From low level perception to high level perception, a coherent approach for visual attention modeling.

O. Le Meur[1,2,*], P. Le Callet[2], D. Barba[2], D. Thoreau[1], E. Francois[1]
[1]-THOMSON multimedia R&D France, 1 Avenue de Belle Fontaine, 35511 Cesson-Sévigné cedex, FRANCE.
[2]-IRCCyN UMR n°6597 CNRS, Ecole Polytechnique de l'Université de Nantes, rue Christian Pauc, La chantrerie, BP50609, 44306 Nantes cedex, FRANCE.

ABSTRACT

The saliency-based or bottom-up model of visual attention presented in this paper deals with still color images. The model we built is based on numerous properties of the human visual system (HVS), thus providing a biologically plausible system. The computation of early visual features such as color and orientation is a key step for any bottom-up model and the way to extract these visual features easily permits to differentiate a model from an other. The novelty of the proposed approach lies on the fact that the computation of early visual features is fully based on a HVS model consisting in projecting the picture into an opponent-colors space, applying a perceptual decomposition, contrast sensitivity and masking functions. Moreover, a strategy essentially based on a center surround mechanism and on the perceptual grouping phenomena underscores conspicuous locations by combining visual feature maps. A saliency map which is defined as a 2D topographic representation of conspicuity is then deduced. The model is applied to a number of natural images. Our results are then compared with the results of a well-know bottom-up model.

Keywords : saliency map, bottom-up model, feature extraction, visibility threshold.

1. INTRODUCTION

The human information processing system is intrinsically a limited system and especially for the visual system. In spite of the limits of our cognitive resources, this system has to face up to a huge amount of information containing in our visual environment. Nevertheless and paradoxically, humans seem to succeed in solving this problem since we are able to understand our visual environment.

It is commonly assumed that certain visual features are so elementary to the visual system that they require no attentional resources to be perceived. These visual features are called preattentive features. As described in [1], there is reasonable consensus about a small number of such features : color, orientation, curvature, motion, …

According to this tenet of vision research, human attentive behavior is shared between preattentive and attentive processing. As explain before, preattentive processing, so-called bottom-up processing, is linked to involuntary attention. Our attention is effortless drawn to salient parts of our view. When considering attentive processing, so-called top-down processing, our attention is linked to a particular task that we have in mind. This second form of attention is thus a more deliberate and powerful one in the way that this form of attention requires effort to direct our gaze towards a particular direction.

In this paper, a coherent approach for modeling visual attention on still color images is proposed. This approach is a purely bottom-up model. In others words, during the initial design of our model, the top-down approach is disregarded although it's obvious that a model of visual attention must include top-down cues. The model we built up is based on the Koch and Ullman's hypothesis [2] in a similar fashion as many models of the bottom-up control of attention. They assume that the maps stemming from the extraction of early visual features are combined into a unique saliency map which is a two-dimensional map representing the conspicuous location. In the section 2, the whole process is briefly described. In the section 3, the extraction of preattentive features from the input picture id described. Section 3 is dedicated to mimic the intrinsic limitations of human eyes. To suppress irrelevant information several actions are necessary. The aim of the section 4 is to put forward these mechanisms. Section 5 refers to the perceptual grouping

* Olivier.le-meur@thomson.net

domain in which several combination of early visual features could be done in order to improve the saliency map. The gaze prediction is discussed in section 6 and compared with a well-known computational model for bottom-up attention. Finally, we conclude with remarks concerning our approach and concerning future works.

2. GLOBAL SCHEME DESCRIPTION

The whole process is shown in figure 2-1. The global scheme is shared into three main parts : visibility, perception and domain of perceptual grouping.

The first one, named visibility, lies on the fact that the HVS has a limited sensitivity. For example, the HVS is not able to perceive with a good precision all signals in your real environment and is insensible to small magnitude stimuli. The goal of this first step has to take into account these intrinsic limitations by using perceptual decomposition, contrast sensitivity functions (CSF) and masking functions.

The second part is dedicated to the perception concept. We remind the perception definition given by the British psychologist David Marr [3] : 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 select relevant information, a center surround mechanism is notably used in accordance with biological evidences (this mechanism will be described and justified later).

The last step concerns some aspects of the perceptual grouping domain. The perceptual grouping refers to the human visual ability to extract significant images relations from lower level primitive image features without any knowledge of the image content and to group them to obtain meaningful higher-level structure. We just focus on contour integration and edge-linking.

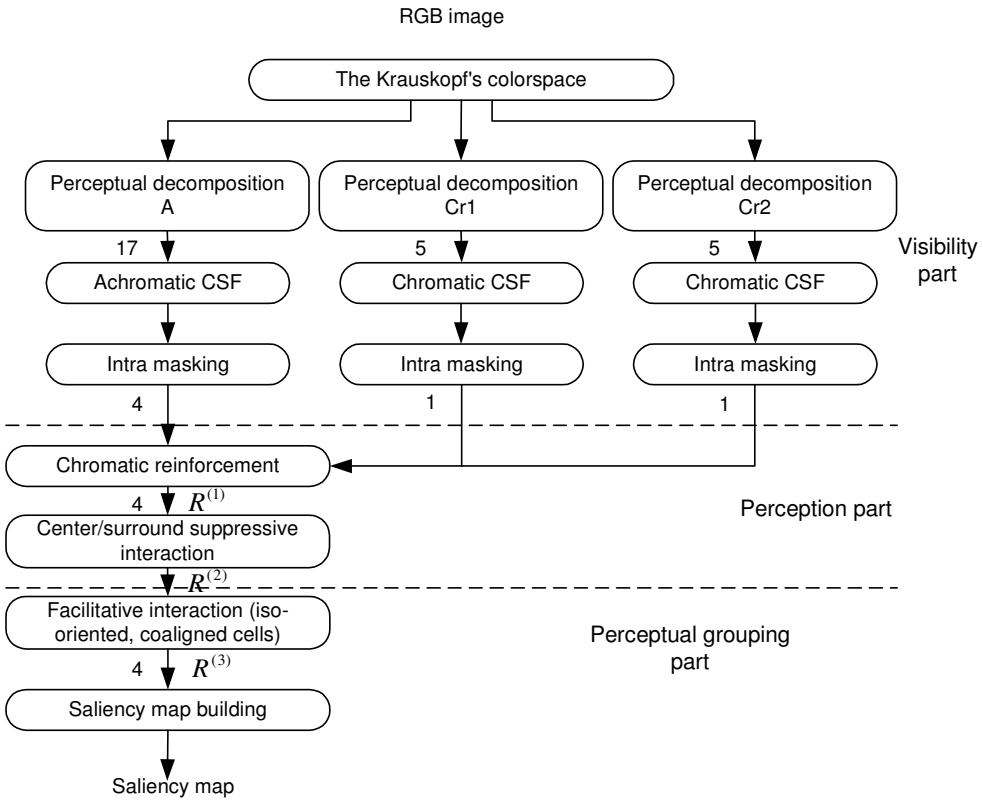


figure 2-1 : general synoptic of our model.

3. VISIBILITY OR PREATTENTIVE COMPUTATION OF VISUAL FEATURES

The first processing stage in any model of bottom-up attention is the computation of early visual features in a massively parallel manner in accordance with the feature integration theory of Treisman and all [4].

3.1. Color space definition

The first step consists in converting the RGB luminances into the Krauskopf's opponent-colors space composed by the cardinal directions A, Cr1 and Cr2 [5].

This transformation to the opponent-colors space is a way to decorrelate color information. In fact, it is believed that the brain uses 3 different pathways to encode information : the first conveys the luminance signal (A), the second the red and green opponent component (Cr1) and the third the blue and yellow opponent component (Cr2). Before this transformation, each of the three components RBG undergoes a power-law nonlinearity (called gamma law) of the form x^{γ} with $\gamma \approx 2.4$ in order to take into account the transfer function of the display system when producing luminance signals.

3.2. Perceptual subband decomposition

A perceptual subband decomposition is then applied to each of these 3 components. This decomposition, largely inspired from the cortex transform stemming from Watson's work [6] and modified by Senane [7] on the basis of different psychophysics experiments, is obtained by carving up the 2D spatial frequency domain both in spatial radial frequency and orientation, as shown for the A component in figure 3-1: the perceptual decomposition of the component A leads to 17 psychovisual subbands distributed on 4 crowns. Concerning the decomposition of the two chromatic components as shown in figure 3-1, the decomposition leads to 5 psychovisual subbands distributed on 2 crowns [8] for each of these components. The shaded region on the figure 3-1 indicates the spectral support of the subband belonging to the third crown and having an angular selectivity of $\pi/6$.

The main properties of these decompositions and the main differences from the cortex transform are a non dyadic radial selectivity and an orientation selectivity which increases with the radial frequency (except for the chromatic components).

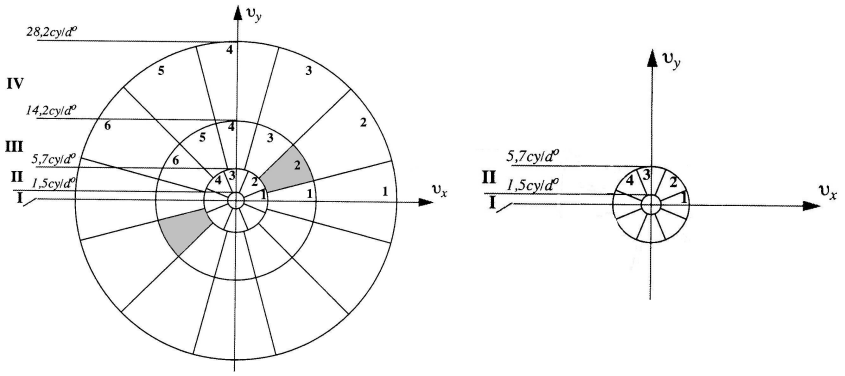


figure 3-1 : psychovisual spatial frequency partitioning for the A (achromatic) component (left) and for the two chromatic components (right).

Each of the resulting subbands may be regarded as the neural image corresponding to a population of visual cells tuned to a range of spatial frequency and a particular orientation. In fact, those cells belong to the primary visual cortex (also called striate cortex or V1 for visual area 1). It consists of about 200 million neurons in total and receives its input from the lateral geniculate nucleus (LGN). About 80 percent of the cells are selective in orientation and in spatial frequency of the visual stimulus.

In a first approach, we just take advantage of the subbands stemming from the second crown of the A component and the low frequency subband of the chromatic components. We indifferently use the terms of subband or channel in order to define a partition of the frequency domain throughout the paper.

3.3. Contrast Sensitivity Function (CSF)

Biological evidences have shown that visual cells respond to stimuli above a certain contrast. The contrast value for which a visual cell response is called the visibility threshold (above this threshold, the stimuli is visible). This threshold varies with many 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.

The CSFs we applied on the images spatial spectrum are multivariate functions mainly depending on the spatial frequency, the orientation and the viewing distance. On the achromatic component, we apply a 2D anisotropic CSF designed by Dally. On the chromatic components, 2D anisotropic CSFs designed by Le Callet [8,10], one for each component, are applied (two low pass filters with cut-off frequency of about 5.5 cpd (cycle per degree) and 4.1 cpd for Cr1 and Cr2 component respectively).

3.4. Intra masking

Intra channel masking is then incorporated as a weighing of the outputs of the CSF function. Masking is a very important phenomenon in perception as it describes interactions between stimuli. Indeed, the visibility threshold of a stimulus can be affected by the presence of another stimulus.

Masking is the strongest between stimuli located in the same perceptual channel or in the same subband. We apply the intra masking function designed by Dally [9] on the achromatic component and the intra masking function on the color component designed by Le Callet [10]. These masking functions consist of non linear transducer as expressed by Legge and Foley [11]

Despite the fact that psychophysical experiments have shown that masking also occurs between channels of different orientations as well as between channels with different spatial frequency and between chrominance and luminance channels, this model only takes into account the intra-channel masking.

4. PERCEPTION OR DETERMINATION OF RELEVANT LOCATIONS

To underscore conspicuous locations in an image, we have to take into account several important phenomenons.

The classical receptive field (CRF) is a crucial concept for systems based on HVS. The concept of CRF, first introduced in 1938 by Hartline [12], permits to establish a link between a retinal image and the global percept of the scene. The CRF is defined as a particular region of visual field within which an appropriate stimulation (with preferred orientation and frequency) gives a relevant response stemming from visual cells. Consequently, by principle, a stimulus in the outer region (called surround) cannot activate the cell directly.

In our proposed model, the CRF behavior of retinal and Lateral Geniculates Nucleus (LGN) cells is achieved by the mechanisms of the visibility part briefly described previously.

In fact, although cells respond directly only to stimuli within their CRFs, the activity of cells may be modulated by contextual stimuli outside (but near) their CRFs [13-16] in primary visual cortex. The contextual influences coming from intracortical interactions can be suppressive or facilitative depending on the stimuli configuration [13-15]. These suppressive or facilitative effects are commonly referred to as nonclassical receptive filed (non-CRF) inhibition.

Moreover, to enhance relevant visual features on achromatic maps, we introduce a chromatic reinforcement since chromatic components are undeniably considered as preattentive features [1] and theirs contributions in a model of bottom-up can be essential.

4.1. Chromatic reinforcement

The existence of areas showing a sharp color (thus with a sufficient saturation) and fully surrounded of areas having quite other colors and rather little saturated, implies a particular attraction of focusing on the very coloured area [17].

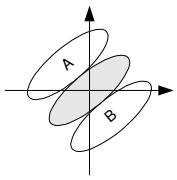
At this stage, the achromatic subbands $R_{i,j}(x, y)$ of the second crown (where i represents the spatial radial frequency band $i \in \{I, II, III, IV\}$, j pertains to the orientation $j \in \{1,2,3,4,5,6\}$ and (x, y) the spatial coordinates) undergo a reinforcement based on the color subbands of the first crown.

Chromatic reinforcement can thus be achieved by the following equation :

$$R_{i,j}^{(1)}(x,y) = R_{i,j}(x,y) \times (1 + |A-B|_{Cr1} + |A-B|_{Cr2})|_{i=II}$$

where,

$R_{i,j}^{(1)}(x,y)$ represents the reinforced achromatic subband and $R_{i,j}(x,y)$ represents an achromatic subband.



$|A-B|_k$ represents the contrast value computed around the current point on the chromatic component k in the preferred orientation of the subband $R_{i,j}(x,y)$, as shown beside. In the example given beside, the sets A and B belong to the subband of the first crown (low frequency subband) of the chromatic component k with an orientation equal to $\pi/4$.

4.2. Modeling center-surround suppressive interactions for neurons in primary visual cortex

To qualitatively reproduce the non-CRF inhibition behavior of cells in primary visual cortex, we have to consider two properties.

The first one is that the non-CRF inhibition is strongest between cells belonging to the same subband (tuned to the same preferred orientation [16] and spatial frequencies [16,18]). The inhibition term is computed in an inhibitory areas adjacent of the CRF flanks, usually termed "side-inhibition". In this first approach, we disregard a second term of inhibition (termed "end-stopping") stemming from the preferred orientation axis. This first property is commonly called anisotropic non-CRF inhibition.

The second property concerns the fact that inhibition depends on the distance from the center [19] : it appears strongest at a particular distance from the center and decreases with the eccentricity.

A two-dimensional Difference-of-Gaussians (DoG) could be favorably used to model the non-CRF inhibition behavior of cells. The $DoG_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y)$ is given by the following equation :

$$DoG_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y) = G_{\sigma_x^{inh},\sigma_y^{inh}}(x,y) - G_{\sigma_x^{ex},\sigma_y^{ex}}(x,y)$$

with $G_{\sigma_x,\sigma_y}(x,y) = \frac{1}{2\pi\sigma_x\sigma_y}\exp(-\frac{x^2}{2\sigma_x^2} - \frac{y^2}{2\sigma_y^2})$ a two-dimensional Gaussian.

Parameters $(\sigma_x^{ex}, \sigma_y^{ex})$ and $(\sigma_x^{inh}, \sigma_y^{inh})$ correspond to spatial extends of the Gaussian envelope along the x and y axis of the central Gaussian (the CRF center) and of the inhibitory Gaussian (the surround) respectively. These parameters have been experimentally determined in accordance with the radial frequency of the second crown (the radial frequency $f \in [1.5,5.7]$ is expressed in cycles/degree). Finally, the non-classical surround inhibition can be modeled by the normalized weighting function $w_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y)$ given by the following equation :

$$w_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y) = \frac{1}{\left\| H(DoG_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}) \right\|_1} H(DoG_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x',y'))$$

with,

$$H(z) = \begin{cases} 0, & z < 0 \\ z, & z \geq 0 \end{cases},$$

(x',y') is obtained by translating the original coordinate system by (x_0, y_0) and rotating it by $\theta_{i,j}$ expressed in radian,

$$\begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} \cos\theta_{i,j} & \sin\theta_{i,j} \\ -\sin\theta_{i,j} & \cos\theta_{i,j} \end{bmatrix} \begin{bmatrix} x - x_0 \\ y - y_0 \end{bmatrix},$$

$\|.\|_1$ denotes the L_1 norm.

The figure 4-1 shows on hand the structure of non-CRF inhibition and on the other hand a profile example of the normalized weighting function $w_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y)$.

The response $R_{i,j}^{(2)}(x,y)$ of cortical cells to a particular subband $R_{i,j}^{(1)}(x,y)$ is computed by the convolution of the subband $R_{i,j}^{(1)}(x,y)$ with the weighting function $w_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y)$ ($H(z)$ has been previously described):

$$R_{i,j}^{(2)}(x,y) = H(R_{i,j}^{(1)}(x,y) - R_{i,j}^{(1)}(x,y) * w_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y))\Big|_{i=II}$$

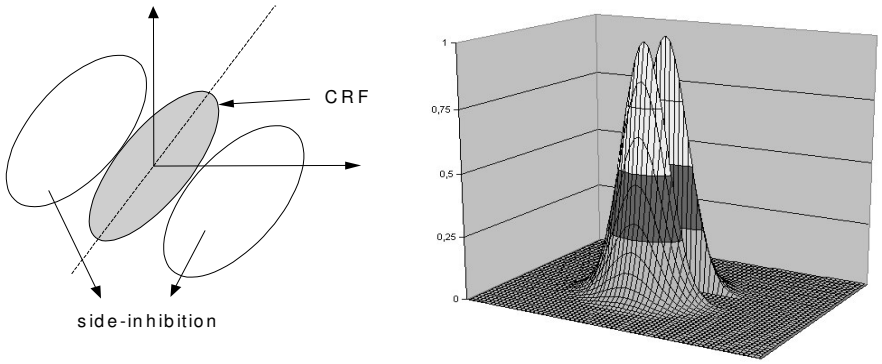


figure 4-1 : On the left : Non-CRF inhibition produced by the adjacent areas of the CRF flanks. On the right : a profile example of the normalized weighting function $w_{\sigma_x^{ex},\sigma_y^{ex},\sigma_x^{inh},\sigma_y^{inh}}(x,y)$ for a particular orientation and radial frequency.

5. DOMAIN OF PERCEPTUAL GROUPING

5.1. Modeling center-surround facilitative interactions for neurons in primary visual cortex

Facilitative interactions have also been reported in a number of studies. In most cases, they appear outside the CRF along the preferred orientation axis. These kinds of interactions are maximal when center and surround stimuli are iso-oriented and co-aligned [20-22]. In other words, as shown by several physiologically observations, the activity of cell is enhanced when the stimuli within the CRF and a stimuli within the surround are bound to form a contour. This facilitative interaction is usually termed contour enhancement or contour integration.

We attempt to simulate contour integration in early visual preprocessing using two halves butterfly filter $B_{i,j}^0$ and $B_{i,j}^1$. The profiles of these filters are shown in the figure 5-1 and they are defined by using a bipole/butterfly filter. It consists of a directional term $D_\theta(x,y)$ and a proximity term generated by a circle C_r blurred by a gaussian filter $G_{\sigma_x,\sigma_y}(x,y)$.

$$B_{\theta_{i,j},\alpha,r,\sigma}(x,y) = D_{\theta_{i,j}}(x,y) \cdot C_r * G_{\sigma_x,\sigma_y}(x,y)$$

with $D_{\theta_{i,j}}(x,y) = \begin{cases} \cos(\frac{\pi/2}{\alpha}\varphi) \text{ if } \varphi < \alpha \\ 0 \text{ otherwise.} \end{cases}$

and $\varphi = \arctan(y'/x')$, where $(x',y')^T$ is the vector $(x,y)^T$ rotated by $\theta_{i,j}$. The parameter α defines the opening angle 2α of the bipole filter. It depends on the angular selectivity γ of the considered subband. We took $\alpha = 0.4 \times \gamma$. The size of the bipole filter is about twice the size of the CRF of a visual cell.

The figure 5-2 shows a plot of a bipole filter for a particular orientation. The two half butterfly filter $B_{i,j}^0$ and $B_{i,j}^1$ are after deduced from the butterfly filter by using appropriate windows.

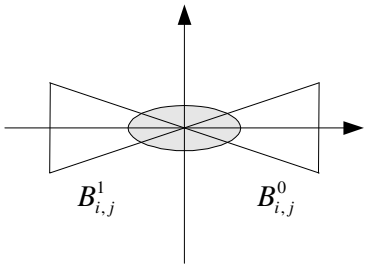


figure 5-1 : profiles filters to model facilitative interactions ($\theta = 0°$).

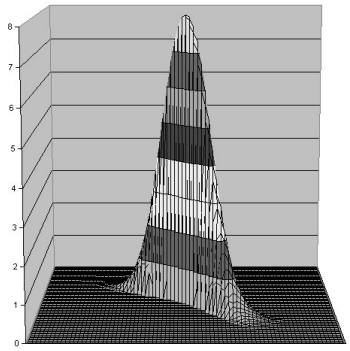


figure 5-2 : profile of a bipole filter for a particular orientation.

For every orientation, subband and location, we compute the facilitative coefficient :

$$f_{i,j}^{iso}(x,y) = \frac{L_{i,j}^1(x,y) + L_{i,j}^0(x,y)}{\max(\beta, \left| L_{i,j}^1(x,y) - L_{i,j}^0(x,y) \right|)}$$

with :

β , a constant

$L_{i,j}^0(x,y) = R_{i,j}^{(2)}(x,y) * B_{i,j}^0(x,y)$ and $L_{i,j}^1(x,y) = R_{i,j}^{(2)}(x,y) * B_{i,j}^1(x,y)$

The subband $R_{i,j}^{(3)}$ issued from the facilitative interaction is finally obtained by weighting the subband $R_{i,j}^{(2)}$ by a factor depending on the ratio of the local maximum of the facilitative coefficient $f_{i,j}^{iso}(x,y)$ and on the global maximum of the facilitative coefficient computed on all subbands belonging to the same range of spatial frequency ($i = II$):

$$R_{i,j}^{(3)}(x,y) = R_{i,j}^{(2)}(x,y) \times (1 + \eta^{iso} \times \frac{\max_{(x,y)}(f_{i,j}^{iso}(x,y))}{\max_{j}(\max_{(x,y)}(f_{i,j}^{iso}(x,y)))} f_{i,j}^{iso}(x,y))$$

From a standard butterfly shape, this facilitative factor permits to improve the saliency of isolated straight lines. η^{iso} controls the strength of this facilitative interaction.

5.2. Saliency map building

All resulting subbands are simply summed into a unique saliency map :

$$S(x,y) = \sum_{i=II,j} R_{i,j}^{(3)}(x,y)$$

Although cortical cells tuned to horizontal and vertical orientations are almost as numerous as cells tuned to other orientations, we do not introduce any weighting. This feature of the HVS is implicitly mimic by the application of 2D anisotropic CSF. figure 5-3 shows an example of saliency map obtained on the Tennis Table picture.

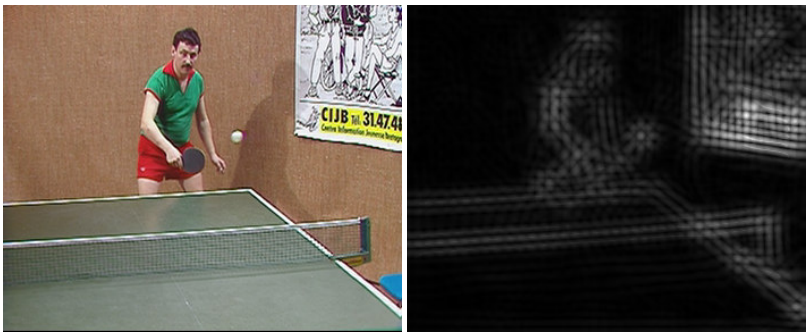


figure 5-3 : the saliency map of the Table picture.

6. RESULTS AND DISCUSSION

The figure 6-1 illustrates the results of applying our model to four images (frame size is 352x288). On these images, we only show the first 20 fixation points. In order to select the most salient locations in the picture, a winner-take-all strategy is used.

In others words, as described by Koch and Ullman [2], we determine the maximum of the saliency map and the location of this maximum is the winner. This location and its neighbors (locations belonging to a circle of radius ψ centered on the current maximum) are inhibited (set to null). As a consequence, the next most conspicuous location will be chosen. In figure 6-1, a red circle represents a fixation point.

These results seem to be good in the sense that there is only few implausible fixation points. Except for the picture Stefan, majority of fixation points is concentrated on or near the region of interest (on the kayak, on the man in the picture tennis table man, on the skaters for the pictures Kayak, Table and Patin respectively). In general, our personal perception of the salient locations is in good agreement with the fixation points generated by our model.

In order to assess our model and to refine our own opinion about the results, a comparison with a purely model of bottom-up visual attention considered as a reference is achieved. This bottom-up model has been developed by Itti, Kock and Niebur and has been shown to be very effective with different kinds of pictures [23-25].

In such model early visual features (color, intensity and orientation) are extracted, in a massively parallel manner, in several multiscale feature maps. A normalization step which could be an iterative within-feature spatial competition based on a Difference-of-Gaussians (DoG) or a simple normalization scheme (described below) is applied on each feature maps. All feature maps are then combined into an unique saliency map. A winner-take-all algorithm coupled with an inhibition of return strategy permits to define a hierarchy of the most salient locations. The results shown in the figure 6-2 were obtained from the latest software version of the Itti's model by using a simple spatial competition mechanism named Maxnorm described in [23]. This normalization mechanism is based on the squared ratio of global maximum over average local maximum. The goal of this step is to promote feature maps with few conspicuous locations to the detriment of maps presenting numerous conspicuous locations.

To be fair in the comparison, we have taken the same number of fixation points and the same radius of the focus of attention (25 pixels close 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, there is a good agreement between the two models if we disregard the saliency ordering. Nevertheless, it's not easy to give an objective measurement which permits to yield a reliable similarity degree between the two sets of fixation points [26-29].

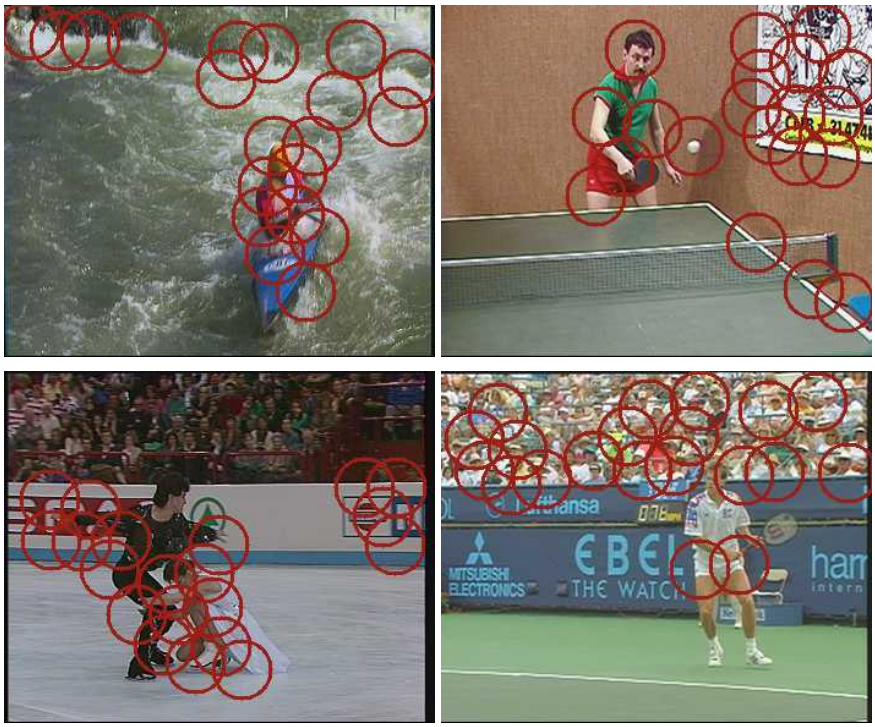


figure 6-1 : the first 20 fixations points on Kayak (top left), Table (top right), Patin (bottom left) and Stefan (bottom right)

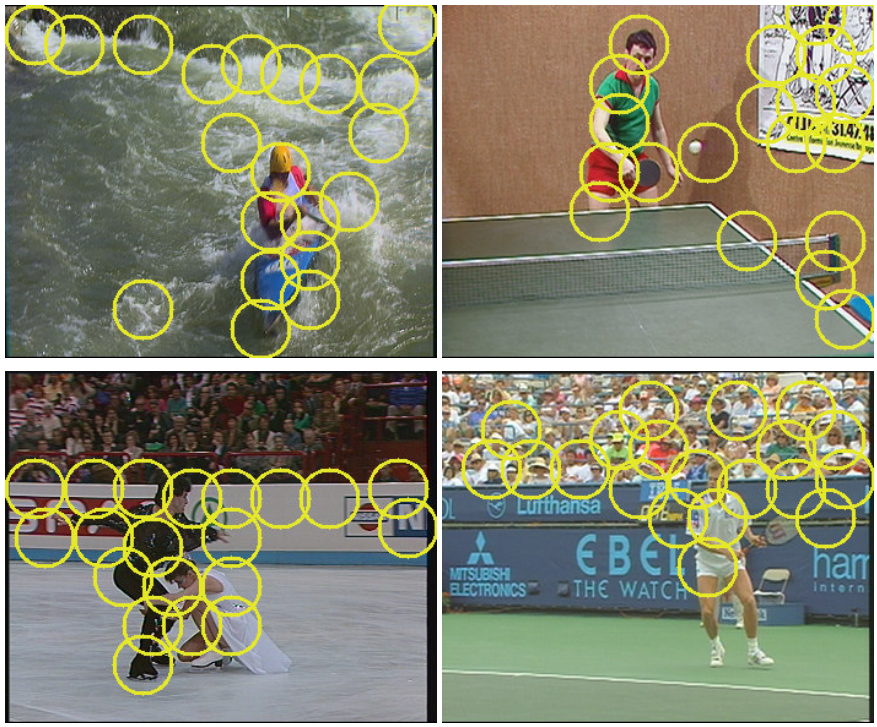


figure 6-2 : the first 20 fixation points obtained by Itti's model.

In this study, to simply assess the similarity of traces, we just count the number of a pair of fixation points having almost the same coordinate. In fact, the method we used is based on the Euclidian distance and is closed to method developed by Privitera and Stark [Pri 00]. For a particular point, we compute the distances between this point and all points belonging to the other set. We find the minimum distance and the analysis consists in setting a threshold (equal to the radius of the circle) below which fixation points were said to be coincident and above which they were not. The results are shown on the first row of the table 6-1. If we disregard the hierarchy of the fixation points, the percentage of overlapping points computed on the whole sets is greater than 75%. These results confirm our subjective assessment in the way that there is a pretty good overlap between the two sets. If we reduced the size of the set to the top 5 locations, the percentage notably decreased.

In fact, in order to take into account the fact that the hierarchy of fixation points could be differ between the two models, we have made a third assessment : we just compared the first 5 locations of the two sets of fixation points with keeping the saliency ordering. The point matching is achieved by taking the first points of each set and by computing the Euclidian distance. If this distance is smaller than the same previous threshold, we consider that the two points are matched (and so on for the remaining points). The results are shown on the third row of the table 6-1. So, when we keep the saliency ordering, the percentage of overlapping points is null. That's mean that the hierarchy of the conspicuous locations differs between the two models. Nevertheless, from the second test (the second row of the table 6-1), we can assert that, on the top 5, the most salient points stemming from the two models are not so far.

	Kayak	Table	Patin	Stef
Percentage of overlapping points without keeping the hierarchy (computed on the whole sets)	80	90	75	85
Percentage of overlapping points without keeping the hierarchy (computed on the top 5)	40	60	0	40
Percentage of overlapping points with keeping the hierarchy (computed on the top 5)	0	0	0	0

table 6-1 : the similarity degree between the two models.

CONCLUSION

In this paper we propose a purely bottom-up model of visual attention working on still color images. A deep understanding of the numerous properties of the HVS and biological evidences leads us to design a bottom-up model fully based on psychovisual models. Nevertheless, the structure of our model does not differ from others bottom-up models in the way that we build it around the Koch and Ullman's hypothesis [2] which is closely related to the feature integration theory of Treisman and colleagues [4]. The architecture of our model tackles three fundamental notions of the HVS: the visibility, the perception and the domain of perceptual grouping.

The visibility expresses the fact that we are not able to perceive all details of our environment with the same accuracy. The following process, called the perception, suppresses all irrelevant information. Several experiments clearly show that a target which differs significantly from its surround easily grabs our visual attention. The last process related to the domain of perceptual grouping is indisputably the more difficult to approach. In order to built a global percept of the visual scene, we have to bind early visual features. In this paper, we only deal with contour enhancement also called contour integration. It consists in enhancing long and thin areas which are considered to be visual attractors.

Simulations of this model have been presented and compared with the results stemming from the Itti's model. There is a pretty good agreement between the two models. To achieve a good assessment of our model, we have now to make a comparison with "real" fixation points obtained from an eye tracker device.

Future works will give us the opportunity to improve our model on still pictures. Nevertheless, the most important modification will concern the integration of the temporal dimension. The velocity of motion is an uncontroversial factor that influences our visual attention. It is intuitively clear that it will be easy to find a moving stimulus among stationary distractors.

ACKNOWLEGMENT

We are grateful to Laurent Itti for his help and for his agreement concerning this performance comparison.

REFERENCES

[1] J. M. Wolfe, *Visual Search*, pp. 13-74, in Pashler, H. editor, Attention, Psychology Press, 1998.

[2] C. Kock and S. Ullman, "Shifts in selective visual attention: towards the underlying neural circuitry", *Human Neurobiology*, **vol. 4(4)**, pp. 219-27, 1985.

[3] D. Marr, "Vision: a computational investigation into the human representation and processing of visual information", W.H. Freeman and Compagny, NY 29-61, 1982.

[4] A. M. Treisman and G. Gelade, "A feature integration theory of attention", *Cognit Psychol*, **vol. 12(1)**, pp. 97-136, 1980.

[5] J. Krauskopf, D.R. Williams, and D.W. Heeley, "Cardinal directions of color space", *Vision Research*, **vol. 22**, pp. 1123-1131 1982.

[6] A. B. Watson, "The cortex transform : rapid computation of simulated neural images", *Computer Vision, Graphics and Image Processing*, vol. 39, pp. 311-327, 1987.

[7] H. Senane, A. Saadane and, D. Barba, "The computation of visual bandwiths and their impact in image decomposition and coding", *International Conference and Signal Processing Applications and Technology*, Santa-Clara, pp. 776-770, 1993.

[8] P. Le Callet, A, Saadane, and D. Barba, "Orientation selectivity of opponent-colour channels", *in Perception*, vol. 28 supplement ECVP'99 abstracts, Trieste, 1999.

[9] S. Daly, "The visible different predictor : an algorithm for the assessment of image fidelity", *in Proc. of SPIE Human vision, visual processing and digital display III*, vol. 1666, pp. 2-15, 1992.

[10] P. Le Callet, and D. Barba, "Frequency and spatial pooling of visual differences for still image quality assessment", *in Proc. SPIE Human Vision and Electronic Imaging Conference*, vol. 3959, San Jose, January 2000.

[11] Gordon E. Legge and John M. Foley, "Contrast Masking in Human Vision", *Journal of the Optical Society of America*, v**ol. 70**, N° 12, pp. 1458-1471, December 1980.

[12] H. K. Hartline, "The response of single optic nerve fibers of the vertebrate eye to illumination of the retina", american *Journal of Physiology*, **vol.121**, pp. 400-415, 1938.

[13] Sillito A.M., Grieve K.L., Jones H.E, Cudeiro J. and Davis, "Visual Cortical Mechanisms Detecting Focal Orientation Discontinuities", *Nature*, **vol. 378**, pp.492-496, 1995.

[14] Knierin J.J. and Van Essen D. C., "Neural responses to static texture patterns in area v1 of the alert macaque monkey", *Journal of Neurophysiology*, **vol. 67**:961-980, 1992.

[15] DeAngelis G.C., Freeman R.D., and Ohzawa I., "Length and width tuning of neurons in the cat's primary visual cortex*", Journal of Neurophysiology*, **vol. 71**, pp.347-374, 1994.

[16] 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.

[17] E. Niebur and C. Koch, "Computational architectures for attention", In R. Parasuraman, editor, The Attention Brain, MIT Press, Cambridge, MA, 1997.

[18] G. A. Walker, I. Ohzawa, R.D. Freeman, "Asymetric suppression outside the classical receptive field of the visual cortex", *Journal Neuroscience*, **vol. 19**, pp. 10536-10553, 1999.

[19] B. Zenger, and D. Sagi, "Isolating excitatory and inhibitory nonlinear spatial interactions involved in contrast detection", *Vision Research*, **vol. 36(16)**, pp. 2497-513, 1996.

[20] Kapadia, M.K and al., "Improvement in visual sensitivity by changes in local context: parallel studies in human observers and in v1 of alert monkeys", Neuron 15, 843-856, 1995.

[21] J. I. Nelson, B.J. Frost, "Intracortical facilitation among co-oriented, co-axially aligned simple cells in cat striate cortex", Exp brain Res, 61, pp. 54-61, 1985.

[22] U. Polat, K. Mizobe, M. W. Pettet, T. Kasamatsu, A.M. Norcia, "Collinear stimuli regulate visual responses depending on cell's contrast threshold", *Nature*, **vol. 391**, pp. 580-584, 1998.

[23] 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.

[24] L. Itti and, C. Koch, "A comparison of feature combination strategies for saliency-based visual attention systems", *Human Vision and Electronic Imaging IV, SPIE*, vol. 3644, pp. 373-382, 1999.

[25] L. Itti and, C. Koch, "A saliency-based search mechanism for overt and covert shifts of visual attention", *Vision Research*, **vol. 40**, 10-12, pp. 1489-1506, 2000.

[26] C.M. Privitera, and L.W. Stark, "Algorithms for Defining Visual Regions-of-Interest : Comparison with Eye Fixation", *IEEE Transactions on Pattern Analysis & Machine Intelligence*, **vol. 22**, pp. 970-982, 2000.

[27] S. Mannan, K.H. Ruddock, and D.S. Wooding, "Automatic control of saccadic eye movement made in visual inspection of briefly presented 2-d images", *Spatial Vision*, **vol. 9**, pp. 363-386, 1995.

[28] M. Pomplum, H. Ritter, and B. Velichkovsky, "Disambiguating complex visual information : Towards communication of personal views of a scene", Perception, 25, pp. 931-948, 1996.

[29] S.A. Brandt, and L.W. Stark, "Spontaneous eye movements during visual imagery reflect the content of the visual scene", *Journal of Cognitive Neuroscience*, **vol. 9**, pp. 27-38, 1997.