Local 3D Symmetry for Visual Saliency in 2.5D Point Clouds

Ekaterina Potapova, Michael Zillich, Markus Vincze

Automation and Control Institute, Vienna University of Technology

Abstract. Many models of visual attention have been proposed in the past, and proved to be very useful, e.g. in robotic applications. Recently it has been shown in the literature that not only single visual features, such as color, orientation, curvature, etc., attract attention, but complete objects do. Symmetry is a feature of many man-made and also natural objects and has thus been identified as a candidate for attentional operators. However, not many techniques exist to date that exploit symmetry-based saliency. So far these techniques work mainly on 2D data. Furthermore, methods, which work on 3D data, assume complete object models. This limits their use as bottom-up attentional operators working on RGBD images, which only provide partial views of objects. In this paper, we present a novel local symmetry-based operator that works on 3D data and does not assume any object model. The estimation of symmetry saliency maps is done on different scales to detect objects of various sizes. For evaluation a Winner-Take-All neural network is used to calculate attention points. We evaluate the proposed approach on two datasets and compare to state-of-the-art methods. Experimental results show that the proposed algorithm outperforms current state-of-the-art in terms of quality of fixation points[1].

1 Introduction

Attention has been studied extensively for many years [1–4]. The value of attention, e.g. for robotic applications, has been demonstrated by Aloimonos *et al.* [5]. Many attentional systems concentrate on pre-attentive features, such as color contrast, orientation, curvature, etc. It has been shown, however, that not only single popping out features attract bottom-up attention, but also complete objects do [6, 7].

Symmetry is one of the characteristics of human-made and natural objects, and thus can be seen as an objectness measure. Kootstra *et al.* [8] showed that human eye fixations can be predicted well by symmetry. Many symmetry operators exist [9–13], which can be divided into two major groups: operators working on 2D data, and operators working on 3D data. Among the 2D operators one well-known operator was developed by Reisfeld *et al.* [9] which detects context-free generalized symmetry based on magnitude orientations of image gradients.

[1] The research leading to these results has received funding from the Austrian Science Fund (FWF) under project TRP 139-N23 InSitu

This symmetry operator was extended by Heidemann *et al.* [10] to a local color symmetry operator. Loy and Zelinsky [11] proposed to detect local radial symmetries in an image using a special transform. Kootstra *et al.* [14] proposed a 2D symmetry saliency operator based on the symmetry operator by Reisfeld *et al.* [9]. This saliency operator is able to detect symmetrical regions at different scales. The basic idea is that symmetries are computed over multiple scales and then summed in across-scale addition manner to obtain a master saliency map. Kootstra *et al.* [14] showed that this approach works better than classical contrast model saliency [15]. Mitra *et al.* [16] gave an extensive overview of the existing methods to detect different types of symmetries in 3D geometries. Methods based on search in oriented histograms [17], spectral analysis [12], feature-graph matching [13] and many others were indicated. The majority of 3D algorithms work on a complete 3D model of an object to detect symmetries. This property limits their use as bottom-up attentional operators working on RGBD images, i.e. partial views of objects.

In this paper, we propose a new 3D symmetry-based saliency operator, calculating a measure of context-free local symmetry from a 3D point cloud. The proposed symmetry operator is used to predict fixation points for further attention-driven segmentation or detailed exploration of the scene. We show that a 3D symmetry-based saliency operator reflects the notion of objectness better than the currently existing 2D symmetry-based saliency operator by Kootstra *et al.* [14] and the classical saliency operator by Itti *et al.* [15]. We extensively evaluated both methods on two databases. The first database consists of images showing table scene, and the second one of scenes of complete rooms. Both in quantity and quality of fixations, the proposed algorithm outperforms previous work.

The paper is structured as follows: In Section 2, we describe our proposed 3D symmetry-based saliency operator. Section 3 presents the evaluation results to demonstrate our approach. We compare our methods to Kootstra *et al.* [14] and Itti *et al.* [15]. Section 4 concludes the paper.

2 Method

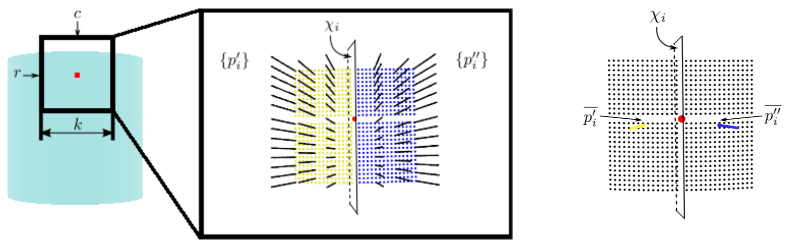
In this section, we describe the method for calculating local symmetries in 3D. The algorithm is based on detecting reflective symmetries using principal axes of Extended Gaussian Images built from patches' normals.

2.1 3D Symmetry Model

The 3D reflective symmetry is calculated from a depth image D (Fig. 1). Based on the depth image D, we create a point cloud P, so that $\forall p\,(r,c) \in P$:

$$p\,(r,c) = (x,y,z) \tag{1}$$

where (r,c) are row-column coordinates in the depth image, (x,y,z) are 3D point coordinates. For each point p a normal $\boldsymbol{n}_p$ is estimated. The normal to a point in



(a) (b)

Fig. 1. The depth image of a cylinder (artificial data) is shown on Figure 1(a), with the point $p(r, c)$ for which the symmetry is calculated (highlighted in red), and the kernel $\Phi(p)$ shown as a black square. The subset of points $\{p\} = \Phi(p) \bigcap P$ is shown in 3D on the right side. Subsets $\{p'_i\}$ and $\{p''_i\}$ are shown in yellow and blue respectively, with the reflective plane χ_i between the two point subsets. Normals n_p are shown as black lines. In Figure 1(b) examples of $\overline{p'_i}, \overline{\mathbf{n'_i}}$ and $\overline{p''_i}, \overline{\mathbf{n''_i}}$ are shown in yellow and blue respectively.

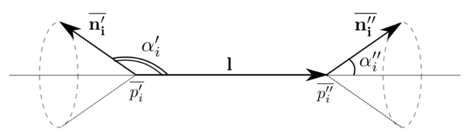


Fig. 2. Visual illustration for the calculation of angles α'_i and α''_i. $\mathbf{l}$ is the line connecting the two mean points $\overline{p'_i}$ and $\overline{p''_i}$. α' is the angle between mean normal $\overline{\mathbf{n'_i}}$ and $\mathbf{l}$, and α'' is the angle between mean normal $\overline{\mathbf{n''_i}}$ and $\mathbf{l}$.

the point cloud is estimated as the normal of a plane tangent to the neighboring surface [18].

$\Phi(p)$ defines the symmetry kernel as a squared patch, centered around p (Fig. 1) with side length k. The amount of 3D symmetry at the given location $p\,(r, c)$ is estimated on the subset of points $\{p\} = \Phi(p) \bigcap P$.

Sun *et al.* [17] proposed to use an Extended Gaussian Image built from point normals to detect symmetries of a model. Minovic *et al.* [19] proved that planes of reflective symmetries are perpendicular to the directions of the principal axes. Thus, to detect planes of reflective symmetries from the patch we build an Extended Gaussian Image from the patch's point normals, and calculate the principal axes $\gamma = \{\gamma_1, \gamma_2.\gamma_3\}$ of the Extended Gaussian Image using Principal Component Analysis (PCA). The corresponding symmetry reflective planes χ_i $(i = 1, 2, 3)$ are defined as planes going through the point $p\,(r, c)$ with the plane normal equal to the corresponding principal axis γ_i.

For a given reflective plane χ_i the point set $\{p\}$ is divided into two subsets $\{p_i'\}$ and $\{p_i''\}$, so that $\forall p \in \{p\}$:

$$p \in \begin{cases} \{p_i'\} & \text{if } d_H\left(p, \chi_i\right) > 0 \\ \{p_i''\} & \text{if } d_H\left(p, \chi_i\right) < 0 \end{cases} \tag{2}$$

where $d_H\left(p, \chi_i\right)$ is the signed Euclidean distance from point p to the plane χ_i (Fig. 1). The signed Euclidean distance is the distance from the plane according to the Hessian normal form.

The amount of 3D reflective symmetry is relative to a given plane χ_i for a given subset of points $\{p\}$ and defined as:

$$\Omega_i\left(\Phi\left(p\right)\right) = \exp\left(- \triangle D_i\right) \cdot \exp\left(- \triangle d_i\right) \cdot \omega_1 \cdot \omega_2 \tag{3}$$

The multiplication of all four components reflects the fact, that we are searching for patches that are symmetrical in all four aspects (see below).

$\triangle D_i$ represents the difference in depth values between mean points:

$$\triangle D_i = \left| D\left(\overline{p_i'}\right) - D\left(\overline{p_i''}\right) \right| \tag{4}$$

$$\overline{p_i'} = \frac{1}{N'} \sum_{p_j \in \left\{p_i'\right\}} p_j \tag{5}$$

$$\overline{p_i''} = \frac{1}{N''} \sum_{p_j \in \left\{p_i''\right\}} p_j \tag{6}$$

where N' and N'' are numbers of points in the subsets $\{p'\}$ and $\{p''\}$ respectively, and $\overline{p'}$ and $\overline{p''}$ are mean points of the respective subsets. $\triangle D_i$ reflects the fact, that we are only interested in symmetries, that are facing our view point.

$\triangle d_i$ represents the difference in distances from mean points $\overline{p'}$ and $\overline{p''}$ to the reflective plane χ_i:

$$\triangle d_i = \left| d\left(\overline{p_i'}, \chi_i\right) - d\left(\overline{p_i''}, \chi_i\right) \right| \tag{7}$$

where $d\left(p, \chi_i\right)$ is the unsigned Euclidean distance from the point p to the plane χ_i. $\triangle d_i$ reflects the fact that we are not only searching for patches with symmetrical orientations, but also for patches that can be divided into two subpatches, which are equally sized and symmetrically positioned in 3D space.

ω_1 is a coefficient measuring the co-planarity between the line $\mathbf{l}$ connecting $\overline{p_i'}$ and $\overline{p_i''}$ and the two mean normals $\overline{\mathbf{n_i'}}$ and $\overline{\mathbf{n_i''}}$ (Fig. 2):

$$\omega_1 = |[\overline{\mathbf{n_i'}} \times \overline{\mathbf{n_i''}}] \times \mathbf{l}| \tag{8}$$

$$\mathbf{l} = \frac{\overline{p_i'} - \overline{p_i''}}{||\overline{p_i'} - \overline{p_i''}||} \tag{9}$$

$$\overline{\mathbf{n_i'}} = \frac{\sum_{p_i \in \{p_i'\}} \mathbf{n_{p_i}}}{|| \sum_{p_i \in \{p_i'\}} \mathbf{n_{p_i}} ||} \tag{10}$$

$$\overline{\mathbf{n_i''}} = \frac{\sum_{p_i \in \{p_i''\}} \mathbf{n_{p_i}}}{|| \sum_{p_i \in \{p_i''\}} \mathbf{n_{p_i}} ||} \tag{11}$$

where $\overline{p_i'}$ and $\overline{p_i''}$ are mean points of the subsets $\{p_i'\}$ and $\{p_i''\}$ respectively, and $\overline{\mathbf{n}'}$ and $\overline{\mathbf{n}''}$ are mean normals.

ω_2 shows the similarity between mean normal directions based on the symmetry operator from Reisfeld *et al.* [9] and is calculated as following:

$$\omega_2 = (1 - \cos(\alpha' + \alpha'')) \cdot (1 - \cos(\alpha' - \alpha'')) \tag{12}$$

where α' is the angle between mean normal $\overline{\mathbf{n_i'}}$ and $\mathbf{l}$, and α'' is the angle between mean normal $\overline{\mathbf{n_i''}}$ and $\mathbf{l}$. Basically this operator gives the largest value to regions, where normals are oriented completely opposite and the smallest value to regions, where normals have the same orientation (i.e. flat surfaces).

Ideally the factors $\triangle D_i$, ω_2 and ω_2 should be calculated on each pair of opposite points and then summed up after multiplication. Due to small errors in the calculation of normals this approach is not very robust. Moreover, it is computationally expensive. Using only the mean points and normals to represent subpatches is a common approximation which proved to be accurate enough for our computations.

The amount of 3D symmetry $s(x, y)$ at a given pixel $p(r, c)$ is equal to:

$$s(r,c) = \begin{cases} 0 & \text{if } D(r,c) = 0 \\ \max_{i=1,2,3}\{\Omega_i(\Phi(p(r,c)))\} & \text{if } D(r,c) > 0 \end{cases} \tag{13}$$

where $D(r, c) = 0$ means that no depth information is available at this point.

Due to the nature of the 3D symmetry operator convex and concave regions will obtain the same symmetry values. While in everyday scenarios the majority of objects, that are claimed to be symmetric by humans, are rarely concave. To eliminate concave regions the following equation is applied:

$$s(r,c) = \begin{cases} 0 & \text{if } (\alpha' > \pi/2 \text{ and } \alpha'' < \pi/2) \text{ or } (\alpha' < \pi/2 \text{ and } \alpha'' > \pi/2) \\ s(r,c) & \text{otherwise} \end{cases} \tag{14}$$

2.2 Multi-Scale Symmetry-Based Saliency Map

To calculate a multi-scale symmetry-based saliency map a Gaussian pyramid of depth images is created. For each depth map the respective point cloud is calculated. 3D based symmetry maps s_l are calculated on every scale l of the pyramid. This results in a pyramid of symmetry maps. A master saliency map S is obtained by across scale addition [15] of the symmetry pyramid:

$$S(r,c) = \bigoplus_{l=L_1}^{L_2} s_l(r,c) \tag{15}$$

where L_1 is the finest scale and L_2 is the coarsest scale. The calculation on different scales allows to detect symmetries of different sizes in a computationally effective manner.

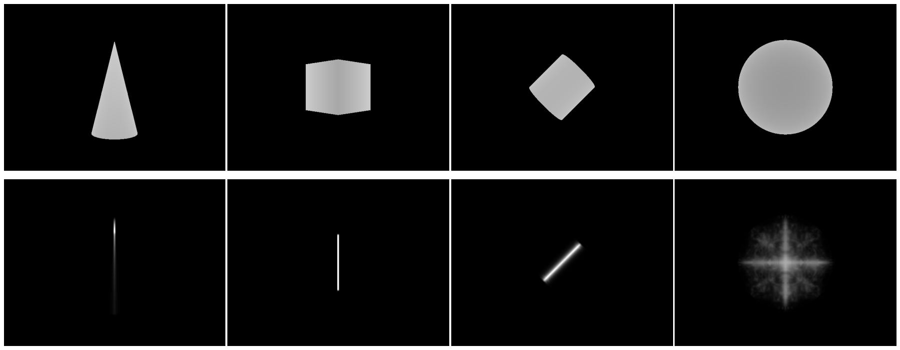


Fig. 3. Examples of symmetry maps calculated on artificial data. In the first row artificially created depth images of a cone, a rotated cube, a rotated cylinder and a sphere are shown in columns (a), (b), (c), (d) respectively. The second row shows the corresponding 3D symmetry-based maps calculated using only one scale $l = 0$ and kernel size $k = 30$.

2.3 Attention Points

The multi-scale symmetry-based saliency map S is used as input to the Winner-Take-All (WTA) neural network [20] to calculate attention points. Attention points can be used as seed points for attention-driven segmentation [21] or as fixations for further investigation of the region like zooming in or foveation. This approach for scene investigation is highly useful in such applications as robot navigation, robot localization and object detection. It can significantly reduce the search space and automatically point to interesting objects or areas, without exhaustive and computationally expensive exploration of the whole scene.

3 Evaluation

The quality of symmetry-based saliency maps was evaluated on artificial data, as well as on real data. Usually saliency operators are evaluated by comparing saliency results with eye-tracking data. This approach is useful when the task on hand is to build a system that tries to explain the human visual attention system. Since our task is to build an attention system that can be useful in robotic tasks, we chose a different approach for evaluation.

Our evaluation consists of several parts. At first, we evaluated the quality of attention points by detecting how many different objects were covered by the attention system with a given number of fixations, the so-called Hit Ratio (HR). Secondly, because our algorithm depends on the size of the kernel, we evaluate the variance of the fixation results under different kernel sizes.

Our experiments were done against the 2D symmetry saliency operator proposed by Kootstra *et al.* [14] and against the orientation-contrast saliency operator proposed by Itti *et al.* [15].

(a)
(b)
(c)
(d)

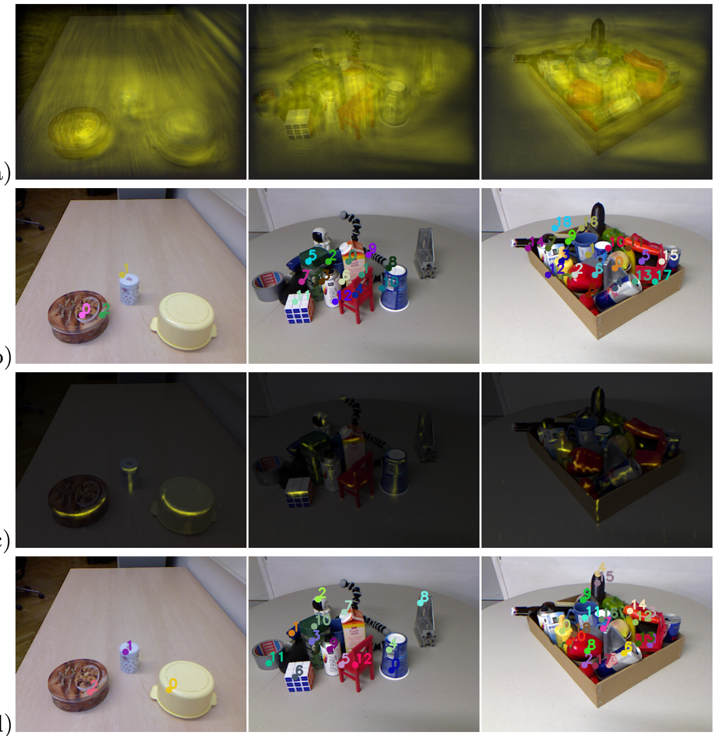


Fig. 4. Examples of images from the Table Objects Scene Database (TOSD) from Vienna University of Technology. Row (a) shows examples of 2D symmetry-based maps [14] overlaid with original images. Respective attention points calculated from 2D symmetry-based maps using WTA are shown in row (b). 2D symmetry-based maps were calculated on scales $l = 1..5$ using an external kernel $k_1 = 11$ and an internal kernel $k_2 = 5$ (for detailed parameters explanation see [14]). Row (c) shows 3D symmetry-based maps calculated using the proposed method and overlaid with original images. Respective attention points are shown in row (d). 3D symmetry-based maps were calculated on scales $l = 0..4$ using kernel $k = 15$.

3.1 Evaluation on Artificial Data

To prove that our 3D symmetry operator is performing as expected, we have tested it on artificially created data. Artificially created data was produced from rendering mathematical models of different objects with known shape (i.e. cylinders, cubes, spheres, cones). Results of symmetry operators are shown in Fig. 3.

From the presented results it is clearly visible that the proposed method works perfectly for synthetic examples. However, the result for the sphere (Fig. 3, (d)) visually does not look perfect, due to artifacts of the visualization process. Symmetry values for the sphere are quite small (note that, as explained in Section 2.1, surface patches, that are rather flat locally, result in small values of ω_2). For visualization in Fig. 3 values were normalized to a visible range, which

(a)
(b)
(c)
(d)

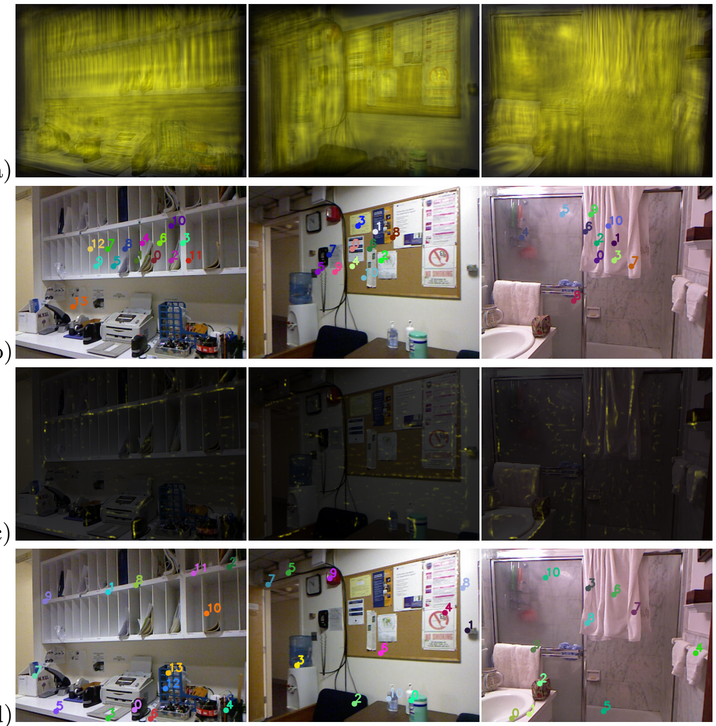


Fig. 5. Examples of images from the New York University Depth Database (NYUDD). Row (a) shows 2D symmetry-based maps [14] overlaid with original images. Respective attention points calculated from 2D symmetry-based maps using WTA are shown in row (b). 2D symmetry-based maps were calculated on scales $l = 1..5$ using an external kernel $k_1 = 11$ and an internal kernel $k_2 = 5$ (for detailed parameters explanation see [14]). Row (c) shows 3D symmetry-based maps calculated using the proposed method and overlaid with original images. Respective attention points are shown in row (d). 3D symmetry-based maps were calculated on scales $l = 0..4$ using kernel $k = 15$.

in this case led to an amplification of small errors from the normal calculation step.

3.2 Evaluation on Real Data

Symmetry-based saliency maps were evaluated on two RGB-D databases. The first database is the Table Object Scene Database (TOSD) from Vienna University of Technology[2]. The database consists of 244 different table scenes including free-standing objects, multiple occluded objects and piles of objects. All objects in the TOSD were hand-labeled with outlining polygons. Examples of images from the database and respective saliency maps are shown in Fig. 4.

[2] https://repo.acin.tuwien.ac.at/tmp/permanent/TOSD.zip

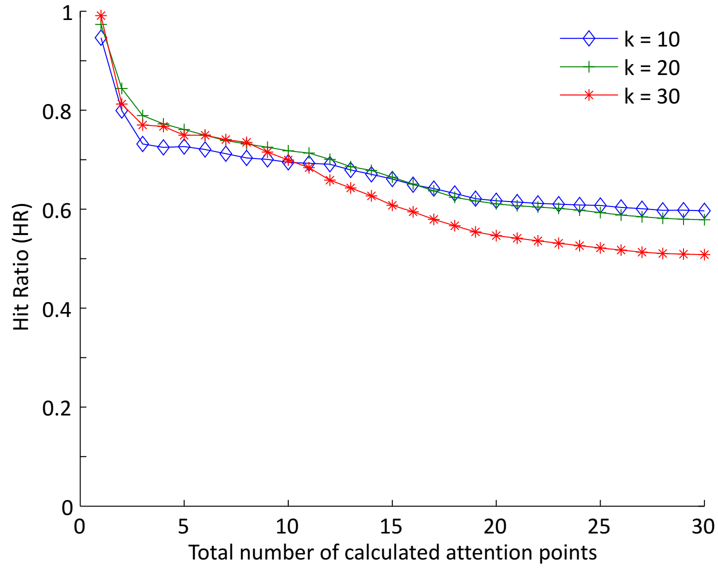


Fig. 6. Hit Ratio (HR) against the total number of calculated attention points for kernel sizes $k = 10$, $k = 20$, $k = 30$. As can be seen from the plot the kernel size does not have a big influence on the performance of the proposed algorithm when the total number of calculated attention points is smaller than 30.

The second database on which we evaluated our results was the New York University Depth Dataset (NYUDD)[3] which consists of more than 1000 densely labeled images of more that 400 different indoor scenes (Fig. 5).

From symmetry-based saliency maps attention points were calculated using the Winner-Take-All neural network. Attention points are evaluated with respect to the Hit Ratio (HR).

The Hit Ratio (HR) shows the percentage of unique attention points being situated inside different objects:

$$HR = \frac{n}{N} \tag{16}$$

where N is the total number of calculated attention points and n is the number of different attended objects. A perfect attention mechanism will hit every object exactly once, resulting in a HR equal to one.

3.3 Choosing the Size of the Kernel

Our algorithm depends on exactly one settable parameter - the kernel size k. We have evaluated the performance of the proposed method with different kernel

[3] http://cs.nyu.edu/ silberman/datasets/nyu_depth_v2.html

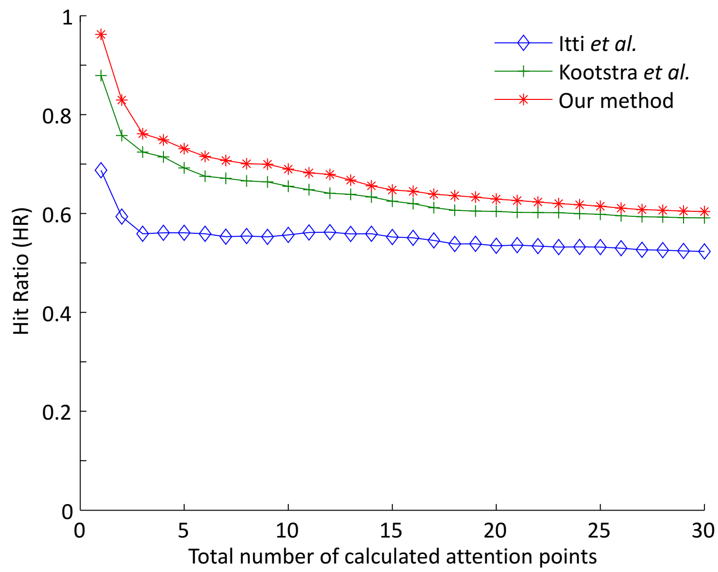


Fig. 7. Comparison plot of the Hit Ratio (HR) against the total number of calculated attention points for the Table Objects Scene Database (TOSD) for different types of saliency maps: the orientation-contrast model [15], the 2D symmetry-based model [14] and our proposed 3D symmetry-based model.

sizes ($k = 10$, $k = 20$ and $k = 30$) on the TOSD. Results are shown on Fig. 6. As can be seen from the plot the size of the kernel does not influence the performance much, when the number of calculated attention points is smaller than 10. This result is expected, because symmetry maps are calculated on different scales. This allows to detect symmetrical objects of different sizes regardless of the kernel size. However, with the kernel size $k = 30$ the performance drops significantly after around 10 attention points. This is easily explained, due to the fact that typical sizes of objects in the TOSD are smaller than 50 pixels. It means that after all relatively big objects were detected in cluttered scenes, smaller objects were ignored. These results suggests that smaller kernels are to be preferred (a value of 15 was chosen in subsequent experiments). And the detection of larger objects can then be done on coarser scales.

3.4 Evaluation on the databases

Fig. 7 and Fig. 8 show comparison plots of the HR against the total number of calculated attention points for the TOSD and for the NYUDD respectively. As can be seen from the plot the use of 3D symmetry-based saliency maps improves the quality of attention points up to 10% starting from the first fixation. An interesting observation can be made from the plots. An improvement for the NYUDD is more noticeable than for the TOSD. While for the TOSD the average

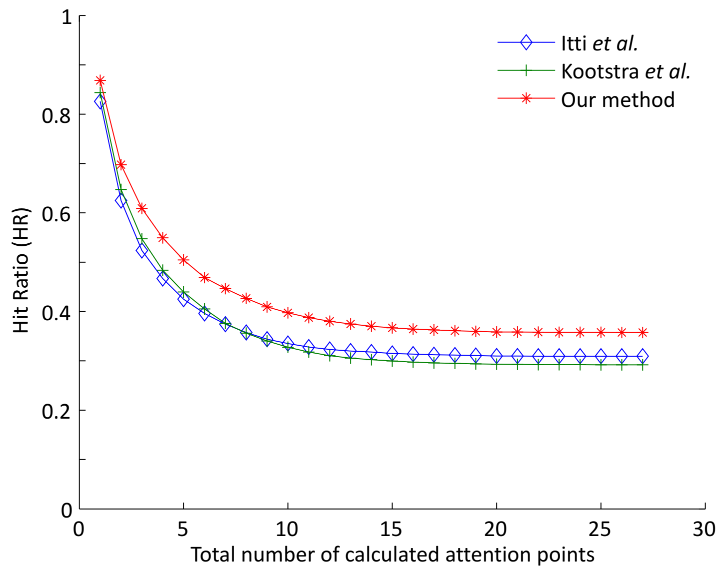


Fig. 8. Comparison plot of the Hit Ratio (HR) against the total number of calculated attention points for the New York University Depth Dataset (NYUDD) for different types of saliency maps: the orientation-contrast model [15], the 2D based symmetry map [14] and our proposed method.

HR is higher. Explanations to these effects lie in the types of scenes. The TOSD consists of crowded table scenes, while the NYUDD presents room scenes. The probability to hit any object in the TOSD just by selecting a random point is much higher than in the NYUDD. It also means that a perfect saliency operator for TOSD should take much more complicated information into account, than only early vision processes, e.g. early segmentation.

4 Conclusion

In the presented paper, we discussed a new algorithm for calculating a 3D symmetry-based saliency map. The proposed algorithm is based on finding local reflective symmetries using Extended Gaussian Images and normal direction dissimilarities. From these saliency maps attention points using a Winner-Take-All neural network were extracted. The quality of attention points was evaluated on two databases (the TOSD and the NYUDD) against the Hit Ratio. Result were compared to two saliency methods: the orientation-contrast method [15] and the 2D symmetry-based method [14]. We showed that the proposed algorithm works better in terms of Hit Ratio (HR). Future work will concentrate on finding a way to combine both 2D and 3D symmetry models to gain better results on a variety of different types of scenes.

References

1. Treisman, A.M., Gelade, G.: A feature-integration theory of attention. Cognitive Psychology **12** (1980) 97–136
2. Koch, C., Ullman, S.: Shifts in selective visual attention: Towards the underlying neural circuitry. Human Neurobiology **4** (1985) 219–227
3. Wolfe, J.M., Cave, K.R., Franzel, S.L.: Guided search: An alternative to the feature integration model for visual search. Journal of Experimental Psychology: Human Perception & Performance **15** (1989) 419–433
4. Tsotsos, J.: Analyzing vision at the complexity level. Behavioral and Brain Sciences **13** (1990) 423–469
5. Aloimonos, J., Weiss, I., Bandyopadhyay, A.: Active vision. International Journal of Computer Vision **1** (1988) 333–356
6. Einhauser, W., Spain, M., Perona, P.: Objects predict fixations better than early saliency. Journal of Vision **8** (2008) 1–26
7. Scholl, B.J.: Objects and attention: the state of the art. Cognition **80** (2001) 1–46
8. Kootstra, G., Nederveen, A., Boer, B.d.: Paying attention to symmetry. In: Proc. of the British Machine Vision Conference, BMVA Press (2008) 1115–1125
9. Reisfeld, D., Wolfson, H., Yeshurun, Y.: Context free attentional operators: the generalized symmetry transform. International Journal of Computer Vision **14** (1995) 119–130
10. Heidemann, G.: Focus-of-attention from local color symmetries. IEEE Trans. on Pattern Analysis and Machine Intelligence (2004)
11. Loy, G., Zelinsky, A.: Fast radial symmetry for detecting points of interest. IEEE Trans. on Pattern Analysis and Machine Intelligence **25** (2003) 959–973
12. Chertok, M., Keller, Y.: Spectral symmetry analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence **32** (2010) 1227–1238
13. Berner, A., Wand, M., Mitra, N.J., Mewes, D., Seidel, H.P.: Shape analysis with subspace symmetries. Computer Graphics Forum **30** (2011) 277–286
14. Kootstra, G., Bergstroem, N., Kragic, D.: Using symmetry to select fixation points for segmentation. In: Proc. of 20th International Conference on Pattern Recognition. (2010) 3894–3897
15. Itti, L., Koch, C., Niebur, E.: A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. on Pattern Analysis and Machine Intelligence **20** (1998) 1254–1259
16. Mitra, N.J., Pauly, M., Wand, M., Ceylan, D.: Symmetry in 3d geometry: Extraction and applications. In: Eurographics State-of-the-art Report. (2012)
17. Sun, C., Sherrah, J.: 3d symmetry detection using the extended gaussian image. IEEE Trans. on Pattern Analysis and Machine Intelligence **19** (1997) 164–168
18. Rusu, R.B.: Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany (2009)
19. Minovic, P., Ishikawa, S., Kato, K.: Symmetry identification of a 3-d object represented by octree. IEEE Trans. on Pattern Analysis and Machine Intelligence **15** (1993) 507–514
20. Lee, D.K., Itti, L., Koch, C., Braun, J.: Attention activates winner-take-all competition among visual filters. Nature neuroscience **2** (1999) 375–381
21. Mishra, A.K., Aloimonos, Y.: Visual segmentation of simple objects for robots. In: Robotics: Science and Systems. (2011)