Paying Attention to Symmetry

Gert Kootstra[1], Arco Nederveen[1] and Bart de Boer[2]

[1]Artificial Intelligence, Univesity of Groningen, Netherlands

[2]Institute of Phonetic Sciences, University of Amsterdam, Netherlands

Abstract

Humans are very sensitive to symmetry in visual patterns. Symmetry is detected and recognized very rapidly. While viewing symmetrical patterns, eye fixations are concentrated along the axis of symmetry or the symmetrical center of the patterns. This suggests that symmetry is a highly salient feature. Existing computational models of saliency, however, have mainly focused on contrast as a measure of saliency. These models do not take symmetry into account. In this paper, we discuss local symmetry as a measure of saliency. We developed a number of symmetry models and performed an eye tracking study with human participants viewing photographic images to test the models. The performance of our symmetry models is compared with the contrast saliency model of Itti *et al.* [1]. The results show that the symmetry models better match the human data than the contrast model. This indicates that symmetry is a salient structural feature for humans, a finding which can be exploited in computer vision.

1 Introduction

Symmetry is a visual stimulus that often occurs in our daily lives. We ourselves, for instance, as well as most other organisms, have a clear left-right symmetrical body. Also flowers and butterflies are a good example. The degree of symmetry is even an indicator of the fitness of the individual. Manipulated images of faces, where the symmetry is enhanced, are judged more attractive than the original faces [2]. Symmetry does not only have value for living organisms. Also in art and decoration, symmetry is usually preferred over asymmetry [3]. Furthermore, symmetry is said to increase the *figural goodness*, since the redundancy in a symmetrical figure makes it simpler to encode and is more ordered than a non-symmetrical one [4].

In the present research, we investigate whether this abundance of symmetry in the environment also means that symmetry plays a role in human visual processing. More specifically, we investigate whether symmetry is a salient feature that guides eye movements and can thus be considered a good predictor for overt visual attention. We propose a new model to calculate saliency on the basis of symmetry, which is able to predict human fixations better than existing saliency models.

It is known that humans are highly sensitive to symmetry. We detect symmetrical patterns very rapidly. When patterns have multiple axes of symmetry, recognition speed increases even more [5]. Similarly, symmetry positively influences recall and discrimination [6]. This suggests that the detection mechanism works preattentively and
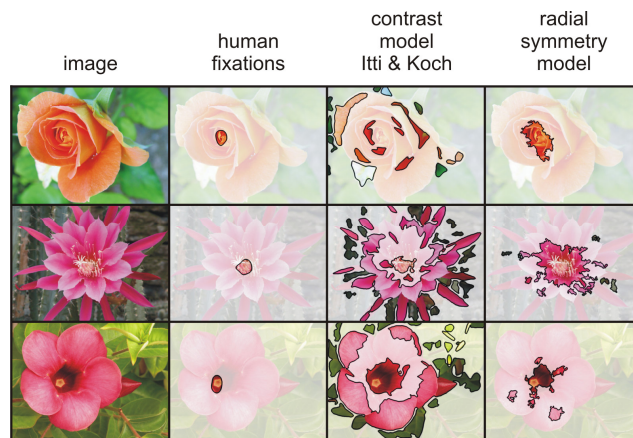
Figure 1: Examples of images containing symmetrical forms. The second column shows the human fixation density map, the third shows the contrast saliency map, and the last shows our symmetry saliency map. The bright regions are the parts of the maps above 50% of its maximum. The preference of humans to fixate on the center of symmetry is correctly reproduced by our symmetry model, whereas the contrast model displays a wide non-specific saliency response.

in a parallel manner [7]. The improvement in performance is thought to be related to the redundancy in the symmetrical forms, which results in more effective and simpler representations [8].

There is also evidence that eye movements are influenced by symmetry. When viewing symmetrical forms, humans tend to fixate on the center of the form [9], or the crossing points of the symmetry axes [10]. Furthermore Locher and Nodine [11] found that fixations are concentrated along the axis of symmetry, and that fixations are more spread out for non-symmetrical images. These studies, however, use relatively simple stimuli with only one pattern presented at a time. In our experiment, we use photographic images to study whether local symmetry also guides eye movements in more complex scenes.

These observations suggest that symmetry is a salient feature. However, most existing saliency models that model overt visual attention do not use symmetry as a measure of saliency. Instead, most models are based on local contrast in the image. The model of Itti *et al.* [1, 12], for instance, is based on contrasts in luminance, color and orientation. Their model is strongly influenced by the feature integration theory of human visual search [13]. The saliency model of Itti *et al.* has been compared to human eye fixations. Parkhurst, Law and Niebur [14] tested the model on photographic images and they showed that the model matches the human fixation points significantly better than expected by chance. Ouerhani *et al.* [15] also found a positive correlation between the model and human fixation data.

Other saliency models, like the model of Le Meur *et al.* [16] are also based on contrast calculations. In their model contrast is calculated in the spatial frequency domain. Le Meur *et al.* compared their model to human data and found a positive correlation, which was slightly higher than the performance of Itti and Koch's model. A set of simpler saliency operators including other features than contrast alone have been proposed by Privitera and Stark [17]. These were also found to predict human fixation

points to some extent. It must be noted that Privitera and Stark also used a basic symmetry operator, which weakly resembled the human data.

   Although most existing models are based on contrast, figure 1 shows that humans have a clear preference to fixate on the center of symmetry. This can neither be explained by Itti and Koch's model, nor by any of the other contrast models. This apparent deficiency in current vision models was the stimulus for the present study. In this paper we therefore investigate the role of local symmetry in overt visual attention. We use photographic images that contain real-world scenes with a complexity that goes beyond the simple forms used in the earlier mentioned psychophysical studies. It will be shown that the symmetry saliency models that we developed, do match the human eye fixation data.

   The paper is organized as follows. We start with a description of the three symmetry saliency models that we developed. These models are based on the isotropic symmetry and radial symmetry operator of Reisfeld, Wolfson and Yeshurun [18], and the color symmetry operator of Heidemann [19]. Furthermore, the eye tracking study that we conducted to test the performance of the saliency models is discussed. Then, the results are shown and compared to the saliency model of Itti and Koch, that is based on contrast [1]. We end with a discussion on the role of symmetry in overt visual attention.

2   Methods

To investigate the role of symmetry in visual attention, we developed a number of symmetry saliency models and compared them with human eye tracking data. To establish a point of reference, the contrast saliency model of Itti *et al.* [1] is also compared with the human data. In this section, the developed symmetry saliency models are explained. Furthermore, we describe the eye tracking studies. The section ends with the description of the methods to compare the models with the human data.

2.1   Symmetry operators

We developed three symmetry saliency models based on the isotropic symmetry and radial symmetry operator of Reisfeld, Wolfson and Yeshurun [18], and the color symmetry operator of Heidemann [19].

   The *isotropic symmetry operator* [18] calculates the amount of symmetry at a given position, x, based upon gradients of the intensity in surrounding pixels. This is done by comparing pairs of pixels i and j at positions x_i and x_j, where $x = (x_i + x_j)/2$ (see fig. 2a). Every pixel pair contributes to the local symmetry by

$$c(i, j) = d(i, j, \sigma) \cdot p(i, j) \cdot m_i \cdot m_j \qquad (1)$$

Where m_i is the magnitude of the gradient at point i, $d(i, j, \sigma)$ is a Gaussian weighting function on the distance between the two pixels with standard deviation σ, and the symmetry measurement $p(i, j)$ is calculated by

$$p(i, j) = \left(1 - \cos(\gamma_i + \gamma_j)\right) \cdot \left(1 - \cos(\gamma_i - \gamma_j)\right) \qquad (2)$$
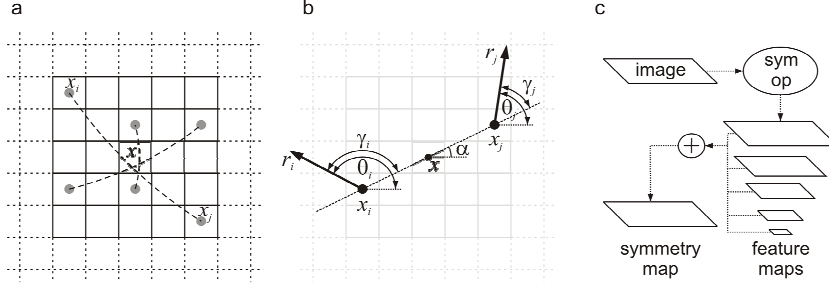
Figure 2: The basis of our symmetry models. (a) gives three examples of pixel pairs whose gradients are compared by the symmetry operator. The geometry of the contribution of a pixel pair is shown in (b) and further explained in the text. (c) gives an overview of the multi scale setup of the symmetry models.

Where $\gamma_i = \theta_i - \alpha$ is the angle between the direction of the gradient angle θ_i and the angle α of the line between p_i and p_j (see fig. 2b). The first term in equation (2) has a maximum value when $\gamma_i + \gamma_j = \pi$, which is true for gradients that are mirror symmetric with respect to p. Using only this term would result in high values for points on a straight edge, which are not considered symmetrical. To avoid this problem, the second term demotes pixel pairs with similar gradient orientation. In this way, the contributions of all pixel pairs, $\Gamma(p)$, within the radius, r, are summed up to give the isotropic symmetry value for p.

$$\mathcal{M}^{\mathrm{iso}}(x,y) = \sum\nolimits_{(i,j)\in\Gamma(p)} c(i,j) \qquad (3)$$

To make the symmetry operator more sensitive to symmetrical patterns with multiple axes of symmetry, Reisfeld *et al.* [18] developed the *radial symmetry operator* as an extension of the isotropic symmetry operator. First, the orientations of the contribution of the pixel pairs are calculated by $\varphi(i,j) = (\theta_i + \theta_j)/2$. Next, the symmetry orientation is determined as $\phi(p) = \varphi(i,j)$ for (i,j) that give the highest contribution $c(i,j)$. This value is then used to promote the contributions of pixels pairs with dissimilar orientations.

$$\mathcal{M}^{\mathrm{rad}} = \sum\nolimits_{(i,j)\in\Gamma(p)} c(i,j) \cdot \sin^2\left(\varphi(i,j) - \phi(p)\right) \qquad (4)$$

The two symmetry operators mentioned above work on intensity values only. Since some color transitions are not detectable in gray-valued images, Heidemann [19] adapted the isotropic symmetry operator to the *color symmetry operator*. This operator uses three color channels, red, green and blue. Equation (3) is adapted so that not only the gradients of pixels in one channel, but also between different channels are compared.

$$\mathcal{M}^{\mathrm{col}}(x,y) = \sum_{(i,j)\in\Gamma} \sum_{(k_i,k_j)\in K} c(i,j,k_i,k_j) \qquad (5)$$

Where K contains all combinations of color channels, and $c(i,j,k_i,k_j)$ is the symmetry contribution calculated by comparing pixel i in color channel k_i with pixel j in color channel k_j. Furthermore, equation (2) is altered to

$$p(i,j) = \cos^2(\gamma_i + \gamma_j) \cdot \cos^2(\gamma_i) \cdot \cos^2(\gamma_j) \qquad (6)$$

Figure 3: Examples of images used in our experiments, one for each image category: flowers, animals, street scenes, buildings and nature.

saliency models operate on multiple scales, and the human visual system is also thought to process on multiple spatial scales. We therefore developed the operators into multi-scale symmetry saliency models, similarly to Itti *et al's* model [1].

The process to calculate the symmetry maps is depicted in figure 2c. First, five spatial scales of the input image are created by progressively applying a Gaussian filter followed by a down scaling of the image by a factor two. The different scales are then processed to symmetry feature maps using the symmetry operators as discussed in the previous section, where we use $r = 24$ and $\sigma = 36$. Next, the five feature maps are normalized using the normalization operator, N, used in [1]. This normalization consists first of scaling the feature map values to the range [0..1], and then multiplying the feature map with $(1-\bar{m})^2$, where $\bar{m}$ is the average value of all local maxima in the map. This normalization promotes feature maps that contain a small number of symmetrical patterns that really stand out, as opposed to feature maps that contain many patterns with similar symmetry values. Finally, the feature maps are combined into a symmetry saliency map by resizing all feature maps to the same size and summing them.

$$\mathcal{S} = \bigoplus_{s=0}^{4} N(\mathcal{M}_s) \qquad (7)$$

Where $\oplus$ is the summation operator that resizes all parts to the same size, and $\mathcal{M}_s$ is the symmetry feature map at scale s. This procedure results in three symmetry saliency maps: $\mathcal{S}^{iso}$ for isotropic symmetry, $\mathcal{S}^{rad}$ for patterns with multiple symmetry axes, and $\mathcal{S}^{col}$ which uses color information.

2.3   Eye tracking experiment

We recorded human fixation data during an eye tracking experiment using the Eyelink head-mounted eye tracking system (SR research). Fixation locations were extracted using the accompanied software. The images were displayed full-screen with a resolution of 1024 by 768 pixels on an 18'' crt monitor of 36 by 27 cm at a distance of 70 cm from the participants. The visual angle was approximately 29º horizontally by 22º vertically. Before the experiment, the eye tracker was calibrated using the Eyelink software. The calibration was verified prior to each session, and recalibrated if needed.

The participants were asked to free view the images. We did not give the participants a task, since we are interested in the bottom-up components of visual attention. A task would give a strong top-down influence on the eye movements. Our approach is similar to [16, 20].

The experiment was carried out by 31 students of the University of Groningen. The participants ranged from 17 to 32 years old, among them 15 females and 16 males with normal or corrected-to-normal vision. In the experiment, 99 images in five different categories were presented, 12 image of animals in a natural setting, 12 images of street

scenes, 16 images of buildings, 40 images of natural environments, and 19 images of natural symmetries, e.g., flowers, plants and butterflies (see figure 3). All these images were taken from the McGill calibrated colour image database [21]. The experiment was split up into sessions of approximately 5 minutes. Between the sessions, the experimenter had a short relaxing conversation with the participants, in order to get them motivated and focused for the next session. Before starting a new session, the calibration of the eye tracker was verified. After each presented images, drift was measured and corrected if needed using the Eyelink software. The participants could decide when to continue and were allowed to take a short break.

2.4 Comparison methods

We used two methods to compare the saliency models with the human data. The first method is a *correlation method* similar to [15, 16]. In this method we correlate the saliency maps with *fixation density maps* calculated from the human fixation data for every single trial. For every single trial, the fixation density map is constructed by placing Gaussian distributions for every fixation with the mean equal to the point of fixation and a standard deviation of 18 pixels. This value is chosen to fit the angular size of the fovea. The resulting value of this comparison method is given by the correlation coefficient, ρ, as calculated by equation (8).

$$\rho = \frac{\sum_{x,y}\left(\left(\mathcal{F}(x,y)-\mu_{\mathcal{F}}\right)\cdot\left(\mathcal{S}(x,y)-\mu_{\mathcal{S}}\right)\right)}{\sqrt{\sigma_{\mathcal{F}}^2 \cdot \sigma_{\mathcal{S}}^2}} \tag{8}$$

Where $\mathcal{F}$ is the fixation density map, $\mathcal{S}$ is the saliency map and μ and σ^2 are respectively the mean and the variance of these maps. The correlation coefficient has a value between -1 and 1. $\rho = 0$ means that there is no correlation between the two maps, which is true when correlating with random fixation density maps. A value for ρ close to zero indicates that a model is a poor predictor of human fixation locations. Positive correlations show that there is a similar structure in the saliency map and the human fixation map.

The second comparison method, the *fixation saliency method*, measures the average saliency at the points of human fixation compared to the average saliency at a large number of randomly chosen points. This method is similar to that used by Parkhurst *et al.* [14]. The method puts an emphasis on the analyses of the fixation points more than on the comparison of the complete saliency maps. The fixation saliency score λ is calculated by
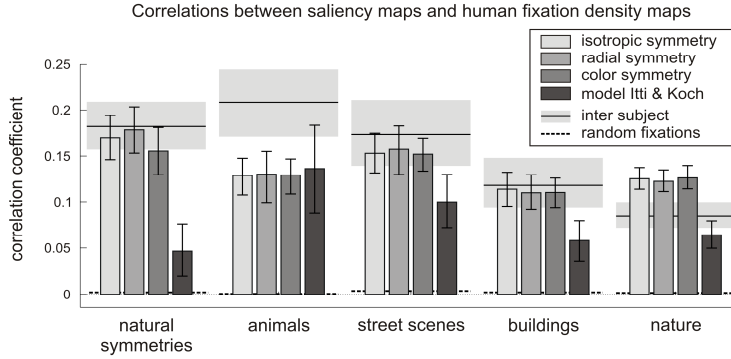
Figure 4: The results of correlation method. The groups show the results for the different image categories. The error bars give the 95% confidence intervals. The horizontal gray bars with the solid line show the mean and 95% confidence interval of the inter subject correlation. The dashed lines show the correlation of the human data with random fixations (close to zero).

$$\lambda = \frac{\sum_{i=1}^{n} s(f_i)}{\sum_{j=1}^{m} s(\text{rnd})} \qquad (9)$$

Where f_i is the ith human fixation location and rnd is a randomly determined location. We used an m of 1000 in our analysis. Furthermore, $s(p)$ is the average saliency value in a patch of the saliency map centered at point p and with a radius $r = 28$. If $\lambda > 1$, the saliency at the human fixation points is higher than in the rest of the image, which means that the given saliency model has predictive power.

3   Results

In figure 4, the results of the correlation method are shown. The five groups contain the results for the different categories. Within each group, the bars show the mean correlation coefficient, as calculated by equation (8), for every method. The error bars give the 95% confidence intervals. To put the scores into perspective, the plot furthermore shows the inter subject correlation, and the correlation of the human data with random fixations. The first is depicted by the horizontal gray bars with a solid line giving the mean and 95% confidence interval. The later is depicted by the horizontal dashed line. All means and confidence intervals are calculated using multi-level bootstrapping analysis.

   The inter-subject correlation is calculated for every image by correlating the fixation density maps of the participants with the maps of all other participants. This correlation shows how well the human fixations can be predicted using the fixations of other participants. This value gives a relative indication of the performance of the models. The random fixation scores are calculated by correlating random fixation density maps containing 15 random fixations with the human fixation density maps. Fifteen fixations are used since this compares to the average number of human fixations per image.
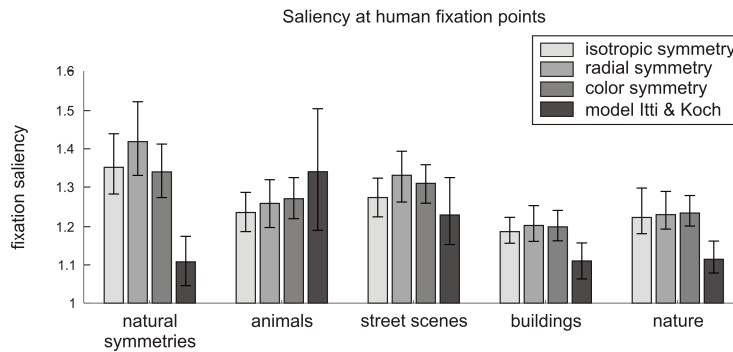
Figure 5: The saliency at human fixation points relative to the average saliency at random points. The bars show the mean fixation saliency for the different saliency models. The error bars are the 95% confidence intervals.

It is important to note that the correlation values are relatively low. The reason for this is that the photographic images are complex stimuli that generate many fixations. Some of these appear to be random, and pull down the correlation. Some, on the other hand re-occur for every subject, and also have high values for symmetric saliency. This causes significantly higher than random correlations. To measure the significance of the correlations between the model data and human data, they are compared with the random case for minimal correlation and the inter-subject case for a top performance.

The difference between the performance of all models and the random fixations is highly significant. Figure 4 furthermore shows that the performance for the images containing natural symmetries is significantly higher for the three symmetry models than for the saliency model of Itti and Koch. The performance of the symmetry models is even comparable to the inter-subject correlation. Not only for the images containing explicit symmetrical forms, also for the other categories have the symmetry models significantly outperformed the contrast model. For these categories too, the performance is similar to the inter-subject correlation.

Among the three symmetry models, there is no significant difference in performance. However, the performance of the radial symmetry model is somewhat higher for the images containing natural symmetries.

Remarkably, for the nature images, the correlations for the symmetry models are higher than the inter-subject correlation. This can be explained by the higher diversity of fixation locations among the participants viewing these images. The images apparently contain many similarly interesting points. However, the higher values for the symmetry models suggest that the participants do primarily attend to the locally symmetrical parts.

For the animal images, there is no significant difference between the symmetry models and the contrast model. Different from other images in our experiment, these images contain objects on low-contrast backgrounds. This explains the higher consensus among the participants, and also explains the higher correlation for Itti and Koch's model.

The results of the fixation saliency analysis are shown in figure 5. The bars show the mean fixation saliency as calculated by equation (9). The error bars are the 95% confidence intervals calculated using multi-level bootstrapping. The results confirm the

previous results with the correlation method. For most image categories, the saliency at human fixation points measured by the symmetry models is significantly higher than that measured by the contrast saliency model. Again, the results are not significantly different for the animal category. Between the symmetry models, no clear differences exist, although the radial symmetry model performs slightly better.

The fact that both methods show a better performance for the symmetry models strengthens the conclusion that local symmetry is a highly significant feature for humans, and that it plays a considerable role in the guidance of eye movements.

4 Discussion

Investigating the role of local symmetry in guiding eye fixations, we developed three symmetry models and compared them to human fixation data. The performance of the models was analyzed using two methods and compared to the performance of the contrast saliency model of Itti and Koch. The results of both analyses show that humans pay attention to local symmetrical patterns more than they pay attention to contrast when viewing complex photographic images. Furthermore, the correlation results of the symmetry models are comparable to the inter-subject correlation, which indicates that symmetry is a good predictor for human eye fixations.

There is no significant difference between the three symmetry models. However, for the natural symmetries, we see a slightly better performance of the radial symmetry model, suggesting a preference of humans for patterns containing multiple symmetry axes. The addition of color processing, in any case, does not improve the performance.

To further investigate the role of symmetry in human vision, we will conduct similar experiments with artificial stimuli. For instance in line drawings, the role of structural features like symmetry is expected to be larger since it contains information about form.

We furthermore plan to investigate the role of symmetry in artificial vision systems. Specifically, in current robotics, robots navigate in man-made environments containing many symmetrical patterns. We therefore hypothesize that our symmetry models are well suited to select interesting visual information for visual navigation.

To conclude, the symmetry saliency models that we developed compare well with human fixations. The results show that humans are sensitive to symmetrical patterns and pay attention to symmetry.

References

[1] 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, pp. 1254-1259, 1998.

[2] K. Grammer and R. Thornhill, Human (Home sapiens) Facial Attractiveness and Sexual Selection: The Role of Symmetry and Averageness, *Journal of Comparative Psychology*, vol. 108, pp. 233-242, 1994.

[3] C. W. Tyler, The Human Expression of Symmetry: Art and Neuroscience, presented at ICUS Symmetry Symposium, Seoul, 2000.

[4] S. E. Palmer, Goodness, Gestalt, Groups, and Garner: Local Symmetry Subgroups as a Theory of Figural Goodness, in *The Perception of Structure. Essays in Honor*

of Wendell R. Garner, G. R. Lockhead and J. R. Pomerantz, Eds. Washington, DC: American Psychological Association, 1991, pp. 23-40.

[5] S. E. Palmer and K. Hemenway, Orientation and Symmetry: Effects of Multiple, Rotational, and Near Symmetries, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 4, pp. 691-702, 1978.

[6] F. L. Royer, Detection of Symmetry, *Journal of Experimental Psychology: Human Perception and Performance*, vol. 7, pp. 1186-1210, 1981.

[7] J. Wagemans, Parallel Visual Processes in Symmetry Perception: Normality and Pathology, *Documenta Ophthalmologica*, vol. 95, pp. 359-370, 1999.

[8] H. B. Barlow and B. C. Reeves, The Versatility and Absolute Efficiency of Detecting Mirror Symmetry in Random Dot Displays, *Vision Research*, vol. 19, pp. 783-793, 1979.

[9] W. Richards and L. Kaufman, "Center-of-Gravity" Tendencies for Fixations and Flow Patterns, *Perception & Psychophysics*, vol. 5, pp. 81-84, 1969.

[10] L. Kaufman and W. Richards, Spontaneous Fixation Tendencies for Visual Forms, *Perception & Psychophysics*, vol. 5, pp. 85-88, 1969.

[11] P. J. Locher and C. F. Nodine, Symmetry Catches the Eye, in *Eye Movements: From Physiology to Cognition*, J. K. O'Regan and A. Lévy-Schoen, Eds. North-Holland: Elsevier Science Publishers B.V., 1987.

[12] L. Itti and C. Koch, Computational Modelling of Visual Attention, *Nature Reviews Neuroscience*, vol. 2, pp. 194-203, 2001.

[13] A. M. Treisman and G. Gelade, A Feature-Integration Theory of Attention, *Cognitive Psychology*, vol. 12, pp. 97-136, 1980.

[14] D. J. Parkhurst, K. Law, and E. Niebur, Modeling the Role of Salience in the Allocation of Overt Visual Attention, *Vision Research*, vol. 42, pp. 107-123, 2002.

[15] N. Ouerhani, R. von Wartburg, H. Hügli, and R. Müri, Empirical Validation of the Saliency-based Model of Visual Attention, *Electronic Letters on Computer Vision and Image Analysis*, vol. 3, pp. 13-14, 2004.

[16] 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, pp. 802-817, 2006.

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

[18] D. Reisfeld, H. Wolfson, and Y. Yeshurun, Context-Free Attentional Operators: The Generalized Symmetry Transform, *International Journal of Computer Vision*, vol. 14, pp. 119-130, 1995.

[19] G. Heidemann, Focus-of-Attention from Local Color Symmetries, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 817-830, 2004.

[20] D. J. Parkhurst and E. Niebur, Scene Content Selected by Active Vision, *Spatial Vision*, vol. 16, pp. 125-154, 2003.

[21] A. Olmos and F. A. A. Kingdom, McGill Calibrated Colour Image Database, http://tabby.vision.mcgill.ca., 2004.