Electronic Letters on Computer Vision and Image Analysis 3(1):13-24, 2004

Empirical Validation of the Saliency-based Model of Visual Attention

Nabil Ouerhani*, Roman von Wartburg+, Heinz Hügli* and René Müri+

* *Institute of microtechnology, University of Neuchâtel, L.-Breguet 2, CH-2000 Neuchâtel, Switzerland*
+ *Department of Neurology, University of Bern, Inselspital, CH-3010 Bern, Switzerland*

Received 28 February 2003 ; accepted 18 December 2003

Abstract

Visual attention is the ability of the human vision system to detect salient parts of the scene, on which higher vision tasks, such as recognition, can focus. In human vision, it is believed that visual attention is intimately linked to the eye movements and that the fixation points correspond to the location of the salient scene parts. In computer vision, the paradigm of visual attention has been widely investigated and a saliency-based model of visual attention is now available that is commonly accepted and used in the field, despite the fact that its biological grounding has not been fully assessed. This work proposes a new method for quantitatively assessing the plausibility of this model by comparing its performance with human behavior. The basic idea is to compare the map of attention - the saliency map - produced by the computational model with a fixation density map derived from eye movement experiments. This human attention map can be constructed as an integral of single impulses located at the positions of the successive fixation points. The resulting map has the same format as the computer-generated map, and can easily be compared by qualitative and quantitative methods. Some illustrative examples using a set of natural and synthetic color images show the potential of the validation method to assess the plausibility of the attention model.

Key Words: Visual Attention, Saliency Map, Empirical Validation, Human Perception, Eye Movements.

1 Introduction

Human vision relies extensively on a visual attention mechanism which selects parts of the scene, on which higher vision tasks can focus. Thus, only a small subset of the sensory information is selected for further processing, which partially explains the rapidity of human visual behavior.

It is generally agreed nowadays that under normal circumstances eye movements are tightly coupled to visual attention [1]. This can be partially explained by the anatomical structure of the human retina, which is composed of a high resolution central part, the fovea, and a low resolution peripheral one. Visual attention guides eye movements in order to place the fovea on the interesting parts of the scene. The foveated part of the scene can then be processed in more detail. Thanks to the availability of sophisticated eye tracking technologies, several recent works have confirmed this link between visual attention and eye movements [2, 3, 4]. Hoffman et *al.* suggested in [5] that saccades to a location in space are preceded by a shift of visual attention to that location.

Correspondence to: nabil.ouerhani@unine.ch

Recommended for acceptance by Walter F. Bischof
ELCVIA ISSN:1577-5097
Published by Computer Vision Center / Universitat Autònoma de Barcelona, Barcelona, Spain

Using visual search tasks, Findlay and Gilchrist concluded that when the eyes are free to move, no additional covert attentional scanning occurs, and most search tasks will be served better with overt eye scanning [6]. Maioli et *al.* agree that "There is no reason to postulate the occurrence of shifts of visuospatial attention, other than those associated with the execution of saccadic eye movements" [7]. Thus, eye movement recording is a suitable means for studying the temporal and spatial deployment of attention in any situation.

Like in human vision, visual attention represents a fundamental tool for computer vision. Thus, the paradigm of computational visual attention has been widely investigated during the last two decades. Numerous computational models have been therefore reported [8, 9, 10, 11, 12]. Most of these models rely on the feature integration theory presented by Treisman *et al.* in [13]. The saliency-based model of Koch and Ullman, which relies on this principle, was first presented in [14] and gave rise to numerous software and hardware implementations [15, 16, 17]. The model starts with extracting a number of features from the scene like color and orientation. Each of the extracted features gives rise to a conspicuity map which detects conspicuous parts of the image according to a specific feature. The conspicuity maps are then combined into a final map of attention named saliency map, which topographically encodes stimulus saliency at every location of the scene. Note that the model is purely data-driven and does not require any a priori knowledge about the scene. This model has been used in numerous computer vision applications including image compression [18], robot navigation [19], and image segmentation [20, 21]. However, and despite the fact that it is inspired by psychological theories, the theoretical grounding of the saliency-based model has not been fully assessed.

This work aims at examining the plausibility of the saliency-based model of visual attention by comparing its performance with human behavior. The basic idea is to compare the map of attention - the saliency map - produced by the computational model with another map derived from eye movement experiments. Practically, a computational saliency map is computed for a given color image. The same color image is presented to human subjects whose eye movements are recorded, providing information about the spatial location of the sequentially foveated image parts and the duration of each fixation. A human attention map is then derived, under the assumption that this human attention map is an integral of single impulses located at the positions of the successive fixation points. Objective comparison criteria have been established in order to measure the similarity of both maps.

The remainder of this paper is organized as follows. Section 2 reports the saliency-based model of visual attention. The recording of human eye movements and the derivation of the human attention map are presented in Section 3. Section 4 is devoted to the comparison methods of the performance of the computational model and human behavior. Section 5 reports some experimental results that illustrate the different steps of our validation method. Finally, the conclusions are stated in section 6.

2 The saliency-based model of visual attention

The saliency-based model of visual attention transforms the input image into a map expressing the intensity of salience at each location: the saliency map. The model is composed of three main steps, as reported in [15] (see Fig. 1).

2.1 Feature maps

First, a number (n) of features are extracted from the scene by computing the so-called feature maps. Such a map represents the image of the scene, based on a well-defined feature. This leads to a multi-feature representation of the scene. This work considers the following features which are computed from an RGB color image.

- Intensity

$$I = (R + G + B)/3 \tag{1}$$

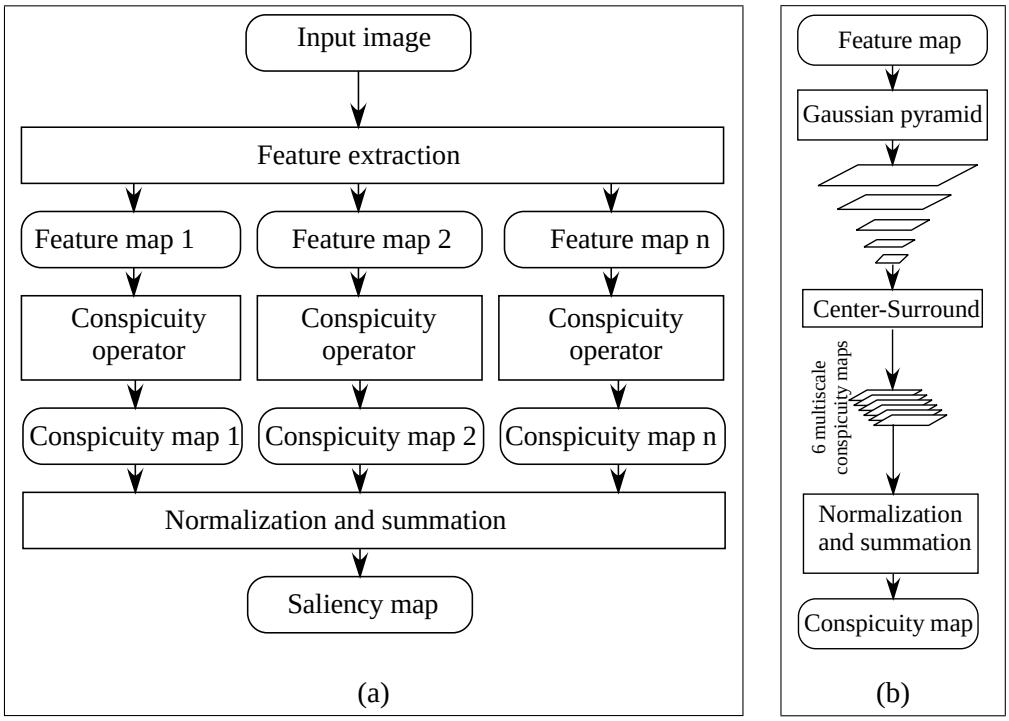


Figure 1: Saliency-based model of visual attention. (a) represents the four main steps of the visual attention model. Feature extraction, conspicuity computation (for each feature) and finally the saliency map computation by integrating all conspicuity maps. (b) illustrates the conspicuity operator in more detail. It computes six multiscale intermediate conspicuity maps. Then, it normalizes and integrates them into the final conspicuity map.

- Two chromatic features based on the two color opponency filtersR^+G^- and B^+Y^- where the yellow signal is defined by $Y = \frac{R+G}{2}$. Such chromatic opponency exists in human visual cortex [22].

$$\begin{aligned} \mathcal{RG} &= R - G \\ \mathcal{BY} &= B - Y \end{aligned} \tag{2}$$

  Before computing these two chromatic features, the color components are first normalized by I in order to decouple hue from intensity.

- Local orientation according to four angles $\theta \in \{0^o, 45^o, 90^o, 135^o\}$ [23]. Gabor filters, which approximate the receptive field impulse response of orientation-selective neurons in primary visual cortex [24], are used to compute the orientation.

2.2 Conspicuity maps

In a second step, each feature map is transformed into its conspicuity map which highlights those parts of the scene that strongly differ, according to a specific feature, from their surroundings. In biologically inspired models, this is usually achieved by using a *center-surround* mechanism. Practically, this mechanism can be implemented with a *difference-of-Gaussians* filter, $\mathcal{D}o\mathcal{G}$ (see Eq.3), which can be applied to feature maps in order to extract local activities for each feature type.

$$\mathcal{D}o\mathcal{G}(x,y) = \frac{1}{2\pi\sigma_{ex}^2} e^{-\frac{x^2+y^2}{2\sigma_{ex}^2}} - \frac{1}{2\pi\sigma_{inh}^2} e^{-\frac{x^2+y^2}{2\sigma_{inh}^2}} \tag{3}$$

A visual attention task has to detect salient objects, regardless of their sizes. Thus, a multi-scale conspicuity operator is required. It has been shown in [10], that applying variable size center-surround filters on fixed size images has a high computational cost. An interesting method to implement the *center-surround* mechanism has been presented in [15]. This method is based on a multi-resolution representation of images. For each feature, nine spatial scales ([0..8]) are created using gaussian pyramids, which progressively lowpass filter and subsample the feature map. Center-Surround is then implemented as the difference between fine and coarse scales. The center is a pixel at scale $c \in \{2, 3, 4\}$ and the surround is the corresponding pixel at scale $s = c + \delta$ and $\delta \in \{3, 4\}$. Consequently, six maps $\mathcal{F}(c, s)$ are computed for each pyramid $\mathcal{P}$ (**Eq. 4**).

$$\mathcal{F}(c, s) = |\mathcal{P}(c) - \mathcal{P}(s)| \tag{4}$$

The absolute value of the difference between the center and the surround allows the simultaneous computing of both sensitivities, dark center on bright surround and bright center on dark surround (red/green and green/red or blue/yellow and yellow/blue for color). For the orientation features, an oriented Gabor pyramid $O(\sigma, \theta)$ is used instead of the gaussian one. For each of the four preferred orientations, six maps are computed according to equation 4 ($P(\sigma)$ is simply replaced by $O(\theta, \sigma)$). Note that Gabor filters approximate the receptive field sensitivity profile of orientation selective neurons in primary visual cortex [24]. A weighted sum of the six maps $\mathcal{F}(c, s)$ results into a unique conspicuity map $\mathcal{C}_i$ for each pyramid and, consequently, for each feature. The maps are weighted by the same weighting function w as described in Section 2.3.

2.3 Saliency map

In the last stage of the attention model, the n conspicuity maps $\mathcal{C}_i$ are integrated together, in a competitive way, into a *saliency map* $\mathcal{S}$ in accordance with equation 5.

$$\mathcal{S} = \sum_{i=1}^{n} w_i \mathcal{C}_i \tag{5}$$

The competition between conspicuity maps is usually established by selecting weights w_i according to a weighting function w, like the one presented in [15]: $w = (M - \overline{m})^2$, where M is the maximum activity of the conspicuity map and $\overline{m}$ is the average of all its local maxima. w measures how the most active locations differ from the average. Indeed, this weighting function promotes conspicuity maps in which a small number of strong peaks of activity is present. Maps that contain numerous comparable peak responses are demoted. It is obvious that this competitive mechanism is purely data-driven and does not require any a priori knowledge about the analyzed scene.
Thus, a computational map of attention - a saliency map - is available. Now we need a human attention map to compare computer and human visual attention.

3 Human map of attention

In the following, the computation of the human attention map will be described. As discussed above, the deployment of visual attention in humans is intimately linked to their eye movements. Under the assumption that attention is guided by the saliency of the different scene parts, the recorded fixation locations can serve to plot a saliency distribution map. The computation of the human attention map is achieved in two steps. First, eye movements of a number of volunteers are recorded while they are viewing a given scene. Second, these measurements are transformed into a map of attention or saliency.

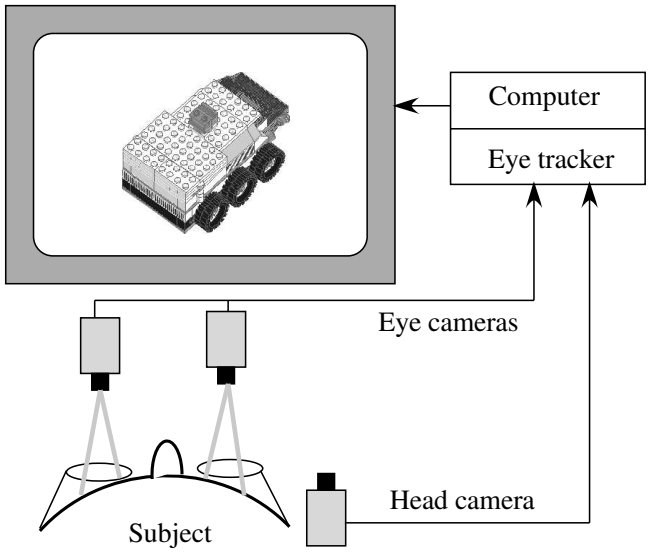


Figure 2: Principle of eye movements recording.

3.1 Eye movement recording

Eye position was recorded with a video-based tracking system (EyeLink^{TM}, SensoMotoric Instruments^{TM} GmbH, Teltow/Berlin). This system consists of a headset with a pair of infrared cameras tracking the eyes (*eye cameras*), and a third camera (*head camera*) monitoring the screen position in order to compensate for any head movements (see Fig. 2). Once the subject is placed in front of a CRT display for stimulus presentation, the position of his pupils is derived from the *eye cameras* using specialized image processing hard- and software. Combining the pupil and head positions, the system computes the gaze position at a rate of 250 Hz and with a gaze-position accuracy relative to the stimulus position of $0.5^o - 1.0^o$, largely dependent on subjects fixation accuracy during calibration. Each experimental block was preceded by two 3x3 point grid calibration sequences, which the subjects were required to track. The first calibration scheme is part of the EyeLink system and allows for on-line detection of non-linearities and correction of changes of headset position. The second calibration procedure is supported by an interactive software in order to obtain a linearised record of eye movement data (off-line).

The eye tracking data are parsed for fixations and saccades in real time, using parsing parameters proven to be useful for cognitive research thanks to the reduction of detected microsaccades and short fixations ($<$ 100 ms). Remaining saccades with amplitudes less than 20 pixels (0.75^o visual angel) as well as fixations shorter than 120 ms were discarded afterwards.

3.2 Computation of the human attention map

This section aims at computing a human attention map based on the fixation data collected in the experiments with human subjects. The idea is that this human attention map is an integral of single impulses located at the positions of the successive fixation points. Practically, each fixated location gives rise to a grey-scale patch whose activity is normally (gaussian) distributed. The width (σ) of the gaussian patch can be set by the user. It should approximate the size of the fovea. We also introduced a parameter α that tunes the contribution of the fixation duration to the gaussian amplitude. If $\alpha = 0$, the amplitude is the same for all fixations regardless of their duration. However, if $\alpha = 1$, the amplitude is proportional to the fixation duration.

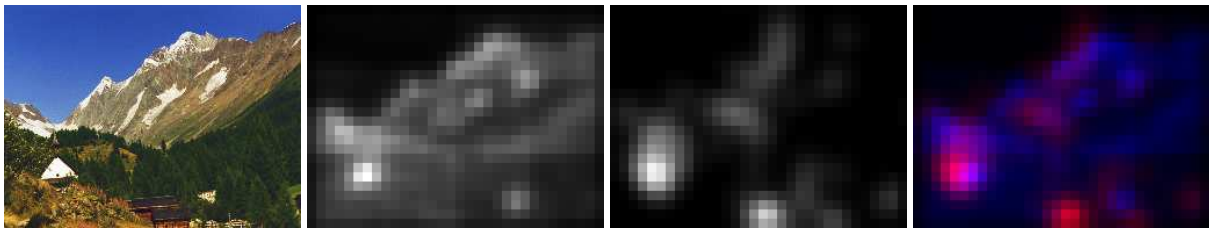
Since numerous human subjects are involved in the experiments, two kinds of human attention maps can be derived. The first category is a mean map of attention that considers all the fixations of all human subjects. The second manner is to compute individual human maps of attention, that is a map of attention for each human subject. In both cases the procedure to compute the map of attention is the same. It is described in Algorithm 1.

Algorithm 1 From fixation points to human attention map

Color image I ($w \times h$)
hM : human attention map (output)
N eye fixations $F_i(x, y, t)$ ((x, y) are spatial coordinates and t is the duration of the fixation)
σ the standard deviation of the gaussian (FOVEA)
$\alpha \in [0..1]$ tunes the contribution of fixation duration to the saliency
$Img1 = Img2 = 0$
$i = 1$
while $i \leq N$ **do**
  $(x, y) = Coord(F_i)$
  $t = Duration(F_i)$
  for $k = 0 \,..\, h - 1$ **do**
    for $l = 0 \,..\, w - 1$ **do**
      $Img1(k, l) = (\alpha.t + (1 - \alpha)).\exp(-\frac{(x-l)^2+(y-k)^2}{\sigma^2})$
    end for
  end for
  $Img2 = Img2 + Img1$
  $Img1 = 0$
end while
$Normalize(Img2, 0, 255)$

4 Comparison of human attention and computational saliency

The idea is to compare the computational saliency map computed by the model of visual attention and the human attention map derived from human eye movement recordings. On one hand, a subjective comparison of both maps gives an approximative evaluation of the correlation of both human and computer attentions. On the other hand, a quantitative correlation measure between the maps provides an objective comparison between human and computer behavior.



Color image *Computational map* *Human map* *Comparison map*

Figure 3: Merging the computational saliency map and the human attention map into a comparison map which better visualizes the correlation between the two maps of attention.

4.1 Subjective comparison

This method is based on a subjective appreciation of the correlation between the human attention map and the computational saliency map by experts. It gives a first and approximative idea about the correlation of both maps.
In order to support the experts in their evaluation, the correlation between the two maps is visually represented. First we assign a different color channel to each map of attention (blue for the computational map and red for the human map). Then we combine the two channel into an RGB color image, which we call *comparison map*, according to equation 6.

$$\begin{aligned} R &= human\ attention\ map \\ G &= 0 \\ B &= computational\ saliency\ map \end{aligned} \tag{6}$$

On this color image, observers can easily see where the two maps correlate and where they do not. Black regions on the color image represent the absence of saliency on both maps, whereas magenta image parts indicate the presence of saliency on both maps of attention. Uncorrelated scene parts are either red (human attention but no computational saliency) or blue (computational saliency but no human attention). An example of this visual representation of correlation is given in Figure 3.

4.2 Objective comparison

The objective comparison of the human attention map and the computational saliency map relies on the correlation coefficient. Let $M_h(x)$ and $M_c(x)$ be the human and the computational maps respectively. The correlation coefficient of the two maps is computed according to equation 7.

$$\rho = \frac{\sum_x \left[(M_h(x) - \mu_h).(M_c(x) - \mu_c)\right]}{\sqrt{\sum_x (M_h(x) - \mu_h)^2 . \sum_x (M_c(x) - \mu_c)^2}} \tag{7}$$

Where μ_h and μ_c are the mean values of the two maps $M_h(x)$ and $M_c(x)$ respectively.

5 Experiments and Results

In this section we report some experimental results that illustrate the different steps of our empirical validation of the saliency-based model of visual attention. The basic idea is to compare computational and human maps of attention gained from the same color images according to objective criteria.

The computational map of attention is computed from the features color (RG and BY), intensity and orientations. The eye recording experiments were conducted with 7 subjects (1 .. 7) between 24 and 34 years, 6 female and 1 male. All of them have normal or corrected-to-normal visual acuity as well as normal color vision. The images were presented to the subjects with a resolution of 800×600 pixels on a 19" monitor, placed at a distance of 70 cm from the subject, which results in a visual angle of approximatively 29×22^o. Each image was presented for 5 seconds, during which the eye movements were recorded. The instruction given to the subjects was "just look at the image".

From the recorded eye movement data, we computed mean as well as individual human maps of attention using the following parameter values:

- $\sigma = 37.0$ for all maps.
- $\alpha = 0$. That means that all fixations have the same importance, regradless of their duration.

Some of the images used in our experiments are illustrated in Fig. 4.

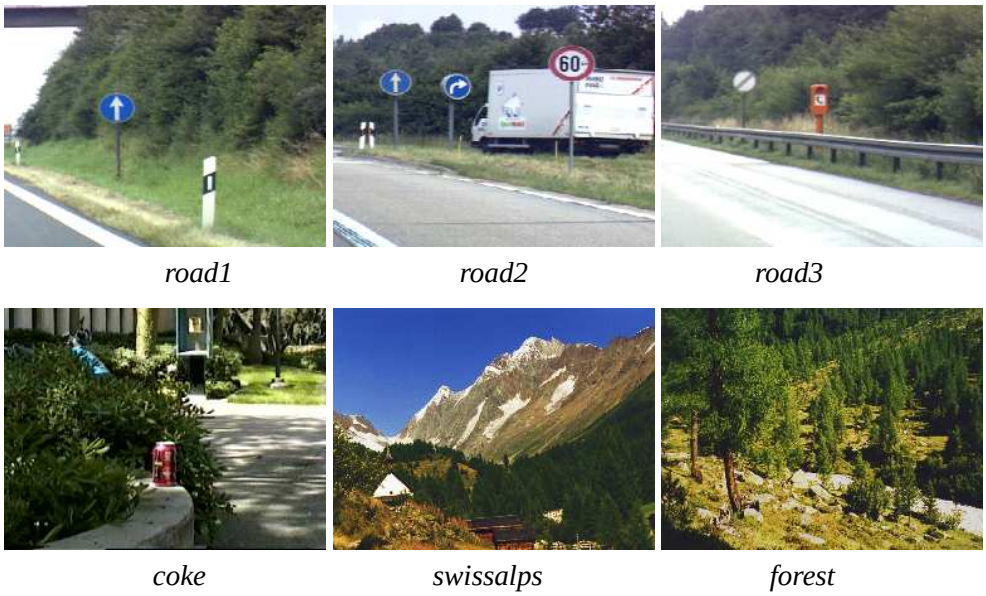


Figure 4: Test images.

Objective comparison

Table 1 represents the correlation coefficients, for the six considered images, of the computational saliency map on one hand and three human attention maps on the other hand. The three human attention maps are: the mean human attention map (first row), the individual human attention map with the highest correlation coefficient (second row), and the individual human attention map with the lowest correlation coefficient (third row). For most of the images the correlation between the computational saliency and the mean human attention map is quite high. The two landscape images (swissalps and forest), where the stimuli are more bottom-up driven, give rise to particularly highly correlated human and computational maps of attention. For the images containing traffic signs, where more top-down information is present, the computational and the human attentional behavior are less correlated. Figure 5 visually illustrates the correlation between the computational map and human maps (mean and individuals) for the image ”swissalps”.

	road1	road2	road3	coke	swissalps	forest
All subjects	0.232	0.3624	0.482	0.4	0.523	0.436
Best subject	0.321	0.348	0.462	0.45	0.608	0.477
Worst subject	0.079	0.194	0.082	0.154	0.134	-0.078

Table 1: Correlation coefficients between the computational map of attention on one hand and mean and individual human maps of attention on the other hand.

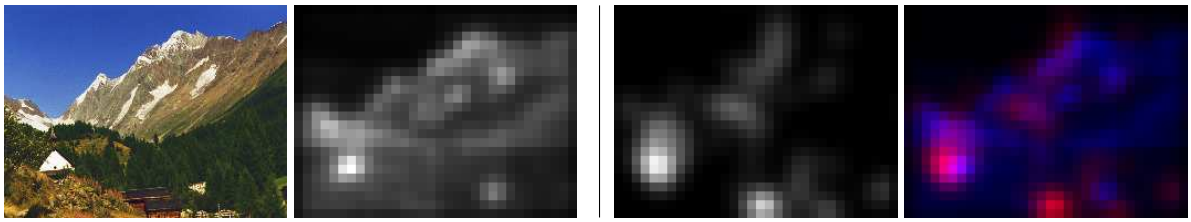
Some human subjects have better correlation with the computational saliency than the mean of all subjects. However, others exhibit a totally uncorrelated behavior with the computational results. This behavior variation among human subjects prompts us to consider also the correlation between subjects themselves. Table 2 illustrates these variations for the ”swissalps” image based on correlation measurements between individual subjects, but also between individuals and the mean of all subjects. In addition, the table includes the correlation between the computational saliency map and all human maps of attention (individual and mean).

To summarize, the results produced by the computational model of visual attention and the human behavior exhibit a satisfactory correlation. Despite the behavior variation among human subject, the ”mean” behavior of

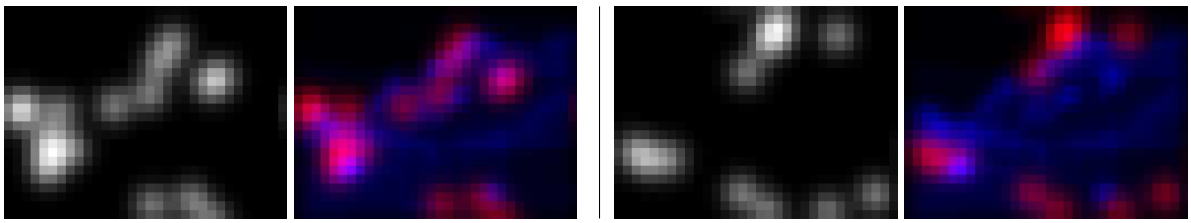
	Model	Mean	1	2	3	4	5	6	7
Model	1	0.523	0.608	0.450	0.613	0.293	0.268	0.271	0.134
Mean		1	0.729	0.752	0.670	0.741	0.674	0.752	0.563
Subj 1			1	0.525	0.661	0.464	0.311	0.333	0.359
Subj 2				1	0.482	0.620	0.220	0.518	0.384
Subj 3					1	0.342	0.300	0.332	0.191
Subj 4						1	0.354	0.439	0.499
Subj 5							1	0.717	0.247
Subj 6								1	0.285
Subj 7									1

Table 2: Inter-subject correlations for the ”swissalps” image.

subjects is still comparable to the model behavior. Although the reduced number of subjects and test images does not allow to draw final conclusions, these preliminary results tend to confirm the bottom-up aspect of the computational model of attention, since higher correlation between computational and human attention is obtained with scenes where less top-down information are present.



Color image *Computational map* *Mean human map* *Correlation* ($\rho = 0.52$)



human map: Subj 1 *Correlation* ($\rho = 0.60$) *human map: Subj 7* *Correlation* ($\rho = 0.13$)

Figure 5: Correlation between computational saliency map and human attention maps. The computational map is compared to 1) the mean human map of attention (top right), 2) the individual human map of attention with the highest correlation (bottom left), and 3) the individual human map of attention with the lowest correlation (bottom right).

6 Conclusion

In this paper, we present a framework for assessing the plausibility of the saliency-based model of visual attention. The basic idea is to compare the results produced by the computational model of attention with a human attention map which is derived from experiments conducted with human subjects under the assumption

that the human visual attention is tightly coupled to eye movements. Therefore, the eye movements of subjects are recorded, using a video-based eye tracker. These eye movements are then transformed into a human map of attention which can be compared to the computational one. A subjective as well as an objective comparison method are investigated in this work. Some experiments which have been carried out to assess the different steps of our validation method gave us preliminary but encouraging results about the correlation of human and computer attention. The full assessment of such a correlation which needs experiments that involves a larger number of human subjects and images will be considered in future works.

Acknowledgments

This work is partially supported by the Swiss National Science Foundation under grant FN 64894. The authors are grateful to Koch Lab (Caltech) and Ruggero Milanese (Uni. Geneva) for making available some of the pictures and the source code of their respective models which represented a source of inspiration for our implementations.

References

[1] A.L. Yarbus. Eye movements and vision. *New York: Plenum Press*, 1967.

[2] A. A. Kustov and D.L. Robinson. Shared neural control of attentional shifts and eye movements. *Nature, Vol. 384, pp. 74-77*, 1997.

[3] D.D. Salvucci. A model of eye movements and visual attention. *Third International Conference on Cognitive Modeling, pp. 252-259*, 2000.

[4] C. Privitera and L. Stark. Algorithms for defining visual regions-of-interest: Comparison with eye fixations. *Pattern Analysis and Machine Intelligence (PAMI), Vol. 22, No. 9, pp. 970-981*, 2000.

[5] J.E. Hoffman and B. Subramaniam. Saccadic eye movements and visual selective attention. *Perception and Psychophysics, 57, pp. 787-795*, 1995.

[6] J.M. Findlay and I.D. Gilchrist. *Eye Guidance and Visual Search.* In G. Underwood (Ed.), Eye Guidance in Reading and Scene Perception, Oxford, Elsevier Science Ltd, pp. 295-312, 1997.

[7] C. Maioli, I. Benaglio, S. Siri, K. Sosta, and S. Cappa. The integration of parallel and serial processing mechanisms in visual search: Evidence from eye movement recording. *European Journal of Neuroscience, Vol. 13, pp. 364-372*, 2001.

[8] B. Julesz and J. Bergen. Textons, the fundamental elements in preattentive vision and perception of textures. *Bell System Technical Journal, Vol. 62, No. 6, pp. 1619-1645*, 1983.

[9] S. Ahmed. *VISIT: An Efficient Computational Model of Human Visual Attention.* PhD thesis, University of Illinois at Urbana-Champaign, 1991.

[10] R. Milanese. *Detecting Salient Regions in an Image: from Biological Evidence to Computer implementation*. PhD thesis, Dept. of Computer Science, University of Geneva, Switzerland, 1993.

[11] J.K. Tsotsos, S.M. Culhane, W.Y.K. Wai, Y.H. Lai, N. Davis, and F. Nuflo. Modeling visual attention via selective tuning. *Artificial Intelligence, Vol. 78, pp. 507-545*, 1995.

[12] G. Backer, B. Mertsching, and M. Bollmann. Data- and model-driven gaze control for an active-vision system. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 23, No. 12, pp. 1415-1429*, 2001.

[13] A.M. Treisman and G. Gelade. A feature-integration theory of attention. *Cognitive Psychology, pp. 97-136*, 1980.

[14] Ch. Koch and S. Ullman. Shifts in selective visual attention: Towards the underlying neural circuitry. *Human Neurobiology, Vol. 4, pp. 219-227*, 1985.

[15] L. Itti, Ch. Koch, and E. Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI), Vol. 20, No. 11, pp. 1254-1259*, 1998.

[16] N. Ouerhani and H. Hugli. Computing visual attention from scene depth. *International Conference on Pattern Recognition (ICPR'00), IEEE Computer Society Press, Vol. 1, pp. 375-378*, 2000.

[17] N. Ouerhani and H. Hugli. Real-time visual attention on a massively parallel SIMD architecture. *International Journal of Real Time Imaging, Vol. 9, No. 3, pp. 189-196*, 2003.

[18] N. Ouerhani, J. Bracamonte, H. Hugli, M. Ansorge, and F. Pellandini. Adaptive color image compression based on visual attention. *International Conference on Image Analysis and Processing (ICIAP'01), IEEE Computer Society Press, pp. 416-421*, 2001.

[19] E. Todt and C. Torras. Detection of natural landmarks through multi-scale opponent features. *ICPR 2000, Vol. 3, pp. 988-1001*, 2000.

[20] N. Ouerhani, N. Archip, H. Hugli, and P. J. Erard. A color image segmentation method based on seeded region growing and visual attention. *International Journal of Image Processing and Communication, Vol. 8, No. 1, pp. 3-11*, 2002.

[21] N. Ouerhani and H. Hugli. Maps: Multiscale attention-based presegmentation of color images. *4th International Conference on Scale-Space theories in Computer Vision, Springer Verlag, Lecture Notes in Computer Science (LNCS), Vol. 2695, pp. 537-549*, 2003.

[22] S. Engel, X. Zhang, and B. Wandell. Colour tuning in human visual cortex measured with functional magnetic resonance imaging. *Nature, Vol. 388, No. 6637, pp. 68-71*, 1997.

[23] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C.H. Anderson. Overcomplete steerable pyramid filters and rotation invariance. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 222-228*, 1994.

[24] A.G. Leventhal. The neural basis of visual function. *Vision and visual dysfunction, Boca Raton, FL:CRC Press, Vol. 4*, 1991.