) are given. Note that for Parvo and Magno filtering, the OPL model (see Fig. 4a) with its included photoreceptors local adaptation is required preprocessing step which is taken into account in the reported values. When dealing with gray level HDR images luminance compression as illustrated in Fig. 8c, the computational cost is the one of the Parvo channel since this part of the model achieves the effect.

The retina model and its parts (local luminance adaptation and OPL) present a linear complexity: their number of operations per pixel does not change when image size and other parameters are modified. The event detector also has such properties. This is dem-


a
Frames Per Second: 43
Input Frame
Corrected Luminance (photoreceptors output)
Contours parvo output
Moving contours magno output
b
Frames Per Second: 9
c
Frames Per Second: 9
d
Human Visual System Toolbox manager
SelectInput
Process commands
Help
General
Retina setup
V1 cortex setup
Motion Analysis
Salienc
Input stream
RGB ->Mx
Color multiplexing (optional)
OPL
Photoreceptors local adaptation
Cph
Photoreceptors low pas filter
Fph
Horizontal cells low pas filter
Fh
Bipolar cells OFF
Bipolar cells ON
IPL
Amacrin cells temporal high pass filter
A
A
Ganglion cells low pass filter
FgM
FgM
Ganglion cells local adaptation
CgP
CgM
CgM
CgP
MAGNO Y
PARVO
Color demultiplexing (optional)
Mx ->RGB
Parvo color output

Input options
☐ Flip image (mirror effect)
☐ color mode Keep RGB
☑ Movie infinite loop mode

Output options
☐ Hide current outputs
☐ Record each output frames
Snapshot: save current outputs

General retina/V1cortex options
Clear Retina Buffers (open the eye)
Reset parameters (set defaults)
Set default tonemapping parameters
☑ use Parvocellular channel
☑ use Magnocellular channel
☐ Use V1 cortex analysis

Additional processing
Clip input histogram
☐ Activate histogram clipping
Highest luminance value 100,00 %
Lowest luminance value 0,00%

Ready, adjust parameters and launch a new process

Fig. 26. Screenshots of the retina demonstration software.

Table 4
Relative computing costs of the retina model demonstrator.

	Measured frame rate (fps) using a single threaded application on a Intel Core 2 Duo processor platform					Relative computing cost in the retina model
Input image size	160 * 120	320 * 240	640 * 480	800 * 600	1024 * 768	
Local luminance adaptation with OPL model (cf. fig. 4)	800	171	33	20	9	45%
Parvo channel (cf. Figs. 4 and 6)	415	89	18	11	5	25% (cf. Fig. 6)
Magno channel (cf. Figs. 4 and 9)	371	72	15	9	4	30% (cf. Fig. 9)
Whole retina model (Parvo and Magno)	260	52	11	7	3	100% (Fig. 11)
Whole retina and log polar spectrum analyzer	156	34	7	4	2	–
Event detector + retina Magno channel	370	71	14	9	4	–

onstrated in Fig. 27 which shows the relation between computational time and image size (number of pixels to process). We can see that whatever processing stage is considered, computational time remains proportional to the number of pixels. Note that when considering high resolution images, background operating system applications and system memory management may disturb models computational time measure and over estimate computation time by 10%.

Then, since the retina model achieves an efficient local features extraction with nonseparable spatial and temporal filtering at a relatively low number of operations per pixel, this approach is well addressed to common computer vision applications which require such low level data extraction. However, for higher analysis level (V1 cortex modeling), the log polar spectrum is higher complexity since a FFT is used. Indeed, its implementation relies on a Fourier transform computed by OpenCV followed by a spectrum log polar transformation. This last step is carried out using pre initialized transformation tables (lookup tables) which ensures high efficiency and low processing cost.

4.2. Parameters

The demonstration parameters have been settled up in order to meet generic contour extraction requirements, taking into account standard problems of image input constraints (noise and back-lit problems). In that sense, the proposed values are efficient for a wide range of images and image sequence input. However, the program allows all the parameters to be freely modified.

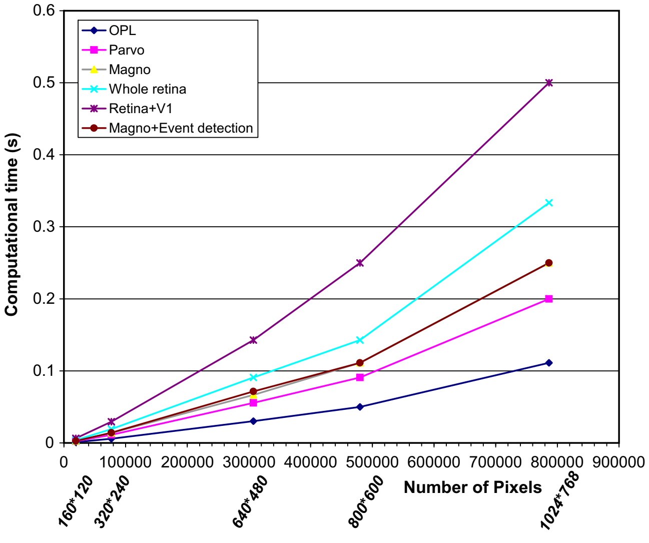


Fig. 27. Computational time of the retina and V1 cortex models (parts and whole) in regard of input image resolution.

Considering retina default setup, the spatial cut frequency α_{ph} of the photoreceptor F_{ph} filter is set to 1.0 (high frequency noise minimization). The second filter F_h has its spatial cut frequency α_h set to 7.0. These two values allow static contours to be extracted with a thickness in the range of (1, 7) pixels. The temporal constants τ_{ph} and τ_h are set to 1 frame which allows the high frequency noise to be minimized.

The temporal constant τ_A of the high pass temporal filter at the IPL Magno level is set to five frames, which allows mainly high frequency changes to be extracted.

Fig. 26a shows the graphical user interface of the demonstrator. Fig. 26b shows an example of standard lighting indoor sequence processing which illustrates the efficiency of the system in standard conditions: noise is minimized, static and moving contours are reliably extracted. Fig. 26c shows the result of the retina processing in the case of a night outdoor city eight bit gray level sequence. The poor lighting induced a dark and noisy acquisition. This video stream is enhanced by retina filtering and its outputs show details which could not be seen on the original frame (mountains and constructions in the background and a moving car with no lights switched on, in the foreground). Finally, Fig. 26d shows the general parameters settings window, it allows all retina and V1 cortex parameters to be adjusted freely by the user.

4.3. Applications and perspectives

Our approach, consisting in taking into account the processing occurring at the retina level, has the main advantage of preparing video data appropriately for high level processing. As a result, it is of great interest to use it for a wide range of applications, from general visual scene analysis (details and motion extraction) to more specific applications (cf. Fig. 28). We have already used it for the purpose of head motion analysis [32,38,39,66] with an application of hypo vigilance analysis to detect and prevent a driver from falling asleep [40,41]. Also, the interest of using our low level image processing modules for head motion analysis in the context of sign language decoding was demonstrated in [42].

In particular, this retina model is well suited for applications such as video surveillance because of its ability to enhance visual information even in the case of back-lit situations and noise. We are currently working on a moving object tracker and identifier able to track objects by considering their motion information at the Magno retina filter output and to identify or classify them by the use of the log polar spectrum of their contour detail (Parvo retina output) given by the V1 cortex model. This last step has already been engaged in [43].

Also, when considering High Dynamic Range images such as the example proposed in Fig. 8a, this models acts similarly as the human retina does in terms of image luminance compression (i.e. “Tone Mapping”): as discussed in [2], since real world scenes are High Dynamic Range (HDR) but neurons cannot code a so wide

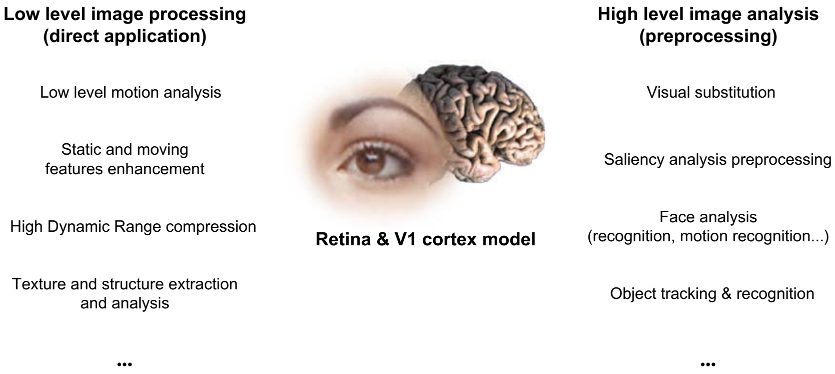

Low level image processing
(direct application)
Low level motion analysis
Static and moving
features enhancement
High Dynamic Range compression
Texture and structure extraction
and analysis
...
Retina & V1 cortex model
High level image analysis
(preprocessing)
Visual substitution
Saliency analysis preprocessing
Face analysis
(recognition, motion recognition...)
Object tracking & recognition
...

Fig. 28. Overview of the possible applications of the retina and V1 cortex models.

variety of values (they can be considered as Lower Dynamic Range coders), the retina compresses this HDR information on a LDR format and keeps all details in the visual scene as shown in Fig. 8c. This particular topic of image tone mapping is discussed in [48].

Another application under consideration is the use of this model as a video preprocessing step for visual prostheses and sensory substitution [44].

As a final remark, this retina model is designed to be used in any context where contours and contrasts constitute important information. This model can be used either on standard camera or other kind of images (infrared, Xray images, etc.) since luminance/source amplitude information is the considered input data. Using the retina model as a generic preprocessing step for input data enhancement can be a very appropriate solution for any kind of application which requires low level features extraction in order to enhance the high level analysis capabilities. This preprocessing mimics human vision: these low level computational steps are required in order to ensure more efficient higher level visual scene analysis and interpretation. As a consequence, applying such an approach to computer vision is expected to allow a new step of reliability and efficiency to be reached. This kind of preprocessing for higher level generic scene recognition or classification is the next step which would show the interest of such an approach. A starting point should be video sequence analysis and interpretation for sequence summary synthesis [45]. Such a topic would allow such models to be evaluated for key frames extraction, objects tracking and recognition and visual scene categorization. Focusing on real life video and animation sequences [46] should constitute a starting point to evaluate the robustness of such an approach against the wide variety of situations which can be addressed.

Fig. 28 summarizes the discussed applications which can potentially benefit from the proposed low level processing. From "classical" computer vision applications to more human vision centered topics, "retina like" preprocessing is expected to enhance target applications performances. On the figure, we distinguish low level image processing applications which can directly be carried out by the proposed models and higher level analysis which would require these tools as preprocessing steps.

5. Conclusion

In this paper, we presented modules for low level data processing. All the modules considered are based on the modeling of specific parts (retina and V1 cortex) of the Human Visual System. The efficiency of the modules proposed for video data structuring before high level processing has been shown.

This model can be considered as an image processing kernel basis which can be extended for more specific applications. From a biological point of view, it shows interesting properties for image processing purposes and for its use in different computer vision applications. Its integration in video-surveillance applications, head motion analysis or image classification shows its qualities in terms of image analysis and fast computing time. This defends our global philosophy of building bio-inspired vision algorithms.

The model presented in this paper focuses on gray level image processing. Color information integration has been proposed in [48] for High Dynamic Range processing application, using an accurate demosaicing algorithm from Chaix de Lavarène et al. [64] . However, since at the retina level, many cells and their actions are still biologically unknown, investigations are still required in order to understand and better model all the vision processing steps which we naturally perform. For instance, the introduction of motion compensation for eye/camera motion handling, proposed in [63] , would enlarge the application spectra. Finally, the next research step is to involve such human retina model in higher level video sequence analysis, taking full advantages of the extracted low level features as shown in [66] in the context of face analysis.

References

[1] 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.

[2] J. Hérault, B. Durette, Modeling visual perception for image processing, in: F. Sandoval et al. (Eds.), IWANN 2007, LNCS 4507, Springer-Verlag, Berlin Heidelberg, 2007, pp. 662–675.

[3] D.J. Jobson, Z. Rahman, G.A. Woodell, A multi-scale Retinex for bridging the gap between colour images and the human observation of scenes, IEEE Transactions on Image Processing 6 (7) (1997) (Special Issue on Colour Processing).

[4] Z. Rahman, D.J. Jobson, G.A. Woodell, G.D. Hines, Image enhancement, image quality and noise, in: Photonic Devices and Algorithms for Computing VII, Proc. SPIE, vol. 5907, 2005, pp. 59070N.1–59070N.15. ISBN: 0-8194-5912.

[5] H. Senan, A. Saadane, D. Barba, Design and evaluation of an entirely psychovisual-based coding scheme, Journal of Visual Communication and Image Representation 12 (4) (2001) 401–421 (21).

[6] C.A. Mead, M.A. Mahowald, A silicon model of early visual processing, Neural Networks 1 (1988) 91–97.

[7] N. Franceschini, A. Riehle, A. Le Nestour, in: D.G. Stavenga, R.C. Hardie (Eds.), Facets of Vision, Springer, Berlin, 1989, pp. 360–390.

[8] R. Van Rullen, S. Thorpe, Surfing a spike wave down the ventral stream, Vision Research 42 (2002) 2593–2615.

[9] T.M. Bernard, P.E. Nguyen, F.J. Devos, B.Y. Zavidovique, A programmable VLSI retina for rough vision, Machine Vision and Applications 7 (1) (1993) 4–11.

[10] T. Allen et al., Orientation selective VLSI retina, in: Proc. SPIE, Visual Communications and Image Processing, vol. 1001, 1988.

[11] W.H.A. Beaudot, P. Palagi., J. Hérault, Realistic Simulation Tool for Early Visual Processing Including Space, Time and Colour Data, International Workshop on Artificial Neural Networks, Barcelona, June 1993.

[12] S. Marcelja, Mathematical description of the responses of simple cortical cells, Journal of the Optical Society of America 70 (1980).

[13] J.G. Daugman, Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters, Journal of the Optical Society of America 2 (1985).

[14] C. Palm, T.M. Lehmann, Classification of color textures by Gabor filtering, Machine Graphics and Vision International Journal 11 (2/3) (2002) 195–219.

[15] A. Chauvin, J. Hérault, C. Marendaz, C. Peyrin, Natural scene perception: visual attractors and image neural computation and psychology, in: W. Lowe, J. Bullinaria (Eds.), Connexionist Models of Cognition and Perception, World Scientific Press, 2002.

[16] N. Guyader, A. Chauvin, C. Massot, J. Hérault, C. Marendaz, A biological model of low-level vision suitable for image analysis and cognitive visual perception, Perception 35 (2006).

[17] L. Itti, C. Koch, Computational modeling of visual attention, Nature Reviews 2 (2001).

[18] D. Walther, Interactions of Visual Attention and Object Recognition Computational Modeling, Algorithms, and Psychophysics, PhD Thesis of the California Institute of Technology, Pasadena, California, 2006.

[19] S. Daly, A visual model for optimizing the design of image processing algorithms, in: Proc. IEEE Int'l Conf. Image Processing, 1994, pp. 16–20.

[20] O. Le Meur, P. Le Callet, D. Barba, D. Thoreau, A coherent computational approach to model bottom-up visual attention, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (5) (2006).

[21] A.B. Torralba, Analogue Architectures for Vision Cellular Neural Networks and Neuromorphic Circuits, PhD Thesis in Computer Science, UJF, Grenoble, France, 1999.

[22] Y. Dan et al., Efficient coding of natural scenes in the lateral geniculate nucleus: experimental test of a computational theory, Journal of Neuroscience 16 (1996) 3351–3362.

[23] J.J. Attick, A.N. Redlich, Towards a theory of early visual processing, Neural Computation 2 (1990) 308–320.

[24] H.B. Barlow, Redundancy reduction revisited, Computation in Neural Systems 12 (2001) 241–253.

[25] D.M. Dacey, Higher order processing in the visual system, in: Ciba Foundation Symposium, vol. 184, Wiley, Chichester, 1994, pp. 12–34.

[26] S.M. Smirnakis, M.J. Berry, D.K. Warland, W. Bialek, M. Meister, Adaptation of retinal processing to image contrast and spatial scale, Nature 386 (1997) 69–73.

[27] F. Werblin, G. Maguire, P. Lukasiewicz, S. Eliasof, S.M. Wu, Neural interactions mediating the detection of motion in the retina of the tiger salamander, Visual Neuroscience 1 (1988) 17–29.

[28] Non-linear spatial summation in cat retinal ganglion cells at different background level, Experimental Brain Research 36 (2) (1979). doi:10.1007/BF00238913.

[29] D.H. Kelly, Motion and vision. II. Stabilized spatio-temporal threshold surface, Journal of the Optical Society of America 69 (1979) 1340–1349.

[30] J. Bullier, Integrated model of visual processing, Brain Research Reviews 36 (2–3) (2001) 96–107.

[31] D.H. Hubel, T.N. Wiesel, Sequence regularity and geometry of orientation columns in monkey striate cortex, Journal of Computational Neurology 158 (1974) 267–293.

[32] A. Benoit, A. Caplier, Biological Approach for Head Motion Detection and Analysis, EUSIPCO 2005, Antalya, Turkey 2005.

[33] A. Benoit, A. Caplier, Motion Estimator Inspired from Biological Model for Head Motion Interpretation WIAMIS2005, Montreux, Switzerland, 2005.

[34] Retina demo at <http://sites.google.com/site/benoitalexandrevision/demonstrations> or access to authors labs, <http://www.polytech.univ-savoie.fr/index.php?id=listic> and <http://www.gipsa-lab.inpg.fr>.

[35] Open Source Computer Vision Library: <www.intel.com/technology/computing/opencv/>.

[36] Simple Direct Media Layer : <www.libsdl.org>.

[37] Industrial Light & Magic: <www.openexr.org>.

[38] A. Benoit, A. Caplier, Hypovigilence Analysis: Open or Closed Eye or Mouth? Blinking or Yawning Frequency? IEEE AVSS 2005, Como, Italy, 2005.

[39] A. Benoit, A. Caplier, Head Nods Analysis: Interpretation of Nonverbal Communication Gestures IEEE ICIP 2005, Genoa, Italy, 2005.

[40] A. Benoit, L. Bonnaud, A. Caplier, P. Ngo, L. Lawson, D.G. Trevisan, Multimodal Focus Attention Detection in an Augmented Driver Simulator, Personal and Ubiquitous Computing, vol. 13(1), pp. 33–41.

[41] A. Benoit, L. Bonnaud, A. Caplier, I. Damousis, F. Jourde, J.-Y.L. Lawson, L. Nigay, M. Serrano, D. Tzovaras, Multimodal signal processing and interaction for a driving simulator: component-based architecture, Journal on Multimodal User Interface 1 (1) (2007) 49–58.

[42] O. Aran, I. Ari, A. Benoit, A.H. Carrillo, F.X. Fanard, P. Campr, L. Akarun, A. Caplier, M. Rombaut, B. Sankur, Sign Language Tutoring Tool, eNTERFACE 2006, The Summer Workshop on Multimodal Interfaces, Dubrovnik, Croatia, 2006.

[43] A. Benoit, N. Guyader, A. Caplier, J. Herault, Emotion Classification and Face Identification, A Bio-inspired Model Perception 36 ECVP Abstract Supplement, 2007.

[44] B. Durette, R. Corvino, S. Mancini, D. Alleysson, J. Hérault. Model of the human retina for sensory substitution and retinal implants, in: Euroconférence Sensory Perception: Basic Mechanisms and Human Diseases, Institut Pasteur, France, Paris, 2006.

[45] B.T. Truong, S. Venkatesh, Video abstraction: A systematic review classification. ACM Transactions on Multimedia Computing, Communications and Applications 3 (1) (2007) 37p (Article 3).

[46] B. Ionescu, D. Coquin, P. Lambert, V. Buzuloiu, A fuzzy color-based approach for understanding animated movie content in the indexing task, EURASIP Journal on Image and Video Processing 2008 (1) (2008) 20–36.

[47] D.G. Lowe, Object recognition from local scale-invariant features, in: Proceedings of the International Conference on Computer Vision 2, 1999, pp. 1150–1157. doi:10.1109/ICCV.1999.790410.

[48] A. Benoit, D. Alleysson, J. Herault, P. Le Callet, Spatio-temporal tone mapping operator based on a retina model, Lecture Notes in Computer Science 5646 (2009) 12–22.

[49] J.J. Attick, A.N. Redlich, What does the retina know about natural scenes?, Neural Computation 4 (1992) 196–210

[50] W.H.A. Beaudot, Sensory coding in the vertebrate retina: towards an adaptive control of visual sensitivity, Network: Computation in Neural Systems 7 (2) (1996) 317–323.

[51] M. Carandini, J.B. Demb, V. Mante, D.J. Tolhurst, Y. Dan, B.A. Olshausen, J.L. Gallant, N.C. Rust, Do we know what the early visual system does?, Journal of Neuroscience 25 (2005) 10577–10597

[52] E.L. Schwartz, Cortical anatomy and size invariance, Vision Research 18 (1983) 24–58.

[53] R. De Valois et al., The orientation and direction selectivity of cells in macaque visual cortex, Vision Research 22 (1982) 531–544.

[54] L.O. Harvey, V.V. Doan, Visual masking at different polar angles in the two dimensional Fourier plane, Journal of Optical Society of America A 7 (1990) 116–127.

[55] Julio C. Martinez-Trujillo, John K. Tsotsos, E. Simine, M. Pomplun, R. Wildes, S. Treue, H.-J. Heinze, J.-M. Hopf, Selectivity for speed gradients in human area MT/V5, Neuroreport 16 (5) (2005) 435–438.

[56] Julio C. Martinez-Trujillo, D. Cheyne, W. Gaetz, E. Simine, J.K. Tsotsos, Activation of Area MT/V5 and the Right Inferior Parietal Cortex during the Discrimination of Transient Direction Changes in Translational Motion Cerebral Cortex Advance Access, September 29, 2006 Cereb. Cortex 2007, vol. 17, pp. 1733–1739.

[57] M.Y. Villeneuve, R. Kupers, A. Gjedde, M. Ptito, C. Casanova, Pattern–motion selectivity in the human pulvinar, NeuroImage 28 (2) (2005) 474–480.

[58] B. Hassenstein, W. Reichardt, Systemtheoretische analyse der zeirreihenfolgen- und vorzeichenauswertung bei der bewgungsperzeptiion des russelkafers chlorophanus, Zeitshchrift fur Naturforschung B, vol. 11, 1956, pp. 513–525.

[59] E.H. Adelson, J.R. Bergen, Spatiotemporal energy models for the perception of motion, Journal of the Optical Society of America A 2 (1985) 284–299.

[60] J.K. Tsotsos, Y. Liu, J.C. Martinez-Trujillo, M. Pomplun, E. Simine, K. Zhou, Attending to visual motion, Computer Vision and Image Understanding 100 (1–2) (2005) 3–40.

[61] Z.-L. Lu, G. Sperling, Three systems theory of human visual motion perception: review and update: errata, Journal of the Optical Society of America A 19 (2002) 413.

[62] G. Sperling, Z.-L. Lu, A systems analysis of visual motion perception, in: T. Watanabe (Ed.), High-level Motion Processing, MIT Press, Cambridge, MA, 1998, pp. 153–183.

[63] S. Marat, T. Ho Phuoc, L. Granjon, N. Guyader, D. Pellerin, A. Guérin-Dugué, Spatio-temporal saliency model to predict eye movements in video free viewing, European Signal Processing Conference (EUSIPCO'2008), Lausanne, Suisse, August 2008.

[64] B. Chaix de Lavarène, D. Alleysson, J. Hérault, Practical implementation of LMMSE demosaicing using luminance and chrominance spaces, Computer Vision and Image Understanding 107 (1) (2007) 3–13.

[65] S.A. Baccus, B.P. Ölveczky, M. Manu, M. Meister, A retinal circuit that computes object motion, Journal of Neuroscience 28 (2008) 6807–6817.

[66] A. Benoit, A. Caplier, Fusing bio-inspired vision data for simplified high level scene interpretation: application to face motion analysis, Computer Vision and Image Understanding 114 (7) (2010) 774–789.