MASKING EFFECT IN VISUAL ATTENTION MODELING

O. LE MEUR[1,2,], P. LE CALLET[2], D. BARBA[2], D. THOREAU[1]*

[1] - THOMSON multimedia R&D, 1 Avenue Belle Fontaine, 35511 Cesson-Sévigné, France.

[2] - IRCCyN UMR n°6597 CNRS, Ecole Polytechnique de l'Université de Nantes, rue Christian Pauc, La Chantrerie, BP50609, 44306 Nantes, France.

ABSTRACT

In this paper, we focus on a bottom-up model of visual attention permitting to determine conspicuous locations on still color pictures. Such a model is useful to define visual regions of interest allowing to drive for instance a video coding scheme. The aim of this paper is not to provide a full description of the model (wholly described in [1]), but specially to assess the influence of a particular property of the human visual system (HVS) inserted in our model. The property that we want to emphasize here refers to the visual masking abilities of the HVS. Such property is a key step in many applications, including watermarking and video quality assessment. We provide evidences on the fact that visual masking influences the determination of conspicuous locations.

1. INTRODUCTION

The HVS is limited by our cognitive resources. Nevertheless, humans have the capacity to understand their visual environment; therefore, this system is able to deal with a huge amount of information.

It's now commonly assumed that human attentive behavior can be shared between a pre-attentive and an attentive processing. The pre-attentive processing, so-called bottom-up processing, refers to some visual features which are able to attract our visual attention toward salient parts of our field of view without requiring attentional resources. The second form of visual attention called top-down is driven by the task that we have in mind. This second form of visual attention was disregarded during the initial design of our model.

A variety of different approaches dealing with the determination of conspicuous locations exists. The most famous is surely the bottom-up model of Itti et al[2]. All of these approaches encompass ideas related to psychophysics experiments. Nevertheless, all of them neglect several low-level HVS's properties: on one hand the intrinsic visual features of a signal (orientation, spatial frequencies,…) modulates our ability to perceive an information. On the other hand, the contextual influences on the signal's visibility are neglected. Our model based on a psychovisual backing takes advantage of low-level properties in order to define a fixation map or a map of region of interest. Among these properties, the visual masking ability is one of the most important. It acts on the signal by increasing or, more often, decreasing its visibility. Numerous applications, such as watermarking and quality assessment, take advantage of such this property.

The model we developed is briefly described in section two. The third part of this paper introduces the visual masking and the elaborate models of masking we used. The fourth section is dedicated to assess the influences of the visual masking on the computation of conspicuous locations obtained from our model.

2. A SYNTHETIC DESCRIPTION OF OUR BOTTOM-UP MODEL

The proposed bottom-up model dealing with still color pictures (wholly described in [1]) is based on numerous properties of the HVS, thus providing a biologically plausible system. This model consists of three parts named visibility, perception and domain of perceptual grouping.

The first part, called visibility, permits to mimic the limited HVS sensitivity. It is clear that our visual system is not able to perceive all information present in our visual field with the same accuracy. To take into account these intrinsic limitations, the visibility part includes the following mechanisms entirely identify from psychophysics experiments (conducted in the IRCCyN's lab):

- The transformation of the RGB luminance into the Krauskopf's colorspace composed by the cardinal direction A, Cr1 and Cr2 [3]. This conversion permits to decorrelate color information.
- A perceptual channel decomposition [4,5], which consists in carving up the image spatial spectrum both in spatial radial frequency and orientation, is applied on each of the three components (A, Cr1, Cr2). Psychovisual spatial frequency partitioning for the achromatic component is shown on the figure 2.1. The shaded region on the figure 2.1 indicates the spectral support of the channel belonging to the third

* Olivier.le-meur@thomson.net

crown and having an angular selectivity equal to $\pi/6$. This channel is termed III,2.

- Then, we apply on each component a contrast sensitivity function (CSF). This is defined later in the paper.
- The last mechanism of the visibility part refers to the visual masking. We describe this mechanism in the next section and assess its influence on our results in the last section.

The second part of our model deals with the perception. As defined by the British psychologist David Marr [6], the perception is a process that produces from images of the external world a description that is useful to the viewer and not cluttered with irrelevant information. To mimic the human perception, we reproduce the behavior of visual cells belonging to the primary visual cortex. In fact, we attempt to reproduce the non-classical receptive field (non-CRF) [7] by using a two-dimensional difference-of-Gaussians [1].

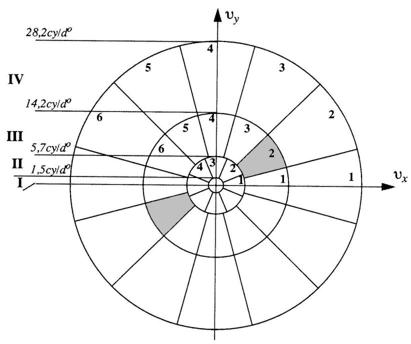

figure 2.1 : psychovisual spatial frequency partitioning for the achromatic component.

The last step focuses on some aspects of the domain of perceptual grouping which refers to the human visual ability to group and to bind visual features to form a meaningful higher-level structure.

Finally, a saliency map which is a two-dimensional map representing the conspicuous location is computed.

3. VISUAL MASKING

3.1. Definition and properties

Biological evidences have shown that visual cells respond to stimuli above a certain contrast. The contrast value for which a visual cell responds is called the visibility threshold (above this threshold, the stimuli is visible). This threshold varies with numerous parameters such as the spatial frequency of the stimuli, the orientation of the stimuli, the viewing distance, … This variability leads us to the concept of the CSF which expresses the sensitivity of the human eyes (the sensitivity is equal to the inverse of the contrast threshold) as a multivariate function. Consequently, the CSF permits to assess the sensitivity of the human eyes for a given stimuli.

Nevertheless, for natural pictures, the sensitivity can be modulated (increased or decreased visibility threshold) by the presence of another stimulus. This modulation of the sensitivity of the human eyes is called the visual masking.

An illustration of masking effect is shown on the figure 3.1. We consider two cues, a target and a masker where C_T and C_M are the contrast threshold of the target in the presence of the masker and the contrast of the masker respectively. Moreover, C_{T0} is the contrast threshold measured by a CSF (without masking effect).

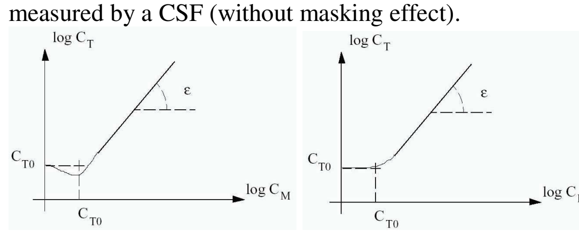

figure 3.1 : illustration of visual masking (left); non-linear transducer model of masking.

When C_M varies, three regions can be defined :

- At low values of C_M , the detection threshold remains constant. The visibility of the target is not modified by the masker.
- When C_M tends toward C_{T0}, the masker eases the detection of the target by decreasing the visibility threshold. This phenomenon is called facilitative or pedestal effect.
- When C_M increases, the target is masked by the masker. His contrast threshold increases.

There are several methods to achieve the visual masking modeling based on psychophysics experiments : a common one is based on the detection of simple signals as sinusoidal patterns. A better method relates to the detection of quantization noise. It is obvious that the first method is a strong simplification in regard to the intrinsic complexity of natural pictures. Nevertheless, numerous applications (watermarking, video quality assessment) are built around such principle with interesting results compared to the complexity. Consequently, we have elected the detection of simple signal as a preferred approach.

In the context of subband decomposition, masking has been intensively studied [9-11] leading to the definition of three kinds of masking : intra-channel masking, inter-channel masking and inter-component masking.

The intra-channel masking occurs between signals having the same features (frequency and orientation) and consequently belonging to the same channel. It is the most important masking effect.

The inter-channel masking occurs between signals belonging to different channels of the same component. The inter-component masking occurs between channels of different components (the component A and the Cr1 component for example). These two last visual masking are put together and are just called inter-masking in the following of this paper.

3.2. Visual masking in our bottom-up model

Table 3.1 summarizes the interactions taken into account in our bottom-up model. Only the significant interactions have been elected in regard to psychophysics experiments conducted in IRCCyN's lab.

The visibility threshold variation $T_{i,j,c}(m,n)$ at a particular location (m,n), for a particular channel (i,j) and for a particular component c can be written as follows :

$$T_{i,j,c}(m,n) = T^{\text{intra}}_{i,j,c}(m,n) \cdot \prod_{i'}\prod_{j'}\prod_{c'} T^{\text{inter}}_{i',j',c' \to i,j,c}(m,n)$$

The modulation of the visibility threshold is obtained by the product between the modulation of the visibility threshold stemming from the intra-channel masking and the set of the modulation coming from the inter-channel masking. The term $T^{\text{inter}}_{i',j',c' \to i,j,c}(m,n)$ expresses a particular interaction provoked by a particular location (m,n) of the channel (i',j') belonging to the component c' on the same location of the channel (i,j) belonging to the component c. For example, if we refer to the table 3.1, the channels II,n of the Cr1 component could be masked by the channel I of the A component. The variation of the visibility threshold due to this inter-channel masking is noted: $T^{\text{inter}}_{I,0,A \to II,n,Cr1}$. We propose hereafter to describe the masking models we used.

3.2.1. Intra component masking

For the achromatic component, we used the function designed by Dally [11] in spite of the fact that this model does not take into account the pedestal effect. The strength of this model lies in the fact that it has been optimized with a huge amount of experimental results.

The variation of the visibility threshold is given by :

$$T^{\text{intra}}_{i,j,A}(m,n) = (1 + (k_1(k_2 \left|R_{i,j}(m,n)\right|)^s)^b)^{1/b}$$

where $R_{i,j}$ is a psychovisual channel stemming from the perceptual channel decomposition (For example, the shaded region on the figure 2.1 leads to the channel $R_{III,2}$). The values k_1, k_2, s, b are given in [11].

For the chromatic components, the model has been developed by IRCCyN's lab [9, 10] from our psychophysics experiments. The analytic form of the model given below takes into account the pedestal effect :

$$T^{\text{intra}}_{i,j,Cr1\&Cr2}(m,n) = \frac{1 + a \cdot \left|R_{i,j}(m,n)\right| + b \cdot \left|R_{i,j}(m,n)\right|^2}{1 + c \cdot \left|R_{i,j}(m,n)\right|}$$

The values a, b, c can be find in [11].

3.2.2. Inter component masking

As the intra component masking, the inter component masking has been defined from experimental results conducted in the IRCCyN's lab [9, 10]. Two models have been defined : a first model (called A model) taking into account the pedestal effect and a second (called B model) which just mimic the classical visual masking.

The analytic form of theses models (A first, B in second place) are given below :

$$T^{\text{inter}}_{i,j,c}(m,n) = \frac{1 + a \cdot \left|R_{i',j',c'}(m,n)\right| + b \cdot \left|R_{i',j',c'}(m,n)\right|^2}{1 + c \cdot \left|R_{i',j',c'}(m,n)\right|}$$

$$T^{\text{inter}}_{i,j,c}(m,n) = a - b\exp(-c \cdot \left|R_{i',j',c'}(m,n)\right|)$$

The reader can find the correct way to apply these models and parameters values in [10].

	"Target signal"		
"masker"	A	Cr1	Cr2
A	Intra masking	I,I / I,IIn IIn,I / IIn,IIn	insignificant
Cr1	I,I / I,IIn / I, IIIn IIn,I / IIn, IIn,	Intra masking	I,I / I,IIn
Cr2	insignificant	I,I / I,IIn	Intra masking

table 3.1 : intra and inter masking considered in our bottom-up model (a couple Y,Z means that the channel Y could decrease or increase the visibility threshold of the signal containing in the channel Z).

4. RESULTS AND DISCUSSION

In order to assess the influence of visual masking in a bottom-up model, we have to compare the results stemming from the whole model with the results stemming from a simplified model using no visual masking. Figure 4.1 illustrates the first 20 fixation points on four test pictures (CIF format). A red circle represents a conspicuous location (the radius of the red circle is equal to one degree of visual angle defined with a distance of 6H (H is the height of the screen)).

From our own opinion and purely subjectively, the visual masking improves the results. In some cases, this enhancement can be very important. For example on Table Tennis, the improvement consists of the man's face detection.

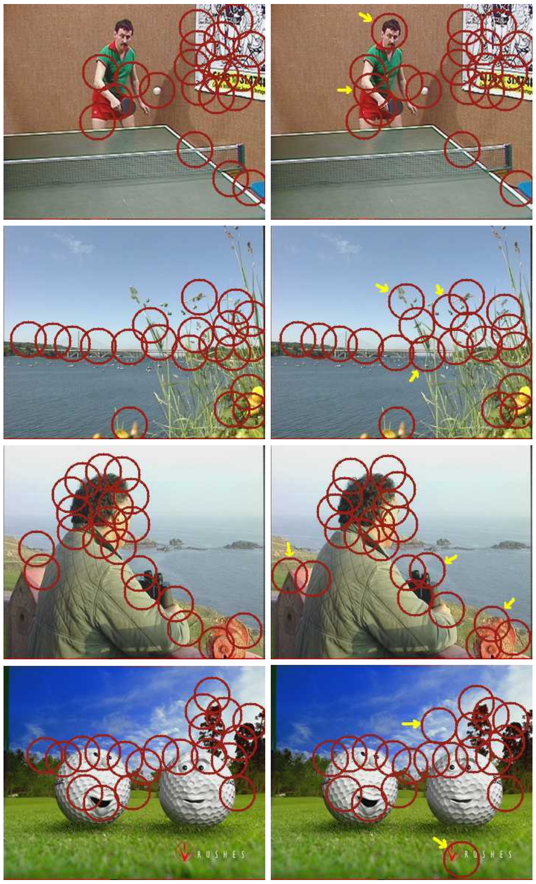

figure 4.1 : the first 20 fixations points on Table (top), on Iroise's bridge (second row), on man in front of the sea (on the third row) and on rushes (last row) for the two configuration (left without visual masking, right with visual masking).

In other cases, it increases the accuracy of the locations of conspicuous points: on the twig on the foreground for the Iroise's bridge picture, on the binoculars and on the bottom-left for the third picture …

5. CONCLUSION

In this paper, we have investigated the importance of visual masking in a computational model of bottom-up visual selective attention. Our conclusions are that visual masking can be very useful in such a model. Indeed, visual masking increases the accuracy and the relevance of the conspicuous locations detection.

Moreover, the proposed visual masking model can be very useful in others applications such as watermarking, image quality assessment, or determination of visual advanced descriptors.

In future works, we will pursue the improvement of our model by integrating new visual phenomena and by integrating the temporal dimension.

6. REFERENCES

[1] O. Le Meur, D. Thoreau, E. François, D. Barba and P. Le Callet, "From Low Level Perception to High Level Perception, a Coherent Approach for Visual Attention Modeling", *in proc. SPIE Human Vision and Electronic Imaging*, San Jose, CA, 2004.

[2] L. Itti, C. Koch and, E. Niebur, "A model of saliency – based visual attention for rapid scene analysis", *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)*, vol. 20, N°11, pp.1254-1259, 1998.

[3] J. Krauskopf, D.R. Williams, and D. W. Heeley, "Cardinal Directions of Color Space", *Vision Research*, vol. 22, pp. 1123-1131, 1982.

[4] A.B. Watson, "The Cortex Transform : Rapid Computation of Simulated Neural Images", *Computer Vision, Graphics and Image Processing*, vol. 39, pp. 311-327, 1987.

[5] H. Senane, A. Saadane, and D. Barba, "The Computation of Visual Bandwiths and their Impact in Image Decomposition and Coding", *Int. Conf. and Signal Processing Applications and Technology*, Santa-Clara, CA, pp. 776-770, 1993.

[6] D. Marr, "Vision: a Computational Investigation into the Human Representation and Processing of Visual Information", *W.H. Freeman and Company*, NY 29-61, 1982.

[7] C. Li, W. Li, "Extensive Integration Field Beyond the Classical Receptive Field of Cat's Striate Cortical Neurons – Classification and Tuning Properties", *Vision Research*, vol. 387, pp. 73-86, 1994.

[8] J.M. Foley, and G.M. Boynton, "A New Model of Human Luminance Pattern Vision Mechanisms : Analysis of the Effects of Pattern Orientation, Spatial Phase and Temporal Frequency", *in proc. SPIE Computational Vision Based on Neurobiology*, vol. 2054, pp. 32-42, 1994.

[9] L. Bedat, "Aspects Psychovisuels de la Perception des Couleurs. Application au Codage d'Images Couleurs Fixes avec Compression de l'Information", *PhD Thesis*, University of Nantes, 1998.

[10] P. Le Callet, "Critères Objectifs avec Référence de Qualité Visuelle des Images Couleurs", *PhD Thesis*, University of Nantes, 2001.

[11] S. Daly, "A Visual Model for Optimizing the Design of Image Processing Algorithms", in *IEEE Int Conference on Image Processing*, pp. 16-20, 1994.