

Research Article

Pop-Out: A New Cognitive Model of Visual Attention That Uses Light Level Analysis to Better Mimic the Free-Viewing Task of Static Images

Makiese Mibulumukini

SVP TV, 7000 Mons, Belgium

Correspondence should be addressed to Makiese Mibulumukini; makiese.mibulumukini@gmail.com

Received 9 October 2014; Revised 3 December 2014; Accepted 3 May 2015

Academic Editor: Djamel Bouchaffra

Copyright © 2015 Makiese Mibulumukini. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Human gaze is not directed to the same part of an image when lighting conditions change. Current saliency models do not consider light level analysis during their bottom-up processes. In this paper, we introduce a new saliency model which better mimics physiological and psychological processes of our visual attention in case of free-viewing task (bottom-up process). This model analyzes lighting conditions with the aim of giving different weights to color wavelengths. The resulting saliency measure performs better than a lot of popular cognitive approaches.

1. Introduction

Saliency models are more and more important in computer vision due to their fundamental contribution in intelligent systems (robotics [1], serious games [2], intelligent video surveillance [3], etc.). Indeed, the mechanisms of visual attention they are supposed to mimic enable selection of relevant contextual information [4, 5] and lead to more autonomous systems.

If it is known that the human visual attention can be deployed unintentionally on the scene (bottom-up mechanism) or subjectively (top-down mechanism), the cognitive and biological development of this process remain the subject of scientific investigation.

In recent decades, many facets of science have been aimed towards answering this question. An early psychological model, which strongly influenced the field of visual attention, is the *Features Integration Theory* (FIT) [6]. Treisman and Gelade suggest that some image features (colors, orientations, etc.) are first processed in parallel in order to build a master map of location that draws our attention to an area (*where*) of the scene. Object recognition takes place after focusing attention on the *where* and requires inhibition of feature maps that do not describe the searched target. This theory also

assumes that our attention is deployed sequentially on each stimulus present in the scene. This finding has been refuted by several studies that found that, during a task (in top-down case), our attention can be deployed to 4 or 5 regions simultaneously [7, 8].

An improved version of the FIT is the *Guided Search Model* of Wolfe [9]. In addition to selecting the features that better describe the target, top-down bias are introduced for highlighting features that better discriminate target from its distractors (similar objects).

Other approaches are subject to connectionist strategies during their processes of visual attention [10–14]. In those models, a neural network describes our visual attention by inhibition and excitation mechanisms that allow the emergence of an area of the scene.

Some models of visual attention use other kinds of processes: *The Gestalt theory* [15, 16] proposed by Christian von Ehrenfels in 1890 [17]. The Gestalt is both psychological and philosophical theory which maintains that perception and mental representation spontaneously treat phenomena as structured sets (forms) and not as a simple addition or juxtaposition of elements (features).

In this paper we propose a bottom-up saliency model which provides a lot of explanations about both biological and

cognitive mechanisms of visual attention. Section 2 provides a state of the art in cognitive saliency models (by cognitive we mean biologically plausible model that describes not only the psychological mechanisms of visual attention but also those physiological processes [18]). We just focus on bottom-up axis and computational models in this topic. Section 3 allows us to introduce a new biologically inspired bottom-up model. Section 4 is reserved to the tests of correlations between cognitive models and human visual attention. Analyses are also made. Section 5 provides discussion and conclusion.

2. State of the Art in Cognitive Models of Bottom-Up Visual Attention

Cognitive models have the advantage of expanding our view of biological underpinnings of visual attention [18].

One of the first computational attention models was proposed by Koch and Ullman [19]. This model is based on the FIT. Different features are filtered and combined to form a final saliency map where a neural network (*winner takes all or WTA*) indicates the more salient region. This model is the basis for several implementations such as Clark and Ferrier [20] where feature maps are summed and weighted according to their saliency.

One of the most popular models inspired by the architecture of [19] is Itti et al. [21]. Gaussian pyramid filtering of each feature (color, intensity, and orientation) is added. Other retinal mechanisms like center periphery [22] are introduced before normalization and combination of resulting filtered maps into a final saliency map.

The model of [21] is very popular because it is well documented and freely available online. It was also the first computational model to yield interesting results in the free-viewing task.

Other models like VOCUS [23] use the architecture of [21]. In fact, VOCUS uses a LAB space where the attention process is described in the same manner as Itti et al.'s architecture.

Another computational model based on the FIT was proposed by Le Meur et al. [24]. Contrast sensitivity functions, perceptual decomposition, visual masking, and center-surround interactions are some of the features implemented in this model. In [24], we have three aspects of the vision: visibility, perception, and perceptual grouping. The visibility part simulates the limited sensitivity of the human eyes and takes into account the major properties of the retinal cells. The perception is used to suppress the redundant visual information by simulating the behavior of cortical cells. And the final saliency map is enhanced by the perceptual grouping.

Later, Le Meur et al. extended this model to the spatio-temporal domain [25] by fusing achromatic, chromatic, and temporal information. The architecture of Koch and Ullman also allowed the implementation of models such as Guironnet et al. [26]. The static component extracts common orientations with Gabor filters. The frame difference allows a temporal component to detect moving objects.

Gestalt principles were first implemented by Zou et al. [27] and then by Kootstra et al. [15, 16, 28].

In the last decade, connectionist theories have not been subject to a computational implementation in case of static images. Some researchers have implemented connectionist theories by using dynamic neural fields [29, 30] on video. The connectionist approaches use a microscopic description of cells involved in visual attention, in particular by using a neural network for modeling visual attention. FIT-based approaches (which typically use filter banks) are macroscopic because they describe visual attention without take care of the details of the neural topography of human cortical cells.

We preferred the FIT-based approach to connectionist ones because the functioning of cortical cells described by macroscopic models is better known than the exact mechanism involved in the interaction of different neurons modeled by connectionist models. The *Features Integration Theory* also provides a better understanding of the physiological process of visual attention compared to The Gestalt theory that does not give enough information about its physiological way to deal with visual attention.

So, in this paper we propose to integrate light level analysis and a better description of visual cortex in the process of attention. Our model is based on the FIT but its main advantage is to evolve visual attention mechanisms with lighting conditions. Another level of vision which gives the essence of a scene (furthermore we call this mode *Perceptual*) is also introduced.

3. Pop-Out: A New Cognitive Model of Visual Attention That Uses Light Level Analysis to Better Mimic the Free-Viewing Task of Static Images

Our vision is impaired in the dark, and we are highly sensitive to the light intensity emitted by colors hue (see Figure 2). Luminance is the feature which contains most of the visual information because the finest area of our eye, called fovea, is mainly photosensitive [32, 33]. The luminous efficiency (or the magnitude of the intensity perceived according to a given wavelength) is strongly affected by the color of the light source [34, 35]. In photopic vision, our fovea is more sensitive to the red wavelength while the blue wavelength influences scotopic vision. Between day and night illumination conditions, there is another light level called *mesopic vision* where red and blue hues are competitive [36] (see Figure 3).

To modelize these findings (or the light and color intensity sensitivities), we introduce an improved $Y C_b C_r$ space which recovers the whole luminance of the scene and remains selective to color sensitivity. Our improved $Y C_b C_r$ space allows us to make a suitable combination of C_b and C_r components according to the result of the light level analysis (this analysis let us know if the picture was taken in photopic, scotopic, or mesopic condition). All $Y_l C_b C_r$ components are filtered by Log-Gabor functions (see Figure 1). In fact, Log-Gabor wavelets better describe the receptive fields (RF) or the simple cells responses of the V1 area of our visual cortex [40–43]. A module of light level analysis detects the luminosity of

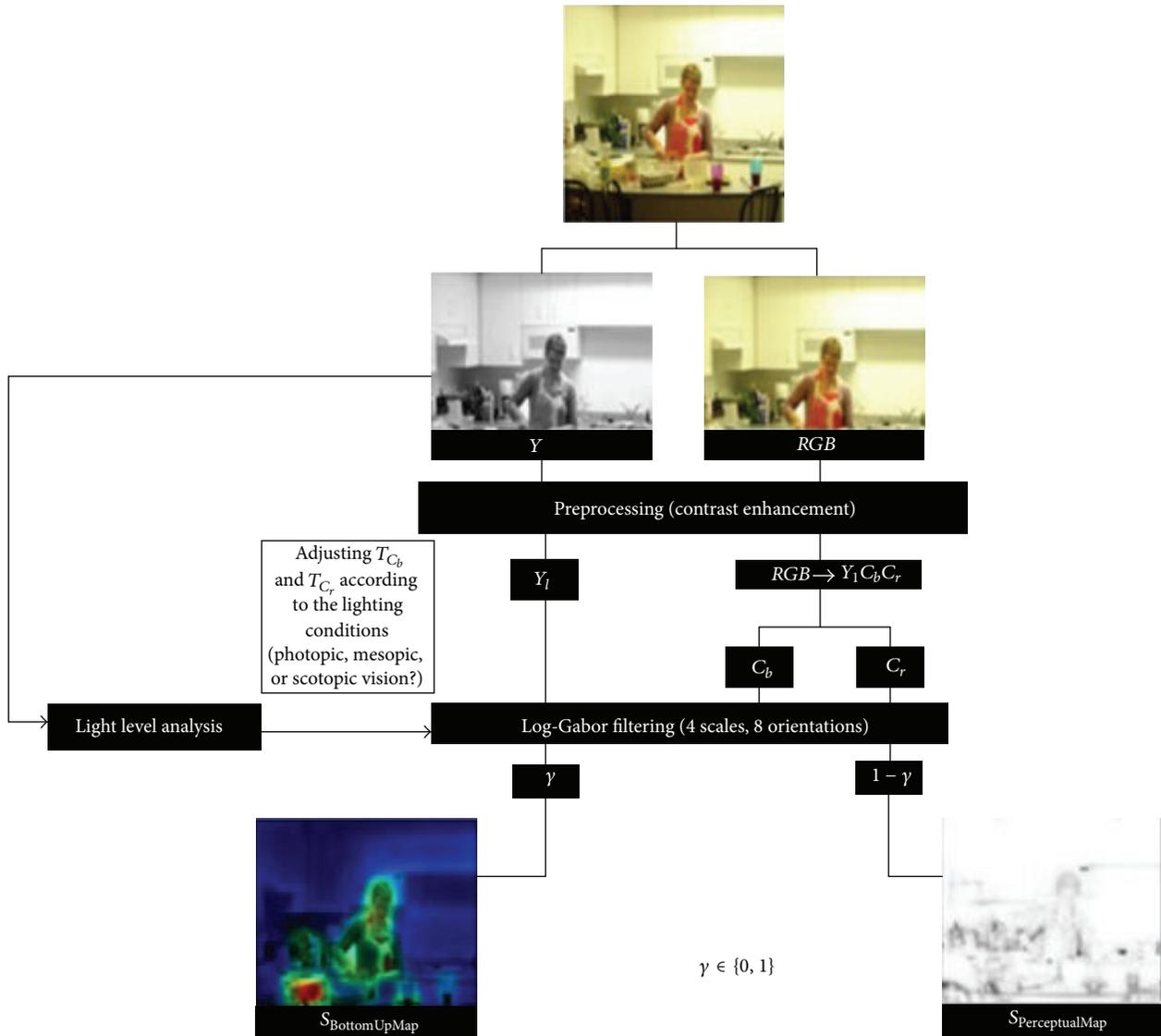


FIGURE 1: The global architecture of pop-out. The system can switch between bottom-up and perceptual fusion processes by changing the value of γ . The initial image is taken from the MIT dataset [31]. The first step consists in extracting the image intensity (Y) and RGB components that will be enhanced during contrast preprocessing. Before contrast enhancement, Y is also used for light level analysis in order to detect in which lighting conditions the initial image was captured (the main goal of this analysis is to know if the image is taken in photopic, mesopic, or scotopic conditions). After contrast enhancement of Y , we obtain Y_l that will be used for Log-Gabor filtering in order to get the texture of the scene. RGB image enhanced is also used to extract C_b and C_r components. After Log-Gabor filtering of C_b and C_r components, we get T_{C_b} and T_{C_r} that are, respectively, the blue and the red color information of the image texture. According to the result of the light level analysis (photopic, mesopic, or scotopic condition detected), different weights are given to T_{C_b} and T_{C_r} . To obtain the $S_{\text{BottomUpMap}}$, T_{C_b} and T_{C_r} are used as a mask that is combined with the luminance texture. The $S_{\text{BottomUpMap}}$ represents the visual attention predicted in case of free-viewing task. $\gamma = 0$ when we just use the luminance texture as visual attention map (case of $S_{\text{PerceptualMap}}$). The $S_{\text{PerceptualMap}}$ is just based on the edge sensitivity or the luminance texture. The $S_{\text{BottomUpMap}}$ takes into account the color components filtered (T_{C_b} and T_{C_r}).

the scene and gives different weights to C_b or C_r component after Log-Gabor filtering. C_b and C_r weighted maps are then used to enlighten a specific region of the Y filtered map.

The system proposed is called “pop-out.” Pop-out is an effect in human vision that only occurs if there is a single target that differs from its surrounding while all distractors (or the rest of the scene) are homogeneous [9]. This mainly

refers to a search task purpose rather than the bottom-up axis of visual attention. Although we proposed a bottom-up architecture in this paper, we choose the name of “pop-out” because a salient area of an image even pop out from the scene to our eyes. The perspective of top-down integration in our architecture also seems to be a part of the reason that led us to call our system “pop-out.”

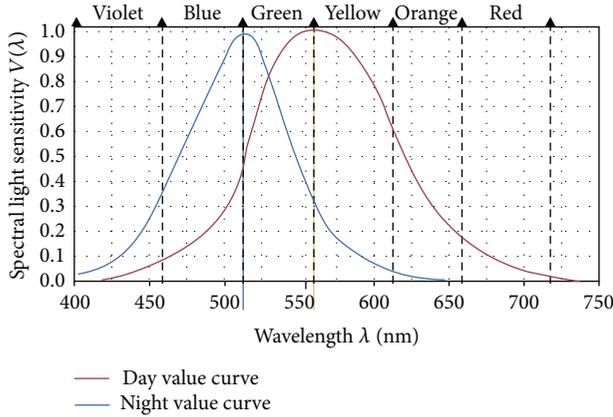


FIGURE 2: Spectral light sensitivity and color wavelength [37]. The red curve corresponds to the luminous efficiency in the day (photopic vision). The night value curve corresponds to the luminous efficiency in the dark (scotopic vision). Between the 2 lighting conditions (photopic and scotopic vision), there is another lighting level (called mesopic vision) where the green wavelength—considered as a merging of enlightened red wavelength (yellow) and blue—is more important than other hues (see Figure 3).

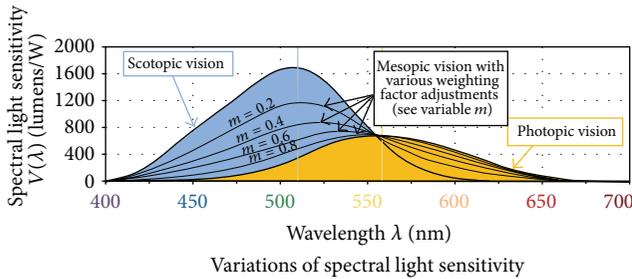


FIGURE 3: Variations of spectral light sensitivity [38]. Several levels of mesopic vision can be found by adjusting the variable m .

3.1. Image Preprocessing. The image preprocessing is used to address a major need: contrasts enhancement. Both psychological and physiological experiments give evidences to the theory of early transformation in the human vision system (HVS) of the L (long wavelength, *which is sensitive to the red part of the visible spectrum*), M (medium wavelength, *sensitive to the green wavelength*), and S (short wavelength, *sensitive to the blue part of the visible spectrum*), signals issued from cones absorption. L-cones, M-cones, and S-cones are mainly located in the central part of the retina, called fovea [24, 36]. HVS is one of the many color spaces that separates RGB color from their intensity. This transformation provides an opponent color space in which the signals are less correlated [24].

There are a variety of opponent color spaces which differ in the way they combine the different cone responses (YC_bC_r , YUV , YD_bD_r , etc.). Here, we use the YC_bC_r space [44, 45] because it is an additional space that allows us to combine intuitively blue and red wavelength for providing all kinds of colors. YC_bC_r is also currently recommended as a standard

definition of digital and high definition television systems [46].

The first step consists of extracting the image intensity (Y). Since we empirically found that traditional value of (Y) in the YC_bC_r standard (ITU-R BT.601-5 and ITU-R BT.709-5) does not get all luminance of the scene, we improve it by using the mean of RGB components:

$$Y = \frac{R + G + B}{3}. \quad (1)$$

This change is very important because, in the conventional YC_bC_r space, Y is more sensitive to the Green component of RGB space. And in accordance with [36] (see Figure 3), this kind of space (where green wavelength is very important) cannot fit with our study since we want to be more flexible to vision conditions (photopic, mesopic, and scotopic). In fact, if we keep the traditional YC_bC_r space this corresponds to be in mesopic conditions where Green component can be more important than any other wavelengths.

We applied a fuzzy mask [47] to enhance the contrast of (Y) and the initial RGB image (we obtain I_f from the enhancement of RGB image). So, we have Y_1 (the intensity extracted from initial image and enhanced by the fuzzy mask) and C_b and C_r components (that we retrieve from I_f). This step highlights image contrast and provides less correlated features to the input of our RF modeling (Log-Gabor functions).

3.2. Log-Gabor Filtering and the Attentive Unit in a New HVS Space. After preprocessing, the image is injected at the input of the Log-Gabor filters. The image at the input of the filter is composed of three elements that are luminance enhanced Y_1 (taken from Y) and C_b and C_r (taken from I_f).

Log-Gabor filter is defined in polar coordinates by the following equation:

$$H_i = H(f, \theta) = H_r(f) \cdot H_\theta(\theta),$$

$$H_r(f) = \exp \frac{-\ln(f/f_0)^2}{2\ln(\sigma_r/f_0)^2}, \quad (2)$$

$$H_\theta(\theta) = \exp \frac{-(\theta - \theta_0)^2}{2\sigma_\theta^2}.$$

H_i or $H(f, \theta)$ describes the Log-Gabor function (of radial component H_r and angular component H_θ) for a frequency f and an orientation θ . Our implementation of Log-Gabor filter is based on Peter Kovessi's work [41], and all parameters (central frequency f_0 , initial orientation θ_0 , radial band-pass σ_r , and angular band-pass σ_θ) are set to have the finest bandwidth (that means a very precise RF modeling). These parameters remain the same for all images. They allow us to avoid any overlap between different Log-Gabor wavelets while providing very fine textures. We use 4 scales and 8 orientations.

Assuming that we are more sensitive to the magnitude of the luminosity perceived (intensity wavelength sensitivity),

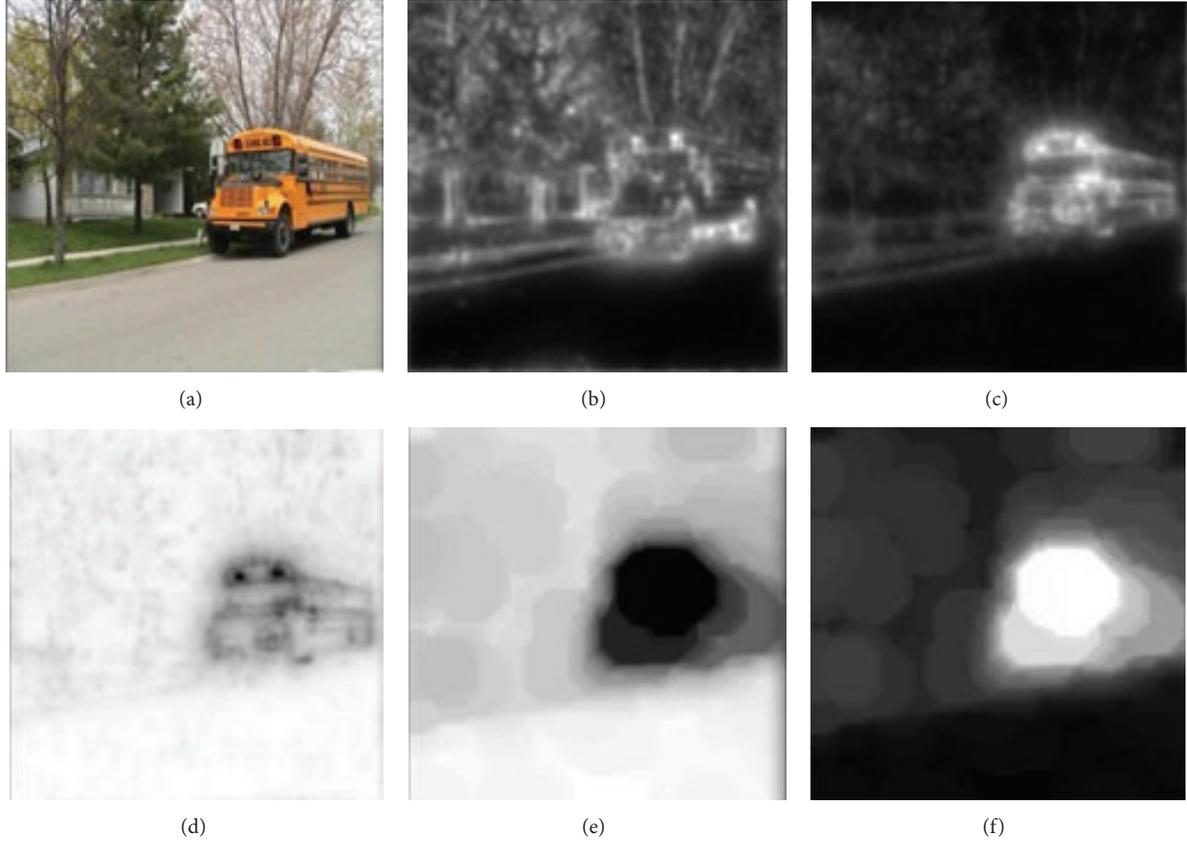


FIGURE 4: (a) Initial image. (b) Luminance texture: T_{Y_i} . (c) “Intuitive $\overline{C_T}$ space” (color texture): $\overline{C_T} = \alpha_b T_{C_b} + \alpha_r T_{C_r}$. We are in photopic conditions; more weights are given to the red wavelength magnitudes (T_{C_r}). $\overline{C_T}$ means the binary complement of C_T . (d) Binary complement of the weighted sum of C_b and C_r components: $C_T = \overline{(\alpha_b T_{C_b} + \alpha_r T_{C_r})}$. (e) C_T after closing. (f) Regions of interest according to spectral light sensitivity $\overline{C_T}$. We note that when we complement the weighted sum of C_b and C_r components (C_T), we obtain a special colorimetric space reflecting the relative spectral sensitivity of each color (see (d)). In this space white area contains low luminous efficiency and describes the wavelength intensity of less bright colors (since we all know that white is the combination of all colors). In our “intuitive $\overline{C_T}$ space” (see (c)), dark red colors combined with the dark blue give black area (which corresponds to the useless white area in (d)) which are less bright than orange, for example (in fact, in photopic vision or when we do not consider C_b component, orange can be considered as an enlightened red wavelength; for a painter, it is also obvious that black color can be seen as the merging of red and blue). In (e), we apply a closing (mathematical morphology) for grouping all whitish regions of the scene (the useless one or the less bright one). Whitish areas in $\overline{C_T}$ are the regions that will attract our attention (f).

we get the amplitude of the filtered result (which is nothing else but the texture T related to a given $Y_i C_b C_r$ components):

$$T_{Y_i, C_b, C_r} = \sum_{i=1}^{s \times o} \text{abs}(\text{IFFT2}(H_i * \text{FFT2}(Y_i, C_b, C_r))). \quad (3)$$

FFT2 and IFFT2 are, respectively, the fast Fourier transform and the inverse fast Fourier transform of a 2D image. s and o are, respectively, the scale and the orientation of the filter. T_{Y_i, C_b, C_r} represents the textures extracted (see Figure 4).

Based on the curve in [34, 38, 48], we analyze the shape of Y (or the intensity of initial image). According to this shape we can know if the picture was taken in photopic, scotopic, or mesopic conditions. In photopic case, we give more weight to T_{C_r} . In scotopic case, T_{C_b} is more important than T_{C_r} . When

we are in mesopic condition T_{C_r} and T_{C_b} are competitive. The weighted sum of T_{C_r} and T_{C_b} is adjusted by α_r and α_b variables (with α_r and $\alpha_b \in [0, 1]$).

The bottom-up saliency map ($S_{\text{BottomUpMap}}$; see Figure 5) is such that

$$S_{\text{BottomUpMap}} = Y_p + \overline{C_T},$$

$$Y_p = \frac{T_{Y_i}}{\sqrt{C_T^2}}, \quad (4)$$

$$C_T = \overline{(\alpha_b T_{C_b} + \alpha_r T_{C_r})}.$$

One finding of our study is that the luminous efficiency perceived can lead us to the most important part of the scene

We ran our model in three modes: automatic (where different weights are automatically given to each color component according to the light level analysis), mesopic (where same weight is given to both C_b and C_r components), and perceptual (edges sensitivity as described in the previous section). One of the goals of this step was to study the contribution of the light level analysis module by using our model in automatic mode and by comparing the results when the same weights are given to C_b and C_r components (case of mesopic vision). Our light level analysis module achieves a performance of 99.17%; just one photopic image is misclassified.

We compared our model with the most popular cognitive saliency measures: Itti-Koch-Niebur [21], VOCUS [23], and Le Meur et al. [24]. Since the Bruce-Tsotsos saliency measure [39] is not considered as a cognitive approach but as an information theoretic model [18], we do not compare our model with it.

Two comparison metrics are used during analysis: Area Under the Receiver Operating Characteristics (AUROC) and Earth Mover’s Distance (EMD). In AUROC score, human fixations are considered as the positive set and some points from the image are sampled, either uniformly or nonuniformly to account for center-bias and form the negative set. The saliency map is then treated as a binary classifier to separate the positive samples from the negatives ones. Perfect prediction corresponds to a score of 1 while a score of 0.5 indicates chance level. While an ROC analysis is useful, it is insufficient to describe the spatial deviation of predicted saliency map from the actual fixation map [49]. If a predicted salient location is misplaced, but misplaced close to or far away from the actual salient location, the performance should be different. To conduct a more representative and selective evaluation, we also use the EMD that indicates the distance between two probability distributions (human gaze versus saliency map) over a region (lower is better).

The model of Itti-Koch-Niebur [21] uses 9 scales and 4 preferred orientations (in total, 42 feature maps are computed: six for intensity, 12 for color, and 24 for orientation).

Pop-out uses 8 scales and 4 orientations: 32 feature maps are thus computed from each $Y_l C_b C_r$ component (32 feature maps from the Y_l component, 32 feature maps from the C_b component, and 32 feature maps from the C_r component, too).

We thus used more feature maps than Itti-Koch-Niebur when color information is taking into account (*case of $S_{\text{BottomUpMap}}$*): Each set, of 32 feature maps obtained, is summed together to give T_{Y_l} , T_{C_b} , and T_{C_r} maps.

Our fusion strategy (see (4)) highlights the color wavelength contained in the luminance texture. This strategy is completely different from the ones used by models [21, 23] which resort to a simple linear combination. In fact, the fusion strategy used by [21] is the linear combination of different feature maps. Like Itti-Koch-Niebur [21], VOCUS [23] also uses a sum of weighted feature maps (linear combination). Thus, by (4), we provide a new fusion strategy that takes into account the color wavelength contained in the luminance texture.

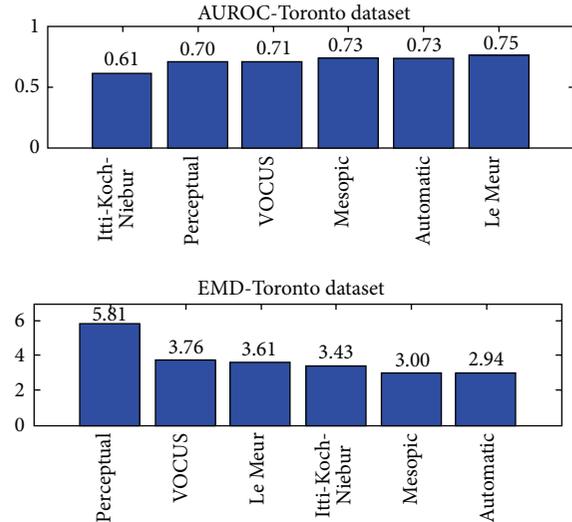


FIGURE 7: Results on Toronto dataset. Our model is used in three modes: perceptual, automatic, and mesopic.

As shown in Figure 7, our model performs better than a lot of cognitive models when it is used in automatic mode (AUROC = 0.73; EMD = 2.94). It is more selective than [24] (see EMD results in Figure 7) because the most enlightened wavelength color of the image is selected without making perceptual grouping of higher-level structures of the scene. Indeed, the model of [24] uses perceptual grouping and some fusion strategies that lead to fuzzy and less selective maps (see Figure 8). We also note the contribution of our model to the FIT when we compare it to the Itti saliency measure. It is mainly caused by the Log-Gabor filters which are more biologically plausible [43] than Gaussian pyramids and Gabor filters used in [21]. The HVS space used also leads us to more accurate architecture than [21]. The bottom-up part of VOCUS [23] is an improvement of the architecture of [21]; features are weighted higher when they are unique in the scene, so, salient objects in the scene are highlighted. However, this performance is close to [21] and it does not give a real improvement of FIT like pop-out in automatic mode.

Both mesopic and automatic mode achieve roughly the same performances. Since the database used just comprises photopic images, it is very difficult to have a difference between the two modes because the curve of mesopic vision encompasses a large portion of the photopic curve (see Figure 3). Besides that, the light level analysis module detects only one mesopic image. Therefore, it is not easy to make a real difference between the mesopic mode and pop-out in automatic mode by using photopic images. Concerning the perceptual mode, it is obvious that the edges sensitivity is less close to eye-tracking experiences in free-viewing task (bottom-up attention).

5. Conclusion and Discussion

We introduced an improved physiological model of the FIT by using light level analysis in order to give different weights

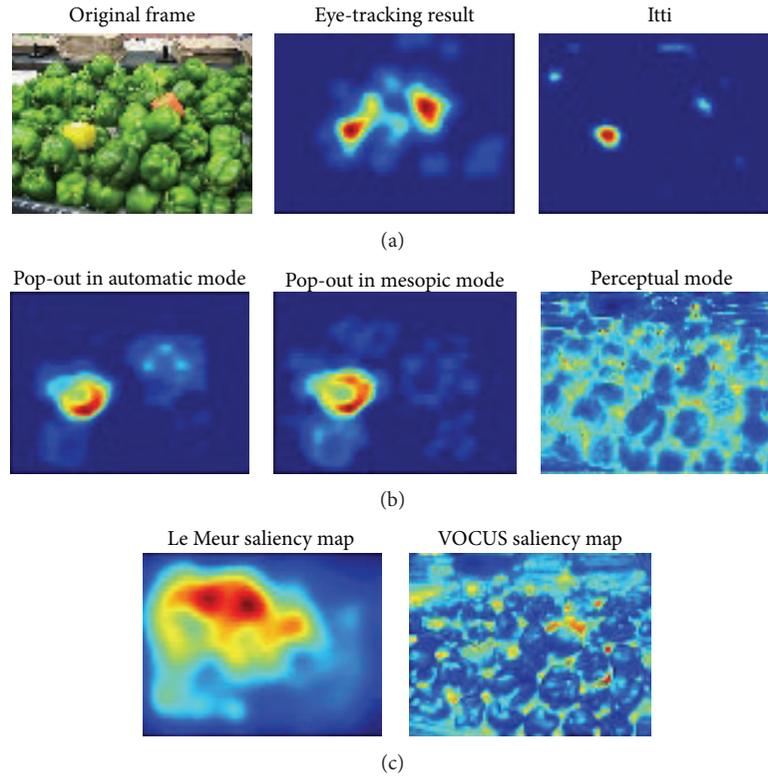


FIGURE 8: Results on Toronto dataset. (a) Initial image, eye-tracking result, and Itti saliency map. (b) Pop-out in automatic mode, mesopic mode of pop-out, and perceptual mode. (c) Le Meur et al. and VOCUS saliency map.

to chrominance components in an enhanced YC_bC_r space. Our saliency model (pop-out in automatic mode) performs better than a lot of popular cognitive approaches [21, 23, 24].

However, as shown in Figure 7, we cannot see its main advantage compared to the traditional FIT-based models because mesopic curve encompasses the photopic curve. Then, the results on the Toronto dataset (which is mainly constituted by photopic images) do not show the real advantage of such light level analysis (in fact, the difference between the performance results in photopic and mesopic mode is not statistically significant). Nevertheless, there are some differences in saliency maps (see Figures 5 and 8) and our approach challenges the eye-tracking experiments which are often made with photopic, mesopic, and scotopic images without make sure to have the same lighting conditions during viewing task.

This latter finding has never been considered before; for instance, when we show a scotopic image (captured by night) in photopic conditions during an eye-tracking experience, we do not have the same lighting conditions as a person who had seen the same image by night, which corresponds to answer to the question: where do people look when it is dark? So, there is a true limit of current eye-tracking databases that should be completely reviewed!

Conflict of Interests

The author declares that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

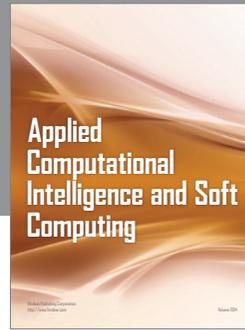
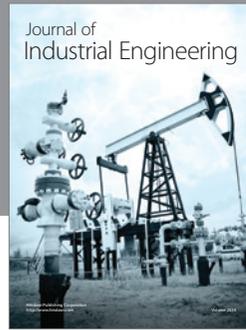
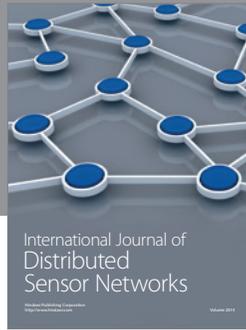
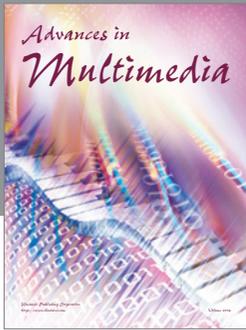
Part of this work is funded by Sebastien Makiesse family and Emmanuel Betukumesu. The author thanks Sophie for being involved in improving the style of the paper and thanks are due to Nathan Salabiaku for his involvement in the validation of the pop-out model.

References

- [1] S. Frintrop, P. Jensfelt, and H. Christensen, "Attentional robot localization and mapping," in *Proceedings of the ICVS Workshop on Computational Attention & Applications*, Bielefeld, Germany, 2007.
- [2] F. Zajega, M. Mancas, R. B. Madhkour et al., "KinAct: the attentive social game demonstration," in *Proceedings of the 11th Asian Conference on Computer Vision*, Daejeon, Republic of Korea, 2012.
- [3] M. Mancas, N. Riche, J. Leroy, and B. Gosselin, "Abnormal motion selection in crowds using bottom-up saliency," in *Proceedings of the 18th IEEE International Conference on Image*

- Processing (ICIP '11)*, pp. 229–232, Brussels, Belgium, September 2011.
- [4] A. Torralba, A. Oliva, M. S. Castelhana, and J. M. Henderson, “Contextual guidance of eye movements and attention in real-world scenes: the role of global features in object search,” *Psychological Review*, vol. 113, no. 4, pp. 766–786, 2006.
 - [5] M. Mancas, “Relative influence of bottom-up and top-down attention,” in *Attention in Cognitive Systems*, vol. 5395 of *Lecture Notes in Computer Science*, pp. 212–226, Springer, Berlin, Germany, 2009.
 - [6] A. M. Treisman and G. Gelade, “A feature-integration theory of attention,” *Cognitive Psychology*, vol. 12, no. 1, pp. 97–136, 1980.
 - [7] Z. W. Pylyshyn and R. W. Storm, “Tracking multiple independent targets: evidence for a parallel tracking mechanism,” *Spatial Vision*, vol. 3, no. 3, pp. 179–197, 1988.
 - [8] S. A. McMains and D. C. Somers, “Multiple spotlights of attentional selection in human visual cortex,” *Neuron*, vol. 42, no. 4, pp. 677–686, 2004.
 - [9] J. Wolfe, “A revised model of visual search,” *Psychonomic Bulletin and Review*, vol. 1, no. 2, pp. 202–238, 1994.
 - [10] M. Mozer, “Early parallel processing in reading: a connectionist approach,” in *Attention and Performance XII: The Psychology of Reading*, M. Coltheart, Ed., pp. 83–104, 1987.
 - [11] B. A. Olshausen, C. H. Anderson, and D. C. Van Essen, “A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information,” *Journal of Neuroscience*, vol. 13, no. 11, pp. 4700–4719, 1993.
 - [12] J. K. Tsotsos, “Analyzing vision at the complexity level,” *Behavioral and Brain Sciences*, vol. 13, no. 3, pp. 423–445, 1990.
 - [13] J. Tsotos, “An inhibitory beam for attentional selection,” in *Proceedings of the York Conference on Spatial Vision in Humans and Robots*, pp. 313–331, 1993.
 - [14] J. Tsotos, “Modeling visual attention via selective tuning,” *Artificial Intelligence*, vol. 78, pp. 507–547, 1995.
 - [15] G. Kootstra, B. de Boer, and L. R. B. Schomaker, “Predicting eye fixations on complex visual stimuli using local symmetry,” *Cognitive Computation*, vol. 3, no. 1, pp. 223–240, 2011.
 - [16] G. Kootstra, A. Nederveen, and B. D. Boer, “Paying attention to symmetry,” in *Proceedings of the British Machine Vision Conference (BMVC '08)*, Leeds, UK, 2008.
 - [17] G. Amy, M. Piolat, and J. Roulin, “L’ecole Gestaltiste: une psychologie allemande de la “forme”,” in *Psychologie Cognitive*, pp. 41–46, Breal, 2006.
 - [18] A. Borji and L. Itti, “State-of-the-art in visual attention modeling,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 185–207, 2013.
 - [19] C. Koch and S. Ullman, “Shifts in selective visual attention: towards the underlying neural circuitry,” *Human Neurobiology*, vol. 4, no. 4, pp. 219–227, 1985.
 - [20] J. J. Clark and N. J. Ferrier, “Modal control of an attentive vision system,” in *Proceedings of the 2nd International Conference on Computer Vision*, pp. 514–523, 1988.
 - [21] L. Itti, C. Koch, and E. Niebur, “A model of saliency-based visual attention for rapid scene analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
 - [22] R. Milanese, *Detecting salient regions in an image: from biological evidence to computer implementation [Ph.D. thesis]*, University of Geneva, Geneva, Switzerland, 1993.
 - [23] S. Frintrop, *VOCUS: A Visual Attention System for Object Detection and Goal-Directed Search*, vol. 3899 of *Lecture Notes in Computer Science*, Springer, Berlin, Germany, 2006.
 - [24] O. Le Meur, P. Le Callet, D. Barba, and D. Thoreau, “A coherent computational approach to model bottom-up visual attention,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 5, pp. 802–817, 2006.
 - [25] O. Le Meur, P. Le Callet, and D. Barba, “Predicting visual fixations on video based on low-level visual features,” *Vision Research*, vol. 47, no. 19, pp. 2483–2498, 2007.
 - [26] M. Guironnet, N. Guyader, D. Pellerin, and P. Ladret, “Static and dynamic features-based visual attention model: comparison to human judgement,” in *Proceedings of the European Signal Processing Conference*, Antalya, Turkey, 2005.
 - [27] Q. Zou, S. Luo, and J. Li, “Selective attention guided perceptual grouping model,” in *Advances in Natural Computation*, vol. 3610 of *Lecture Notes in Computer Science*, pp. 867–876, Springer, Berlin, Germany, 2005.
 - [28] G. Kootstra and D. Kragic, “Fast and bottom-up object detection, segmentation, and evaluation using gestalt principles,” in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA '11)*, pp. 3423–3428, Shanghai, China, May 2011.
 - [29] J. Vitay and N. Rougier, “Using neural dynamics to switch attention,” in *Proceedings of the International Joint Conference on Neural Networks*, Québec, Canada, 2005.
 - [30] J. Fix, N. Rougier, and F. Alexandre, “A dynamic neural field approach to the covert and overt deployment of spatial attention,” *Cognitive Computation*, vol. 3, no. 1, pp. 279–293, 2011.
 - [31] T. Judd, K. Ehinger, F. Durand, and A. Torralba, “Learning to predict where humans look,” in *Proceedings of the IEEE International Conference on Computer Vision*, Kyoto, Japan, 2009.
 - [32] R. Ihaka, “Human Vision,” <https://www.stat.auckland.ac.nz/~ihaka/120/Notes/ch04.pdf>.
 - [33] O. L. Meur, *Attention sélective en visualisation d’images fixes et animées affichées sur écran: modèles et évaluation de performances—applications [Ph.D. thesis]*, University of Nantes, Nantes, France, 2005.
 - [34] J. A. Kinney, “Comparison of scotopic, mesopic, and photopic spectral sensitivity curves,” *Journal of the Optical Society of America*, vol. 48, no. 3, pp. 185–190, 1958.
 - [35] M. Daniel, “La luminance de la CIE,” <http://www.profil-couleur.com/ec/110b-luminance.php>.
 - [36] J. Decuyper, J. L. Capron, T. Dutoit, and M. Renglet, “Implementation of a retina model extended to mesopic vision,” in *Proceedings of the 27th Session of the CIE*, pp. 871–880, Sun City, South Africa, 2011.
 - [37] Technical basics of light (OSRAM), <http://www.imageled.com/Technologies...html>.
 - [38] J. Decuyper, J. L. Capron, T. Dutoit, and M. Renglet, “Mesopic contrast measured with a computational model of the retina,” in *Proceedings of CIE Lighting Quality and Energy Efficiency*, pp. 77–84, Hangzhou, China, 2012.
 - [39] N. D. B. Bruce and J. K. Tsotsos, “Saliency based on information maximization,” in *Advances in Neural Information Processing Systems*, vol. 18, pp. 155–162, 2006.
 - [40] D. J. Field, “Relations between the statistics of natural images and the response properties of cortical cells,” *Journal of the Optical Society of America*, vol. 4, no. 12, pp. 2379–2394, 1987.

- [41] P. Kovesi, “What Are Log-Gabor Filters and Why Are They Good?” <http://www.csse.uwa.edu.au/~pk/research/matlabfns/PhaseCongruency/Docs/convexpl.html>.
- [42] M. Makiese, “De la perception des images à l’algorithme Log-Gabor PCA,” in *Workshop sur les Technologies de l’Information et de la Communication (WOTIC ’11)*, Casablanca, Morocco, 2011.
- [43] M. Makiese, N. Riche, M. Mancas, B. Gosselin, and T. Dutoit, “Biologically plausible context recognition algorithms,” in *Proceedings of the IEEE International Conference on Image Processing (ICIP ’13)*, Melbourne, Australia, 2013.
- [44] Recommendation ITU-R BT.601-5 (1982–1995).
- [45] Recommendation ITU-R BT.709-5, (1990–2002).
- [46] ITU-R Recommendations and Reports, Editions 2, 2012.
- [47] A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, “Context-based vision system for place and object recognition,” in *Proceedings of the International Conference on Computer Vision (ICCV ’03)*, Nice, France, 2003.
- [48] Photopic and Scotopic lumens: when the photopic lumen fails us, <http://www.visual-3d.com/>.
- [49] T. Judd, F. Durand, and A. Torralba, “A benchmark of computational models of saliency to predict human fixations,” Tech. Rep., MIT Computer Science and Artificial Intelligence Laboratory, 2012.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

