

Using Symmetry to Select Fixation Points for Segmentation

Gert Kootstra, Niklas Bergström, and Danica Kragic
CAS/NADA, KTH, Stockholm
[kootstra, nbergst, dani]@kth.se

Abstract

For the interpretation of a visual scene, it is important for a robotic system to pay attention to the objects in the scene and segment them from their background. We focus on the segmentation of previously unseen objects in unknown scenes. The attention model therefore needs to be bottom-up and context-free. In this paper, we propose the use of symmetry, one of the Gestalt principles for figure-ground segregation, to guide the robot's attention. We show that our symmetry-saliency model outperforms the contrast-saliency model, proposed in [4]. The symmetry model performs better in finding the objects of interest and selects a fixation point closer to the center of the object. Moreover, the objects are better segmented from the background when the initial points are selected on the basis of symmetry.

1. Introduction

To have a robot understand an unknown visual scene, it needs to pay attention to the objects in the scene. However, when the scene and the objects are unknown, no top-down knowledge can be used to detect the objects and segment them from the background. This paper deals with the problem of bottom-up detection and segmentation of unknown objects. We focus in particular on the development of a visual-attention model for the selection of fixation points to initiate segmentation.

Existing segmentation methods, such as methods based on graph cuts [8], and our current work using 3D information and belief propagation [2], presume a fixation point from where to start the segmentation. In these approaches, this is provided by the human operator. We aim to automate this process. It can be seen in Fig. 1 that the quality of segmentations improves when fixation points lie near the center of the object. Fixations selected by our visual-attention model should therefore a) be on the object, and b) be close to the center of the object to facilitate segmentation. We propose a visual-

attention model based on local symmetry in the image.

Many visual-attention models determine the saliency in an image. This is often calculated by center-surround contrasts of *basic features*, like in the well-known model of Itti *et al.* [4] that utilizes brightness, color, and orientation contrasts. We will refer to this model henceforth as the *contrast-saliency model*. However, when interpreting a scene, humans pay attention to objects, not so much to basic features. Hence, *configural features* play an important role in human visual attention. A configural feature is a higher-level feature that integrates different parts of a figure. Symmetry, for instance, can be a stronger visual attractor than basic features [9]. Predictions of human eye fixations based on symmetry also outperform predictions using the contrast-saliency model [5]. Contrast focuses mainly near corners and edges, whereas symmetry highlights symmetrical configurations which often coincide with objects in the scene.

Since we focus on the segmentation of previously unseen objects in unknown scenes, no top-down knowledge about the object can be used. The Gestalt theory studies this topic in human visual processing. A number of cues have been established that are used for figure-ground segregation, among which symmetry is one [3]. Symmetrical regions are more likely to originate from objects than less symmetrical regions, since symmetry is non-accidental.

The contributions our work are i) a visual-attention model based on symmetry to select fixation points near

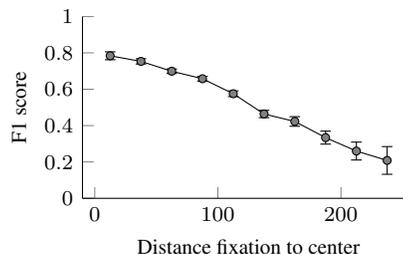


Figure 1: Segmentation performance as a function of the distance of the fixation point to the center of the object.

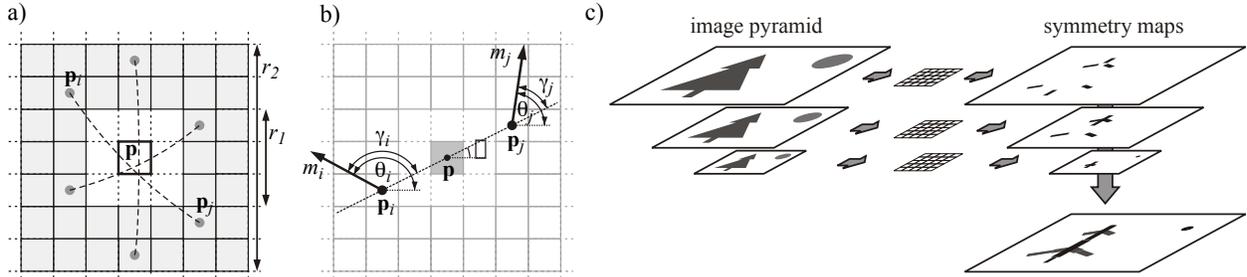


Figure 2: The symmetry model. a) The symmetry kernel that is applied to all pixels. b) The symmetry contribution of a pixel pair is based on brightness gradients. c) The symmetry responses on different scales are combined to the symmetry-saliency map.

the center of objects, and ii) the use of this model for automatic object segmentation. We furthermore present a new object-segmentation database. The results show that our symmetry-saliency model outperforms the contrast-saliency model in detecting unknown objects in the scene and selecting fixation points near the center of the objects. Our model furthermore results in better segmentation performance.

2 Methods

2.1 Symmetry-saliency model

The work that we present here extends our prior work on the prediction of human eye fixations [5] and the selection of visual landmarks for robotic SLAM [6]. The model builds upon the symmetry operator of Reissfeld *et al.* [10] and is extended to a multi-scale model similar to the contrast-saliency model [4].

The amount of local symmetry at a given pixel, $\mathbf{p} = (x, y)$, is calculated by applying a symmetry kernel to the pixel. Pixels pairs in the symmetry kernel contribute to the local symmetry value. A pair consists of two pixels, \mathbf{p}_i and \mathbf{p}_j , so that $\mathbf{p} = (\mathbf{p}_i + \mathbf{p}_j)/2$ (see Fig. 2a). The contribution of the pixel pair to the local symmetry of \mathbf{p} is calculated by comparing the intensity gradient g_i at \mathbf{p}_i and gradient g_j at \mathbf{p}_j :

$$c(i, j) = (1 - \cos(\gamma_i + \gamma_j)) \cdot (1 - \cos(\gamma_i - \gamma_j)) \quad (1)$$

where $\gamma_i = \theta_i - \alpha$ is the angle between the orientation of the gradient, θ_i , and the angle, α , of the line between \mathbf{p}_i and \mathbf{p}_j (see Fig. 2b). The first term in Eq. (1) has a maximum value when $\gamma_i + \gamma_j = \pi$, which is true for gradient orientations that are mirror symmetric in \mathbf{p} . Using only this term would also respond to two pixels that have the same gradient orientation and thus lie on a straight edge. Since we want to find 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:

$$s(i, j) = d(i, j, \sigma) \cdot c(i, j) \cdot \log(1 + m_i) \cdot \log(1 + m_j) \quad (2)$$

where m_i is the magnitude of the gradient, and $d(i, j, \sigma)$ is a Gaussian weighting function on the distance between \mathbf{p}_i and \mathbf{p}_j with a standard deviation of σ . We used $\sigma = 16$ in our experiments. 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 total symmetry value at \mathbf{p} is calculated by summing the contributions of all symmetrical pixel pairs in the kernel, $\Gamma(\mathbf{p})$. The symmetry kernel is defined by an inner and an outer square centered around \mathbf{p} . The size of the sides of the squares are respectively r_1 and r_2 . All pixels that lie inside the outer square, but outside the inner square are considered (gray-color pixels in Fig. 2a). In our experiments we used $r_1 = 7$ and $r_2 = 17$. Lower values for r_1 results in many spurious salient points. Higher values for r_2 make the model computational more expensive and more susceptible to small deviations from perfect symmetry, which often occurs in photographic images. Larger symmetric patterns are detected in less detail on higher scales. The total amount of local symmetry at $\mathbf{p} = (x, y)$ is then:

$$S_l(\mathbf{p}) = \sum_{(i,j) \in \Gamma(\mathbf{p})} s(i, j) \quad (3)$$



Figure 3: Examples of image in the databases. First row: MSRA Saliency Object Database, second row: KOD Database [1].

where S_l is the resulting symmetry map at scale l .

The multi-scale symmetry-saliency model is depicted in Fig. 2c. Similar to the contrast-saliency model [4], the scale space consists of an image pyramid that is built by progressively applying a Gaussian filter, followed by a downscaling of the image by a factor of two ($l = 0$ is the original resolution). Applying the symmetry operator results in a pyramid of symmetry maps. These symmetry maps are resized to the size of the first scale, and summed up to result in the overall symmetry map:

$$S(x, y) = \bigoplus_{l=L_1}^{L_2} S_l(x, y) \quad (4)$$

where L_1 is the first, and L_2 is the last scale. The operator \bigoplus rescales all maps to the first scale, and subsequently sums the values of the different scales.

The fixation points are selected from the saliency map with inhibition of return to spread the fixations over the image. The fixation points are selected in an iterative process: 1) All local maxima in the map are found. 2) The highest local maximum is selected as the next fixation point. 3) The salient region belonging to that point is found by region growing, so that all connected pixels are included that have a saliency value greater than 60% of the fixation’s saliency. 4) All local maxima in this region are devaluated using Gaussian weighting on the distance to the fixation. Step 2–4 are repeated.

The attention model runs at approximately 10Hz on a 2.53 GHz Intel processor.

2.2 Segmentation method

We use the active segmentation method of Mishra *et al.* [8], which gives a segmentation based on a fixation point. The method combines monocular and stereo cues. The edges in the image are reweighted based on depth. If the depth on either side of the edge is different, the edge is likely to be an object-boundary edge. Otherwise, the edge is more likely to be an internal edge. Using the fixation point, the image is transformed into



Figure 4: Segmentation examples. First row: symmetry model, second row contrast model. The fixation point is marked with white cross.

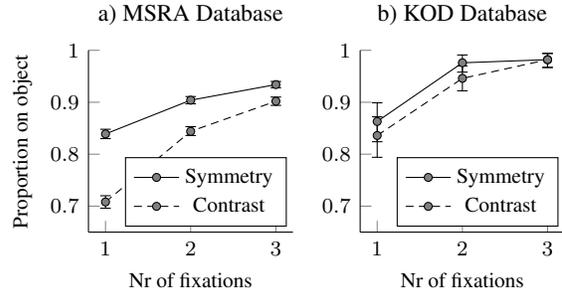


Figure 5: The object-detection performance

polar coordinates and the optimal segmentation is found by a graph cut using the depth-reweighted edges and color information. This method has been shown to outperform other segmentation methods.

In the experiments, we use the saliency method to select fixation points to initialize the segmentation.

3 Experiments and Results

The performance of our symmetry-saliency model is compared to the contrast-saliency model [4]. We investigate the performance to detect the object in the scene, the ability to find fixations near the center of the object, and the quality of the segmentation based on the selected fixation points.

The performance is tested on two datasets. The MSRA Salient Object Database [7] contains 5.000 images and the bounding boxes of the most salient object in the images labeled by nine participants. The images are diverse and objects are generally on very cluttered backgrounds. Our KTH Object and Disparity (KOD) Database [1] contains images and disparity maps of 25 objects with different light conditions, backgrounds, and object poses. Detailed ground-truth segmentations of the objects are included. Examples of images from the databases are given in Fig. 3.

Object detection In Fig. 5, the proportion that any of the fixations is on the object is plotted as a function of the number of fixations. It shows that the object-detection performance of the symmetry model is higher than that of the contrast model. The difference in Fig. 5(a) is highly significant. In Fig. 5(b), performance for the symmetry model is also higher, but the difference is not significant and the performance is high for both. It can be explained by the less complex backgrounds in this database. The results suggest that symmetry is particularly advantageous for object detection in cluttered scenes. The performance increases when more fixations are chosen, but is already high for the first, most salient, fixation.

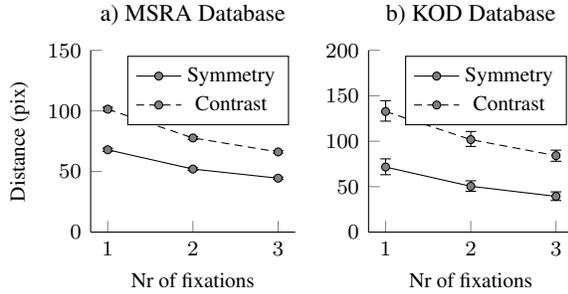


Figure 6: The center-fixation performance

Center fixation The distance of the closest fixation to the center of the object is plotted in Fig. 6. The plots show that the symmetry model selects fixation points that are much closer to the object’s center. This is an important property, as we have seen in Fig. 1 that the segmentation performance is much improved when the fixation point is closer to the center. The performance increases when more fixations are chosen, but is already good for the first fixation.

Segmentation Finally, we investigate whether the superior performance of the symmetry model in the previous experiments also leads to a better segmentation performance. This is tested on the KOD Database. The quality of the segmentations is measured with the F1 score. The results are shown in Fig. 7. Fig. 7(a) shows the F1 score of the best segmentation resulting from any of the fixations. The performance is significantly higher when the fixations are selected by the symmetry model. Fig. 7(b) shows the F1 score per fixation. This illustrates that the first fixation (the most salient) results in the best segmentations for the symmetry model. The segmentations with the contrast model are not better for the first than for later.

4 Discussion and Conclusion

We presented our symmetry-saliency model for the selection of fixation points. All experiments show improved performance over the contrast-saliency model [4]. The symmetry model has more success in detecting unknown objects in unknown scenes. The model furthermore selects fixation points that are closer to the center of the object, which is a good property of a visual-attention model, since the quality of object segmentation is strongly effected by the distance from the fixation to the object’s center. The experiments indeed show that the segmentation performance is significantly higher when the symmetry model is used to select fixation points.

The experiments show that the performance of the symmetry model is already good for the first fixation.

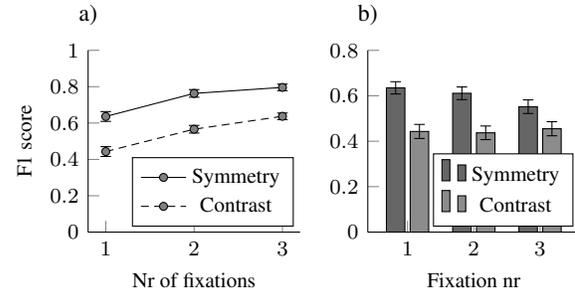


Figure 7: The segmentation performance. (a) shows the best segmentation of any of the fixation. (b) shows the performance per fixation.

This is an important property on, for instance, a robot that physically needs to focus its cameras.

We conclude that the proposed symmetry-saliency model is a successful model for the selection of fixation points for segmentation.

References

- [1] N. Bergström and G. Kootstra. <http://www.csc.kth.se/~kootstra/kod>.
- [2] M. Björkman and D. Kragic. Active 3d scene segmentation and detection of unknown objects. In *International Conference on Robotics and Automation (ICRA)*, 2010.
- [3] J. Driver, G. C. Baylis, and R. D. Rafal. Preserved figure-ground segregation and symmetry perception in visual neglect. *Nature*, 360:73–75, 1992.
- [4] 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*, 20(11):1254–1259, 1998.
- [5] G. Kootstra and L. R. B. Schomaker. Prediction of human eye fixations using symmetry. In *Cognitive Science Conference (CogSci)*, Amsterdam, The Netherlands, 2009.
- [6] G. Kootstra and L. R. B. Schomaker. Using symmetrical regions-of-interest to improve visual SLAM. In *the International Conference on Intelligent Robots and Systems (IROS)*, St. Louis, USA, October 11-15 2009.
- [7] T. Liu, J. Sun, N.-N. Zheng, X. Tang, and H.-Y. Shum. Learning to detect a salient object. In *Int. Conf. on Computer Vision and Pattern Recognition (CVPR '07)*, 2007.
- [8] A. Mishra, Y. Aloimonos, and C. L. Fah. Active segmentation with fixation. In *Proceedings of the International Conference on Computer Vision (ICCV)*, 2009.
- [9] J. R. Pomerantz. Colour as a gestalt: Pop out with basic features and with conjunctions. *Visual Cognition*, 14(4-8):619–628, 2006.
- [10] D. Reisfeld, H. Wolfson, and Y. Yeshurun. Context-free attentional operators: The generalized symmetry transform. *International Journal of Computer Vision*, 14:119–130, 1995.