Journal of Integrative Neuroscience, Vol. 6, No. 3 (2007) 477–509

Imperial College Press
www.icpress.co.uk

Special Report

BIO-INSPIRED MODEL OF VISUAL INFORMATION ENCODING FOR LOCALIZATION: FROM THE RETINA TO THE LATERAL GENICULATE NUCLEUS

THIBAUD DEBAECKER* and RYAD BENOSMAN†
UPMC — ISIR, Groupe SIMA, 3 rue Galilée, 94200
Ivry sur Seine, France
**thibaud.debaecker@lisif.jussieu.fr*
†*benosman@ccr.jussieu.fr*

Received 29 March 2007
Accepted 3 August 2007

In this study, a bio-inspired approach for extracting efficient features prior to the recognition of scenes is proposed. It is highly inspired from the model of the mammals visual system. The retina contains many levels of neurons (bipolar, amacrine, horizontal and ganglion cells) accurately organized from cones and rods to the optic nerve up till the lateral geniculate nucleus (LGN) which is the main thalamic relay for inputs to the visual cortex. This structure probably eases other brain areas tasks in preprocessing the visual information. This paper is focusing on the study of these specific structures, relying on a bottom up approach to propose a comprehensive mathematical model of the low level image processing performed within the eye. The presented system takes into account the foveolar structure of the retina to produce a low-resolution representation of observed images by decomposing them into a local summation of elementary gaussian color histograms. This representation corresponds to the LGN biological organization. It has been thought that due to short timings, some very quick localization tasks involving particularly fast information processing pathways cannot be provided by the classical ones passing through higher level cortical areas. This work proposes a model of retinal coding and LGN-visual representation that we show provides reliable and sufficient early features for scenes recognition and localization. Experiments on real scenes using the developed model are presented showing the efficiency of the approach on localization.

Keywords: Lateral-inhibition; retina; foveolar vision; GMM; image features; image coding; LGN; lateral geniculate nucleus.

1. Introduction

Visual analysis of scenes is a cognitive task that generally involves a scan of observed environments to find targets among distractors. It is suggested that the brain extracts features from visual flows in order to perform localization [21]. Among these features [30], shapes, color [55], motion can be found [1]. In general, location is directly linked to find the conjunction of features. The conjunctions are intended to describe the scene in a unique manner in order to enable further comparisons.

Many approaches to vision based spatial memory can be found in the literature visual search [21, 23], sometimes also combined with saccadic eye movements [25, 24], visual attention [27, 26], and finally some approaches that develop computational models [28–33].

This paper uses the model of the eye as a starting point to inquire on the possible image processing work derived from its structure. The retina contains many levels of neurons (bipolar, amacrine, horizontal and ganglion cells) accurately organized from cones and rods to the optic nerve [19]. This internal structure of the eye apart from all the known foveolar advantages probably preprocesses the visual information introducing a pre-coding of the acquired images [7]. It provides an adaptive retinotopic visual representation in the LGN [3], which facilitates the tasks performed by the downstream visual cortex and other brain areas. There exists a large amount of work dealing with the issue of studying the processing activity in the early retina. It is interesting to notice that approaches related to information theory in the literature fail to give all the possibilities involved by the lateral inhibition mediated by retinal horizontal cells [11, 5]. Most of the existing works deal with the hypothesis that accounts for the extentions of bell-shape behavior, trying to explain its general functioning. In general, these theories use high-order statistics (autocorrelation or power spectrum) in natural images [40, 47–49]. The main idea behind these assumptions is connected to the fact that due to intensity decrease, the relative noise caused by random photon absorption increases. It then becomes essential to increase the lateral-inhibition extent to average out the noise. Very few papers set the hypothesis that the lateral-inhibition adaptation in the early retina is part of a much more complex system enabling the extraction of several image attributes, such as edges, the amount of contrast, and in a general way the radiometric information.

It seems that there is a change in the hypothesis between the new and early theories that deal essentially around what is really extracted from images [47, 48]. It has been believed that the extraction is based on collecting as much nonredundant information from acquired images as possible. Recently, it has been suggested that it may be relevant to consider that essential information may be extracted too [46]. It seems that the visual system eliminates specific information very early in the processing chain. As explained in Balboa [45], two possible answers can be found. The first is connected to the idea that if one reduces the amount of types of signals to be transmitted then large quantities can be emitted. The second considers that due to the limited capacity of the central nervous system, it might be better dealing with fewer types of signals essential for the entity survival rather than wasting energy among nonrelevant information. One may wonder how information is inhibited and why?. It is suggested in the predictive coding theory [40] that there is a reduction of information intensity at the points with the same intensity as their neighbors. The goal is to reduce the wide range of intensities of acquired scenes in the narrow dynamic range of the second-order neurons (bipolar cells). There are two contradictory ideas concerning the main goal of lateral inhibition. On one hand,

October 8, 2007 18:53 WSPC/179-JIN 00158

it is suggested that the main goal of lateral inhibition is not the coding of intensity [40, 43, 44]. On the other hand, it is proposed that the nonchromatic goals of the OPL (optical plexiform layer of the retina) lateral inhibition is meant to ease the optimization of detection [42], contrast [41] and intensity [39]. The idea of a possible link between lateral inhibition and edges has been proposed in Ratliff [50]. It can also be pointed out that there might be different channels as contrast is particularly important at edges, whereas intensity is important away from them. The segmentation performed within the retina probably provides nonaccurate information, but quick enough to be used efficiently by reactive loops [37, 38].

Very few papers propose to give a visual representation of the information within the LGN. Studies on neural information transmission (See Ref. [1] for an overview), point out its significant complexity and diversity. The approaches using information theory and spikes show that a huge information content can be transmitted using a wide variety of stimuli [5, 6]. The presented work is proposing a bioinspired method combining both feature extraction and noise minimization. Its functioning is inspired by the circuitry of the primate visual system, using statistical methods to achieve both tasks of coding and noise reduction. The method combines both the lateral inhibition mechanism and the foveolar structure of the retina. This sets it apart from existing work which is more focused on local information theory within cells. The properties of the retina seem to have many interesting uses [7]. Among them, there are two major ones: its nonlinear resolution [56], inducing a directional field of view, starting from a lower to a higher resolution; and its pyramidal implicit structure, decreasing drastically the amount of information to be processed. These observations tend to show that there might be a deep connection between the need of a directional resolution and a decrease of the amount of information collected.

The experimental observations and theoretical considerations exposed in Zugaro [8] involve a simple mechanism, underlying the capacity to rapidly reorient in a familiar environment independent of high brain areas. The very short latencies observed are consistent with the fact that the anterodorsal thalamic nucleus receives direct projections from the retina as well as indirect projections from the visual cortex via other areas. In addition, the visual stimulation of the retina evokes field potentials in the primary visual cortex with delays as brief as 40 msec [9, 10]. The primary goal of this paper is to inquire on the possible effects of combining variable image resolution and a chain of image coding in order to propose a model of early retinal image processing. The proposed method is a complete approach (from retina's image acquisition to LGN retinotopic representation) that we will show sufficient to perform quick localization tasks and provide a robust and quick input for a spatial localization algorithm.

The paper is organized as follows. Section 2 briefly presents the structure of the retina, going from rods and cones to the optic nerve. In Sec. 3, foveolar vision presented followed by the architecture of the preprocessing based on a local decomposition of histograms. Mathematical results are introduced in Sec. 4 and LGN

representation is exposed in Sec. 5. Spatial localization using our bio-inspired technique is given in Sec. 6 where a spatial localization model based on scene recognition is presented. Finally, conclusions and future work are included in Sec. 7.

2. The Retina Structure

The neural structure of the retina inspired the developed visual system. This section will briefly introduce its architecture and properties. The retina is a seven-layered structure involved in signal transduction as seen in Fig. 1.

Light enters from the ganglion cell layer (GCL) side first, and must penetrate all cells before reaching the rods and cones. The rods and cones transform light into a signal and send it through the cells of the outer nuclear layer (ONL). In the outer plexiform layer (OPL) photoreceptor axons contact the dendrites of bipolar cells and horizontal cells. Horizontal cells provide interneuron connections. The bipolar cells in the inner nuclear layer (INL) process input from photoreceptors and horizontal cells, and transmit the signal to their axons. In the inter plexiform layer (IPL), bipolar axons contact ganglion cell dendrites and amacrine cells. The ganglion cells of the GCL send their axons through the optic fiber layer (OFL) to the optic nerve.

It is important to notice three particular points, implicit in the above description.

(1) This structure is constituted of two kinds of cells: the interneurons (horizontal and amacrine cells), which aid in signal processing, and the informative neurons, which contain and transmit the information.

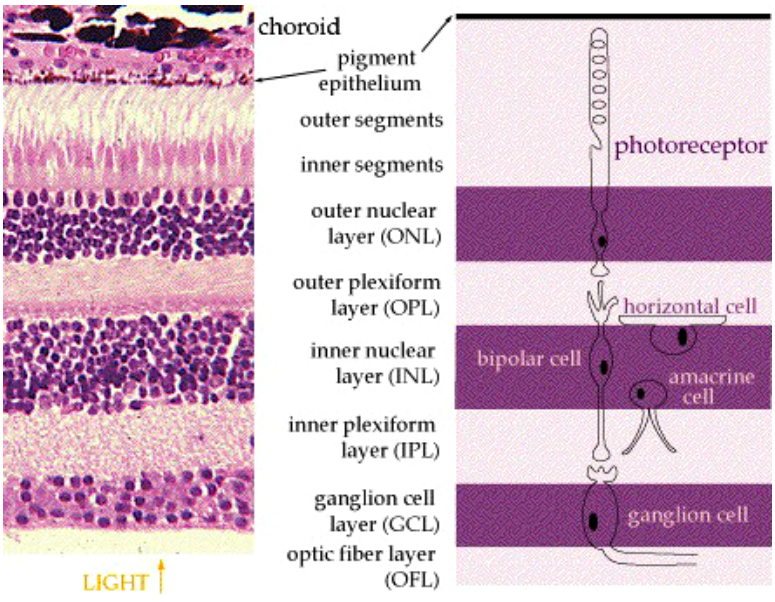


Fig. 1. A section of the retina: The architecture of layers.

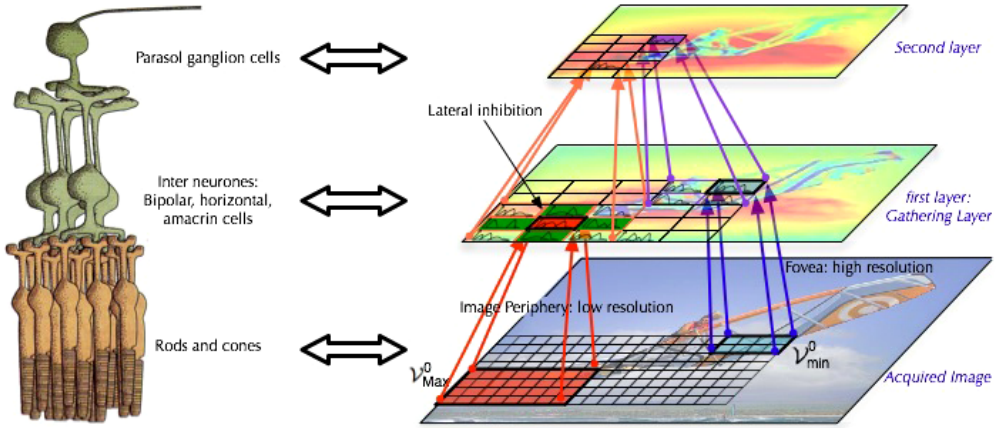


Fig. 2. The convergent aspect of retina neurons: The biological on the left and the proposed model on the right.

(2) As shown in Fig. 2, the amount of standard neurons decreases when closer to the optic nerve.
(3) Despite the reduction in the amount of data, the organization of the cells is not purely convergent. There is statistically one ganglionic cell for five bipolar cells and about twenty photoreceptors, but one bipolar cell is connected to more than one ganglionic cell. The structure is then overall convergent and locally divergent, which brings in information redundancy.

Equating the scene representation perceived by the eyes to a numerical image, the information contained in the photoreceptors is equivalent to pixels. This means that (in a view based approach), there are two other particular image representations: one at the INL and one at the GCL. The specific gathering of the information from one level to the next is due to the combination of the three items explained above. Therefore it appears that the information contained in a bipolar cell (of the INL) concerns the information of many photoreceptors (of the ONL), and in a similar manner, the information contained in a ganglion cell accounts for the information of many bipolar cells.

3. Foveolar Image: Spatial Geometry and Preprocessing

3.1. *Foveolar image decomposition*

Similar to the visual system of the mammals, the developed method takes into account the foveolar vision which can be introduced in two ways. The cylindrical development is given in Fig. 3(B). Starting from the original acquired image $I(i, j)$ with i, j expressing the coordinates of pixels [Fig. 3(A)], a cylindrical development

October 8, 2007 18:53 WSPC/179-JIN 00158

[Fig. 3(B)] remaps the image to a cylindrical image using polar coordinates as $I_c(r, \theta)$ (θ in degrees) with the relation,

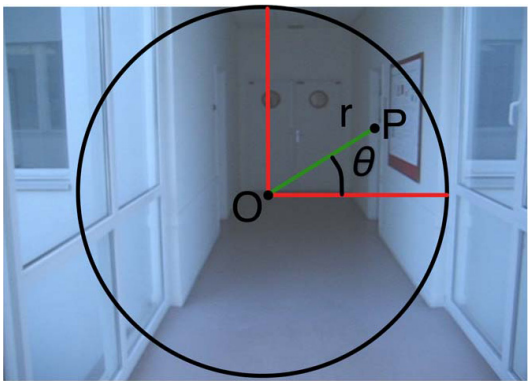
$$i = r\cos\theta \quad \text{and} \quad j = r\sin\theta, \tag{3.1}$$

where

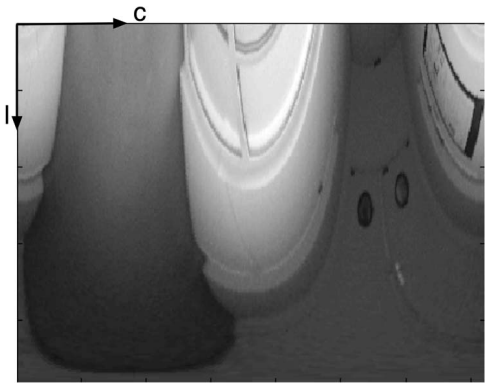
$$0 \leq r \leq r_{\max} \quad \text{and} \quad 0 \leq \theta \leq 360. \tag{3.2}$$

The size of I_c is then $r_{\max}$ rows, and the number of columns is set in this case to 360. The number can be chosen smaller if a thinner angular step is chosen. As shown in Fig. 3(B), each column of I_c is the corresponding radius of image I: to reduce information at the periphery of the image giving a foveolar distribution, each row of I_c is locally averaged:

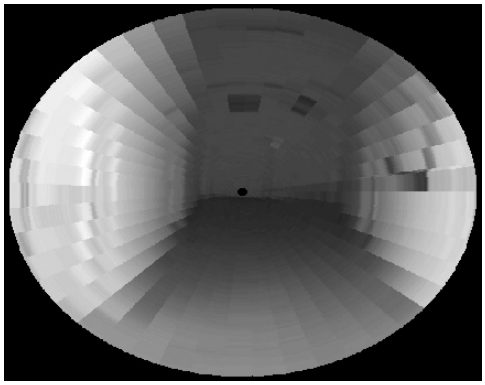
$$I_c(1 \leq r \leq r_{\max}, \theta) = \text{mean}(I_c(1 \leq r \leq r_{\max}, \theta - 2 \leq \theta \leq \theta - 2)). \tag{3.3}$$



(A) Original image with polar settings.



(B) Cylindric image of (A).



(C) Foveolar image of (A) from (B).

Fig. 3. Foveolar images generated by a cylindrical development, averaging, and inverse remapping.

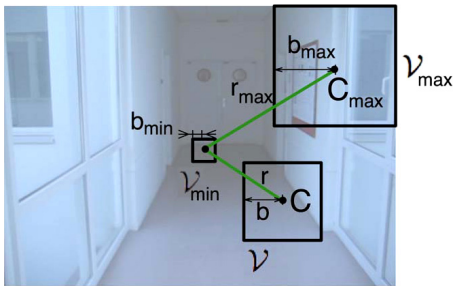


Fig. 4. Radius influence on the size of subimages to ensure the foveolar structure of the image.

Finally, an inverse mapping from polar to cartesian coordinates provides a foveolar image as shown in Fig. 3(C).

Another possible way to produce foveolar images is by merging contiguous variable sets of pixels using variable neighborhood windows as shown in Fig. 4. This technique is different from the previous one but has the property of being very close to the mechanisms involved within the retina preprocessing architecture.

Let $\mathcal{V}$ be a square subimage and b be the half-width of $\mathcal{V}$, as shown in Fig. 4. The bounds of b are $b_{\min}$ and $b_{\max}$ with $b_{\min} \leq b \leq b_{\max}$. It is essential that the values of b depend strictly on the distance between C (the center of $\mathcal{V}$) and the center of the whole image. This distance is actually the r component of the polar coordinate system.

The relation between b and r is then given by:

$$b = b_{\min} + \frac{b_{\max} - b_{\min}}{\log(2)} \log\left(\frac{r}{r_{\max}} + 1\right), \tag{3.4}$$

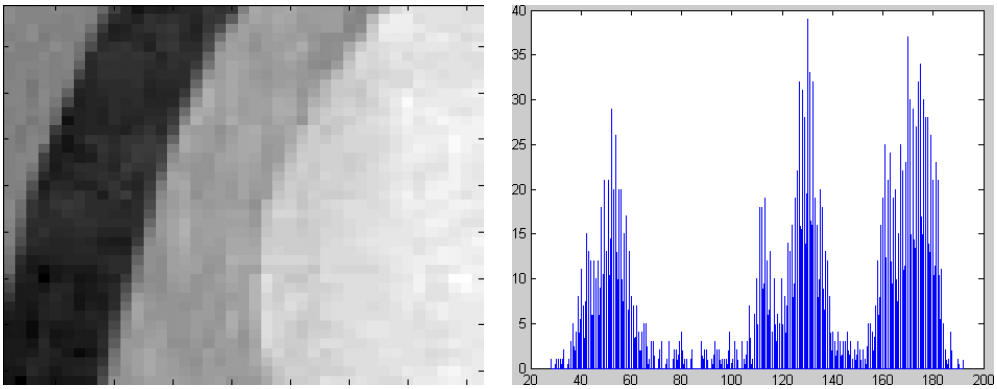
with $r_{\max}$ the upper bound of r.

The foveolar structure is implicit. The use of large sized subimages in the macula (at the periphery) and smaller size in the fovea (at the center) provides a (nonlinear) data amount reduction and a variable resolution.

3.1.1. *Image coding of foveolar areas*

Different kinds of visual features can be extracted from the foveolar image. The geometric features often used in the literature as angles, lines, corners can still be retrieved from foveolar images. Due to the nonlinear resolution, the accuracy of the detection can be altered significantly according to the location of the feature within the image.

The choice made here consists of using statistical tools (local histograms) inside foveolar areas in order to start an image coding and implicitly extract features as will be shown further.



(A) 40 × 40 subimage. (B) Image histogram.

Fig. 5. Example of a decomposition: A real subimage and its histogram.

3.1.2. *Gaussian decomposition of histograms*

Given the image presented in Fig. 5(A), its corresponding histogram is shown in Fig. 5(B).

In order to reduce the amount of information of the histogram (256 gray-levels which means a 256-dimension vector), it is decomposed as a Gaussian Mixture Model (GMM, [51, 52]) with expectation maximization algorithm (EM, [53]). This algorithm models a signal as a sum of normal distribution (ND), as described in Fig. 6.

Obviously, Fig. 5(B) shows that this subimage contains four classes (even though two of them are close). The decomposition is performed with a variable number of ND. The best match between the real histogram and the mixture of ND is shown in Fig. 6(D), i.e., corresponding to four ND. The bayesian information criterion (BIC [54]) is used to find the right number of ND to correctly characterize an histogram, which it is not possible to know *a priori*. This algorithm seeks a compromise between accuracy and model complexity depending on the number of parameters. This step is applied to each RGB-level of the color image, enabling the algorithm to keep color information.

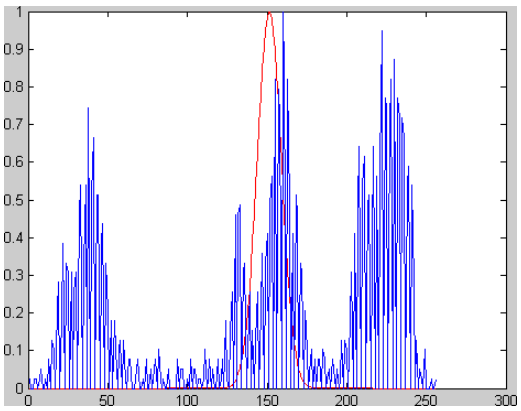
3.1.3. *Optimization of histogram representation*

Consider a subimage $\mathcal{V}$. After GMM decomposition, its information is defined as an histogram $H_{\mathcal{V}}(x)$:

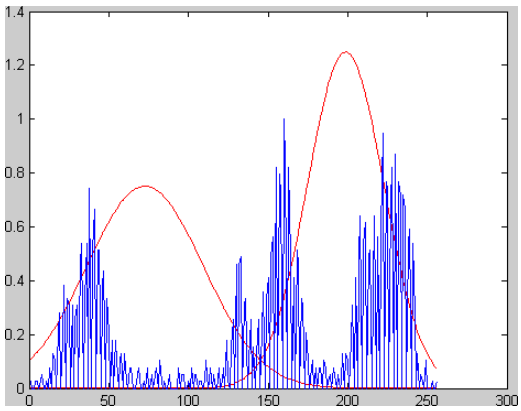
$$H_{\mathcal{V}}(x) = \sum_{n=1}^{Nbg} m_n \mathcal{N}_{(\mu_n,\sigma_n)}(x), \tag{3.5}$$

where $\mathcal{N}_{(\mu,\sigma)}(x)$ is the normal distribution whose standard deviation is σ and whose mean is μ, namely

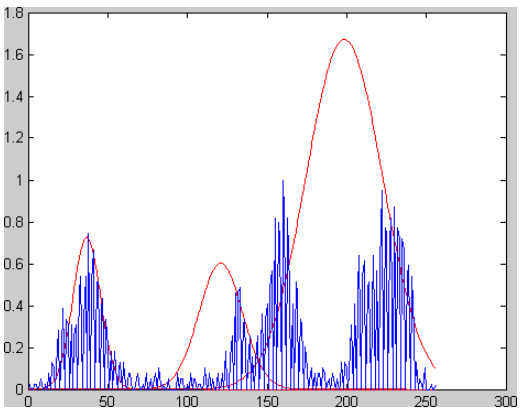
$$\mathcal{N}_{(\mu,\sigma)}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\frac{x-\mu}{\sigma})^2}. \tag{3.6}$$



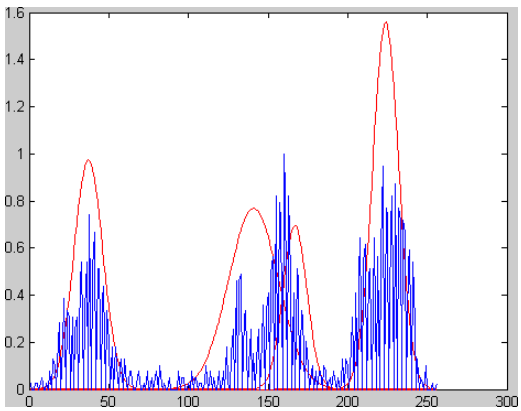
(A) 1 ND modelization.



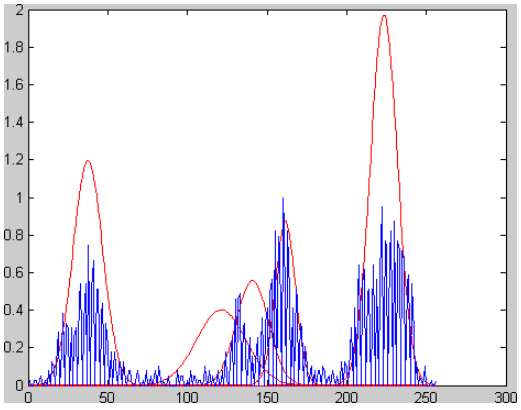
(B) 2 ND modelization.



(C) 3 ND modelization.



(D) 4 ND modelization.



(E) 5 ND modelization.

Fig. 6. Five different steps of the gaussian decomposition of an histogram.

In Eq. (3.6), m_n is the corresponding weight of the normal distribution n, and $H_{\mathcal{V}}(x)$ is composed by Nbg distributions. It is obvious that $0 \leq \mu \leq 255$ in the usual case of 256 gray-values. One ND fits with one class of pixels existing in the neighborhood $\mathcal{V}$. Therefore it is important to put together the similar classes (i.e., distribution with close means) and to keep aside distributions which are not representative (i.e., distributions whose weight are insignificant). Finally, the most representative NDs are sorted according to their weights.

3.2. *Multi-layer image coding*

3.2.1. *From acquired image to first-layer: The gathering layer*

In the following sections, $\mathcal{V}_i^j$ represents a subimage, with i corresponding to its index, and j the layer which it belongs to. The acquired image is set to be the initial layer corresponding to $j = 0$, with $\mathcal{V}^j$ implicitly representing the whole layer. Following a biological approach discussed in Sec. 2, and taking into account the different assumptions presented in the previous section (particularly the nonlinear resolution and the convergent/divergent), an implementation of the foveolar coding using specific layers is introduced. As shown in Fig. 7, subimages of $\mathcal{V}^0$ are gathered according to the foveolar decomposition. The cells of the gathering layer receive the information of variable size pixel sets, depending on the distance from the image center. Following the foveolar decomposition, a subimage $\mathcal{V}^0_{\max}$ located at the periphery encodes more pixels than the more central one $\mathcal{V}^0_{\min}$.

For a better understanding, assuming that the acquired image is taken from a linear camera (one rank of pixels), Fig. 8 shows a planar representation of the mechanism from the original image to the gathering layer. Four subimages are represented: $\mathcal{V}^0_{m_3}$, $\mathcal{V}^0_{m_2}$, $\mathcal{V}^0_{m_1}$ and $\mathcal{V}^0_{m_0}$. Their sizes are variable according to the foveolar

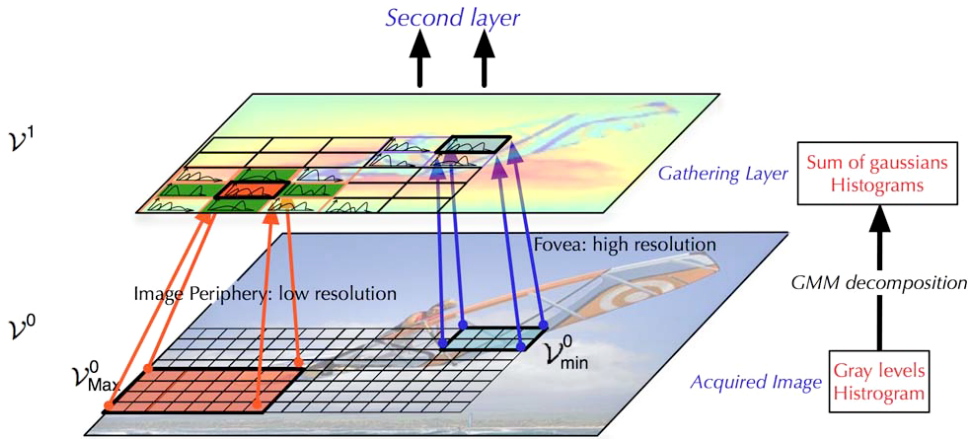


Fig. 7. Data reduction from an acquired image $\mathcal{V}^0$ to the first layer $\mathcal{V}^1$.

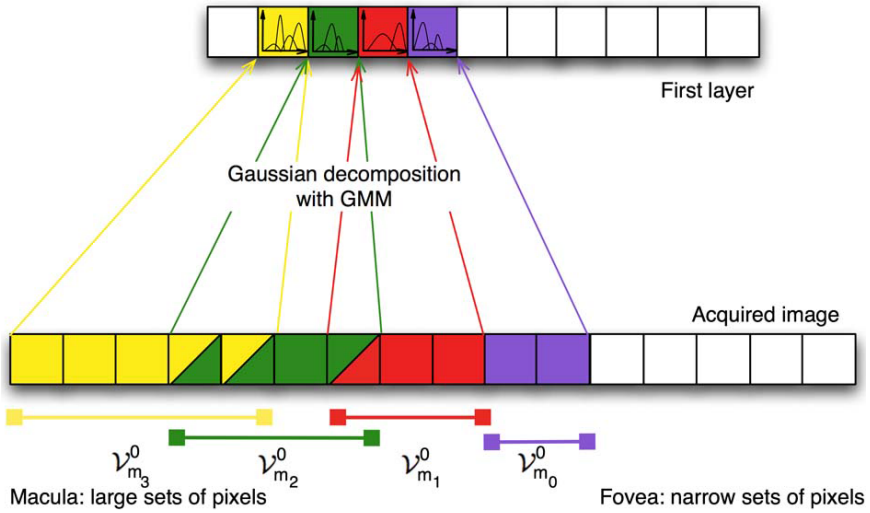


Fig. 8. Layer-1 data reduction equivalent to bipolar cells: Case of a linear camera. According to the foveolar decomposition, variable size subimages $\mathcal{V}^0_{m_3}$, $\mathcal{V}^0_{m_2}$, $\mathcal{V}^0_{m_1}$ and $\mathcal{V}^0_{m_0}$ enable to take into account the convergent/divergent properties.

decomposition. Two contiguous subimages have common pixels introducing redundancy into the system. Each subimage can be defined by its histogram. This histogram is then GMM-decomposed to provide the information of the cells of layer-1.

3.2.2. *From the gathering layer to the second layer*

As explained in Sec. 2, the retina structure consists of two layers of informative neurons, controlled by different kinds of interneurons. The system proposed here and described in Fig. 9 introduces also a second layer similar to the GCL, whose information is controlled by a process acting as interneurons.

These interneurons consider the information of many neurons to inhibit or facilitate the transmitted data to the next neuronal layer. As shown in Fig. 9, the GMM

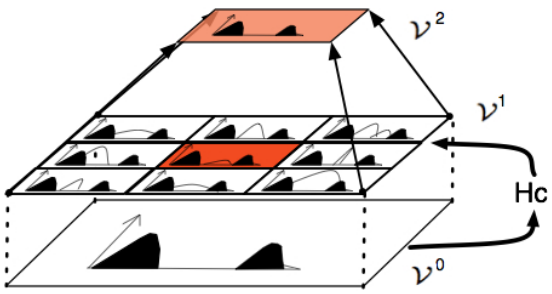


Fig. 9. Data reduction from layer-1 to layer-2: The common normal distribution are only the two in black. After the control process, they are the only ones kept in the higher level.

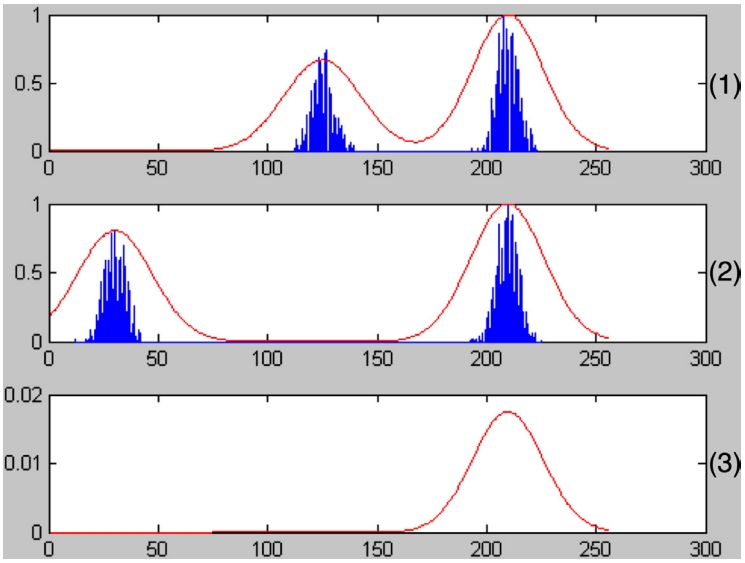


Fig. 10. An example of a histograms processing in layer-2. Histograms (1) and (2) are merged using (2) as a control information to provide (3) that contains only the ND that they both have in common.

histogram of the subimage of $\mathcal{V}^0$ corresponding to the original image of the neighborhood of $\mathcal{V}^1$ is used as a control to inhibit or ease the transmitted data to the next layer. The information is stored in the GMM-decomposed H_c. The set of histograms and the control ones share the shaded ND, therefore only these two ND are transmitted to the following layer.

A theoretical example of this control is presented in Fig. 10. To illustrate the principle, two histograms (1) and (2) are considered. The ND of (1) are controlled by the ND of (2). The result (3) shows that the common ND centred around a mean value of 200 will be selected.

4. Multi-Layer Decomposition: Mathematical and Visual Results

4.1. *Mathematical results*

Aiming to obtain a functional model, we will set in this section the mathematical framework of the decomposition operated by layer-2 in order to give a predictable model and provide a better understanding of the process.

4.1.1. *Comparison using Bhattacharyya proximity*

Let X and Y be two normalized vectors of same size:

$$\text{card}(X) = \text{card}(Y) \quad \text{and} \quad \sum_i X(i) = \sum_i Y(i) = 1. \tag{4.7}$$

The Bhattacharyya distance [20] is given as:

$$\mathcal{D}_{\mathcal{B}}(X,Y) = 1 - \sum_i \sqrt{X(i).Y(i)}. \tag{4.8}$$

It is easy to verify that $\mathcal{D}_{\mathcal{B}} = 0$ in the case of two identical vectors ($X(i) = Y(i), \forall i$, so $\mathcal{D}_{\mathcal{B}} = 1 - \sum_i \sqrt{(X(i))^2} = 0$ from Eq. (4.7)), and $\mathcal{D}_{\mathcal{B}} = 1$ in the case of two orthogonal vectors. Hence,

$$0 \leq \mathcal{D}_{\mathcal{B}}(X,Y) \leq 1. \tag{4.9}$$

To characterize the similarity instead of dissimilarity, Bhattacharyya proximity is introduced as

$$\mathcal{P}_{\mathcal{B}}(X,Y) = 1 - \mathcal{D}_{\mathcal{B}} = \sum_i \sqrt{X(i).Y(i)}, \tag{4.10}$$

with

$$0 \leq \mathcal{P}_{\mathcal{B}} \leq 1. \tag{4.11}$$

Before transmitting its own information (histogram of few normal distributions) to the next layer, each normal distribution is weighted by setting its Bhattacharyya proximity coefficient.

4.1.2. *Mathematical expression of the encoding of layer-2*

The information contained in an informative neuron of the second layer can be expressed as:

$$H_i(x) = \sum_{n=1}^{Nbg} m_{i_n} \mathcal{N}_{(\mu_{i_n}, \sigma_{i_n})}(x), \quad i = 2, \tag{4.12}$$

where i is the layer index, n is the index of the ND and Nbg is their total number.

The question is now how to express the layer-2 parameters as functions of the first layer parameters. Modifiying the influence of a Gaussian histogram does not mean changing the averages and standard deviations but rather the weights of the normal distribution. The information of a layer-2 neuron is linked to a layer-1 neighborhood. Let $\mathcal{V}_1^1$ be a neighborhood of layer-1. We then set:

$$\mu_{2_n} = \frac{\sum_{\mathcal{V}_1^1} \mu_{1_n}}{\text{card}(\mathcal{V}_1^1)} \tag{4.13}$$

and

$$\sigma_{2_n} = \frac{\sum_{\mathcal{V}_1^1} \sigma_{2_n}}{\text{card}(\mathcal{V}_1^1)}. \tag{4.14}$$

The weights m_2 are provided by the Bhattacharyya proximity:

$$m_{2_n} = \sum_{\mathcal{V}_1} \mathcal{P}_{\mathcal{B}}(m_{1_n} \mathcal{N}_{(\mu_{2_n}, \sigma_{2_n})}(x), H_c(x)) \tag{4.15}$$

with $H_c(x)$ defined in Eq. (4.12):

$$H_c(x) = \sum_{n=1}^{Nbg} m_{c_n} \mathcal{N}_{(\mu_{c_n}, \sigma_{c_n})}(x) \tag{4.16}$$

(c being the reserved dash for control parameters) and replacing each term, it comes that:

$$m_{2_n} = \sum_{\mathcal{V}_1} \mathcal{P}_{\mathcal{B}}(m_{1_n} \mathcal{N}_{(\mu_{2_n}, \sigma_{2_n})}(x), m_{c_n} \mathcal{N}_{(\mu_{c_n}, \sigma_{c_n})}(x)). \tag{4.17}$$

We set

$$P = \mathcal{P}_{\mathcal{B}}(m_{1_n} \mathcal{N}_{(\mu_{2_n}, \sigma_{2_n})}(x), m_{c_n} \mathcal{N}_{(\mu_{c_n}, \sigma_{c_n})}(x)). \tag{4.18}$$

Moreover, these expressions are valid for each normal distribution of the histogram, therefore the dash n will be voluntarily missing (to simplify notation):

$$P = \int_{\mathbb{R}} \sqrt{\frac{m_1}{\sigma_2\sqrt{2\pi}} e^{-\left(\frac{x-\mu_2}{\sigma_2}\right)^2} \frac{m_c}{\sigma_c\sqrt{2\pi}} e^{-\left(\frac{x-\mu_c}{\sigma_c}\right)^2}} \mathrm{d}x \tag{4.19}$$

$$P = K \int_{\mathbb{R}} e^{-\frac{1}{2}\left(\frac{x-\mu_2}{\sigma_2}\right)^2 - \frac{1}{2}\left(\frac{x-\mu_c}{\sigma_c}\right)^2} \mathrm{d}x \tag{4.20}$$

with

$$K = \sqrt{\frac{m_1 m_c}{2\pi\sigma_2\sigma_c}}. \tag{4.21}$$

The numerator of the exponential function is quadratic then it can be written as

$$P = K \int_{\mathbb{R}} e^{-(\alpha(x-\beta)^2+\gamma)} \mathrm{d}x = K' \int_{\mathbb{R}} e^{-\alpha(x-\beta)^2} \mathrm{d}x \tag{4.22}$$

with

$$\alpha = \frac{1}{2}\left(\frac{1}{\sigma_2^2} + \frac{1}{\sigma_c^2}\right), \quad \beta = \frac{(\mu_c\sigma_2^2 + \mu_2\sigma_c^2)}{\sigma_c^2 + \sigma_2^2}, \quad K' = Ke^{-\gamma} \tag{4.23}$$

and also

$$\gamma = \frac{1}{\sigma_2^2\sigma_c^2}\left((\mu_c\sigma_2)^2 + (\mu_2\sigma_c)^2 - \frac{1}{2}\frac{(\mu_c\sigma_2^2 + \mu_2\sigma_c^2)^2}{\sigma_c^2 + \sigma_2^2}\right). \tag{4.24}$$

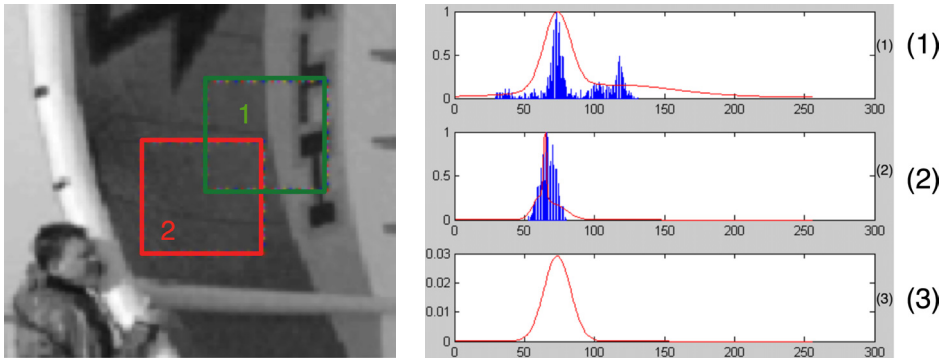
Considering the result:

$$\int_{\mathbb{R}} e^{-x^2} \mathrm{d}x = \sqrt{\pi}, \tag{4.25}$$

after the variable substitution $X = \sqrt{\alpha}(x - \beta)$, P in Eq. (4.22) becomes

$$P = K'\sqrt{\frac{\pi}{\alpha}}. \tag{4.26}$$

This last equation means that the information of a layer-2 neuron is completely computable and predictable from the values of layer-1. To illustrate the explanation



(A) Image and histogram areas.

(B) Corresponding histograms (test (1) and control (2)), and results of layer-2 coding (3).

Fig. 11. An example of a histogram coding of layer-2: Case of a real gray-level image. Fig. 11(B) shows the envelops of the entire histograms being the sum of all ND extracted from histograms.

of Sec. 3.2 and the theoretical example of Fig. 10, a real example of this control is shown in the following section.

4.1.3. *Example of layer-2 encoding*

The real histogram Fig. 11(B)(1) corresponding to the area 1 in Fig. 11(A) is defined as a sum of ND according to the decomposition introduced in Eq. (4.12). The gaussian parameters of each ND of this histogram are presented in Table 1.

The parameters of the histogram of Fig. 11(B)(2) corresponding to area 2 in Fig. 11(A) are given in Table 2.

The new weights provided by the Eq. (4.26) are computed and are given in Table 3. These weights are expressed on the normalized histogram shown in Fig. 11(B)(3).

Table 1. Normal distribution parameters of Fig. 11(B)(1) histogram.

n	m_{2_n}	μ_{2_n}	σ_{2_n}
1	1	70	10
2	0, 5	120	10
3	0, 1	100	8

Table 2. Normal distribution parameters of Fig. 11(B)(2) histogram.

n	m_{c_n}	μ_{c_n}	σ_{c_n}
1	1	60	2
2	0, 5	70	12

Table 3. Layer-2 weights calculation.

n	P	P Normalized
1	$0,34$	1
2	$0,035$	0,103
3	$1,7.10^{-6}$	0

4.2. *Multi-layer image processing*

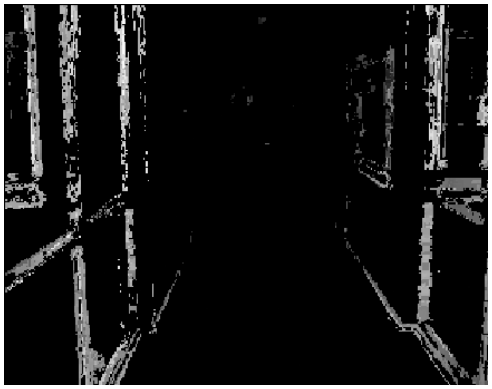
4.2.1. *First layer: Scene decomposition*

Figure 12 shows the decomposition of the image appearing in Fig. 12(A) representing an indoor scene. Each processed image 12(B), 12(C), 12(D), corresponds to the ND averages of respectively the first (1-ND), the second (2-ND) and the third (3-ND) ND provided by the GMM. We notice that the first ND image [Fig. 12(B)] is very

(A) Original image before processing.

(B) Mean image of 1-ND.

(C) Mean image of 2-ND.



(D) Mean image of 3-ND.

Fig. 12. The mean images of the first four must weighted normal distribution of a GMM decomposed image in layer-1.

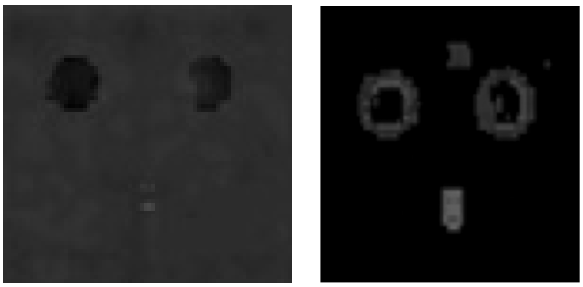
October 8, 2007 18:53 WSPC/179-JIN 00158

close to the original image as most details can still be perceived even with the decrease of resolution. In the second ND image [Fig. 12(C)], uniform areas are set to zero. Actually this second image provides elementary edges. The third and fourth gaussian images provide more and more complex images regions corresponding to three or four (at least) pixel classes. These ND images can correspond in some cases to textured areas. It is important to notice that the whole visual features seen in Fig. 12 are still complex and nonaccurate. The edges defined by a wide variety of gray-levels make them difficult to use at this step.

The influence of the foveolar decomposition can be observed in Fig. 13, where (1) and (2) are zoomed details located at the center and periphery of the image. Figures 14 and 15 are the corresponding 1-ND and 2-ND images. If the processes were linear and not foveolar, their coding would have been equivalent. However the actual codings are very different. The door edges are detected because it is encoded using subimages close to the size of $\mathcal{V}^0_{\min}$ (defined in Sec. 3.1) as shown in Fig. 14.

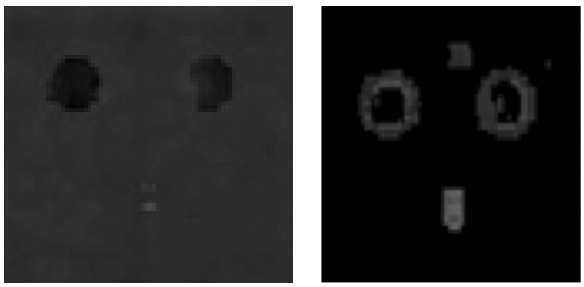
In this case, color information is also not lost. Figure 15 shows the opposite case, in which edges are not preserved. Still, color information is not completely lost. Figure 15(A) does not show the window handle because its white pixels are not the

Fig. 13. Zooms of particular zones of the scene.



(A) Mean image of 1-ND. (B) Mean image of 2-ND.

Fig. 14. ND images comparison of centred details.



(A) Mean image of 1-ND. (B) Mean image of 2-ND.

Fig. 15. ND images comparison of peripheric details.

most representative in comparison to more gray pixels in this area, but the white color is the second most important ND as shown in Fig. 15(B).

4.2.2. *Second layer: Scene segmentation*

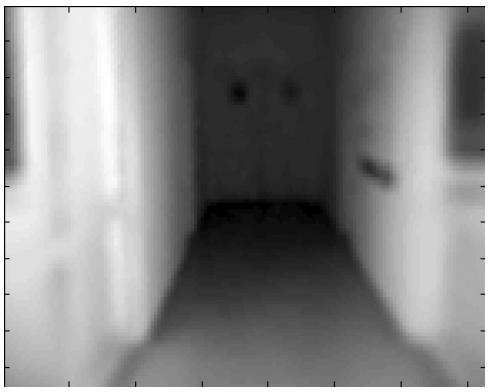
After the layer-2 reduction data, the information contained by the informative cells are shown in Fig. 16 where the control influence can be observed. Comparing all images of Figs. 12 and 16, the images of Figs. 16(B) and 16(C) look indeed like images of Fig. 12, but the gray level of edge portions are more homogeneous. Moreover, the edges seem to be simpler, with filled gaps. The data reduction combined to the foveolar decomposition fill up the textured areas. This means that using the image of Fig. 16(B) as a mask for the original image provides an automatic segmentation of the interesting scene areas (shown Fig. 17).

5. Visual Information in LGN

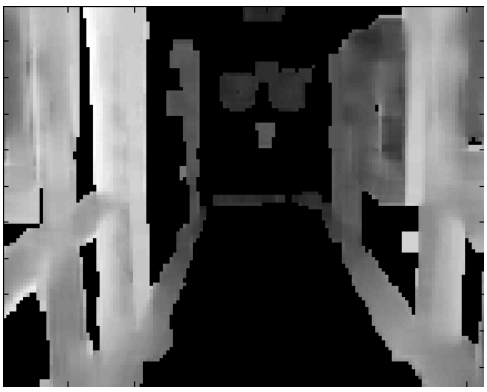
5.1. *Biological LGN architecture*

The main function of both LGN (left and right) is to transmit stimuli from the sensory systems to cortex. The relay cells in LGN receive visual signals from retinal ganglion cells and transmit processed information to the primary visual cortex. The LGN has a laminar (6-layer) architecture. Layer 1 and 2 are composed of magnocellular cells (M1 and M2), whereas layer 3 to 6 are composed of parvocellular cells (P3 to P6). The eye on the same side of one LGN sends information to layers 2, 3 and 5 while the eye on the opposite side sends information to layers 1, 4 and 6 [1].

In the human brain, little is known about the functional anatomy and response properties of subcortical nuclei containing visual maps such as the LGN. Its retinotopic structure has been widely studied using functional magnetic resonance imaging (fMRI) [14, 19] or using positron emission tomography [16], and it supposes that visual data are organized as images.



(A) Image of must weighted ND.



(B) Image of second must weighted ND.

(C) Image of third must weighted ND.

Fig. 16. The mean images of the first four ND of second layer-2.

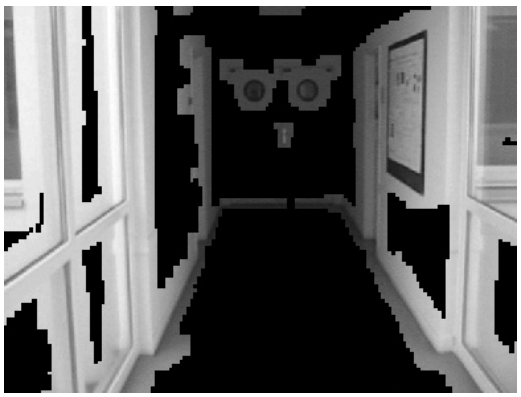


Fig. 17. Automatic segmentation of discriminant scene areas.

5.2. *LGN visual modeling from retina's multi-layer processing input*

Very few works deal with the visual representation of data in LGN. Original stimuli-images have been constructed from LGN monitoring [12], but no direct visual representation of the information contained in the LGN has been given. Each codified image using the method exposed in Eq. (3.2) can be represented by a set of ND-images. As shown in Fig. 18, each LGN-layer image is a ND-image generated by the retinal decomposition. It is known that each LGN layer is divided into

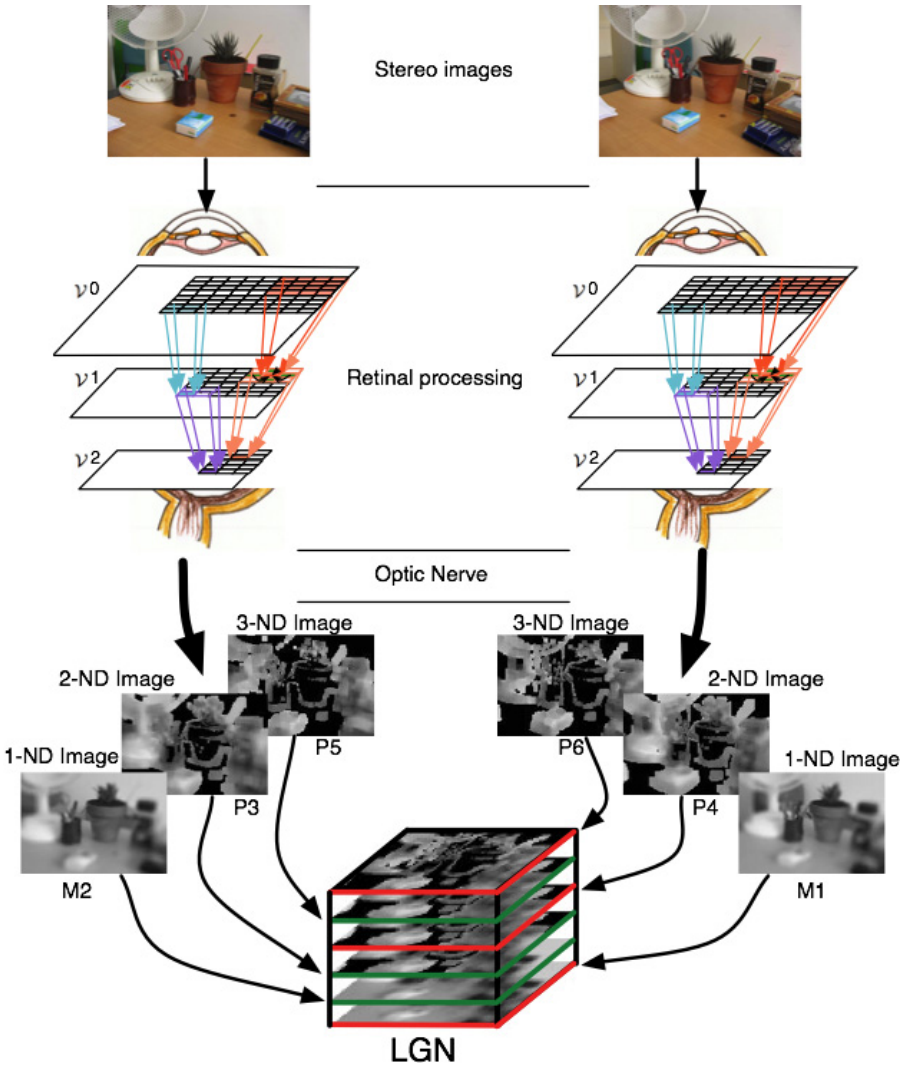


Fig. 18. Visual Information transmission from eyes to LGN: From a stereo pair of images, the multi-layer retinal encoding provides to the LGN 3 ND-images for each eye which contain the entire original images, edges and textures. This visual representation is already composed of efficient features for rapid localization tasks.

two sublayers [19] which means that a particular features in the visual space is characterized with three pairs of data per eye. We may notice that the introduced model fits particularly well with this organization because since each ND is regulated by two parameters: its mean value, and its standard deviation.

5.3. *Noise independence and information content*

5.3.1. *Gray level influence over noise reduction*

As introduced in Sec. 1, one of the roles of the retina neural coding seems to be the noise reduction [40, 47–49]. A polar striped image of increasing gray levels (Fig. 19) has been used to study the influence of noise. Three kinds of noise are used: gaussian additive [AG — Fig. 19(B)], Salt and Pepper [SP — Fig. 19(C)], and multiplicative [MU — Fig. 19(D)]. Layer decomposition are applied to each image.

For each stripe corresponding to a particular gray-level, it is possible to compute the standard deviation of the noise, using the difference between the original and the noisy image. This process provides us the curves of Figs. 20 and 21. Figure 20 shows that before processing, all the standard deviations are very different. SP noise decreases with the gray-level, AG noise is almost constant, and MU noise increases. Moreover, each standard deviation mean value differs from the others. Figure 21 shows that all these differences disappear after processing the image: standard deviation mean value are almost the same, and the different curves get nearly the same shape.

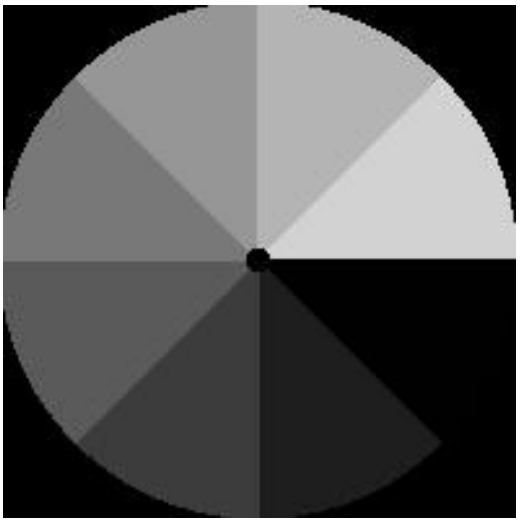
Table 4 shows the standard deviation mean values for each type of noise. The right column displays the noise reduction rate.

It is important to notice that the noise standard deviation decreases very strongly for all noises (see ordinate axis units of Figs. 20 and 21).

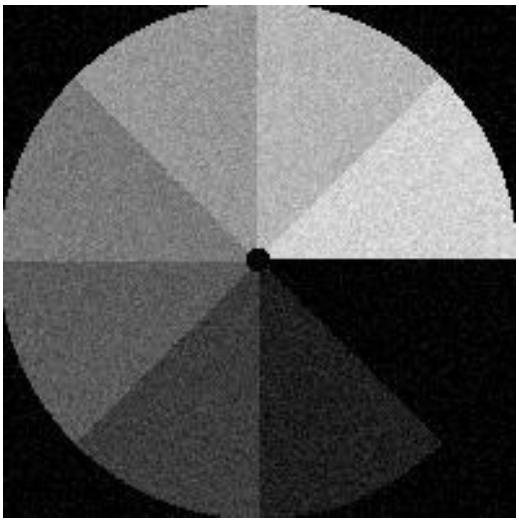
5.3.2. *Noise reduction for a real case*

The issue of noise reduction has been tested on images of Fig. 22. Figure 22(A) is the original image, whereas Fig. 22(B) contains additional gaussian noise whose standard deviation is 10. Both images have been processed by the decomposition/encoding exposed in the previous section, and the LGN-images M1 are respectively shown in Fig. 23.

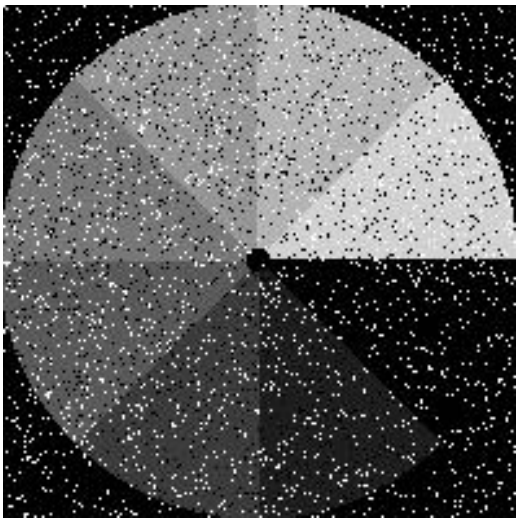
One can see that both result images are very close, which means that the noise seems to have been filtered out. This observation can be illustrated and confirmed by Fig. 24. Noise image of Fig. 24(A) is the difference between Figs. 22(A) and 22(B), and noise image of Fig. 24(B) is the difference between Figs. 23(A) and 23(B). Also, Fig. 25 represents the histogram's superposition of both noisy images, before and after encoding, in (A) the original image, in (B) after encoding. The standard deviation of the noise decreased from 9.3 to 1.5 (84%).



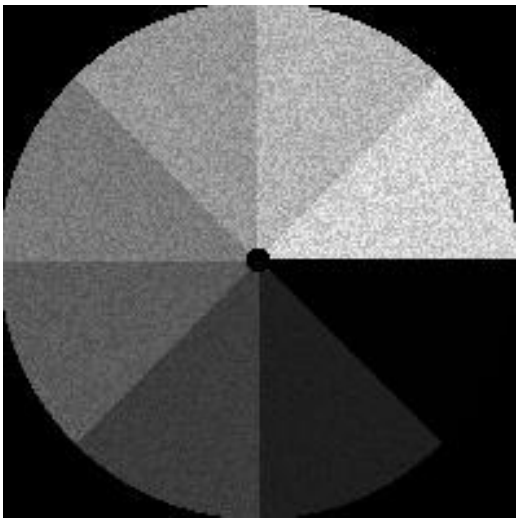
(A) Original polar image of variable gray-levels.



(B) Image noised by an additive gaussian (AG) noise of 20.



(C) Image noised by a salt and pepper (SP) noise of 10.



(D) Image noised by a multiplicative noise (MN) of 30%.

Fig. 19. Original and noisy polar images.

5.3.3. *Information content*

It is interesting to study the impact of the method on the amount of information contained in both original foveolar image [Fig. 26(A)] and the processed one [Fig. 26(B)]. In order to extract the influence of the foveolar gathering, an entropy measure is computed on 18 annular rings R_k, $k \in [1, 18]$ following a log-polar coverage as shown

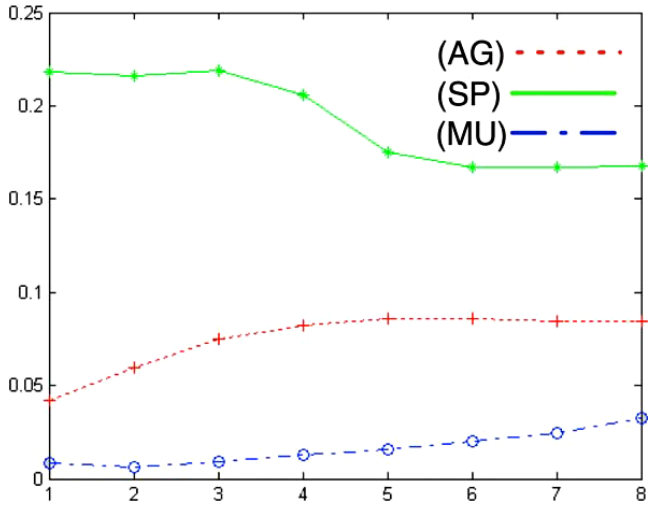


Fig. 20. Standard deviation of original noised images according to gray values. The maximal value of ordinate axis is $0,22$.

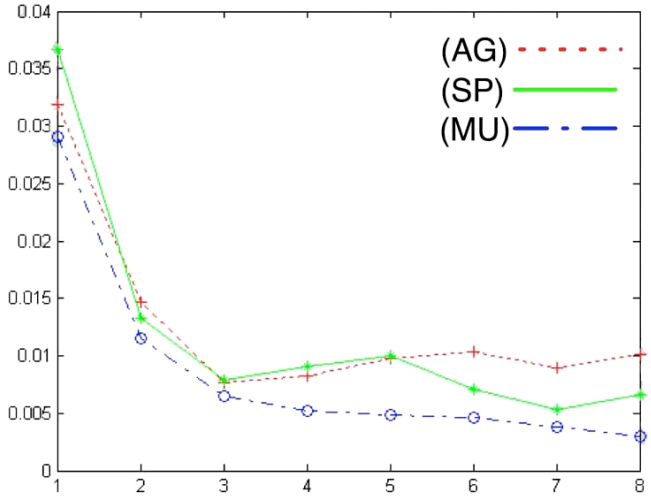


Fig. 21. Standard deviation of processed noised images at layer-2 according to gray values. The maximal value of ordinate axis is $0,037$.

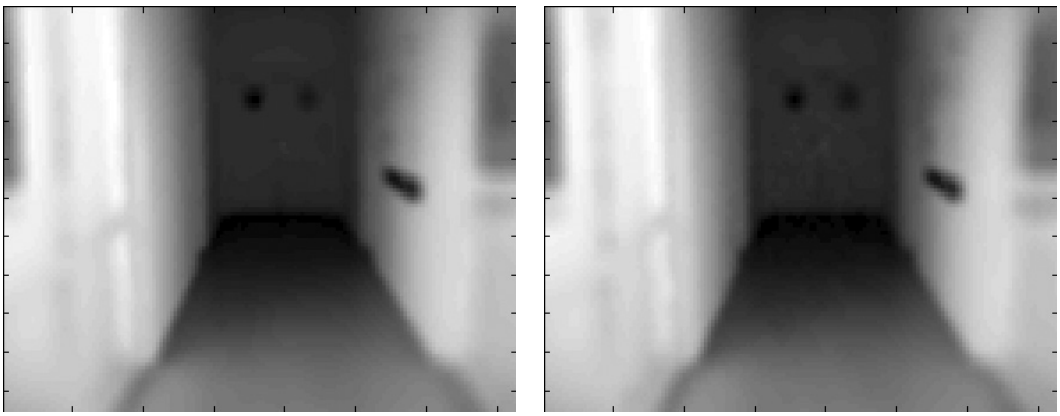
Table 4. Mean value of each noise standard deviation.

	Before Coding	After Coding	Noise Reduction Rate
AG	$0,0750$	$0,0127$	83%
SP	$0,1920$	$0,0120$	94%
MU	$0,0162$	$0,0086$	47%

(A) Original image.

(B) Original image with a additional gaussian noise of 10.

Fig. 22. Original image and original + noise.



(A) LGN-image M1 of 22(A).

(B) LGN-image M1 of 22(B).

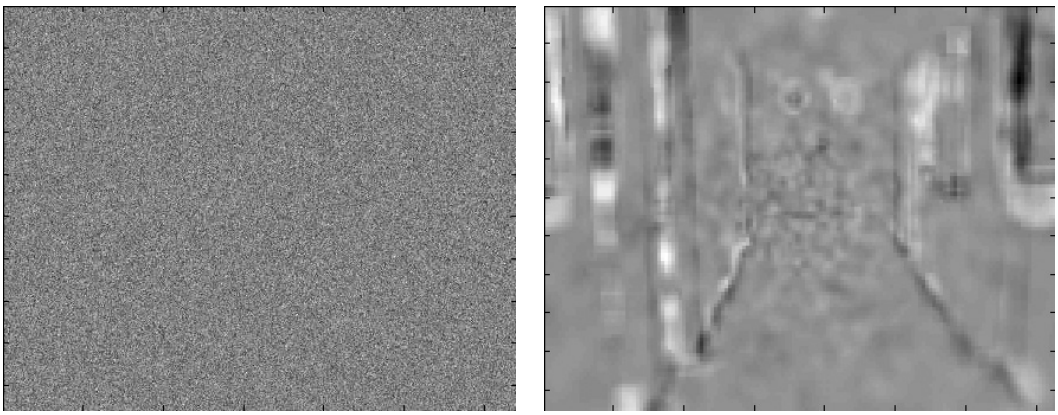
Fig. 23. Non-noisy and noisy result images.

in Fig. 26(A). This quantity of information is given for a ring R_k by:

$$E(R_k) = -\sum_{c=0}^{c=255} Occ(R_k = c) \log P(R_k = c), \tag{5.27}$$

with $Occ(R_k = c)$ the occurrence of c in R_k and $P(c)$ is the probability of appearance of the gray value c within R_k. The entropy values for the whole image [considering I instead of R_k in Eq. (5.27)] is presented in Table 5. It can be noticed that the information content value of the original image and the 1-ND image are very close with a difference of 0.5%. The results using rings R_k are given in Fig. 26(C).

October 8, 2007 18:53 WSPC/179-JIN 00158



(A) Difference image of Figs. 22(A) and 22(B). (B) Difference image of Figs. 23(A) and 23(B).

Fig. 24. Noise images.

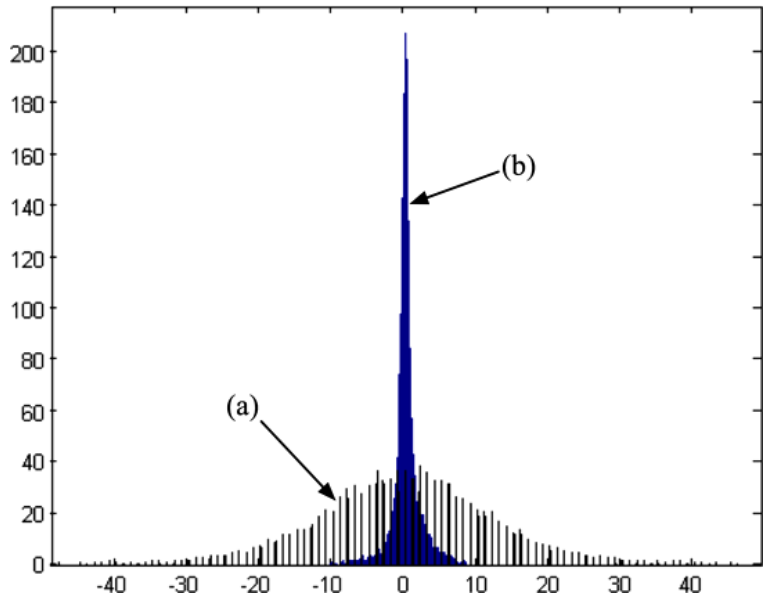


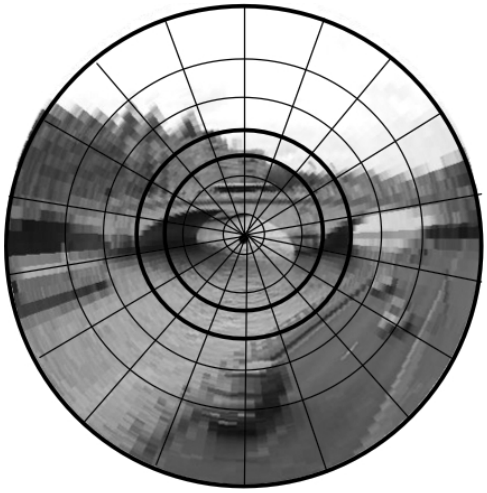
Fig. 25. Histogram superposition of residual noises: (A) for Figs. 22(B) and (B) for Fig. 23(B).

A very large amount of information is conserved after processing, it is mostly located within the 1-ND images.

6. Spatial Orientation and Scene Recognition

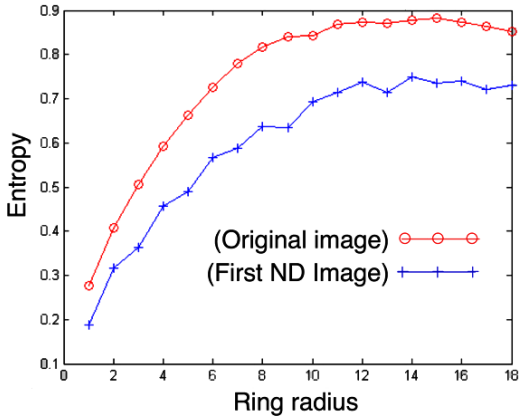
6.1. *The issue of fast localization tasks*

The head-direction (HD) cells found in the brain of freely moving rats have remarkable properties. They signal the instantaneous head direction of the animal in the



(A) Foveolar mapping: (480 × 480 pixels).

(B) LGN-image M1: (117 × 117 pixels).



(C) Images' entropy.

Fig. 26. Original and codified images and the entropy of radius-variable size ring images.

Table 5. Entropy values.

Original	1-ND	2-ND	3-ND	4-ND
7, 12	7, 08	5, 57	3, 32	1, 01

horizontal plane regardless of the location of the animal in the environment. This system has "cognitive" aspects as it can use familiar landmarks to reset or calibrate the internal representation of the direction body of the animal [35]. Even if the actual input-output characteristics of the HD cells are still partially unknown, LGN have been mentioned as one of the most potential input [34]. Neural network

simulations predict that after changes in the visual scene, the firing rates of the newly activated cells reach their maximum at latencies on the order of several hundreds of milliseconds [34]. Whether this activity profile responds to rapid visual reorientation by traveling toward a new preferred direction or by jumping abruptly to it has been unresolved until now [36]. Recent studies about rapid transient response (80 ± 10 msec) for a 90 reorienting signal have supported the abrupt shift model [8]. As early visual areas, LGN seem able to transmit information very quickly to HD cells, forming a rapid localization processing pathway. In this case, LGN must already contain efficient and reliable features allowing HD cells to perform quick reorientation takes.

A simple localization algorithm using the retinal coding and visual LGN representation described above will show that the features provided here are enough to perform these kinds of tasks. The localization algorithm chosen is inspired by computer vision. It does not pretend to be a biological actual model, even though it takes into account some spatial orientation issues specifications and particularly the directional field of view.

6.2. *Spatial localization*

At this stage, we now consider using the information provided by the presented visual encoding system to perform spatial localization. The aim is to show that the retrieved information is sufficient to feed a localization algorithm. In general, if the correct information is extracted from scenes, then the algorithm does not need to be very complex as the features are discriminant enough.

Consider a sequence of images describing different locations. This sequence has been processed by the two layers of the system. Considering color images, the encoding algorithm has been applied to each color plane. The learning of a particular scene is performed using a 3D color statistical distribution. Starting from Eq. (4.12), each color of the second layer $I_{2\gamma}$, $\gamma = \{R, G, B\}$ is described as a set of normal distributions. For each RGB component:

$$I_{2\gamma} = \sum_{n=1}^{Nbg} m_{2\gamma_n} \mathcal{N}_{(\mu_{2\gamma_n}, \sigma_{2\gamma_n})}(x), \quad \gamma = \{R, G, B\}, \quad \text{with} \tag{6.28}$$

$$I_2 = (I_{2R}, I_{2G}, I_{2B}). \tag{6.29}$$

Let $\mathbf{S}$ be the volume of coordinates (i, j, k) representing the three levels of color RGB contained in the image areas expressed by more than two ND, (i.e., nonuniform areas). As shown in Fig. 27, the three color planes of an acquired color scene image 27(A) are encoded and given in Fig. 27(B). Each point of the 3-plane contains at most 4-ND and at least 2-ND. Information extraction is encoded by the three gray-level values, one for each RGB color component and for each weighted ND of the area. The color signature $\mathbf{S}(i, j, k)$ is then constructed [Fig. 27(C)] as a volume:

$$\mathbf{S}(i, j, k) = \text{card}((l, c) \in \mathbb{R}^2, I_{2R}(l, c) = i, I_{2G}(l, c) = j, I_{2B}(l, c) = k) \tag{6.30}$$

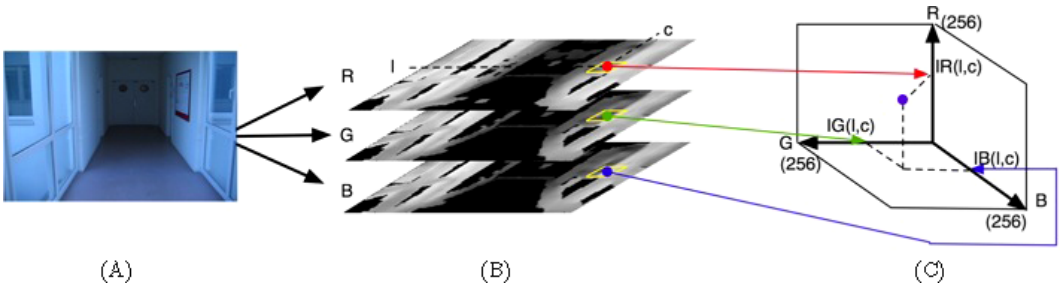


Fig. 27. Construction of the color signature of the object, example for one patch of the image.

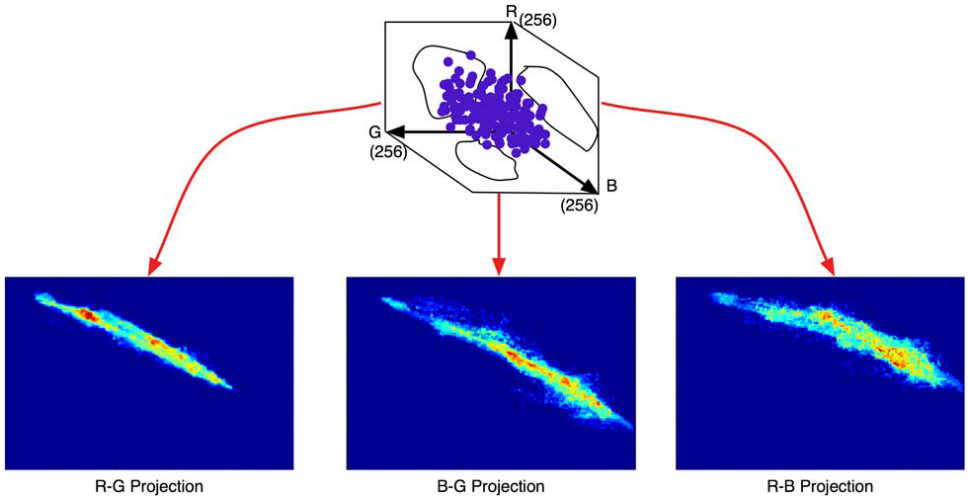


Fig. 28. Construction of the color signature of the object, global signature and projection on each two-colors plane.

where (l, c) are the coordinates in the image space, and:

$$0 \leq i, j, \quad k \leq 255. \tag{6.31}$$

This construction is applied to all images, providing a global signature presented in Fig. 28. It is then possible to project this volume on the three bi-color planes, introducing a new characterization of the spatial location. The signature of a particular scene O is noted $\mathbf{S}_O$, and the three signature projections on the 3-color planes R-G, G-B, and B-R are denoted, respectively, by $P_{(1,O)}$, $P_{(2,O)}$, $P_{(3,O)}$.

6.3. *Scene recognition using color-plane residual error*

It is possible to estimate the efficiency of the extracted features using a simple process. As previously shown in Fig. 28, it is possible to characterize a scene or an object by their two-color planes, as these planes are the statistical distribution of pixels within the color space RGB. Two locations L_m and L_n can then be compared based

on from the respective 3-plane projections thanks to the residual error $\Delta(L_m, L_n)$ defined as:

$$\Delta(L_m, L_n) = \sum_{\gamma=1}^{3} \left(\sum_i \sum_j |P_{(\gamma,L_m)}(i,j) - P_{(\gamma,L_n)}(i,j)| \right). \tag{6.32}$$

where $P_{(\gamma,L_m)}(i,j)$, $\gamma \in [1,3]$ is the value of the color plane RGB of the object L_m on the coordinates (i,j).

6.4. *Results for real sequences*

An eye-viewpoint sequence of a man walking through corridors, gallery, stairs, along a trajectory is shown in Fig. 29. N different locations characterize this sequence (here, $N = 16$), a location is defined using three successive images every fifteen images. A scene recognition using color-plane residual error is tested on this sequence with 76 randomly chosen images. Let F_k be the frame k of this sequence, k is randomly chosen, and L_i, $i \in [1 \leq i \leq N]$ the N learned locations. Ground truth assumes that,

$$F_k \in [L_n, L_{n+1}]. \tag{6.33}$$

Let L_n be called the lower bound and L_{n+1} the higher bound. Let R_1 be the first responding reference, that is to say, the location whose residual error to F_k is the lowest. In mathematical terms:

$$R_1 = \{L_p, \Delta(F_k, L_p) = \min(\Delta(F_k, L_i),\ 1 \leq i \leq N)\}. \tag{6.34}$$

In the same way, R_2 is the second responding reference. There is in fact four cases of recognized objects:

(1) The first responding reference is L_n and the second is L_{n+1}: $(R_1 = L_n) \cap (R_2 = L_{n+1})$.
(2) The first responding reference is L_n and the second is not L_{n+1}: $(R_1 = L_n) \cap (R_2 \neq L_{n+1})$.

Fig. 29. Validation sequence, each image shown here is a reference learned locations.

Table 6. Scene recognizing results for a real indoor/outdoor sequence.

Errors	Case 1	Case 2	Case 3	Case 4	Total	Rate
12/76	12	36	10	6	64/76	84%

(3) The first responding reference is L_{n+1} and the second is L_n: $(R_1 = L_{n+1}) \cap (R_2 = L_n)$.
(4) The first responding reference is L_{n+1} and the second is not L_n: $(R_1 = L_{n+1}) \cap (R_2 \neq L_n)$.

It can be noticed that in this kind of sequence, due to the directional field of view, a well classified scene is not necessarily represented by its R_1 response as the camera might be spatially close to R_1 but facing R_2. In this case R_2 is visually the best response and the correct answer.

The results for the sequence are shown in Table 6. It can be noticed in Fig. 29 that some learned locations are very similar (outdoor course or stairs, white uniform walls). It is then an expected result to find some aliasing between certain very similar locations. This effect can be easily removed by adding memory for locations by storing previous recognized locations that will inevitably remove outliers. Log polar transform is usually very sensitive to translations, in the present experiment the combined use of 2-ND images and statistical information introduced robustness even in case of large motions. This can be explained by the fact that 2-ND images being very close to saliency maps, image statistical information extracted around interest areas tend to remain stable even in case of large translations around known locations (approximatively two meters). It is also interesting to point out that the experiment relied on color images in order to increase the statistical information to be extracted from scenes, gray level images would provide closer results depending of course on the complexity of observed locations.

7. Conclusions, Perspectives, and Future Work

This paper presented a biologically-inspired implementation of the retina preprocessing and the LGN neural coding applied to the case of spatial recognition. A functional model has been mathematically obtained validating theoretical results confirmed by experiments. It is interesting to notice that the system provides image encoding giving implicitly valuable features that were used in the present case for localization tasks. These features are reliable as proved by the localization tests, due to the robustness of the method to different kinds of noise. It is interesting to notice that the retinal coding and the LGN retinotopic characterization of images in this case produce in an automatic way features that are enough for a fast localization. Also, they are suitable for a wide variety of applications because beside interesting features (edges and textures), the original image is still present in the first most weighted gaussians which corresponds to M1 or M2 depending on the eye.

Further application using this encoding schemes are being investigated. At the moment, the method is used to segment objects, and preliminary results show that it provides a quick and efficient segmentation of objects even in the presence of complex backgrounds. It may be interesting to inquire about the possibility of integrating feedback loops into the system. This will allow the system to emphasize the wanted features that may be urgently needed to perform a given task.

References

[1] Kandel ER, Schwartz JH, Jessel TM, *Principles of Neural Sciences*, 4th edn., 2000.
[2] Bell J, Behavior of some models of myelinated axions, *IMA J Math Appl Med Biol* **1**:149–167, 1984.
[3] Tanaka S, Miyashita M, Ribôt J, Roles of visual experience and intrinsic mechanism in the activity-dependent self organization of orientation maps: Theory and experiment, *Neural Netw*, 2004.
[4] Ullman S, Vidal-Naquet M, Sali E, Visual features of intermediate complexity and their use in classification, *Nat Neurosci* **5**(7):1–6, 2002.
[5] Thorpe S, Gautrais J, Rapid visual processing using spike asynchrony, in Mozer MC, Jordan MI, Petsche T (eds.), *Neural Information Processing Systems*, Vol. 9, MIT Press, Cambridge, pp. 901–907, 1997.
[6] Dan Y, Alonso JM, Martin W, Usrey R, Reid C, Coding of visual information by precisely correlated spikes in the lateral geniculate nucleus, *Nat Neurosci* **1–6**:501–507, 1998.
[7] Tistarelli M, Sandini G, On the advantages of polar and log-polar mapping for direct estimation of time-to-impact from optical flow, *T Patt Anal Mach Intel* **15**(4): 1993.
[8] Zugaro MB, Arleo A, Berthoz A, Wiener SI, Rapid spatial reorientation and head direction cells, *J Neurosci* **23**(8):3478–3482, 2003.
[9] Itaya SK, Van Hoesen GW, Jenq CB, Direct retinal input to the limbic system of the rat, *Brain Res* **226**:33–42, 1981.
[10] Ahmed AK, Guison NG, Yamadori T, A retrograde fluorescent-labeling study of direct relationship between the limbic (anterodorsal and anteroventral thalamic nuclei) and the visual system in the albino rat, *Brain Res* **729**:119–123, 1996.
[11] Thorpe S, Delorme A, Van Rullen R, Spike based strategies for rapid processing, *Neural Netw* **14**:715–726, 2001.
[12] Stanley GB, Li FF, Dan Y, Reconstruction of natural scenes from ensemble responses in the lateral geniculate nucleus, *J Neurosci* **19**(18):8036–8042, 1999.
[13] Riesenhuber M, Poggio T, Hierarchical models of object recognition in cortex, *Nat Neurosci* **2**(11):1019–1025, 1999.
[14] Schneider KA, Richter MC, Kastner S, Retinotopic organization and functional subdivisions of the human lateral geniculate nucleus: A high-resolution functional magnetic resonance imaging study, *J Neurosci* **24**(41):8975–8985, 2004.
[15] Tarr MJ, Bulthoff HH, Image-based object recognition in man, monkey and machine, *Cognition* **67**(1–2):1–20, 1998.
[16] Mizoguchi S, Suzuki Y, Kiyosawa M, Mochizuki M, Kawaski T, Ishii K, Senda M, Detection of visual activation of lateral geniculate nucleus by positron emission tomography, *Graef Arch Clin Exp* **241**: 2004.

[17] Marr D, Nishihara HK, Representation and recognition of the spatial organization of three-dimensional shapes, *P Roy Soc Lond B* **200**:269–294, 1978.
[18] Hummel JE, Biederman I, Dynamic binding in a neural network for shape recognition, *Psychol Rev* **99**(3):480–517, 1992.
[19] Rodieck WR, *La vision*, de Boeck Universitée, première édition, 2003.
[20] Kailath T, The divergence and Bhattacharyya distance measures in signal selection, *IEEE Trans Commun Technol COM* **15**:52–60, 1967.
[21] Treisman A, Gelade G, A feature-integration theory of attention, *Cogn Psyc* **12**(1):97–136, 1980.
[22] Wolfe J, Visual search, *Attention*, University College London Press, 1998.
[23] Goodale M, Milner A, Separate visual pathways for perception and action, *Neurosciences* **15**(1):20–25, 1992.
[24] Leigh R, Zee D, *The Neurology of Eye Movements*, 3rd edn., Oxford University Press, New York, 1999.
[25] Carpenter R, *Movements of the Eyes*, 2nd edn., Plon Limited, 1988.
[26] James W, *The Principles of Psychology*, Harvard University, 1980.
[27] Moore T, Fallah M, Control of eye movements and spatial attention, *PNAS* **98**(3):1273–1276, 2001.
[28] Koch C, Ullman S, Shifts in selective visual attention: Towards the underlying neural circuitry, *Hum Neurobiol* **4**(4):219–227, 1985.
[29] Tsotsos J, Culhane S, Lai W, Davis N, Modeling visual attention via selective tuning, *Artif Intel* **78**:507–545, 1995.
[30] Wolfe J, Visual attention, *Seeing: Handbook of Perception and Cognition*, 2nd edn., De Valois KK, pp. 335–386, 2000.
[31] Itti L, Koch C, Computational modeling of visual attention, *Nature Rev Neurosci* **2**(3):194–203, 2001.
[32] Hamker F, A dynamic model of how feature cues guide spatial attention, *Vis Res* **44**:501–521, 2004.
[33] Posner M, Cohen Y, Components of visual orienting, in Bouma H, Bouwhuis DG (eds.), *Attention and Performance*, Erlbaum, Hilldale, NJ, Vol. 10, pp. 531–556, 1984.
[34] Zhang K, Representation of spatial orientation by the intrinsic dynamics of the head-direction ensemble: A theory, *J Neurosci* **16**:2112–2126, 1996.
[35] Taube JS, Muller RU, Ranck Jr JB, Head-direction cells recorded from the postsubiculum in freely moving rats. Effects of environmental manipulations, *J Neurosci* **10**:436–447, 1990.
[36] Taube JS, Head direction cells and the neurophysiological basis for a sense of direction, *Prog Neurobiol* **55**:1–32, 1998.
[37] Adelson EH, Perceptual organization and the judgment of brightness, *Science* **262**:2042–2044, 1993.
[38] Wishart KA, Frisby JP, Buckley D, The role of 3D surface slope in a lightness/brightness effect, *Vis Res* **37**:467–473, 1997.
[39] Stevens SS, Neural events and the psychophysical law, *Science* **170**:1043–1050, 1970.
[40] Srinivasan MV, Laughlin SB, Dubs A, Predictive coding: A fresh view of inhibition in the retina, *P Roy Soc Lond B* **216**:427–459, 1982.

[41] Campbell FW, Robson JG, Application of Fourier analysis to the visibility of gratings, *J Physiol (Lond)* **197**:551–566, 1968.

[42] Shapley RM, Tolhurst DJ, Edge-detectors in human vision, *J Physiol (Lond)* **229**:165–183, 1973.

[43] Baylor DA, Lamb TD, Yau KW, Responses of retinal rods to single photons, *J Physiol* **288**:613–634, 1979.

[44] Fain G, Dowling JE, Intracellular recordings from single rods and cones in the mudpuppy retina, *Science* **180**:1178–1181, 1973.

[45] Balboa RM, Grzywacz NM, The minimal local-asperity hypothesis of early retinal lateral inhibition, *Neural Comp* **12**:1485–1517, 2000.

[46] Kamermans M, Kraaij DA, Spekreijse H, The cone/horizontal cell network: A possible site for color constancy, *Vis Neurosci* **15**:787–797, 1998.

[47] McCarthy ST, Owen WG, Preferential representation of natural scenes in the salamander retina, *Invest Ophthalmol Vis Sci* **37**:S674, 1996.

[48] Atick JJ, Redlich AN, What does the retina know about natural scenes? *Neural Comp* **4**:196–210, 1992.

[49] Field DJ, What is the goal of sensory coding? *Neural Comp* **6**:559–601, 1994.

[50] Ratliff F, *Mach Bands: Quantitative Studies on Neural Networks in the Retina*, Holden-Day, San Francisco, 1965.

[51] Everitt BS, Hand DJ, *Finite Mixture Distributions*, Chapman and Hall, London, UK, 1981.

[52] Hastie T, Tibshirani R, Discriminant analysis by Gaussian mixtures, *J R Stat Soc B* **58**:155–176, 1996.

[53] Dempster AP, Laird NM, Rubin DB, Maximum likelihood from incomplete data via the EM algorithm, *J R Stat Soc B* **39**:1–38, 1977.

[54] Schwartz, Estimating the dimension of a model, *Ann Stats* **6**:461–464, 1978.

[55] Herault J, Model of color processing in the retina of vertebrates: From photoreceptors to color opposition and color constancy phenomena, *Neurocomputing* **12**(2–3):113–129, 1996.

[56] Beaudot W, Oliva A, Herault J, Retinal model of the dynamics of X and Y pathways: A neural basis for early coarse-to-fine perception, *Perception* **24**:93b, Tuebingen, Germany, 1995.

October 8, 2007 18:53 WSPC/179-JIN 00158