Assessing the contribution of color in visual attention

Timothée Jost[1], Nabil Ouerhani[1], Roman von Wartburg[2], René Müri[2], Heinz Hügli[1]

[1] Institute of microtechnology
University of Neuchâtel
CH-2000 Neuchâtel
Switzerland

[2] Department of Neurology
University of Bern
Inselspital, CH-3010 Bern
Switzerland

Abstract

Visual attention is the ability of a vision system, be it biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus. The saliency-based model of visual attention represents one of the main attempts to simulate this visual mechanism on computers. Though biologically inspired, this model has only been partially assessed in comparison with human behavior. Our methodology consists in comparing the computational saliency map with human eye movement patterns. This paper presents an in-depth analysis of the model by assessing the contribution of different cues to visual attention. It reports the results of a quantitative comparison of human visual attention derived from fixation patterns with visual attention as modeled by different versions of the computer model. More specifically, a one-cue gray-level model is compared to a two-cues color model. The experiments conducted with over forty images of different nature and involving twenty human subjects assess the quantitative contribution of chromatic features in visual attention.

1. Introduction

Visual attention is the ability of a vision system, be it biological or artificial, to rapidly detect potentially relevant parts of a visual scene, on which higher level vision tasks, such as object recognition, can focus.

It is generally agreed nowadays that under normal circumstances human eye movements are tightly coupled to visual attention. 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 information 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 [Kus97, Sal00, Pri00]. Hoffman *et al.* suggested in [Hof95] that saccades to a location in space are preceded by a shift of visual attention to that location. 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 [Fin97]. 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" [Mai01]. Thus, eye movement recording is a suitable means for studying the temporal and spatial deployment of visual attention in most situations.

Like in human vision, visual attention can play a fundamental role in computer vision, given the high computational complexity of typical tasks [Tso90]. Thus, the paradigm of computational visual attention has been widely investigated during the last two decades, and numerous computational models of visual attention have been suggested [Jul83, Ahm90, Mil93, Culh92, Tso95, Bac01]. For a more complete overview on existing computational models of visual attention, the reader is referred to [HeHu].
Most of these models rely on the feature integration theory presented in [Tre80]. The saliency-based model, which relies on this principle, has first been presented in [Koc85], and has given rise to numerous software and hardware implementations [Itt98, Oue00, Oue03c]. The model starts

with extracting a number of features from the scene, such as color, intensity, and orientation. Each of the extracted features gives rise to a conspicuity map which highlights conspicuous parts of the image according to this 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 of the scene. This model has been used in a number of computer vision applications, including image compression [Oue01b], color image segmentation [Oue03a], and object tracking in dynamic environments [Oue03b].

However, and despite the fact that it is inspired by psychophysical studies, only few works have addressed the biological plausibility of the saliency-based model [OuWa]. Recently, Parkhurst *et al* [Par02] presented for the first time a quantitative comparison between the computational model and human visual attention. Using eye movement recording techniques to measure human visual attention, the authors report a relatively high correlation between human attention and the saliency map, especially when the images are presented for a relatively short time of few seconds. Although the contribution of different cues in visual attention was also addressed in that paper, the presented results did not allow a general conclusion regarding the contribution of chromatic features.

The work presented in the present paper goes further and provides an in-depth analysis of the saliency-based model by quantitatively assessing the contribution of different visual cues in computing visual attention. More specifically, it is aimed at assessing the contribution of the chromatic channels to the control of visual attention. Our hypothesis is that a model including luminance and chrominance based feature channels fares better in predicting where human observers foveate than a model based only on those features derived from luminance.
The basic idea is to compare human fixations derived from eye movement experiments with the computational maps of attention - the saliency maps - produced by two different versions of the saliency-based model. To this end, color images were presented to human subjects while their eye movements were recorded, providing information about the spatial locations of foveated image parts, as well as the duration of each fixation.
Then, two computational saliency maps were computed for the same image: a grayscale-based map and a color-based one. For computing the former, only intensity-based features like intensity itself and orientations were considered, whereas for the color-based saliency map, chromatic features were used additionally.
Another contribution of this work is the use of different metrics for comparing human and computational visual attention. The first comparison metric is a correlation coefficient computed for two maps of attention: the human map of attention, which is computed as the integral of the recorded human fixations, and the computational saliency map. The second comparison metric is a saliency difference measure between randomly picked values of a saliency map on the one hand, and fixation-guided values of the same map on the other hand.

The remainder of this paper is organized as follows. Chapter 2 recalls basics of the saliency models. Then, chapters 3 and 4 present the experimental workflow considered in this research. Both human fixation measurement methods and comparison methods will be exposed. Finally, chapter 5 presents the results, and a general conclusion follows in chapter 6.

2. Saliency models

The saliency-based model of visual attention was proposed by Koch and Ullman in [Koch85]. It is based on four major principles: visual attention acts on a multi-featured input; saliency of locations is influenced by the surrounding context; the saliency of locations is represented on a scalar map (the saliency map); and the winner-take-all and inhibition of return mechanisms are suitable to provide the locations for consecutive attentional shifts.

Several works have dealt with the realization of this model [Mil93, Itt98]. In our work, we used an implementation of the saliency-based model of visual attention that was inspired by these works. The different steps of the model are detailed below (Fig. 1).

2.1. Feature maps

First, a number of features (1..j..n) are extracted from the scene by computing the so-called feature maps F_j. Similar to Itti *et al* seven different features are considered in this work. They are computed from an RGB color image and belong to two main cues, namely intensity and color.

- Intensity feature

$$F_1 = I = 0.3 \cdot R + 0.59 \cdot G + 0.11 \cdot B \qquad \textbf{(1)}$$

- Four local orientation features $F_{4..7}$ according to the angles $\theta \in \{0^o, 45^o, 90^o, 135^o\}$. Gabor filters, which approximate the receptive field impulse response of orientation-selective neurons in primary visual cortex [Lev91], are used to compute the orientation features.
- Two chromatic features based on the two color opponency filters R^+G^- and B^+Y^- where the yellow signal is defined as $Y = \frac{R+G}{2}$.

$$F_2 = \frac{R-G}{I} \qquad F_3 = \frac{B-Y}{I} \qquad \textbf{(2)}$$

Note that such chromatic opponency exists in human visual cortex [Eng97] and that the normalization of the opponency signals by I decouples hue from intensity.

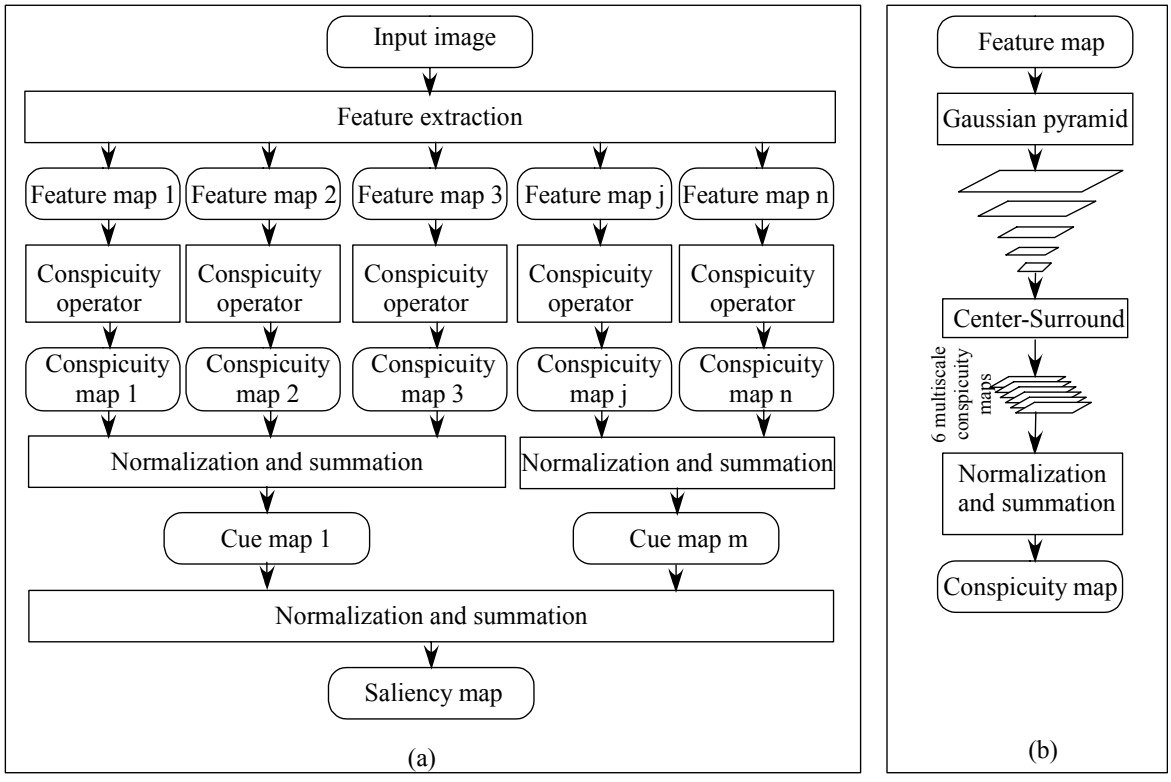

Figure 1. Saliency-based model of visual attention. (a) depicts the different steps of the model. (b) illustrates the conspicuity operator.

2.2. Conspicuity maps

In a second step, each feature map is transformed into its conspicuity map which highlights the parts of the scene that strongly differ, according to the feature specificity, from their surroundings. The computation of the conspicuity maps relies on three main components:

- The center-surround mechanism, implemented with a difference-of-Gaussians-filter, DoG is used to extract local activities for each feature type.

- A multiscale approach in order to detect conspicuous regions, regardless of their sizes. The solution used in this work is based on a multi-resolution representation of images [Itt98] and computes, for each feature *j*, a set of conspicuity maps $M_{j,k}$ at different resolutions *k*.
- A normalization and summation step during which, for each feature *j,* the multiscale maps $M_{j,k}$ are combined, in a competitive way, into a unique feature-related conspicuity map C_j in accordance with **Eq.3**

$$C_j = \sum_{k=1}^{K} N(M_{j,k}) \qquad \textbf{(3)}$$

where $N(.)$ is a normalization operator which simulates the competition between the different scales. A detailed description of the normalization strategy is given below. Note that the summation of the multiscale maps is achieved at the coarsest resolution. Maps of finer resolutions are lowpass filtered and downsampled to the required resolution.

2.3. Cue maps

For the comparison purposes of the present work, we group together several features $j \in J_{cue}$ and we define cue conspicuity maps $\hat{C}_{cue}$, according to **Eq.4**.

$$\hat{C}_{cue} = \sum_{j \in J_{cue}} N(C_j) \qquad \textbf{(4)}$$

2.4. Saliency map

In the third step of the attention model, the cue-related conspicuity maps $\hat{C}_{cue}$ are integrated , in a competitive manner, into a saliency map *S* in accordance with **Eq.5**

$$S = \sum_{cue=1}^{m} N(\hat{C}_{cue}) \qquad \textbf{(5)}$$

where *m* is the number of the considered cues. The normalization operator $N(.)$ is described below.

2.5. Normalization for map combination

In order to integrate several conspicuity maps into a unique map, the normalization strategy $N(.)$ used in this work consists of the following [Itt98]:

1) Scale all maps to the same dynamic range in order to eliminate across-modality amplitude differences due to dissimilar extraction mechanisms.
2) For each map, compute the global maximum *M* and the average $\overline{m}$ of all other local maxima.
3) Globally multiply the map by a weight $w = (M - \overline{m})^2$. Thus, $N(.)$ normalizes a conspicuity map *C* in accordance with **Eq.6**

$$N(C) = w \cdot C \qquad \textbf{(6)}$$

In fact, *w* measures how the most active locations differ from the average of local maxima of a conspicuity map. Thus, this normalization operator promotes conspicuity maps in which a small number of strong peaks of activity are present and demotes maps that contain numerous comparable peak responses.

3. Human visual attention

Under the assumption that under most circumstances, visual attention and eye movements are tightly coupled, the deployment of human visual attention is experimentally derived from the spatial pattern of fixations.

3.1. Eye movement recording

Eye movements were recorded with a infrared video-based tracking system (EyeLink™, SensoMotoric Instruments GmbH, Teltow/Berlin). This system consists of a headset with a pair of infrared cameras tracking the eyes, and a third camera monitoring the screen position in order to compensate for any head movements. It has a temporal resolution of 250 Hz, a spatial resolution of 0.01º, and a gaze-position accuracy relative to the stimulus position of 0.5° - 1.0°, largely dependent on subjects' fixation accuracy during calibration. As the system incorporates a head movement compensation, a chin rest was sufficient to reduce head movements and ensure constant viewing distance.
The images were presented in blocks of 10. Each image block was preceded by a 3 × 3 point grid calibration scheme.

The images were presented in a dimly lit room on a 19" CRT display with a resolution of 800 × 600, 24 bit color depth, and a refresh rate of 85 Hz. Active screen size was 36 × 27 cm and viewing distance 70 cm, resulting in a viewing angle of 29 × 22°. Every image was shown for 5 seconds, preceded by a center fixation display of 1.5 seconds. Image viewing was embedded in a recognition task.

Eye monitoring was conducted on-line throughout the blocks. The eyetracking data was 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 angle) as well as fixations shorter than 120 ms were discarded afterwards [War03].
For every image and each subject *i*, the measurements yielded an eye trajectory T^i composed of the coordinates of the successive fixations $\boldsymbol{f}_k$, expressed as image coordinates (x_k, y_k):

$$T^i = \left(\boldsymbol{f}_1^i, \boldsymbol{f}_2^i, \boldsymbol{f}_3^i, ...\right) \qquad (7)$$

3.2. Human saliency map

As a global representation of the set of all fixations $\boldsymbol{f}_k^i$, a human saliency map $H(\boldsymbol{x})$ was computed, under the assumption that this map is an integral of weighted point spread functions $h(\boldsymbol{x})$ located at the positions of the successive fixations. It is assumed that each fixation gives rise to a normally (gaussian) distributed activity. The width σ of the activity patch was chosen to approximate the size of the fovea. A weighting of $h(\boldsymbol{x})$ as a function of the fixation duration or position k in the trajectory was not considered .
Formally, $H(\boldsymbol{x})$ is computed according to **Eq.8**.

$$H(\boldsymbol{x}) = H(x,y) = \sum_{i=1}^{N_{subj}} \sum_{\boldsymbol{f}_k \in T^i} \exp\left(\frac{(x_k - x)^2 + (y_k - y)^2}{\sigma^2}\right) \qquad \textbf{(8)}$$

where (x_k, y_k) are the spatial coordinates of fixation $\boldsymbol{f}_k$ in image coordinates. The right part of figure 3 shows an example of human fixations superimposed on the corresponding image and the created human saliency map.

4. Comparison metrics

Two different metrics are considered in order to compare human fixations and computer saliency maps: a correlation and a score measurement. An extensive description of each metrics follows.

4.1. Correlation of human and saliency maps

The first metric is defined by the correlation ρ between the computational and human saliency maps.

Let $H(\boldsymbol{x})$ and $S(\boldsymbol{x})$ be the human and the computational maps, respectively. The correlation coefficient ρ of the two maps is defined by **Eq.9.**

$$\rho = \frac{\sum_{\boldsymbol{x}} (H(\boldsymbol{x}) - \mu_H).(S(\boldsymbol{x}) - \mu_S)}{\sqrt{\sum_{\boldsymbol{x}} (H(\boldsymbol{x}) - \mu_H)^2 . \sum_{\boldsymbol{x}} (S(\boldsymbol{x}) - \mu_S)^2}} \qquad \textbf{(9)}$$

where μ_H and μ_S are the mean values of the two maps $H(\boldsymbol{x})$ and $S(\boldsymbol{x})$, respectively.

The value of ρ lies in the [-1, 1] interval. A value of 1 indicates that both maps are exactly similar, a value of 0 indicates that both maps are totally different and a value of -1 indicates that the two maps are anti-correlated, i.e. that a salient feature in one map is not salient in the other one.

4.2. Score s

The score s, also called chance-adjusted saliency by Parkhurst *et al.* [Par02] is shown in figure 2 and can be written according to **Eq.10**.

$$\mathrm{s} = \bar{s}_{fix} - \bar{s}_{ran} \qquad \textbf{(10)}$$

It corresponds to the difference of average values of two sets of samples from the computer saliency map $S(\boldsymbol{x})$; $\bar{s}_{fix}$ refers to the set of N samples taken at the recorded human fixation locations, while $\bar{s}_{ran}$ refers to N random samples.

Considering N fixations $\boldsymbol{f}_k$ from an eye trajectory T^i, the value of $\bar{s}_{fix}$ is computed according to **Eq. 11**.

$$\bar{s}_{fix} = \frac{1}{\mathrm{N}} \sum_{\boldsymbol{f}_k \in T} S(\boldsymbol{f}_k) \qquad \textbf{(11)}$$

The average value of N random fixations in a saliency map *S(x)* is Gaussian distributed and centered at the mean value of the saliency map, μ_S, with an associated standard error of $\sigma_{\mathrm{N}} = \frac{\sigma_S}{\sqrt{\mathrm{N}}}$, where σ_S is the standard deviation of the saliency map. For simplicity, we take $\bar{s}_{ran} = \mu_S$.

Another possibility could have consisted in taking $\bar{s}_{ran} = \mu_S + \sigma_{\mathrm{N}}$ to consider the effective distance between the fixation driven samples and the "core" of the average random sample distribution. Practically, as long as the considered number of fixation N is high enough, σ_{N} can be neglected.

All in all, the score s of equation 10 can be expressed by

$$\mathrm{s} = \frac{1}{\mathrm{N}} \sum_{\boldsymbol{f}_k \in T} S(\boldsymbol{f}_k) - \mu_S \qquad \textbf{(12)}$$

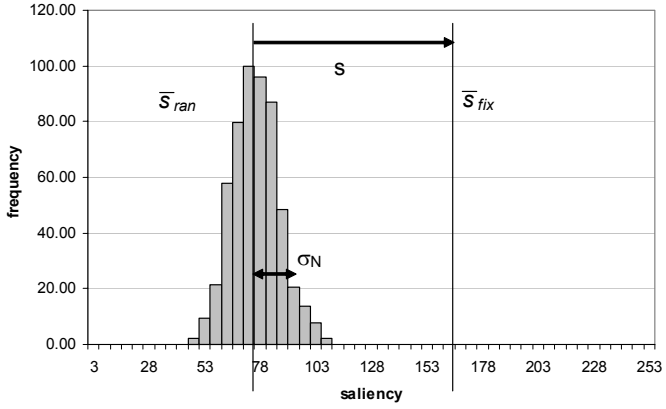

Figure 2. Average values of the sampled saliency map: $\bar{s}_{fix}$ and distribution of $\bar{s}_{ran}$.

If the human fixations are focused on the more salient points in the saliency map, which we expect, the score should be positive. Furthermore, the better the model, the higher the similarity and the higher this score should be.

The principal difference between the s score and the correlation metric is that the former avoids relying on parameters (i.e. σ as used to create the human map) while the later is more "global" than the score s and is independent regarding the scaling of the considered saliency map.

5. Experiments and results

The experimental image dataset consisted in 41 color images containing a mix of natural scenes, fractals, and abstract art images. Most of the images (36) were shown to 20 subjects; the remaining 5 were viewed by 7 subjects. We deem this not to be crucial for the ideas presented here. As stated above, these images were presented to the subjects for 5 seconds apiece, resulting in an average of 290 fixations per image.

For each image of the dataset, the human map $H(\boldsymbol{x})$ was created with the parameter σ = 37 pixels, which was chosen to approximate the fovea in our experimental system, and all fixations were taken into account.

For all images, we also created a color saliency map $S_{col}(\boldsymbol{x}) = N(\hat{C}_{chrom}) + N(\hat{C}_{\text{int}})$ and a grayscale saliency map $S_{gray}(\boldsymbol{x}) = N(\hat{C}_{\text{int}})$, according to equation 5. Both saliency maps are also normalized to the same dynamic range [0..255]. Then, a comparison of these 2 models with the human fixations was performed, following the metrics defined in chapter 4. Figure 3 shows an example of the different measurements and maps involved .

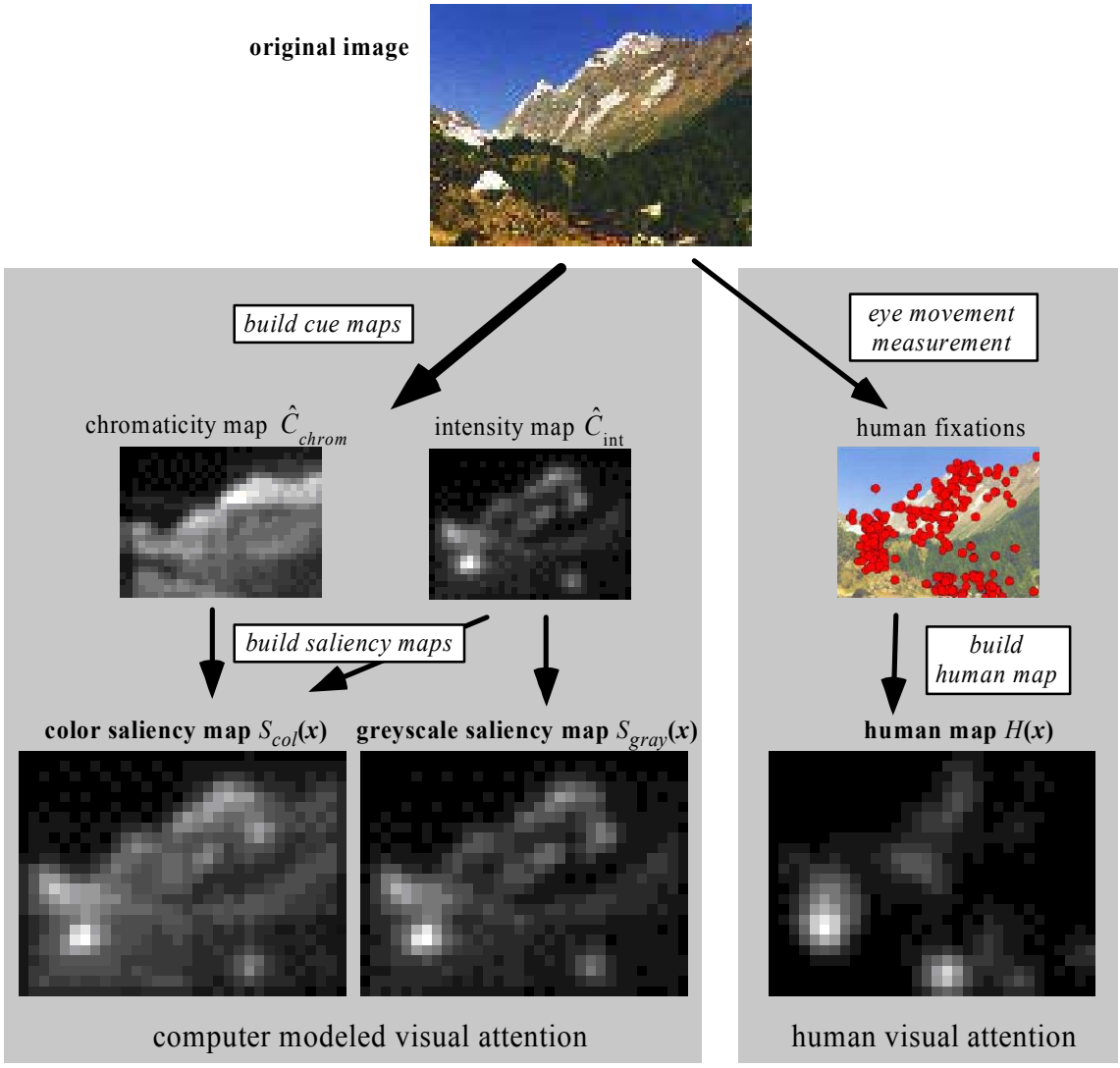

Figure 3. Overview of the different measurements and maps

5.1. Color vs. grayscale, general results

Table 1 presents the average scores s and correlation coefficients ρ over the whole experimental dataset, for both the color and grayscale models. The score s was computed taking the first 5 fixations of each subject into account, since it has been suggested that, with regard to human observers, initial fixations are controlled mainly in a bottom-up manner.

The main observation is that, based on both evaluation methods, the color model fares better than the grayscale one. More specifically, the color model yields an average score approximately 30% higher than the grayscale model. This underlines the usefulness of the color channel in the model and goes toward assessing that colors have a considerable influence on visual attention.

	Score s	ρ
grayscale model $S_{gray}(\boldsymbol{x})$	25.4	.26
color model $S_{col}(\boldsymbol{x})$	32.8	.34

Table 1. Similarity measurements of the two computer models $S(\boldsymbol{x})$ vs. human behavior

5.2. Influence of the number of fixations

Figure 4 presents the average score s taking different numbers of fixations into account for the calculation. Four cases were considered: (1) taking only the first fixation of each subject into account, (2) considering the average of the first three and then (3) five fixations respectively, (4) taking all fixations made over the whole viewing duration into account. From a temporal point of view, these four cases correspond to the first 0.5, 1, 2 and 5 seconds of observation, approximately. We observe a general decrease of scores with the considered number of fixations. This suggests that those features calculated by the model as the most salient ones are also foveated first by human observers. There is one exception to this trend, in the values based on the first fixation only. This might be due to the experimental design: As the subjects had to fixate the center of the screen before any image, the location of the first fixation might have been influenced by this starting point.

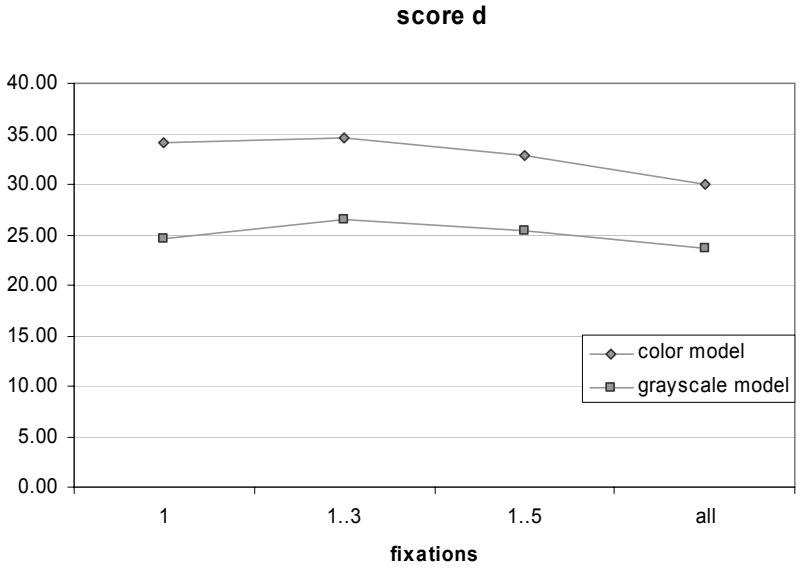

Figure 4. Average score s vs. number of fixations

5.3. Model performance on individual images

Figure 5 presents a few typical images of our database, sorted by their score s values on the color model. The images yielding the best results are found in the upper left; the ranking goes from left to right and top to bottom. As can be seen, the ranking based on the correlational metric would be very similar.

The principal observation here is that the resulting scores and correlation coefficients are widely spread in the value range. It is apparent that the images found on the top row generally contain a few very salient features, such as the fish, the small house or the water lily, and yield the best results. On the other hand, images that lack highly salient features, such as the abstract art or the fractal images on the bottom row, result in much lower values. Nonetheless, there is only one image (out of 41) that yields a negative value (see figure 6); it is shown in the bottom right position in figure 5.

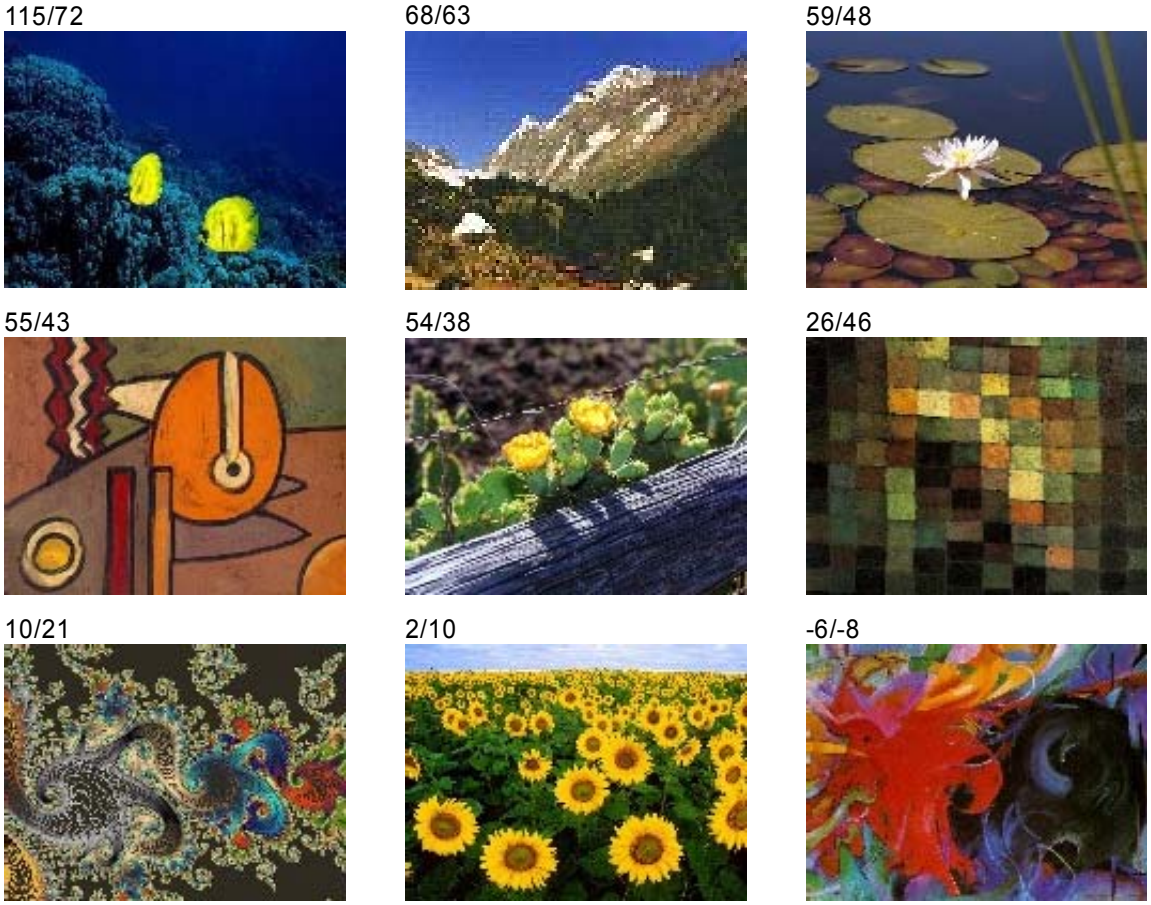

Figure 5. Some typical images from the experimental set, ranked by average values for the color model. Metrics values are given in the form: score s / map correlation in %

5.4. Color vs. grayscale, individual results

Figure 6 shows the correlation values ρ for the color and grayscale models, while figure 7 depicts the relative difference of score s between the color and grayscale models for each image. The relative difference is given by:

$$\frac{s_{col} - s_{gray}}{\max\left(\left|s_{col}\right|, \left|s_{gray}\right|\right)} \qquad \textbf{(13)}$$

As seen on both figures 6 and 7, the color model yields better results than the grayscale one in the large majority of the images. These results confirm the tendency showed by the average results of sections 5.1 and 5.2 about the importance of the use of the chromatic channels in the saliency model.

However, based on our hypothesis and in the perfect case, we would expect the results for the color model to be at least as good as the ones of the grayscale model, for *all* images. We can note that it is not the case for about 12% to 20% of the images (depending on the considered metric, as seen on figure 6 and 7), where the results for the color model are worse than the ones of the grayscale model. However, the difference is generally not very large and might be due to the fact that, in a few images, the color channel adds nothing but "noise" to the saliency map.

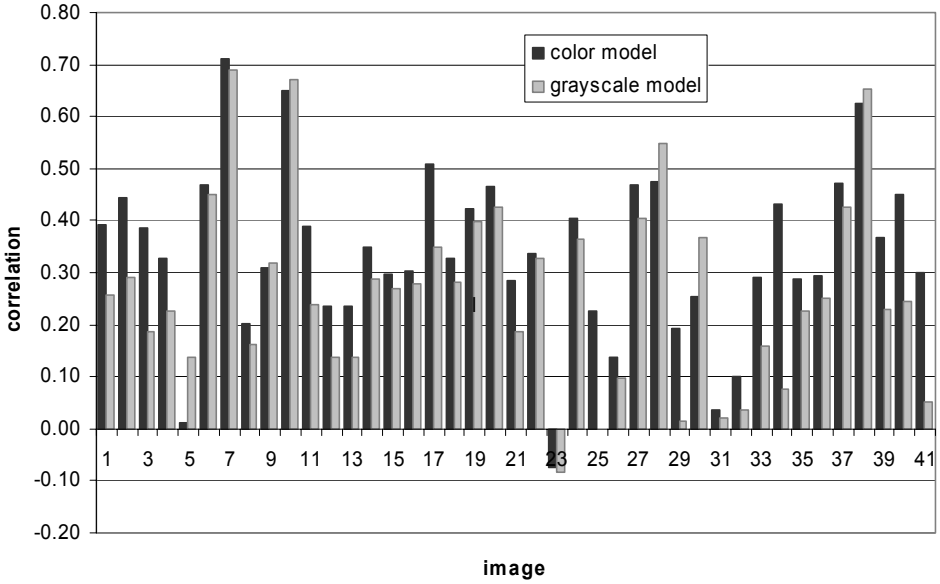

Figure 6. Correlation coefficients for individual images

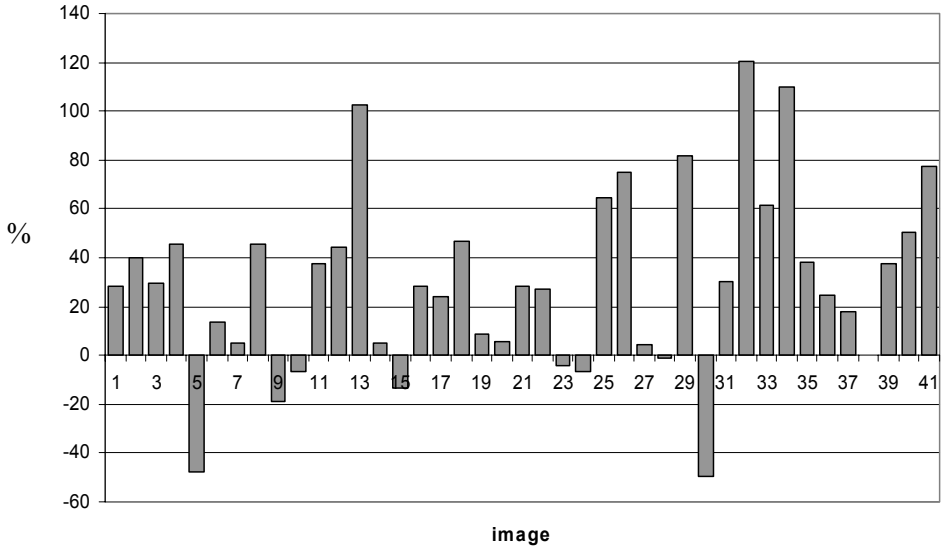

Figure 7. Relative gain of scores s for the color model vs. grayscale model for each image

5.5. Representative cases

Figure 8 presents three specific images with their corresponding computational and human saliency maps. Basically, one of the images fits better with the color model (road sign, 41), one corresponds best with the grayscale model (a fractal, 5), and one results in low values in both cases (the sunflowers, 32). The first image is representative of the large majority of the images which show a better performance with the color model. The other two represent the fewer cases where this is not true.

The road sign image permits to clearly see the influence of color in the visual attention process. Indeed, the blue road sign is not detected as very salient in the grayscale model, as opposed to the color model. When looking at human behavior, it is evident that the panel is clearly the most salient feature of the image, thus the importance of colors.

The sunflower field image is also interesting: It results in very low values with all metrics, even if one might expect the results to be much better at first sight. In fact, on such an image, subjects are mainly focusing onto the horizon and forget about the flowers, maybe because there are a lot of

them and they all look the same. Another downside of the saliency model is that it tends to favor the circumference of the flowers, while humans seem to focus on their center. This is a case where the computer saliency model totally fails to reproduce human behavior, due to the human tendency to focus on the horizon.
Finally, in the fractal example, the grayscale model is superior to the color model. In fact, at first sight the maps are pretty close to each other, but the color component seems only to add "noise" to the saliency map in this case. This is especially well visible in the lower central part of the image.

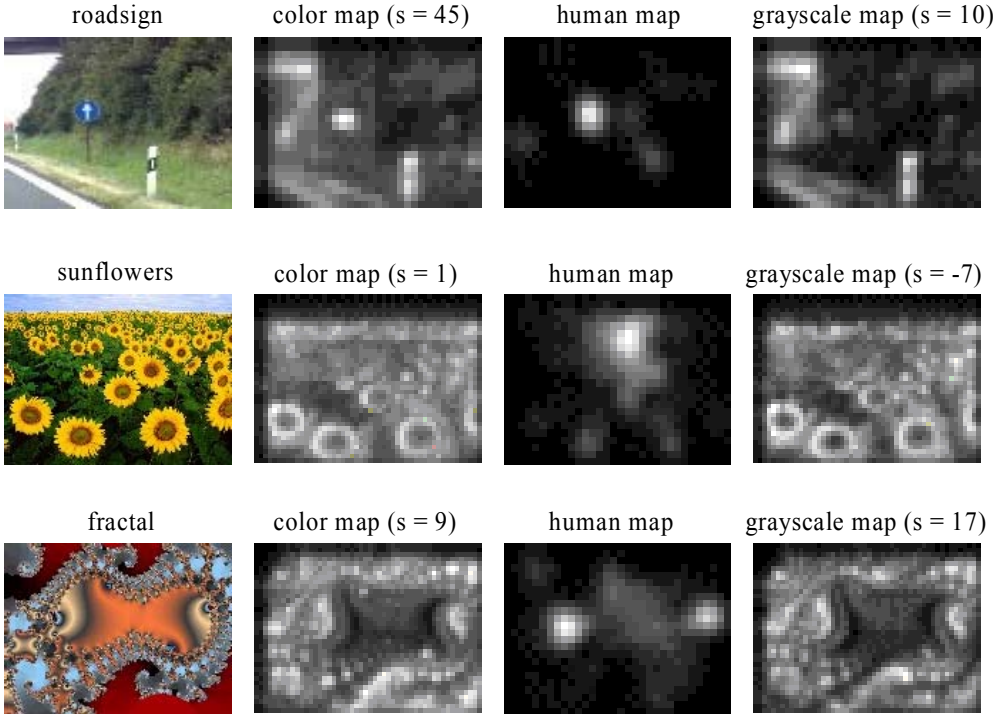

Figure 8. Three representative examples with their corresponding human and saliency maps, s calculated with the first 5 fixations

All in all, these results confirm the overall usefulness of color in the visual attention process and explain the few cases with a different behavior.

6. Conclusion

The work reported in this paper performs comparisons of computer models of visual attention with human attention as measured by recording eye movements of human subjects. The results follow from experiments involving over 40 images of different kinds and nature observed by 20 human subjects in most cases.
The reported comparisons rely on the comparison of a computer saliency map with the set of fixation points extracted from the eye movements. For measuring their similarity, two different metrics were used - the correlation coefficient ρ and the score s - which each have their own advantages.
The contribution of color in visual attention was quantitatively measured as the increase in similarity when the one-cue computer model for grayscale is replaced by the two-cues computer model for color. The similarity improvement, measured as the average on the whole dataset, is from ρ= .26 to .34 for the correlation coefficient, and from S= 25 to 33 for the score s. This assesses the average quantitative contribution of adding the chromaticity cue to a monochrome computer model of visual attention.
Notwithstanding their different nature, both considered metrics yielded very similar results, be it while comparing color and grayscale models or when looking at the image rankings.
A comparison of model performance when the number of considered fixation points is modified shows that on average, the few first fixations are better explained by the model than the set of all fixations of the complete eye movement record.

Finally, a more detailed analysis of the model performance shows a rather large variation of results, depending on the kind of images. On the other hand, all but one images yield positive scores and correlation coefficients, which speaks for the quality of the model itself when compared to human vision. When compared to the grayscale model, the color one also lead to better results on a large majority of the image of the dataset. All in all, the results assess the usefulness of the chromatic cue in the model of visual attention and speak for the considerable influence of color on human visual attention.

Acknowledgements

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

[Ahm90] S. Ahmed. "VISIT: An Efficient Computational Model of Human Visual Attention." *PhD thesis, University of Illinois at Urbana-Champaign*, 1991.

[Bac01] 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.

[Culh92] S.M. Culhane and J.K. Tsotsos. "A prototype for data-driven visual attention." ICPR92, Vol. 1, pp. 36-40, 1992.

[Fin97] J.M. Findlay. "Saccade target selection during visual search". *Vision Research, 37, pp. 617-631*, 1997.

[Gre94] H. Greenspan, S. Belongie, R. Goodman, P. Perona, S. Rakshit, and C.H. Anderson. "Overcomplete steerable pyramid fillters and rotation invariance." *Proc. IEEE Computer Vision and Pattern Recognition (CVPR)*, pp. 222-228. 1994.

[HeHu] D. Heinke and G.W. Humphreys. "Computational Models of Visual Selective Attention: *A Review." In Houghton, G., editor,* Connectionist Models in Psychology*,* In press*.*

[Hof95] J.E. Hoffman and B. Subramaniam. "Saccadic eye movements and visual selective attention". *Perception and Psychophysics*, 57, pp. 787-795, 1995.

[Itt98]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.

[Itt99] L. Itti and Ch. Koch. "A comparison of feature combination strategies for saliency-based visual attention systems." *SPIE Human Vision and Electronic Imaging IV (HVEI'99), San Jose, CA, Vol. 3644, pp. 373-382*, 1999.

[Jul83] B. Julesz and J. Bergen. "Textons, the fundamental elements in preattentive vision and perception of textures". *Bell System Technical Journal, Vol. 62, pp. 1619-1645*, 1983.

[Koc85] Ch. Koch and S. Ullman. "Shifts in selective visual attention: Towards the underlying neural circuitry." *Human Neurobiology, Vol. 4, pp. 219-227*, 1985.

[Kus97] A. A. Kustov and D.L. Robinson. "Shared neural control of attentional shifts and eye movements". *Nature* 384, pp. 74-77, 1997.

[Lev91] A.G. Leventhal. "The neural basis of visual function." *Vision and visual dysfunction Vol. 4, Boca Raton, FL:CRC Press*, 1991.

[Mai01] 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.

[Mil93] 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.

[Oue00] N. Ouerhani and H. Hugli. "Computing visual attention from scene depth". *International Conference on Pattern Recognition (ICPR'00), Vol. 1, pp. 375-378*, 2000.

[Oue01] 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), pp. 416-421*, 2001.

[Oue03a] 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) 2695, pp. 537-549*, 2003.

[Oue03b] N. Ouerhani and H. Hugli. “A model of dynamic visual attention for object tracking in natural image sequences.” *International Conference on Artificial and Natural Neural Network (IWANN), Springer Verlag, Lecture Notes in Computer Science (LNCS) 2686, pp. 702-709*, 2003.

[Oue03c] N. Ouerhani and H. Hugli. ”A real time implementation of visual attention on a SIMD architecture”. *International Journal of Real Time Imaging, Vol. 9, pp. 189-196*, 2003.

[OuWa] N. Ouerhani, R. von Wartburg, H. Hügli, R.M. Müri, " Empirical validation of the Saliency-based model of visual attention", *under review.*

[Par02] D. Parkhurst, K. Law, E. Niebur, “Modeling the role of salience in the allocation of overt visual attention”, *Vision Research*, vol. 42, pp. 107-123, 2002.

[Pri00] 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.

[Sal00] D.D. Salvucci. “A model of eye movements and visual attention”. *Third International Conference on Cognitive Modeling,* pp. 252-259, 2000

[Tri80] A.M. Treisman and G. Gelade. “A feature-integration theory of attention”.*Cognitive Psychology, pp. 97-136*, 1980.

[Tso90] J.K. Tsotsos. “Analyzing vision at the complexity level.” *Behavioral and Brain Science, Vol. 13, pp. 423-469*, 1990.

[Tso95] J.K. Tsotsos. “Toward computational model of visual attention.” *In T. V. Papathomas, C. Chubb, A. Gorea & E. Kowler, Early vision and beyond, pp. 207-226. Cambridge, MA: MIT Press*, 1995.

[War03] R. von Wartburg. "Visuo-motor behaviour during complex image viewing: The influence of colour and image type". *Licentiate paper, Dept. of Psychology, University of Bern, Switzerland,* 2003.