Fast and Automatic Detection and Segmentation of Unknown Objects

Gert Kootstra, Niklas Bergström and Danica Kragic

Abstract— **This paper focuses on the fast and automatic detection and segmentation of unknown objects in unknown environments. Many existing object detection and segmentation methods assume prior knowledge about the object or human interference. However, an autonomous system operating in the real world will often be confronted with previously unseen objects. To solve this problem, we propose a segmentation approach named Automatic Detection And Segmentation (ADAS). For the detection of objects, we use symmetry, one of the Gestalt principles for figure-ground segregation to detect salient objects in a scene. From the initial seed, the object is segmented by iteratively applying graph cuts. We base the segmentation on both 2D and 3D cues: *color*, *depth*, and *plane* information. Instead of using a standard grid-based representation of the image, we use *super pixels*. Besides being a more natural representation, the use of super pixels greatly improves the processing time of the graph cuts, and provides more noise-robust color and depth information. The results show that both the object-detection as well as the object-segmentation method are successful and outperform existing methods.**

I. INTRODUCTION

As humans, robots need the ability to autonomously learn and explore new environments by exploring and manipulating unknown objects. The difficulty with unknown objects in unknown environments is that no prior knowledge of the object is available. Top-down search and segmentation methods can therefore not be used. We focus on the bottom-up detection and segmentation of objects. Real-time processing is furthermore of great importance for a robot interacting with the environment. In this paper, we present a method for fast and automatic detection and segmentation of objects. This is in contrast to methods not developed for a robotic application, with run times in the order of minutes [1].

Fig. 1 shows the humanoid head used in our lab along with two images from the left foveal and wide-field cameras. In previous work we dealt with object detection in the wide-field views [2], while here we present object detection and segmentation using the foveal cameras.

Many existing segmentation methods, for instance methods based on graph cuts [1], [3], [4], and our current work using 3D information and belief propagation [5], presume a selection of foreground or background parts by a human operator from where to start the segmentation. Other methods use prior knowledge about the object to perform a top-down segmentation, e.g., [6], [7]. A robot dealing with unknown objects, however, cannot rely on either human input, or top-down information. Instead, we propose a bottom-up detection and segmentation method using a saliency method based on local symmetry in the image to initiate the segmentation.

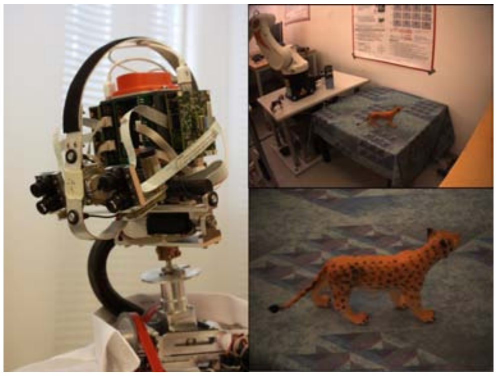

Fig. 1. A close up of our humanoid head. The upper lenses capture the wide-field view (upper right image) and the lower lenses the foveal view (lower right image)

Many saliency method calculate saliency by center-surround contrasts of *basic features*, like the well-known model of Itti *et al.* [8] that utilizes brightness, color, and orientation contrasts. We will refer to this model henceforth as the *contrast-saliency model*. This model has been used, for instance, in [2] to guide the attention of the robot for the sake of segmentation. 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]. Human eye fixations can also be predicted based on symmetry in the image [10]. Contrast focuses near

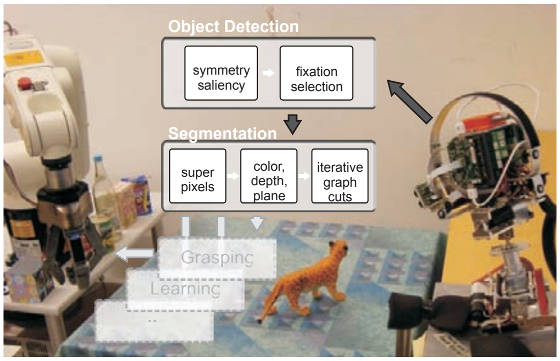

Fig. 2. To let our robot deal with unknown objects in its environment, we propose a bottom-up method for the fast and automatic detection and segmentation of objects. The robot perceives the scene with its stereo cameras. Objects are detected based on symmetry, and fixation points are generated to initialize segmentation. Based on an oversegmentation of the image into super pixels, an iterative graph-cut procedure segregates foreground from the background using color, depth, and plane information.

All authors are with the Centre for Autonomous Systems (CAS) and the Computer Vision and Active Perception Laboratory (CVAP), KTH, Stockholm. Contact: kootstra@kth.se

corners and edges in the image, whereas symmetry highlights symmetrical configurations which often coincide with objects in the scene. The ability of symmetry to detect objects is also supported by the Gestalt theory, which assumes symmetry to be one of the principles for figure-ground segregation. In [11], symmetry has shown to outperform the contrast-saliency model in detecting salient objects in the scene.

Once a hypothesis of an object is made in the detection step, it should be segmented from the background. Most current segmentation methods are based on Markov Random Fields and use graph cuts to efficiently find the segmentation of the image that minimizes a particular energy function [3]. The computational complexity of the segmentation process heavily depends on the number of points in the graph. To obtain real-time segmentation, we propose a representation of the image by super pixels, thereby greatly reducing the number of points, while preserving the important boundaries in the image. Moreover the use of super pixels provides more noise-robust color and depth information and enables us to estimate local plane information. Super pixels has previously been used for segmentation using graph cuts [12], belief propagation [6], and the detection of learned object information [7]. In [13], super pixels are clustered into image segments using texture and brightness similarities and the smoothness of the contours.

The utilization of disparity information apart from color information has been shown to be beneficial in robotic vision [1], [5], [12]. The advantage of using disparities is that it enables to separate color differences within the object from color difference between the object and the background. This makes the segmentation of both homogeneously and heterogeneously colored objects possible. A problem however arises at object points close to its supporting surface. The disparities of the object and the surface are similar at these points. We propose to solve this problem by making additional use of local and global plane information. Based on differences in orientation and depth of the local planes, the object parts and the surface parts can be separated.

The contribution of this paper is a fast and automatic method used in a robotic scenario for 1) detection of unknown objects in realistic environments using local symmetry, and 2) real-time segmentation of the object by iteratively applying graph cuts based on a super-pixel representation of the image using color, depth, and plane information. The results show that our object-detection method is successful and outperforms the saliency model of Itti, Koch, and Niebur [8]. Furthermore, our object-segmentation method outperforms the method proposed in [1] in terms of the quality of segmentation and above all in terms of computation time.

II. Automatic Detection And Segmentation

Our Automatic Detection And Segmentation (ADAS) method consists of two parts, the object-detection method and the segmentation method. The object-detection method selects a number of most salient point in the image as a fixation points, f_i, using the Gestalt principle of Symmetry. Based on a fixation point, the segmentation method iteratively segments the figure from its background using color, binocular disparity, and plane information. By utilizing a super-pixel representation of the image we achieve real-time performance while reducing the influence of noise. The object-detection method that we use has been presented in [10] and will therefore not be further detailed. The next paragraphs describs the object-segmentation method. We first discuss the image information. Next, the different processing steps of the method are outlined and discussed in detail.

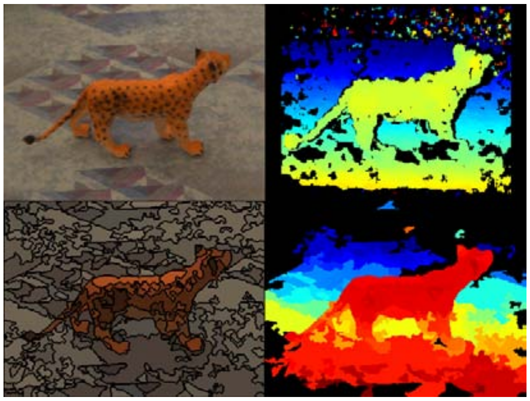

Fig. 3. An example of results of the super-pixel segmentation. The first row shows the original image on the left and disparity map on the right. The second row shows the super-pixel image and the super-pixel disparity map. The number of pixels is reduced from 307200 regular pixels to 202 super pixels. It can also be appreciated that the noise in the original disparity map is reduced

Image information: The object segmentation is based on *color*, *distance*, and *plane* information. The color distribution of every super pixel s is represented as a two-dimensional histogram, H_s, in ab-space, where a and b are the corresponding dimensions in the CIE Lab color space. This color space is perceptually uniform, which is preferable when comparing colors of different super pixels.

The distance information of every super pixel is represented by $\{\mu_s^d, \sigma_s^d\}$, where μ_s^d is the mean and σ_s^d is the standard deviation of the binocular disparities, $\delta_i \in [0, 1]$, of all points in the super pixel. For many points in the images the disparity cannot be calculated due to insufficient texture, or due to the fact that the point is occluded in one of the images (see Fig. 3).

Finally, we assume that the objects are placed on a planar surface. To detect this surface, we estimate the dominant

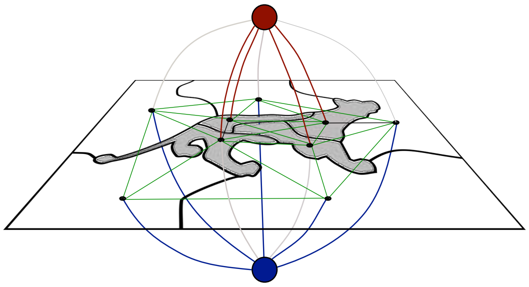

Fig. 4. An example of the graphical model used fro segmentation. Initially, all super pixels are connected to the source (red) and sink (blue). In the final segmentation, the points are either connected to the source or to the sink (blue and red lines). The connections between neighboring super pixels are shown with green lines.

plane based on the disparities in the total image. To test if the super pixels are part of the dominant plane, we also estimate the plane of every super pixel based on its disparity information. A plane is defined as $d = \alpha x + \beta y + \gamma$. Using the method that we proposed in [5], we efficiently estimate the plane parameters for the dominant plane, $\{\alpha_{\mathrm{D}}, \beta_{\mathrm{D}}, \gamma_{\mathrm{D}}\}$ and for every individual super pixel, $\{\alpha_s, \beta_s, \gamma_s\}$.

A. Processing steps

We use a graphical model to solve the figure-ground segmentation. The layout of the graph is based on a super-pixel representation of the image. A graph-cut method is used to minimize the energy in the model [14], [3]. The segmentation method consists of the following steps:

1) Initial segmentation of the image into super pixels.
2) Initial segmentation of fore- and background.
3) Update of fore- and background information.
4) Repeat until convergence of the figure-ground labeling:
    a) Graph-cut segmentation.
    b) Update of fore- and background information.
5) All super pixels connected to the super pixel fixated on are foreground.

The different steps are detailed in the following subsections.

B. Super pixels (step 1)

To achieve real-time performance, we segment the camera image into super pixels by clustering regions in the image that are more or less homogeneous in color. To deal with noise in the image, we use an image-pyramid based segmentation method that makes an initial segmentation on a low-resolution version of the image that is then successively refined on higher resolutions [15]. We use the method as implemented in OpenCV. Since we are not interested in too small super pixels, we merge all super pixels that contain less than 50 pixels with the neighboring super pixel that is closest in color.

This super-pixel segmentation method typically transforms an image of $640 \times 480 = 307200$ pixels to $50 - 200$ super pixels. This yields a vast improvement of the computational complexity of the Markov Random Field. Another advantage of the super-pixel representation is that the image information can be collected over the whole super pixel, making our method less susceptible to noise. Fig. 3 shows an example.

C. Initial segmentation (step 2)

Based on the fixation point from the object-detection method, the initial set of foreground pixels becomes $\mathcal{F} = \{f\}$, where f is the index of the super pixel containing the fixation point. The set of background pixels then becomes $\mathcal{B} = \mathcal{S} - \mathcal{F}$, where $\mathcal{S}$ is the total set of super pixels.

D. Updating fore-/background information (step 3)

Based on this initial figure-ground segmentation, the information of the foreground, $\mathcal{I}_{\mathrm{F}}$, and background, $\mathcal{I}_{\mathrm{B}}$, is set. The foreground information consists of the color histogram of the foreground and the disparity information of the foreground: $\mathcal{I}_{\mathrm{F}} = \{C_{\mathrm{F}}, \mathcal{D}_{\mathrm{F}}\}$. The background information consists of the color histogram of the background and information about the dominant plane in the image: $\mathcal{I}_{\mathrm{B}} = \{C_{\mathrm{B}}, \mathcal{P}_{\mathrm{B}}\}$. The color histograms of respectively the fore- and background become:

$$C_{\mathrm{F}}(a,b) = \sum_{s \in \mathcal{F}} C_s(a,b), \quad C_{\mathrm{B}}(a,b) = \sum_{s \in \mathcal{B}} C_s(a,b) \qquad (1)$$

The mean and standard deviation of the disparities in the foreground, $\mathcal{D}_{\mathrm{F}} = \{\mu^d_{\mathrm{F}}, \sigma^d_{\mathrm{F}}\}$, become:

$$\mu^d_{\mathrm{F}} = \frac{1}{\sum_{s \in \mathcal{F}} |\mathcal{D}_s|} \sum_{s \in \mathcal{F}} \sum_{i \in \mathcal{D}_s} \delta_i \qquad (2)$$

$$\sigma^d_{\mathrm{F}} = \sqrt{\frac{\sum_{s \in \mathcal{F}} \sum_{i \in \mathcal{D}_s} (\delta_i - \mu^d_{\mathrm{F}})^2}{\sum_{s \in \mathcal{F}} |\mathcal{D}_s| - 1}} \qquad (3)$$

where $\mathcal{D}_s$ is the set of all valid disparity values $\delta_i \geq 0$ of points contained in super pixel s. The mean and variance can be efficiently computed from the already computed μ^d_s and σ^d_s values. Finally, the background information is completed by adding the parameters of the dominant plane in the image, which is likely to be the surface on which the object is placed: $\mathcal{P}_{\mathrm{B}} = \{\alpha_{\mathrm{D}}, \beta_{\mathrm{D}}, \gamma_{\mathrm{D}}\}$.

E. Graph-cut segmentation (step 4a)

The graph is setup based on the super-pixel representation of the image. All super pixels are connected to the source and the sink (t-links) and horizontal connection are made between all neighboring super pixels (n-links) (see Fig. 4). The goal of the graph cut is to find a labeling l that assigns a label $l_s \in \{0,1\}$ to every super pixel $s \in \mathcal{S}$ so that the energy function $E(l)$ is minimized. Adopting the notation used in [14], the energy function is defined as:

$$E(l) = \overbrace{\sum_{s \in \mathcal{S}} D_s(l_s)}^{\text{data}} + \overbrace{\alpha \sum_{\{s,t\} \in \mathcal{N}} V_{s,t}(l_s, l_t) \cdot T(l_s \neq l_t)}^{\text{smooth}} \qquad (4)$$

where $\mathcal{N}$ is the set of neighboring super pixels s and t. The smoothing term has the role to obtain a smooth segmentation. The term penalizes for cuts in the segmentation between two super pixels that are similar. Discontinuities in the image are preserved by using the indicator function $T(\cdot)$, which is 1 if the argument is true (i.e., if the labels are different) and 0 otherwise. The data term measures the disagreement between the labeling and the back- and foreground information. α is the relative influence of the smoothing term with respect to the data term. We used $\alpha = 0.1$ in our experiments.

The energy of the data term, $D_s(l_s)$, is defined as follows:

- if $s = f$, that is the fixated super pixel,

$$D_f(l_f) = \begin{cases} 1000 & \text{if } l_s = 0 \\ 0 & \text{if } l_s = 1 \end{cases} \qquad (5)$$

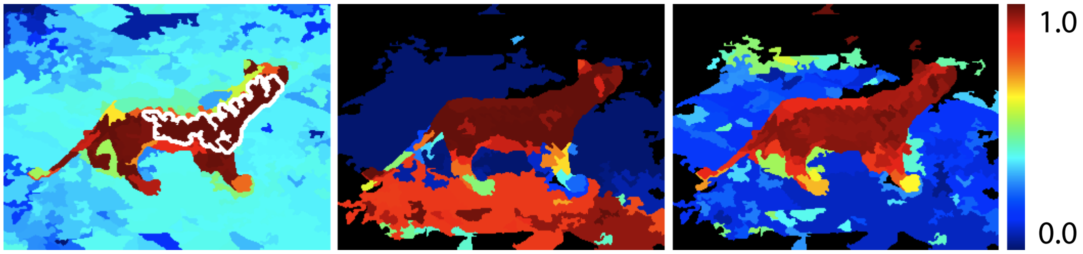

Fig. 5. Illustration of the probabilities for every super pixel to belong to the foreground. Left: $p^1_{C_\text{F}}$, based on color. Middle $p^1_{\mathcal{D}_\text{F}}$, based on disparity. Right: $1 - p^1_{\mathcal{P}_\text{B}}$, the probabilities of not belonging to the dominant plane. The super pixel fixated on is marked with a white boundary. Black super pixels have no valid disparity information. The image is best viewed in color.

- else, if μ^d_s is valid,

$$D_s(l_s) = \begin{cases} (\frac{1}{2}p^1_{C_\text{F}} + \frac{1}{2}p^1_{\mathcal{D}_\text{F}}) \cdot (1 - p^0_{\mathcal{P}_\text{B}}) & \text{if } l_s = 0 \\ \\ \frac{1}{2}(p^0_{C_\text{B}} \cdot (1 - p^1_{\mathcal{D}_\text{F}})) + \frac{1}{2}p^0_{\mathcal{P}_\text{B}} & \text{if } l_s = 1 \end{cases} \tag{6}$$

- else,

$$D_s(l_s) = \begin{cases} p^1_{C_\text{F}} & \text{if } l_s = 0 \\ p^0_{C_\text{B}} & \text{if } l_s = 1 \end{cases} \tag{7}$$

where $p^{l_s}_I$ can be read as the probability of the label of super pixel s being l_s given the information I. Note, however, that it is not a probability in the strict sense. The different $p^{l_s}_I$ values are defined as follow. For color:

$$p^1_{C_\text{F}} = \rho(C_s, C_\text{F}) \tag{8}$$

$$p^0_{C_\text{B}} = \rho(C_s, C_\text{B}) \tag{9}$$

where $\rho(C_1, C_2)$ is the correlation coefficient calculated by the Pearson's correlation between color histogram C_1 an C_2. Then, for disparity and plane information:

$$p^1_{\mathcal{D}_\text{F}} = e^{(\mu^d_s - \mu_\text{F})^2/(\sigma^d_\text{F})^2} \tag{10}$$

$$p^0_{\mathcal{P}_\text{B}} = \frac{1}{2}e^{-10\cdot\theta} + \frac{1}{2}e^{-10\cdot\Delta d} \tag{11}$$

where θ is the angle between the dominant plane in the image and the estimate plane of the super pixel. Δd is the distance between the centroid of the super pixel and the dominant plane. $1 - p^0_{\mathcal{P}_\text{B}}$ becomes high if both θ and Δd are low. This results in low costs in eq (6) for labeling a super pixel as background that has both the same orientation and disparity as the dominant plane, as it is probably part of the dominant plane. Super pixels that differ in either the orientation or in depth are more likely to be labeled as foreground. This helps differentiating parts of the object close to the surface from the background.

Eq. (5) results in a hard cut assuring that the super pixel fixated on ends up in the foreground. Eq. (6) causes super pixels to be labeled as foreground if they are similar to the foreground information in color *or* in disparity *and* are not part of the dominant plane. This results in the segmentation of homogeneously colored objects based on color as well as of heterogeneously colored objects based on the disparity, while discarding pixels belonging to the dominant plane in the image. If no valid disparity data is available for the super pixel, the energy in Eq. (7) is calculated using color only.

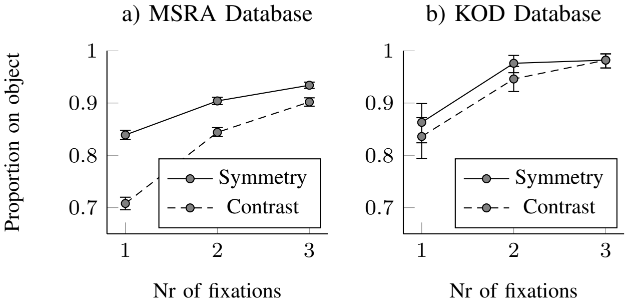

Fig. 6. The object-detection performance

Fig. 5 gives an illustration of the probabilities that super pixels belong to the foreground. It can be observed from the probability maps that the object cannot be completely segmented from the background based on color, disparity, or plane information alone. Based on color, parts of the object have low probability, whereas parts in the background have high probability. Disparity assigns high probabilities to large parts of the table plane close to the object. Based on plane information, parts of the background get labeled as foreground. However, when combining the three measures, the object can be nicely segmented from the background.

We define the smoothing term, $V_{s,t}(l_s, l_t)$, as:

$$V_{s,t}(l_s, l_t) = \begin{cases} \frac{1}{n}(\alpha_c p^{st}_C + \alpha_d p^{st}_\mathcal{D})p^{st}_\mathcal{P} & \text{if valid disp.} \\ \\ \frac{1}{n}\alpha_c p^{st}_C & \text{otherwise} \end{cases} \tag{12}$$

where $n = \min(|\mathcal{N}_s, \mathcal{N}_t|)$ is the minimal number of neighbors of either super pixel s and t, α_c and α_d are weights. In our experiments we obtain good results with $\alpha_c = 0.9$ and $\alpha_d = 0.1$.

$$p^{st}_C = \rho(C_s, C_t) \tag{13}$$

$$p^{st}_\mathcal{D} = 1 - |\mu^d_s - \mu^d_t| \tag{14}$$

$$p^{st}_\mathcal{P} = 1 - |p^{l_s=1}_{\mathcal{P}_B} - p^{l_t=1}_{\mathcal{P}_B}| \tag{15}$$

Equation (12) is applied only when the labels l_s and l_t are different, see Eq. (4). If the disparity information is valid, the costs to split two super pixels in the segmentation are high if the two super pixels are similar in color *or* in disparity *and* have a similar probability to be part of the dominant plane. This promotes the consistent labeling of super pixels with similar color or disparity that are either both part of the dominant plane or are both not part of that plane. If

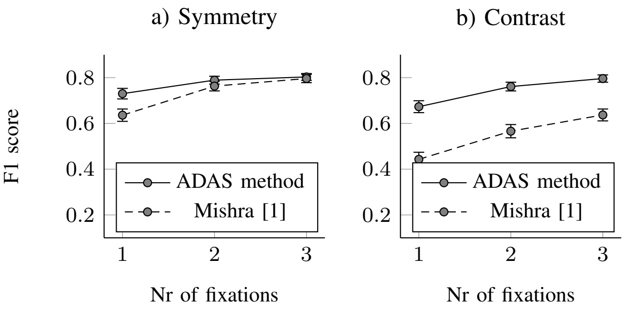

Fig. 7. The object-segmentation performance showing the best segmentation of any of the fixations selected by respectively symmetry and contrast.

valid disparity information is available, the costs are based on color only.

Our implementation is based on GC_optimization [14], [3], [16] using α/β swap to minimize the energy.

F. Updating fore-/background information (step 4b)

The sets of fore- and background super pixels are updated to $\mathcal{F}_{t+1} = \{s|l_s = 1\}$ and $\mathcal{B}_{t+1} = \{s|l_s = 0\}$. $\mathcal{I}_\mathrm{F}$, and $\mathcal{I}_\mathrm{B}$ are then updated according to eq. (1) and (3). Based on this updated information, a new graph-cut segmentation is performed until convergence, that is until $\mathcal{F}_t = \mathcal{F}_{t+1}$.

G. Connected foreground component (step 5)

The graph-cut segmentation can result in the selection of multiple objects in the foreground. We are interested in segmenting the fixated object only and therefore select all super pixels that are connected to the fixated super pixel as foreground. This gives the final segmentation.

The total algorithm is shown in Algorithm 1.

III. EXPERIMENTS AND RESULTS

The performance of our symmetry-based object-detection method is compared to the contrast-saliency model of Itti *et al.* [8]. The performance of the object-segmentation method is compared to the segmentation method of Mishra *et al.* [1]. We investigate the performance to detect the object in the scene, and the quality of the segmentation based on the selected fixation points.

The performance is tested on two datasets. The MSRA Salient Object Database [17] 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 [18] presented in [11] 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.

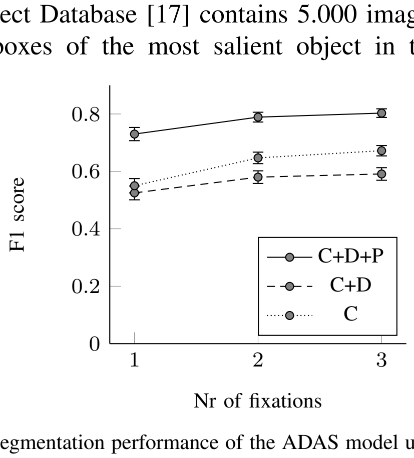

Fig. 8. The segmentation performance of the ADAS model using different types of information: color (C), disparity (D), and plane (P).

Algorithm 1 Automatic Detection And Segmentation.

```
$S \leftarrow$ getSymmetrySaliencyMap($I$)
$SP \leftarrow$ getSuperPixels($I$)
**for** $i = 1$ to nrFixations **do**
    $f_i \leftarrow$ getFixatedSuperPixel($S$, $SP$)
    $\mathcal{F}_0 \leftarrow \{f_i\}$, $\mathcal{B}_0 \leftarrow \mathcal{S} - \mathcal{F}_0$
    $\mathcal{I}_F \leftarrow$ getForegroundInfo($\mathcal{F}_0$, $SP$)
    $\mathcal{I}_B \leftarrow$ getBackgroundInfo($\mathcal{B}_0$, $SP$)
    **repeat**
        $l \leftarrow$ getGraphCutLabeling($SP$, $\mathcal{I}_F$, $\mathcal{I}_B$)
        $\mathcal{F}_{t+1} \leftarrow \{s|l_s = 1\}$, $\mathcal{B}_{t+1} \leftarrow \{s|l_s = 0\}$
        $\mathcal{I}_F \leftarrow$ getForegroundInfo($\mathcal{F}_{t+1}$, $SP$)
        $\mathcal{I}_B \leftarrow$ getBackgroundInfo($\mathcal{B}_{t+1}$, $SP$)
        $t \leftarrow t + 1$
    **until** $\mathcal{F}_{t-1} = \mathcal{F}_t$
**end for**
```

Object detection: In [11], we presented different experiments to investigate the object-detection performance of our symmetry model. To give a overview of the performance of the complete ADAS system, we present here one of the experiments. In Fig. 6, the proportion that any of the selected 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. 6(a) is highly significant. In Fig. 6(b), the performance for the symmetry model is also higher, but the difference is not significant and the performance is high for both. This 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. The implementation of the model on a GPU (Nvidia GTX 480) runs in 5-10 ms.

Segmentation: We test the performance of our object-segmentation method on the KOD Database. The quality of the segmentations is measured with the F1 score by comparing it to the ground-truth segmentation. The performance of our ADAS method is compared to the active segmentation method [1] using the same symmetry fixation points as initialization. Furthermore, both models are tested using fixation points selected by the contrast-saliency model.

Fig. 7 shows the F1 score of the best segmentation resulting from any of the fixations as a function of the number of fixations. The performance of our ADAS method is significantly higher for the first two symmetry fixations. The improvement of our method when contrast fixations are used is large. This can be explained by the fact that the contrast-saliency model often selects fixations close to the border of the object. Mishra *et al.*'s method is more

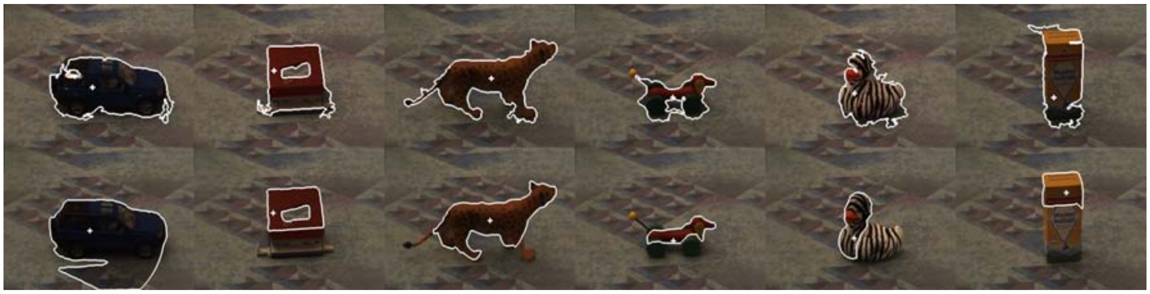

Fig. 9. Some segmentation examples. Fixation points are select using symmetry. The first row shows the segmentations resulting from our ADAS method. The second row shows the results of Mishra *et al.*'s segmentation method [1].

sensitive to fixation points close to the border than the proposed method. The figure furthermore shows that the use of the symmetry-saliency method for object detection outperforms the contrast-saliency method. See Fig. 9 for some segmentation examples of both methods.

The improvement in computation time of our ADAS model is enormous. Our object-segmentation method runs between 50-100 ms, whereas Mishra *et al.*'s method runs in the order of a couple of minutes. This is mainly due to the use of a complex edge-detection method, but also the recalibration of edge strengths using disparity information and the graph cuts on a regular pixel representation of the image slows down the process.

Fig. 8 shows the performance of the proposed method using different types of information. The use of color, disparity, and plane information clearly outperforms the method using color and disparity or color only. The performance when using color and disparity is lower than when color is used only. This can be explained by the fact that using disparity adds many super pixels on the table plane to the foreground. However, disparity also includes the heterogeneously colored parts of the objects. Adding the plane information to the model gets rid of the points on the table plane, while preserving all correctly labeled super pixels on the object.

IV. DISCUSSION AND CONCLUSION

We presented our Automatic Detection And Segmentation (ADAS) method. It is a fast and autonomous method to detect and segment objects previously unknown to the robot. The method consists of two parts. In the object-detection part, a salient object is detected using local symmetry in the image. Based on the selected fixation point, the object-segmentation part iteratively segments the object from its background using color, disparity, and plane information.

The results show that the our symmetry-saliency method outperforms the contrast-saliency method [8] in detecting salient objects. The object-segmentation method outperforms the active segmentation method [1] on the quality of the segmentation, and above all in computation time. The later method processes an image in a couple of minutes, wheras our ADAS method runs real-time, which is important for robotics. Using symmetry to select the fixation points furthermore results in better segmentation than using contrast.

The use of super pixels results in a great speed up, but comes with the disadvantage of early commitment. If super pixels cross object boundaries, it is impossible to recover. To minimize this problem, we chose conservative parameters.

REFERENCES

[1] A. Mishra, C. Fermuller, and Y. Aloimons, "Active segmentation for robotics," in *Proceedings of the Interation Conference on Intelligent Robots and Systems (IROS)*, 2009.

[2] B. Rasolzadeh, M. Björkman, K. Huebner, and D. Kragic, "An active vision system for detecting, fixating and manipulating objects in the real world," *The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 133–154, 2010.

[3] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.

[4] A. Blake, C. Rother, M. Brown, P. Perez, and P. Torr, "Interactive image segmentation using an adaptive gmmrf model," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2004.

[5] M. Björkman and D. Kragic, "Active 3d scene segmentation and detection of unknown objects," in *International Conference on Robotics and Automation (ICRA)*, 2010.

[6] J. Reynolds and K. Murphy, "Figure-ground segmentation using a hierarchical conditional random field," in *Proc. of the Fourth Canadian Conference on Computer and Robot Vision (CRV 2007)*, 2007.

[7] E. Borenstein and S. Ullman, "Combined top-down/bottom-up segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 12, pp. 2109–2125, 2008.

[8] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[9] J. R. Pomerantz, "Colour as a gestalt: Pop out with basic features and with conjunctions," *Visual Cogn.*, vol. 14, no. 4-8, pp. 619–628, 2006.

[10] G. Kootstra and L. R. B. Schomaker, "Prediction of human eye fixations using symmetry," in *Cognitive Science Conference (CogSci)*, Amsterdam, The Netherlands, 2009.

[11] G. Kootstra, N. Bergström, and D. Kragic, "Using symmetry to select fixation points for segmentation," in *Internation Converence on Pattern Recognition*, 2010.

[12] M. Björkman and J.-O. Eklundh, "Foveated figure-ground segmentation and its role in recognition," in *Proceedings of the British Machine Vision Conference (BMCV)*, 2005.

[13] X. Ren and J. Malik, "Learning a classification model for segmentation," in *Proceedings of the Internation Converence on Computer Vision (ICCV '03)*, vol. 1, Nice, 2003, pp. 10–17.

[14] Y. Boykov, O. Veksler, and R. Zabih, "Efficient approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 12, pp. 1222–1239, 2001.

[15] B. Jaehne, *Digital Image Proc.*, 3rd ed. Berlin: Springer-Verlag, 1995.

[16] V. Kolmogorov and R. Zabih, "What energy functions can be minimized via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[17] 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 Recogntion (CVPR '07)*, 2007.

[18] N. Bergström and G. Kootstra. http://www.csc.kth.se/˜kootstra/kod. [Online]. Available: http://www.csc.kth.se/ kootstra/kod