A bio-inspired model of central and peripheral vision for scene categorization

Raluca Vlad-Debusschere, Nathalie Guyader, Anne Guérin-Dugué

GIPSA-lab, 11 rue des Mathématiques, BP46, F - 38402 Saint Martin d'Heres Cedex, France
{first_name.family_name}@gipsa-lab.grenoble-inp.fr

Résumé – La vision périphérique a été souvent négligée dans les modèles, qui simulent pour une grande partie la vision centrale, et dans les expériences comportementales, qui étudient majoritairement les capacités de la vision centrale. A partir d'un modèle de vision centrale déjà existant, nous testons si l'ajout d'un filtrage spatialement variant modélisant la diminution d'acuité visuelle en périphérie nous permet de répliquer des performances humaines de catégorisation de scènes présentées à différentes excentricités. Les résultats obtenus correspondent aux résultats comportementaux pour une sous-base d'images sans être généralisables à la totalité de la base. Nous expliquons ces résultats par le fait que notre modèle ne prend pas en compte l'information de couleur et utilise uniquement l'information de luminance pour décrire les images.

Abstract – Peripheral vision has been left aside both in modeling approaches and in psychological experiments, with the majority of existing studies mainly dedicated to our central vision exclusively. Using an existing model of central vision, we investigate whether adding a spatially variant filter simulating the decrease of visual acuity toward periphery allows us to reproduce human performances in a task of categorization of scenes positioned at different eccentricities in the visual field. The results we obtain fit human performances for a subset of images of natural scenes, however, for the moment these results cannot be extended to the entire database. We explain this result by the fact that the human performances were measured using colored scenes and the proposed model only processed the luminance information.

1 Introduction

1.1 Context

Most models of the human visual system (HVS) only simulate the central vision. However, a better understanding and quantification of the capacities of our peripheral vision is crucial, for example, for people suffering from age-related macular degeneration (AMD), who are subject to partial or complete loss of their central vision and use their peripheral vision in everyday life. Thus, understanding the ability of performing various tasks in function of the eccentricity of the visual stimulus could give comprehensive insights on how visually impaired persons perform their everyday life activities and, more interestingly, on how new devices could be developed for facilitating these activities.

In this paper, we present a simplified bio-inspired model of the HVS, adapted to both central and peripheral vision. We illustrate the results obtained with this model compared to behavioral data on a visual task of scene categorization at various eccentricities.

1.2 Existing models

The decomposition of visual content with cortical-like filters using a bank of Gabor filters has already been extensively used for reproducing the performances of the HVS on various tasks, from texture segmentation [1], to more complex image categorization [2], [3]. Other models have gone further on the modeling of the HVS by adding the pre-processing performed by retinal cells [4], [5]. However, these models have only focused on the central vision and are thus only able to reproduce the performances of the HVS for stimuli visualized in the center of the visual field.

In the present article, we propose to extend these previous results, by taking into account the fact that our central vision has a high acuity whereas our peripheral vision has a lower visual acuity. Note that in this model we only investigate the processing of static stimulus without studying the peripheral vision particularly efficient to detect moving objects. Hence we modeled the visual processing made by our HVS for a stimulus appearing anywhere in the visual field.

2 The proposed model

The proposed model is organized in two main layers, the *retina* and the *cortex*. Before these two layers, we include a *spatially variant (SV)* processing layer to model the loss of visual acuity with the increasing eccentricity of the visual stimulus. This pre-processing step takes into ac-

count the position of the visual stimulus in the visual field and consequently establishes the degree of precision with witch the stimulus is further processed. The structure of the proposed model is schematized in Figure 1.

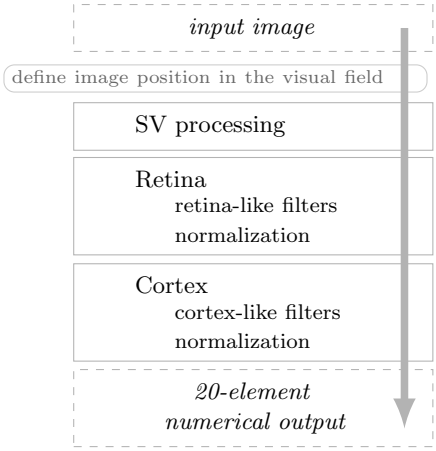

FIGURE 1 – The structure of the proposed model.

2.1 Input

The input of the model has to be an image in a *gray-level* format, since for the moment we only focus on the luminance information, without taking into account the chrominance. Along with the input image, its visualization distance must be provided, since this parameter determines directly its size in the visual field and therefore some of the parameters of the subsequent processing.

2.2 SV processing

The *SV processing* layer not only models the non-uniform density of retinal photoreceptors but also the ratio of retina output cells for input cells that dramatically decreases with eccentricity. The aim is to take into account the loss in visual acuity from the central retina towards the periphery of the visual field [6].

It has already been shown that the retinal and cortical cells distribution is reflected by experimental data on the sensitivity to contrast of the HVS [7]. Thus, the SV processing consists of applying the contrast sensitivity curve of [8], which was obtained by modeling the data in [9] and [7]. More precisely, we use this curve as it was reproduced in [10]:

$$\textit{contrast threshold}\,(e) = A\frac{\alpha}{\alpha + e}, \qquad (1)$$

where e is the eccentricity in degrees of visual angle at which the image is visualized. The A constant related to the visualization distance is set to 1 and the α parameter controlling the decrease in contrast is set to 2.3°. This law thus explains the contrast to be perceived when visualizing an image at a given eccentricity. Its graphical illustration is given in Figure 2.

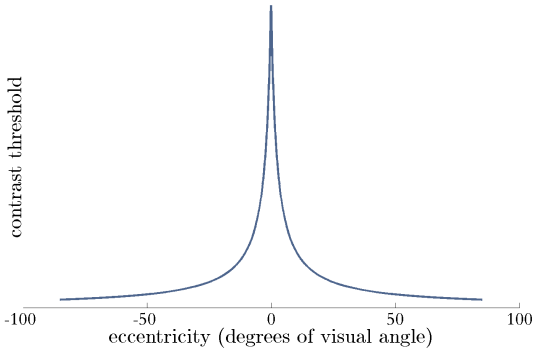

FIGURE 2 – The contrast sensitivity curve of [8] that explains the decrease in visual acuity in peripheral vision. The 0° eccentricity point corresponds to the center of the retina and subsequently to the fixation point.

By using this contrast sensitivity curve and by considering the parameter giving the position of the input image in the visual field, we are able to filter this image accordingly, leading to a visual content similar to what an observer would see if the image was displayed in his or her field of view.

2.3 Retina

The *retina* layer reproduces the *functioning of the principal types of cells in the human retina*. Hence, the retina processing step corresponds to a series of digital filters, as explained in [11]. Such processing leads to local contrast enhancement and spectral whitening as illustrated by Figure 3.

2.4 Cortex

The *cortex* layer models the *functioning of some cortical cells*. These cells have been found to preferentially respond to precise orientations and spatial frequencies of the visual stimulus [12]. Moreover, it has been shown that the response of these cells can been successfully modeled by Gabor filters [13]. A specific type of such filters, the log-Gabor filters, which have the profile of Gabor filters viewed on a logarithmic scale, was found to be particularly adequate for processing natural images.

Therefore, in order to simulate a column of cortical cells sensitive to multiple orientations and spatial frequencies, we used a set of log-Gabor filters with predefined parameters. The frequency response of each of these filters is defined as:

$$G(u,v) = \exp\left[-\left(\frac{\log(u'/f_0)^2}{2\sigma_u^2} + \frac{v'^2}{2\sigma_v^2}\right)\right], \qquad (2)$$

where f_0 is its central frequency, (u', v') are obtained by

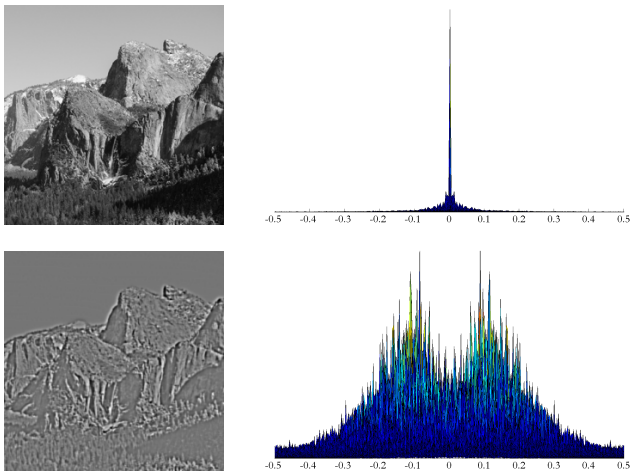

FIGURE 3 – An example of photographed image (left) and the profile of its amplitude spectrum (right) before (top) and after (bottom) the retina processing.

rotating the filter axes, (u, v), with an angle θ, and σ_u and σ_v are the standard deviation values of the Gaussian in the u' and v' directions, determined as:

$$\sigma_u = \frac{log\, 2^{B_{\xi_b}}}{4\alpha} \text{ and } \sigma_y = \frac{f_0 \tan\left(\frac{\Omega}{2}\right)}{2\alpha}, \qquad (3)$$

with B_{ξ_b}, the radial band in octave, set to 1.

The filter bank configuration we use is that of 20 log-Gabor filters, capturing 4 different orientations and 5 different spatial frequencies. The maximum central frequency is 0.25 and the central frequencies of the following k^{th} filters are computed using $f_k = 0.25/2^k$. This configuration is illustrated in Figure 4.

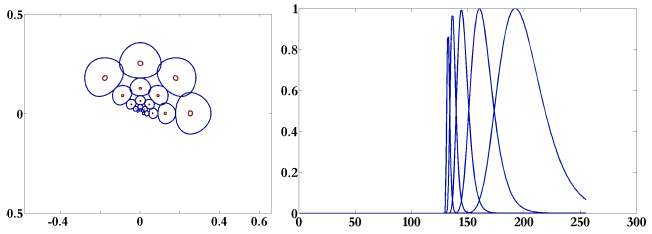

FIGURE 4 – The contours of the filter banks at half-height and 99% height of the amplitude spectrum (left) and their profile along the horizontal spatial frequency axis (right).

2.5 Output

To summarize, for any input image, the model produces the output of the layers of filters presented. Thus, as a last step, by applying the series of 20 cortex-like filters to the retina-processed image, a decomposition into 20 cortical features is obtained. As a consequence, for an input image, the model allows the computation of a 20-element numeric descriptor, where each value represents the total energy of a cortical feature.

3 Scene categorization

Scene descriptors as tho ones provided by the proposed model might be efficient to predict the category of a scene or at least the gist of a scene [2], [3].

The aim of our model is to test whether such results might be extended to scenes presented in the periphery. To validate our model we used behavioral results on large eccentricity scene categorization described below.

3.1 Experimental data

The experimental data that we consider as ground-truth reference for our computational model is the data collected during Experiment 1 in [14]. This data corresponds precisely to a *natural vs. urban* categorization task. In this experiment, observers were attributed a *target category*, *i.e. natural* or *urban*, then watched random *natural-urban* or *urban-natural* pairs of images, with one image displayed to the left and the other to the right of the central fixation point that observers were supposed to fixate, at the same eccentricity. For each pair of images, they voted *left* or *right*, in function of the side where they saw the *target category* scene. The images were displayed at 10°, 30°, 50°, and 70° of horizontal eccentricity. The results obtained in this study consist of mean percent correct responses obtained in function of the horizontal eccentricity at which the visual stimuli have been displayed. They are plotted in Figure 5.

3.2 Test data set

For comparing the performances of our model to those obtained experimentally, we use the same image data set as in [14]. Hence, we consider 200 images visualized at a distance of 2.1 m and covering $20° \times 20°$ of eccentricity in the visual field. Half of the images are *natural* and half are *urban*.

3.3 Method

We perform the scene categorization using a machine learning approach with a *comparison of distances* between 20-element vectors (our image descriptors). We perform the training procedure on the scenes processed as visualized at 0° of eccentricity and the testing procedure on the same images, but processed as visualized at the eccentricities used in the original experiment: 10°, 30°, 50°, and 70°.

During the training procedure we computed two mean descriptors, one for the *natural* images and one for the

urban images, using the 20-element descriptors obtained on the training images of each of these two classes.

Then, during the testing procedure, we test the same trial configurations as those used for the 12 participants of the study. We thus consider the target category attributed to each participant and, for each natural-urban pair of images visualized by this participant, we apply the decision rule relative to that target category. This rule decides which among two images belongs to the *target* category by choosing the shortest euclidean distance between the mean descriptor of the target category and each of the two image descriptors.

3.4 Results

The results are shown in Figure 5.

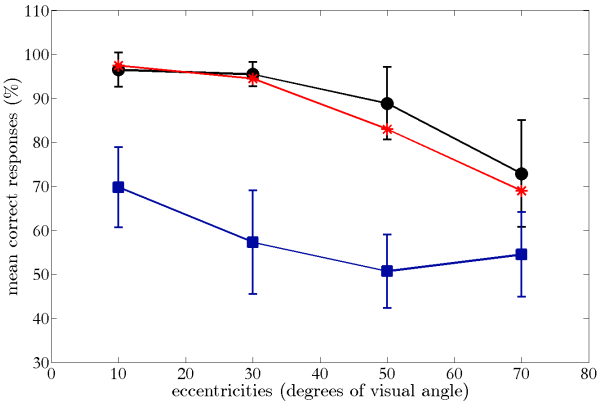

FIGURE 5 – ● – the experimental results in [14]; ■ and * – the computational results obtained with our model on the 200 images and 20 images data sets, respectively.

Despite the fact that the mean percent correct responses obtained with our model are inferior to the experimental ones, the decrease in performance with eccentricity approaches the slope of the experimental results. The major factor influencing the low percentages of correct responses obtained seems to be the large variability of the images in the data set under study, visible in the large variability of the spectral features inside of each category. The data set from [14] contains, especially for the *urban* category, images of various contents, which are to be argued whether representative for the category they represent. To illustrate this point, we also show in Figure 5 the results obtained with our model, by following a similar procedure, on a selection of only 20 images from the initial data set. This small test data set was selected so that the 10 images of each category would unambiguously represent the two categories. The results of this test illustrate how, for images that are characteristic of the two categories, the bio-inspired model of the HVS that we propose reaches very good categorization performances.

4 Conclusion

The study presented in this article is work in progress.

At present, we could show that the decrease in performance of the HVS with eccentricity can be reproduced with the bio-inspired model proposed, however more insight is necessary on the influence of the data set content on the results.

An interesting perspective envisaged is the extension of the current model by taking into account the chromatic information as well and a more in-depth analysis of the luminance approach *vs.* the chrominance approach.

Références

[1] A. Guérin-Dugué and P. Palagi, "Texture segmentation using pyramidal gabor functions and self-organising feature maps," *Neural Processing Letters*, vol. 1, no. 1, pp. 25–29, 1994.

[2] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[3] N. Guyader, A. Chauvin, C. Peyrin, J. Hérault, and C. Marendaz, "Image phase or amplitude? rapid scene categorization is an amplitude-based process," *Comptes Rendus Biologies*, vol. 327, no. 4, pp. 313–318, 2004.

[4] S. Marat, T. H. Phuoc, L. Granjon, N. Guyader, D. Pellerin, and A. Guérin-Dugué, "Modelling spatio-temporal saliency to predict gaze direction for short videos," *International Journal of Computer Vision*, vol. 82, no. 3, pp. 231–243, 2009.

[5] A. Benoit, A. Caplier, B. Durette, and J. Hérault, "Using human visual system modeling for bio-inspired low level image processing," *Computer vision and Image understanding*, vol. 114, no. 7, pp. 758–773, 2010.

[6] B. Sere, C. Marendaz, and J. Herault, "Nonhomogeneous resolution of images of natural scenes," *Perception London*, vol. 29, no. 12, pp. 1403–1412, 2000.

[7] M. S. Banks, A. B. Sekuler, and S. J. Anderson, "Peripheral spatial vision: limits imposed by optics, photoreceptors, and receptor pooling," *Journal of the Optical Society of America*, vol. 8, no. 11, pp. 1775–1787, Nov 1991.

[8] W. S. Geisler and J. S. Perry, "A real-time foveated multiresolution system for low-bandwidth video communication," in *Human Vision and Electronic Imaging III, SPIE Proceedings*, 1998, pp. 294–305.

[9] J. Robson and N. Graham, "Probability summation and regional variation in contrast sensitivity across the visual field," *Vision Research*, vol. 21, no. 3, pp. 409 – 418, 1981.

[10] T. Ho-Phuoc, *Développement et mise en œuvre de modèles d'attention visuelle*, Ph.D. thesis, University of Grenoble, 2010.

[11] J. Hérault, *Vision: Images, Signals and Neural Networks Models of Neural Processing in Visual Perception*, World Scientific Publishing Co., Inc., 2010.

[12] D. H. Hubel and T. N. Wiesel, "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex," *The Journal of physiology*, vol. 160, no. 1, pp. 106, 1962.

[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 A*, vol. 2, no. 7, pp. 1160–1169, Jul 1985.

[14] M. Boucart, C. Moroni, M. Thibaut, S. Szaffarczyk, and M. Greene, "Scene categorization at large visual eccentricities," *Vision Research*, vol. 86, no. 0, pp. 35 – 42, 2013.