Predicting Human Eye Fixations by Local Symmetry
Abstract

Most bottom-up models that predict human eye fixations are based on contrast features. The saliency model of Itti et al. (1998) is an example of such a contrast-saliency model. Although the model has been successfully compared with human eye fixations, we show that it lacks accuracy in the prediction of fixations on symmetrical forms. The contrast model gives high response at the borders of the forms. However, human observers consistently look at the symmetrical center of these forms. We propose a saliency model that predicts eye fixations using local symmetry. To test the model, we performed an eye-tracking experiment with participants viewing complex photographic images, and compared the data with our symmetry model and the contrast model. The results show that our symmetry model significantly better predicts eye fixations on a wide variety of images including many that are not selected for their symmetrical content. Moreover, our results show that especially early fixations are on highly symmetrical areas of the images. Our results show that symmetry is a strong predictor of human eye fixations, and that it can be used as a predictor of the order of fixation.

A modified version of this chapter is submitted as:


Parts of the chapter have been published as:


3.1 Introduction

Humans continuously make eye movements to investigate the visual environment in an efficient manner. Interesting parts of the visual field are focused on, and inspected with high acuity. Eye movements are influenced both top-down, for instance based on the task at hand or past experiences, and bottom-up, based on properties of the stimulus. Although both influences play a role, we are only interested in the role of the stimulus in guiding eye fixations. The questions that are address in this chapter are: what are properties of the stimulus that attract overt visual attention, and can human eye fixations be predicted with bottom-up models?

More specifically, the role of local symmetry as an alternative to contrast for the prediction of eye fixations is investigated. We propose saliency models that calculate the conspicuousness in an image on the basis of symmetry, and discuss the results of comparing these models to human eye fixations recorded in an eye-tracking experiment. The main result shows that local symmetry is a better predictor of human gaze than contrast.

This chapter is organized as follows. The backgrounds of the presented research are discussed first. Then, the symmetry-saliency models are presented, along with the performed eye-tracking experiment and the methods to compare the models with the human data. Next, the experiments and results are presented, and the chapter ends with a discussion on these results.

3.2 Background

As discussed in Section 2.2, human eye movements are controlled both top down as well as bottom up. Although it is clear that both influences play a role, this chapter focuses on the bottom-up influences. We are interested in the role of the stimulus in the guidance of eye movements, specifically in the visual features that can be used to predict human eye fixations. This gives insights in the inherent properties of the stimulus that attract attention. To investigate this, a saliency model is proposed that determines the salient regions in an image, which is then compared to human eye fixations on the same images. Whereas most existing saliency models focus on contrast features to determine parts of the image that stand out from their local environment, the use of local symmetry to predict the eye movements is advocated in this dissertation.
3.2.1 Saliency Models

A saliency model is a computational model that determines the conspicuous parts of an image based on specific image features. The results of a saliency model, a saliency map, can be compared with actual human eye fixations to determine how well the model correlates with the human data. A good correlation is a strong suggestion that the used image features play a role in the control of eye movements. Not only does it give more insight in visual processing in natural systems, the model can also be used to predict overt visual attention and to improve computer and machine vision systems. Although dynamic features are undeniably of importance for the prediction of eye fixations, we focus on the static features in this paper.

Most existing bottom-up saliency models use contrast features to determine the saliency in an image. The influential saliency model of Itti and Koch, for instance, calculates the saliency of an image on the basis of contrast in three different feature channels: intensity, color, and orientation (Itti & Koch, 2001; Itti et al., 1998) (see Appendix A for details). The model is based on a biologically-plausible architecture for visual attention (Koch & Ullman, 1985), and is an implementation of the feature-integration theory of human visual search (Treisman & Gelade, 1980). It can correctly predict human behavior in visual pop-out experiments (Itti & Koch, 2000). The model has also been compared to human eye fixations on complex photographic images by Parkhurst et al. (2002). They showed that the saliency at the points of human fixation, as measured by the model, is significantly higher than expected by chance. Similarly, Ouerhani et al. (2004) found a positive correlation between the resulting saliency maps and human fixations.

Other saliency models, like the model of Le Meur et al. (2006) are also based on contrast calculations. They found a positive correlation between their model and human data that was slightly higher than the performance of Itti and Koch's model. Privitera & Stark (2000) investigated a set of simpler contrast-saliency operators. These operators were also found to predict human fixation points to some extent. Besides the contrast operators, Privitera and Stark also tested some other operators, including a simple symmetry operator, which also resembled the human data to some extent. The saliency model of Bruce & Tsotsos (2009) compares the distribution of features in the center to the surround, and defines the saliency based on the contrast between the two. The center-surround structure also emerged as the most representative receptive fields when fitting a non-parametric model to human eye-fixation data Kienzle, Franz, Schölkopf,
& Wichmann (2009). However, the model used in this experiment could not result in the concept of symmetry, as we propose in this chapter.

Although contrast has been the dominant feature for saliency models, a clear deficiency in the current visual attention models can be seen in Figure 2.5. The figure shows the results of the eye-tracking experiment presented in this chapter, as well as the results of two saliency models. For the images that are shown in the first column, the participants had a clear preference to fixate on the center of these symmetrical objects (last column). The response of the contrast-saliency model (Itti et al., 1998), shown in the second column, however, is much more spread out and not focused so much on the center of the object, but on the borders where the object contrasts with the backgrounds. The saliency model based on local symmetry that is proposed in this chapter, on the other hand, does more specifically predict the fixations on the center (third column). The results of this chapter show that this is true not only for photographic images that are selected explicitly to contain symmetrical objects as shown in the figure, but more generally for a wide variety of images containing natural and man-made content. Local symmetry calculations can thus be used to predict human gaze.

### 3.2.2 Symmetry in Vision

Symmetry is an abundant visual feature with esthetic properties. Humans very rapidly detect symmetrical patterns (Palmer & Hemenway, 1978) and recognition performances increase for symmetrical patterns (Royer, 1981). The perception of symmetry is suggested to be pre-attentive (Wagemans, 1999). Fixations on symmetrical forms are concentrated at the center of the form, or at the crossing points of the symmetry axes (Kaufman & Richards, 1969), or near the axis of symmetry (Locher & Nodine, 1987). This is similar to the tendency of eye saccades to land at the geometric center of a target object or target configuration (Findlay, 1982; He & Kowler, 1989; Ottes et al., 1984; Bindemann et al., 2009). The center-of-gravity of a pattern is approximately its center of symmetry. We propose to predict the tendency to fixate on the geometric center on the basis of symmetry. The advantage of using symmetry is that the symmetrical center can be determined without the need of prior segmentation of the objects in the scene. Since symmetry is a context-free cue for figure-ground segregation (Driver et al., 1992), it can be used for bottom-up object detection.

This chapter investigates whether symmetry can be used to predict the eye fixations of
humans watching complex photographic images with natural and man-made scenes. A more detailed discussion on the role of symmetry in vision can be found in Section 2.6.

### 3.2.3 Fixation Sequence

When humans view an image for a couple of seconds, they make a sequence of saccades to investigate the interesting regions of the image. According to the scanpath theory of Noton & Stark (1971a,b), humans make the same sequence of fixations every time they view the same pattern. This suggests that the sequence is remembered as part of the pattern representation. This spatial memory is then used as top-down guidance for the eye movements. Since the focus is on the bottom-up components of eye movements, scanpaths are not considered in this chapter.

Parkhurst et al. (2002) compared human eye fixations in a free-viewing experiment with the contrast-saliency model (Itti et al., 1998). Investigating the amount of contrast near the point of fixation, they found that it drops over the fixation sequence. Earlier fixations are on parts of the image containing more contrast than the later fixations. Similarly, this chapter shows that the amount of local symmetry at the point of fixations also gradually drops over the fixation sequence. This effect is even larger for local symmetry than for contrast. The reason for the drop of contrast and symmetry at the points of fixation might be that the early fixations are more stimulus-driven than the later, since context then plays a larger role in the guidance of the eyes. However, it is also possible that all attended parts of the scene have above-average contrast and local symmetry, and the sequence is based on the strength of these features. Local symmetry, and to a lesser extent contrast, can then be used to predict the sequence of fixations. It must be noted, however, that this is only true in free-viewing conditions with no particular target. When participants are engaged in a search task, bottom-up saliency is not a good predictor of overt visual attention (Foulsham & Underwood, 2007).

### 3.3 Methods

In this section, we first present the symmetry-saliency model, and give a short overview of the contrast-saliency model of Itti et al. (1998) with which we compare the results
3.3. Methods

Figure 3.1: The multi-scale symmetry-saliency model. a) shows the basic symmetry operator. All pixel pairs in the symmetry kernel contribute to the local symmetry value of the central pixel (a1). The contribution of a pixel pair is calculated using the intensity gradients at the pixel locations (a2). b) gives the layout of the multi-scale symmetry model. A Gaussian image pyramid of five scales is constructed. The symmetry operator is applied to all images in the pyramid, resulting in symmetry maps at different scales. The maps are normalized and added to form the symmetry-saliency map.

3.3.1 Symmetry-Saliency Model

We developed three saliency models based on local symmetry calculations. The models are built upon the isotropic and radial symmetry operator of Reisfeld, Wolfson, & Yeshurun (1995), and the color symmetry model of Heidemann (2004). These three symmetry operators are all based on the same basic symmetry operator. We extended the operators to multi-scale symmetry-saliency models. We first describe the basic symmetry operator, followed by the multi-scale symmetry models.

3.3.1.1 Basic Symmetry Operator

The basic or isotropic symmetry operator calculates the amount of local symmetry at a given pixel, \( p = (x, y) \), in an image by applying a symmetry kernel to this pixel. Using a sliding window approach, the symmetry is calculated for all pixels. The amount of local
symmetry at \( p \) is calculated based on the intensity gradients of the surrounding pixels in the kernel. Pixels pairs in the symmetry kernel contribute to the local symmetry value. A pixel pair consists of two pixels, \( p_i \) and \( p_j \), so that \( p = (p_i + p_j)/2 \) (see Figure 3.1a1). In other words, the two pixels forming a pair are point symmetric in the center of the kernel. The contribution of the pixel pair to the local symmetry of \( p \) is calculated by comparing the intensity gradient \( g_i \) at \( p_i \) and gradient \( g_j \) at \( p_j \). The intensity gradients are obtained by approximating the image derivatives in the horizontal, \( I_x \), and vertical, \( I_y \), direction using Sobel filters:

\[
I_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} * I, \quad I_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} * I, \tag{3.1}
\]

The gradient vector \( g_i = (I_x(p), I_y(p))^T \), with the magnitude, \( m_i \), and orientation, \( \theta_i \), determined as:

\[
m_i = \sqrt{I_x(p)^2 + I_y(p)^2} \tag{3.2}
\]

\[
\theta_i = \arctan2(I_y(p), I_x(p)) \tag{3.3}
\]

Based on the orientation of the gradients at point \( i \) and \( j \), the symmetry is measured by:

\[
c(i,j) = (1 - \cos(\gamma_i + \gamma_j)) \cdot (1 - \cos(\gamma_i - \gamma_j)), \tag{3.4}
\]

where \( \gamma_i = \theta_i - \alpha \) is the angle between the orientation of the gradient, \( \theta_i \), and the angle, \( \alpha \), of the line between \( p_i \) and \( p_j \) (see Figure 3.1a2). The first term in Equation (3.4) has a maximum value when \( \gamma_i + \gamma_j = \pi \), which is true for gradient orientations that are mirror symmetric with respect to \( p \). Using only this term would also respond to symmetry values for two pixels that have the same gradient orientation and thus lie on a straight edge. Since we are not interested in detecting edges, but in finding the centers of symmetrical patterns, the second term in the equation demotes pixels pairs with similar gradient orientations.

The symmetry measurement is weighed by a distance function and the magnitudes of
the gradients to get the local symmetry contribution of the pixel pair:

\[ s(i, j) = d(i, j, \sigma) \cdot c(i, j) \cdot \log(1 + m_i) \cdot \log(1 + m_j), \]

\[ (3.5) \]

where \( m_i \) is the magnitude of the gradient, and \( d(i, j, \sigma) \) is a Gaussian weighting function on the distance between \( p_i \) and \( p_j \) with a standard deviation of \( \sigma \). The multiplication with the gradient magnitudes assures that only strong edges contribute to the local symmetry value, since these are likely to belong to objects in the scene. The logarithm is used to attenuate the influence of large magnitude values.

The total symmetry value at point \( p \) is calculated by summing the contributions of all symmetrical pixel pairs in the kernel, \( \Gamma(p) \). The symmetry kernel has a size of \( r \times r \) (see Figure 3.1a2). We used \( r = 24 \) in our experiments. The amount of local symmetry calculated by the isotropic symmetry operator is then:

\[ S_{l}^{iso}(p) = \sum_{(i, j) \in \Gamma(p)} s(i, j), \]

\[ (3.6) \]

where \( S_{l}^{iso} \) is the resulting isotropic symmetry map at scale \( l \). The use of different scales to acquire a multi-scale symmetry-saliency model is discussed in the next section.

Based on this isotropic symmetry operator, Reisfeld et al. (1995) developed a radial symmetry operator that is extra sensitive to patterns containing multiple axes of symmetry. Due to the summation in Equation (3.6), the isotropic operator has already a higher activation for patterns with multiple axes of symmetry. However, the radial operator extra promotes these kinds of patterns. To achieve this, the orientation of the symmetry contribution of every pixel pair is calculated by:

\[ \phi(i, j) = (\theta_i + \theta_j)/2. \]

\[ (3.7) \]

Next, the pixel pair that contributed most to the symmetry value at point \( p \) is determined by:

\[ (i', j') = \arg\max_{(i, j) \in \Gamma(p)} s(i, j) \]

\[ (3.8) \]

and the symmetry orientation at point \( p \) is established:

\[ \psi(p) = \phi(i', j'). \]

\[ (3.9) \]
This orientation is then used to promote the contributions of pixel pairs with dissimilar orientations:

$$S_{\text{rad}}(p) = \sum_{(i,j) \in \Gamma(p)} s(i,j) \cdot \sin^2(\phi(i,j) - \psi(p)).$$

Both the isotropic and the radial symmetry operator are based on the intensity of the pixels only. Heidemann (2004) extended the basic operator to a color symmetry operator. This operator compares pixels in three color channels, red, green, and blue, to determine the symmetry value:

$$S_{\text{col}}(p) = \sum_{(i,j) \in \Gamma(p)} \sum_{(k_i,k_j) \in K} c(i,j,k_i,k_j),$$

where $K$ contains all combinations of two color channels, $K = \{(R,R), (R,G), \ldots, (B,B)\}$ $c(i,j,k_i,k_j)$ is the symmetry contribution calculated by comparing pixel $p_i$ in color channel $k_i$ with pixel $p_j$ in color channel $k_j$. Besides the addition of color, Equation 3.4 is altered so that the function gives the same results for gradients that are rotated by $180^\circ$ in order to account for patterns on gradually changing background:

$$c_{\text{col}} = \cos^2(\gamma_i + \gamma_j) \cdot \left(\cos^2(\gamma_i) \cdot \cos^2(\gamma_j)\right).$$

The first term in the equation is a $180^\circ$-periodic symmetry term. The second term has a similar role as the second term in Equation 3.4, to discount for pixels that lie on an edge.

### 3.3.1.2 Multi-Scale Symmetry Model

The three basic symmetry operators discussed in the previous section calculate the symmetry response on one scale. Although a larger kernel size could in theory be able to detect larger symmetrical structures, there are two problems with that approach. Firstly, since two pixels at opposite sides of the kernels center are compared, the pattern needs to be perfectly symmetrical to have matching gradients at pixels far from the center. This will cause problems when using complex stimuli of real-world scenes like we do in our study. Secondly, larger symmetry kernels greatly increase the computational load of the algorithm.

To be able to detect larger symmetrical patterns and to allow for small deviations from
3.3. Methods

perfect symmetry and speed-up of calculation, we apply a multi-scale approach using Gaussian image pyramids (see Figure 3.1b), similarly to (Itti et al., 1998).

The image, $I_0$, at scale zero is at its original resolution ($1024 \times 768$ pixels in our experiments). At subsequent scales, the image is first convolved with a Gaussian kernel, $G$, for low-pass filtering, and then downsampled to obtain an image that is half the width and height of the previous scale:

$$I'_{l-1} = I_{l-1} \ast G \quad (3.13)$$
$$I_l(x,y) = I'_{l-1}(2x,2y) \quad (3.14)$$

In our experiments, we used five different scales ($L = 5$). The resolution of the first scale, $I_0$, was $1024 \times 768$ pixels, and that of the highest scale, $I_4$, $64 \times 48$.

To determine the saliency map, the symmetry operator is applied to all Gaussian images in the pyramid. This results in $L$ symmetry maps at different scales. These maps are combined by first normalizing the maps, then resizing them to the same scale ($l = 2$), and finally adding the different maps:

$$S = \bigoplus_{l=0}^{L-1} N(S_l), \quad (3.15)$$

where $\bigoplus$ is the summation operator that first resizes all elements to the same scale, and then sums the maps pixel wise.

The normalization function, $N$, is adopted from (Itti et al., 1998) and has the purpose to promote symmetry maps at scales with only a few outstanding points, as opposed to symmetry maps that contain many similarly symmetrical patterns. The normalization function first scales the values in the map to the range $[0,1]$, so that the global maximum has a value of 1.0, and then multiplies all values in the map with $(1 - \bar{m})^2$, where $\bar{m}$ is the average value of all local maxima in the map that have a value greater or equal than 0.10. If there are many similarly symmetrical patterns, $\bar{m}$ will be large, and the map will thus be multiplied by a small value. If, on the other hand, there is one clear global maximum, $\bar{m}$ will be small, and the map will be weighed more strongly in calculating the total saliency map. Finally, the resulting saliency map will be normalized so that the total sum of all its elements is 1.0. Another normalization procedure based on lateral inhibition is discuss in (Itti & Koch, 2000). However, in our experience, that procedure results in too few salient locations. We try to predict eye fixations in a
free-view experiment with complex photographic stimuli where participants have many potentially interesting locations to focus on.

We designed our multi-scale symmetry-saliency model similarly to the multi-scale implementation of the contrast-saliency model (Itti et al., 1998) in order to provide a fair comparison of both methods.

### 3.3.2 Contrast-Saliency Model

We compare our symmetry-saliency model with the contrast-saliency model (Itti et al., 1998). In this section, a short overview of the contrast model is given to give the reader an idea of the mechanisms. For a full description, we refer to Appendix A.

The contrast-saliency model calculates saliency based on contrast in three different feature channels: intensity, color and orientation. Contrast is calculated by center-surround operations. The center is excited by the presence of a given feature, whereas the surround is inhibited, or vice versa. In the intensity channel, this corresponds to bright on dark or dark on bright. In the color channel, contrast is calculated using chromatic double-opponency channels, red on green, blue on yellow, or vice versa. Both color and intensity contrasts are implemented by using Gaussian image pyramids. The center-surround calculations are done by subtracting the image at different scales. The center is then taken as a pixel on a certain scale and the surround as the corresponding pixel on a coarser scale. For the calculation of orientation contrast, the Gaussian intensity images are convolved with Gabor filters in four different orientations. Again an image pyramid is constructed, and the center-surround orientation contrast is calculated by subtracting the Gabor-filtered images at different scales.

To obtain a multi-scale contrast-saliency model, contrast is calculated on three different scales, 2, 3, 4 (0 being the original resolution) and with a difference of both 3 and 4 scales between the center and the surround scales. The resulting feature maps on the different scales are normalized and combined similar to Equation 3.15, to form three conspicuity maps, for intensity, color, and orientation. To calculate the total contrast-saliency map, the conspicuity maps are first normalized using the earlier discussed normalization method, and then the average over the three maps is taken. Different from Itti, Koch, and Nieburs implementation, the resulting saliency map is at scale two, so that it is comparable with our symmetry-saliency map.
3.3. Methods

Figure 3.2: Examples of saliency maps produced by the three symmetry models and the contrast model as a response to the artificial stimuli. The color map goes from white (no response) to dark red (highest response). The contrast model has high response for the complete form. For the circle and square the highest points of activation are respectively near the edges and corners. The symmetry models, on the other hand, respond more specifically to the symmetrical center of the form, with the highest specificity for the radial symmetry model.

Itti et al. (1998) discuss a procedure to select a fixation location using winner-takes-all and inhibition-of-return operators. These operators are useful for modeling visual search, or to integrate bottom-up and top-down influences. However, since we are interested in the influences of saliency per se, we do not use this selection procedure, but rather compare the human fixations with the full saliency maps.

Some examples of saliency maps resulting from the symmetry models and the contrast model for artificial stimuli are given in Figure 3.2. There is a large difference between the symmetry and the contrast responses. Whereas the symmetry models specifically highlight the center of the objects, the contrast model gives a much more spread-out activation. For the circle and the square, the most salient points are even near the corners of the forms instead of at the center. The saliency map of the radial symmetry model...
is a little more focused on the center than those of the other symmetry models. Apart from that, the differences among the three symmetry models are relatively modest.

3.3.3  Eye-Tracking Experiment

To test the performance of both the symmetry and the contrast saliency model, we conducted an eye-tracking experiment to record eye fixations while participants viewed complex photographic images. The experiment is discussed in this section.

3.3.3.1  Participants

31 students (15 female, 16 male) of the University of Groningen took part in the experiment for credit points. The age of participants ranged from 17 to 32. All had normal or corrected-to-normal vision.

3.3.3.2  Stimuli

A total of 99 photographic images in five different categories were presented to the participants. 19 images were in the natural-symmetry category. These images were selected explicitly for containing symmetrical natural objects. To test if our methods are not only valid for scenes containing explicit symmetrical forms, but more generally for a wide range of images, we included four other categories in the image set: 12 images of animals in a natural setting, 12 images of street scenes, 16 images of buildings, and 40 images of natural environments. Figure 3.3 gives examples of the different categories included in the dataset. The five categories span a wide variety of images, containing natural symmetries and natural and cultural scenes, with organic and rectilinear shapes. All these images were taken from the McGill calibrated colour image database (Olmos & Kingdom, 2004).

The images were displayed full-screen with a resolution of 1024 × 768 pixels on an 18” CRT monitor of 36 by 27 at a distance of 70 cm from the participants. The visual angle was approximately 29° horizontally by 22° vertically.
3.3. Methods

Figure 3.3: Image examples for all five categories used in the experiment. In total, 99 images were used: 19 images of natural symmetries, 12 of animals, 12 of street scenes, 16 of buildings, and 40 of natural scenes.

3.3.3.3 Experimental Setup

Since we are interested in the bottom-up components of visual attention, the participants were asked to freely view the images. We did not give them a task, since that would give a strong bias on the eye movements. Still, the eye movements are likely to be also controlled top-down, by interests and experiences of the participants. We will discuss our method to capture the consensus among participants in the next subsection.

The images were presented in random order to the participants. Each image was displayed for five seconds. After each presented image, the participant could decide when to continue. The experiment was split up in sessions of approximately five minutes. Between the sessions, the participants had a short break, in which the experimenter had a relaxing conversation to keep the participants motivated and focused.
Figure 3.4: The distribution of human eye fixations for the different categories. The contour plots show the normalized kernel-density estimation of the fixations of all participants for all images in the category. The distributions of the natural-symmetry and animal category are biased to the center, whereas the fixations on the street scenes and buildings are more uniformly distributed.

3.3.3.4 Eye Tracker and Data Acquisition

We used the Eyelink I head-mounted eye-tracking system (SR research) to record the gaze of the participants. Fixations were extracted using the accompanying software. At the beginning of the experiment, the eye tracker was calibrated using the SR-research software. Before every session, the calibration was verified and the experiment continued when the system was correctly calibrated. If not, the eye tracker was recalibrated. Before every trial, i.e., before every presentation of an image, drift was measured by letting the participant focus on a cross displayed in the center of the screen, and the estimation corrected if necessary. Because of the drift correction method, the first fixation was strongly biased. We therefore eliminated this fixation from the data. Using the eye tracker, we acquired 99 trials of five seconds for all 31 participants. A few trials were not used in the data analysis due to interruptions or other incidents.

3.3.3.5 Eye-Tracking Data

On average, the participants made 15.6 (±3.6) fixations while viewing an image for five seconds. The normalized distributions of fixations for the five categories are given in Figure 3.4. The figure shows that the fixations are not uniformly distributed of the image, but biased towards the center. The figure shows that the fixations are not uniformly distributed over the image, but biased towards the center. The standard deviation of the angular distance from the fixations to the center is respectively 8.0°, 8.2°, 9.0°, 9.1° and 8.6° for the images of natural symmetries, animals, street scenes, buildings, and natural scenes. This center bias is expected to occur in free-viewing experiments.
(Tatler, 2007), and might be a result of both the tendency of photographers to place the important objects near the center, and the tendency of humans to center the eyes. In our data, the center bias is stronger for the natural symmetry and animal images, and weaker for the images of street scenes and buildings. In the center-bias experiment, discussed in the results section, we incorporate different strengths of center bias in the saliency maps to investigate the influence of a center bias in predicting human eye fixations.

3.3.3.6 Center Bias

As can be seen in Figure 3.4, the human eye fixations are biased towards the center of the image. To investigate the role of a center bias on the comparison between the saliency models and the human data, we include a center bias in the models similar to (Parkhurst et al., 2002). To do so, the values in the saliency map, $S$, are weighted with a two-dimensional Gaussian distribution with its mean at the center of the image, and a standard deviation, $\sigma_b$, that determines the strength of the center bias, with small values corresponding with strong center bias:

$$S'(p) = S(p)e^{-\|p-\mu\|^2/2\sigma_b^2},$$  

(3.16)

where $p$ is the location of a pixel in the map and $\mu = (512.5, 384.5)$ is the center of the image. The resulting central-biased saliency map, $S'$, is normalized so that the total sum is 1.0.

3.3.4 Comparison Methods

We used two methods to compare the human eye-fixation patterns with the predictions from the saliency models. A correlation method similar to that used in (Le Meur et al., 2006; Ouerhani et al., 2004) and a fixation-saliency method, similar to that used in (Parkhurst et al., 2002). Both methods are discussed in this section.

3.3.4.1 Correlation Method

To correlate the human data with the output of the saliency models, we transform the eye-fixation data to fixation-distance maps (see Figure 3.5). These fixation-distance
maps give the probability that a fixation lands on a certain location based on the human data. Similarly, the saliency maps can be seen as giving the probability of a fixation on that location based on the saliency models. To construct a fixation-distance map from an eye-fixation pattern, the inverse distance transform of the fixation data is calculated. The distance transform, $F'$, gives the distance to the nearest fixation for all pixels in the image. This results in values of zero at the points of fixation with a linear increase at pixels further away from the fixations:

$$F'(\mathbf{p}) = \|\mathbf{p} - \mathbf{f}_n\|, \quad (3.17)$$

where $\mathbf{p} = (x, y)$ is the pixel location, $\mathbf{f}_n = (x_n, y_n)$ is the location of the nearest human-fixation point, and $\|\|$ is the Euclidean distance between the two. Next, the fixation-distance map, $F$, is obtained by subtracting all values from the maximum value in the distance transform:

$$F(\mathbf{p}) = \max(F') - F'(\mathbf{p}) \quad (3.18)$$
3.3. Methods

Figure 3.6: The correlation method. The fixation-distance map obtained from the human eye fixations is correlated with the saliency map calculated from the same image. The correlation results in a correlation coefficient that shows how well the saliency model predicts the human data.

\[ F \text{ is normalized so that the sum of its elements is } 1.0. \] This results in a map with high values at the points of fixations, and lower values further from these points. This is slightly different from the approach in (Kootstra et al., 2008a; Le Meur et al., 2006; Ouerhani et al., 2004), where a fixation density map is calculated using Gaussian kernels. Our method puts emphasis on the location of fixations, rather than on their density. Moreover, this correlation method is parameter free, i.e., there is no width of the kernel to be set.

In Figure 3.6, the correlation method to compare the saliency maps with the fixation-distance maps is depicted. The two maps are correlated with each other to get the correlation coefficient, \( \rho \):

\[
\rho = \frac{\sum_{p \in P} ((F(p) - \mu_F)(S(p) - \mu_S))}{\sqrt{\sigma_F^2 \sigma_S^2}},
\] (3.19)
where \( P \) is the set of all pixel coordinates and \( \mu \) and \( \sigma^2 \) are respectively the mean and the variance of the values in the maps. The correlation coefficient has a value between 1 and 1. A \( \rho \) of 0 means that there is no correlation between the two maps, which is true when correlating with random fixation-distance maps. Values for \( \rho \) close to zero indicate that a model is a poor predictor of human fixation locations. Positive correlations show that there is similar structure in the saliency map and the human fixation map.

In the above described correlation method, the predictions of the saliency models are compared to the fixation-distance maps of individual participants. However, the photgraphic images viewed by the participants are highly complex stimuli that generate many fixations, with substantial variation among the participants. Because of this variation, the correlations of individual fixation-distance maps with the saliency maps will be low. However, some of the fixations are shared by all participants, and are more likely to be caused by bottom-up factors. Because we are interested in general models and not in models that predict visual attention of specific persons, we want to test how well the saliency models predict the consensus among participants as well. To test this, we calculate the correlation coefficient for the combined fixation-distance maps (Figure 3.5). These combined maps are calculated by summing the individual fixation-distance maps:

\[
F_c = \sum_{i=1}^{N} F_i,
\]

where \( F_i \) is the individual fixation-distance map for participant \( i \), \( F_c \) the combined fixation-distance map showing the consensus, and \( N = 31 \). \( F_c \) is normalized so that the elements sum up to 1.0. The saliency maps are compared to the combined fixation-distance maps using Equation (3.19).

### 3.3.4.2 Fixations-Saliency Method

The fixation-saliency method tests how the saliency at the points of human fixation according to the saliency methods compares to the average saliency for the image (see Figure 3.7). Whereas the correlation methods looks at both the presence and the absence of fixations and saliency, the fixation-saliency method focuses on the locations where there actually are eye fixations. With the method, we can investigate whether the local symmetry and contrast at the fixation points are above average. Moreover, it gives
3.3. Methods

Figure 3.7: The fixation-saliency method. The saliency, as calculated by the saliency models, is measured in a patch around the fixation points. The mean saliency for the human fixations is divided by the average saliency in the image, resulting in the fixation-saliency score.

The fixation-saliency score, $\lambda$, is calculated by calculating the average saliency according to the saliency model in a patch around the points of fixation divided by the average saliency over a large number of random points, $K = 1000$. For a given participant and image, that is:

$$\lambda = \frac{\frac{1}{M} \sum_{i=1}^{M} s(f_i)}{\frac{1}{K} \sum_{j=1}^{K} s(r_j)},$$

(3.21)

where $f_i$ is the $i$th fixation of a total of $M$ fixations, $r_j$ is a randomly generated point in the image, and $s()$ gives the average saliency in a patch around the point:

$$s(x,y) = \frac{1}{(2R+1)^2} \sum_{j=-R}^{R} \sum_{i=-R}^{R} S(x + i, y + j),$$

(3.22)

where $R = 28$ pixels. When $\lambda > 1$, the saliency is higher at the fixation points than
average, showing that the given saliency model can predict the fixations to some extent. The above method calculates the average fixation saliency for all fixations in the sequence. To investigate the progression in saliency over the sequence, we calculate the fixation saliency for individual fixations, $\lambda_i$:

$$\lambda_i = K \cdot s(f_i) / \sum_{j=1}^{K} s(r_j).$$

(3.23)

3.4 Results

In this section we discuss the results of the comparison of the symmetry and contrast saliency models with human eye fixations. We firstly show the results of the correlation and fixation-saliency methods on the fixation patterns of individual participants viewing an image. Secondly, we discuss the results of the correlation comparison with the fixations of all participants combined. Next, the saliency over the fixation sequence is shown. Finally, an analysis of the center bias is discussed.

3.4.1 Individual Fixation Patterns

3.4.1.1 Correlation

In Figure 3.8 the results of the correlation between the individual fixation-distance maps and the saliency maps are given. The five groups of bars contain the results for the different image categories. The bars show the mean correlation coefficients, $\rho$, over all participants and images in the category for the different saliency models. The error bars give the 95% confidence intervals on the mean. The scores of the saliency methods are plotted along with the inter-participant correlation, and the correlation of the human data with random fixations. The first, which indicates how well one persons fixations correlate with those of the others, is depicted by the horizontal gray bar with a solid mid-line, giving the mean and 95% confidence interval. The correlation with random fixations is depicted by the horizontal dashed line, which is, as expected, virtually zero for all categories. All means and confidence intervals in this paper are calculated using multi-level bootstrapping. Significant differences can be appreciated by looking at the 95% confidence intervals.
3.4. Results

Figure 3.8: Correlation between the saliency maps and the individual fixation-distance maps. The groups of bars relate to the different image categories. The bars give the mean correlation coefficients. The error bars are the 95% confidence intervals. The horizontal gray bars with the solid line show the mean and 95% confidence interval of the inter-participant correlation. The correlation of the human data with random fixations are given by the dashed lines, which are close to zero. It can be appreciated that the symmetry models significantly outperform the contrast model, not only on the natural-symmetry category, also on the other categories.

The inter-participant correlation is calculated for every image by correlating the fixation-distance maps of every participant with those of all other participants, resulting in a similarity measure among participants. The plot shows that there is variability among the participants. The saliency methods are also faced with this variability, which pulls down the correlation values. The inter-participant correlation can therefore be used to put the scores of the saliency methods into perspective. It must be noted that the correlation scores of the models can be higher than the inter-participants scores when the variation among participants is high. The models can then predict the consensus among the participants better than the participants themselves can.

Figure 3.8 clearly shows that the symmetry models compare significantly better with the human data than the contrast models for the images containing natural symmetries. This is as expected, since the images were selected on the basis of symmetry. Moreover, also for the other categories the correlation scores are significantly higher for the symmetry models than for the contrast model. This suggests that the symmetry models
3.4.1.2 Fixation Saliency

If we look at the fixation-saliency value, $\lambda$, in Figure 3.9, we see that both contrast and symmetry are higher at the points of fixation than in the rest of the images ($\lambda > 1$), showing that both can predict eye fixations to some extent. Especially for the natural-symmetry category, the symmetry models score significantly better than the contrast model. Also for the other categories, except for the animal category, symmetry scores have general validity. The performance of the symmetry models is in the same range as the inter-participant correlations. The performance of the contrast model correlates with the inter-participant score. High inter-participant scores reflect that the individual fixation patterns are more similar, presumably because there are fewer interesting locations for the participants to focus on. The contrast model scores better in these cases than it does when there is more variability among the participants. The performance of the symmetry models, on the other hand, is significantly better for all image categories, and they seem to predict the consensus among participants better even when there is more variability. Among the three symmetry models, isotropic, radial, and color, we do not see significant differences in performance.

**Figure 3.9:** The fixation-saliency results. The bars give the mean saliency near the eye fixations relative to the average saliency in the complete image. The 95% confidence intervals are given by the error bars. The local symmetry at human points of fixation is significantly higher than the contrast at these points for most of the categories, except for the animal images.
significantly better, with an exception for the isotropic model in the street-scene category, which scores better, but not significantly ($\alpha = 0.05$). The animal category gives a different result. There the contrast model scores better than the symmetry model, although not significantly. This result might be explained by the fact that, in contrast with the images in the other categories, many images in this category contain objects animals that are highly distinguishably and sharply depicted on an out-of-focus background. The fore- and backgrounds in the other images are less distinct and more cluttered. In the animal images, there are fewer interesting locations and the background also has less contrast. Among the different symmetry-saliency models, there are no clear differences with the exception of the radial symmetry model on the natural-symmetry category, which scores somewhat better than the others.

The results of both the correlation method and the fixation-saliency model show that, in general, the symmetry-saliency model predicts the eye fixations significantly better than the contrast-saliency model.

### 3.4.2 Combined Fixation Patterns

In Figure 3.8, the saliency maps are correlated with the individual fixation-distance maps. Because there is much variety in the fixation patterns among the participants, the correlation scores are relatively low. Some of the locations in the images, however, are attended by most participants. To investigate how well this consensus is predicted by the saliency models, we combined the fixation-distance maps of the individual participants. The correlation coefficients, $\rho$, of this analysis are given in Figure 3.10. The bar plots show a similar structure as that in 3.8: the symmetry models significantly outperform the contrast model. However, the correlation coefficients went up from around 0.4 to around 0.7 for the symmetry models. This shows that the symmetry models do a good job in predicting the fixation consensus among the participants. Again, this is not only true for the images containing explicit symmetrical forms, but for all categories. This shows that the common fixations of the participants are well captured by the symmetry-saliency models.
Figure 3.10: Correlation between the saliency maps and the combined fixation-distance maps, representing the consensus among the participants. The bars and error bars give the mean and 95% confidence intervals on the mean of the correlation coefficients. The results show the same pattern as for individual fixation-distance maps, with significantly higher scores for the symmetry models. However, the correlation coefficients are much higher, showing a better fit of the models with the participants consensus.

3.4.3 Fixation sequence

In the above, we compared the complete five seconds fixation sequence with the saliency models. In Figure 3.11, the progression of fixation saliency, $\lambda_i$, as a function of the fixation number is shown. It can be appreciated that the symmetry is especially high for the first fixation, and gradually drops for later fixations. This shows that the participants first attend highly symmetrical parts of the image. The contrast at the points of fixation, however, is lower, and runs much more stable over the sequence, except for the animal condition. The difference between the symmetry models and the contrast model is significant for the first fixations for all categories except for the animal images. For later fixations, the difference is less apparent, but still in favor of the symmetry models, and significant for the nature category. The results for the animal condition are different. The fixation saliency of the models is not significantly different and higher for contrast. Contrast is here also high for the first fixations, and lower for later. This might again be explained by the different style of the photos compared to the other categories.
3.4. Results

Figure 3.11: Fixation saliency, i.e., the saliency as measured by the models near the human fixation points, over the fixation sequence. The fixation saliency is plotted as a function of time measured by the fixation number. The lines give the mean fixation-saliency scores, and the error bars the 95% confidence intervals on the mean. The symmetry at fixation points is especially high for early fixations, and drops for later, showing that the fixations can be order on the basis of symmetry. The contrast at fixation points is lower than symmetry and is more constant over the sequence, except for the animal category, where the contrast model shows a similar result as the symmetry models.

3.4.4 Central Bias

In order to test whether the performance of the models is influenced by the center bias that the participants displayed Figure 3.4, we added a center bias to the saliency maps as explained in the methods section. Figure 3.12 shows the correlation coefficients as a function of the center-bias strength, $\sigma_b$, where the combined fixation-distance maps are compared with the center-biased saliency maps. The curves of the contrast-saliency model show a maximum correlation value for $\sigma_b$ between $6^\circ$ and $9^\circ$. The maxima are at $6^\circ$, $7^\circ$, $9^\circ$, $8^\circ$, and $7^\circ$ for respectively the natural-symmetry, animal, street-scene, building, and natural-scene category. This is similar to what is reported in (Parkhurst et al., 2002). The curves of the symmetry-saliency models, on the other hand, do not show a maximal value. They gradually grow when the center-bias is weakened and reach an asymptote between $12^\circ$ and $15^\circ$. The maximally performing central-bias standard deviation for the contrast model is somewhat lower than the standard deviations on the distance to the center as observed in the human data (see subsection about eye-tracking data). The results show that the contrast model can be improved using a center bias, whereas the symmetry models give better results without such a bias. The performances of the symmetry models are significantly better than that of the contrast
model, even when the center bias is applied. There is virtually no difference among the different symmetry models. The fact that the performance drops for the contrast model when the center bias is weakened, suggests that the model incorrectly predicts eye fixations at the periphery of the images. The symmetry models, however, seem to predict valuable fixations in the periphery, since the performance increases, even for standard deviations higher than those observed in the human data. It is therefore better not to apply a center bias to the predictions of the symmetry models.

### 3.5 Discussion

In this chapter, three saliency models for the prediction of human eye fixations based on local symmetry were presented. The models were compared to a saliency model that is based on contrast features (Itti et al., 1998). To test the models, we conducted an eye-tracking experiment using a wide variety of different images. The results show that the symmetry-saliency model compares substantially better with the human data than the contrast-saliency model.

The analysis of the correlation between the models predictions and human fixations shows significantly better performance for the symmetry models, not only for the im-
ages containing explicit symmetries, but for all image categories. The comparison with
the combined fixation-distance maps shows that the models capture the fixation con-
sensus among the participants particularly well. This suggests that local symmetry can
be used as a general model for the prediction of human eye fixations.

The analysis of the fixation saliency gives similar results. The amount of local symme-
try at the points of human eye fixation is well above average and exceeds the contrast
at fixation points for most image categories except for the animal images. Moreover,
the amount of symmetry at the points of fixation is especially high for the first fixations
with a gradual drop for later fixations. The contrast saliency shows a flat curve over the
fixation sequence. This suggests that humans attend to locally symmetrical regions in
an image, and moreover that symmetry can be used to order the fixation sequence.

The distribution of the human fixation data is a little biased towards the center. The
addition of a center bias results in a maximum performance for the contrast model at
a slightly stronger bias than in the human data. The performance of the symmetry
models, on the other hand, does not have a maximum, but grows for weaker center
biases. This suggests that the symmetry models find valuable salient points in the
periphery, which are attended to by the human observers. The contrast model, on the
other hand, suggests salient points in the periphery that do not correspond to human
fixations.

The fixation saliency of the contrast model is different for the images in the animal cat-
egory than for the other categories. The main difference between the categories is that
the images in the animal category depict the animals clearly and sharply on blurred and
out-of-focus backgrounds. This results in high contrast between foreground and back-
ground. The other categories contain images with sharper and more cluttered back-
grounds.

The experiments reveal no significant difference among the three symmetry models,
whereas the radial symmetry model was expected to perform better since humans are
also more sensitive to patterns with multiple axes of symmetry. However, the isomet-
ric symmetry model already results in higher activation for these kinds of patterns.
The extra promotion of multiple symmetry axes apparently only slightly changes the
symmetry saliency maps. Similarly, the addition of color does not result in substantial
changes in performance as well.

Although the performance of the contrast models in the presented experiments is less
than that of the symmetry models, contrast obviously also plays a role in visual atten-
tion. Both the correlation and the fixation saliency of the contrast model are well above chance levels, conform the findings of for instance (Le Meur et al., 2006; Parkhurst et al., 2002; Parkhurst & Niebur, 2003). Moreover, the symmetry models also exploit contrasts in the image gradients to determine symmetry. The main difference between the symmetry and contrast model is the specificity, as can be seen in Figure 2.5 and Figure 3.2. The contrast model gives a more spread-out activation less focused on the centro of objects. This reduces the similarity to the human data. It makes sense to combine the symmetry and contrast model to further improve the prediction of eye fixations.

Both analysis methods show a correlation between local symmetry and human eye fixations. However, that does not prove that there is a causal relation between symmetry and overt visual attention. However, we think that a causal relation is likely, especially considering that symmetry can be used for figure-ground segregation. We discuss this further in Chapter 5.

In (Findlay, 1982; He & Kowler, 1989; Kaufman & Richards, 1969; Ottes et al., 1984) eye fixations are reported to land at the center of gravity of objects. A center of gravity is strongly correlated to the center of symmetry of an object. Our research therefore suggests that the center-of-gravity effect is not only true for simple artificial stimuli like the ones used in the above-mentioned studies, but also for complex photographic images of natural and man-made scenes.

In reality, fully symmetrical forms are almost never observed. When an object is viewed from a nonorthogonal angle, its appearance in the two-dimensional projection is not perfectly symmetrical. This is termed skewed symmetry. Although this is true, still the amount of local symmetry as calculated by our model is higher for the skewed symmetry than for other random configurations in the image. On a side note: Humans have more difficulties perceiving skewed symmetry, but are capable of doing so with sufficient depth cues Wagemans (1993).

As pointed out in (Land & Hayhoe, 2001; Schumann, Einhäuser-Treyer, Vockeroth, Bartl, Schneider, & König, 2008), natural human behavior might be different from the behavior observed in lab experiments. We therefore think it is interesting to study the role of symmetry in overt visual attention during natural behavior such as playing cricket (Land & McLeod, 2000) or making tea (Land & Hayhoe, 2001). In a dynamic setting, symmetry in motion can also be used.
To conclude, our results suggest that symmetry plays a role in the guidance of eye movements, either directly or indirectly by being a cue for the presence of objects. We advocate the study of the role of symmetry in human vision.
Chapter 3. Predicting Human Eye Fixations by Local Symmetry