An In Depth View of Saliency

Arridhana Ciptadi
arridhana@gatech.edu
Tucker Hermans
thermans@cc.gatech.edu
James M. Rehg
rehg@gatech.edu

Center for Robotics and Intelligent Machines
School of Interactive Computing
Georgia Institute of Technology

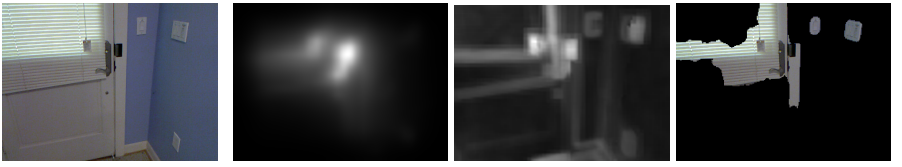
Abstract

Visual saliency is a computational process that identifies important locations and structure in the visual field. Most current methods for saliency rely on cues such as color and texture while ignoring depth information, which is known to be an important saliency cue in the human cognitive system. We propose a novel computational model of visual saliency which incorporates depth information. We compare our approach to several state of the art visual saliency methods and we introduce a method for saliency based segmentation of generic objects. We demonstrate that by explicitly constructing 3D layout and shape features from depth measurements, we can obtain better performance than methods which treat the depth map as just another image channel. Our method requires no learning and can operate on scenes for which the system has no previous knowledge. We conduct object segmentation experiments on a new dataset of registered RGB-D images captured on a mobile-manipulator robot.

1 Introduction and Motivation

Determining objects of interest in unknown environments is an important task for both humans and machines. Visual saliency methods are computational processes which guide the attention of an agent to potentially relevant locations in an image or scene. Standard approaches to saliency use color, gradient, and intensity differences to distinguish unique regions from the rest of the visual field. We propose a novel method which incorporates depth measurements into the computation of visual saliency. Human subject studies have shown that depth is an important cue in determining salient regions in human visual processing [16, 23]. Depth measurements make it possible to separate objects which may be similar in appearance. In addition, shape information can be recovered from the depth channel and used to improve the discriminability of scene elements.

One motivation for saliency research is the development of generic object segmentation and detection capabilities [6, 10, 11, 24]. As an example, consider a personal robot that can move through a home environment and manipulate objects. During a clean-up task, the robot should be able to handle unfamiliar objects for which it has no prior experience. Saliency provides a basic mechanism for characterizing an unfamiliar scene and generating hypotheses about potential object locations. We use the task of generic object segmentation to quantify the effectiveness of our depth-based saliency method. We present a novel segmentation approach that incorporates saliency in an MRF model defined over superpixels. We demonstrate improved performance over several previous approaches.



(a) Color Image (b) [7] Saliency Map (c) Our Saliency Map (d) Resulting Segmentation

Figure 1: An example saliency map and resultant segmentation produced from the color and depth image pair. Note the two light switches in the upper right of the image, which our approach highlights as salient regions.

Consider the images presented in Figure 1 captured on a robot operating in a home environment. The scene has multiple foreground objects of potential interest. Our saliency method scores all the relevant objects (door handle, light switches, window blind) more highly than the background elements of the scene. The resulting saliency-based segmentation successfully separates the foreground objects from the background producing a useful set of proto-objects for a robot to explore or manipulate.

We have collected a new dataset of color and depth images, which form the basis for our experimental evaluation. The dataset was collected using a mobile-manipulator robot in a real-world home environment. The dataset includes ground truth pixel-level segmentations of salient objects that we will release publicly. In summary, this work makes four contributions:

- We introduce a new method for estimating visual saliency, which combines color and depth measurements.
- We demonstrate that explicit 3D layout and shape features from depth measurements produce more informative saliency maps than approaches which simply treat depth as another channel of the image.
- We present an approach to saliency-based segmentation of generic objects based on a superpixel MRF and show promising results.
- We introduce a new dataset for depth-based saliency, including ground truth pixel-level segmentations of salient objects

The paper is organized as follows. Related work on visual saliency and generic object segmentation is reviewed in Section 2. Section 3 introduces our novel saliency method and our approach to saliency-based segmentation. We present qualitative and quantitative experimental results in Section 4. We conclude with a discussion of the work in Section 5.

2 Related Work

We review two bodies of related work: previous methods for visual saliency computation, including earlier attempts to incorporate depth, and previous works on generic object segmentation and detection using depth. We believe we are the first to present a bottom-up saliency framework which incorporates both appearance and 3D depth-based features, and which addresses the use of saliency for generic object segmentation.

2.1 Visual Saliency

Koch and Ullman gave the first comprehensive explanation of how saliency could be computed neurophysical [15]. This work was quite influential in the development of computational models of saliency in the computer vision community. In an early example of such

work Itti et al. compute a saliency map by combining conspicuity maps based on color, intensity, and orientation information, where locally-different patches are found through center-surround computations [12].

A recent review of computational approaches to visual saliency can be found in [2]. We briefly review several state-of-the-art methods. Goferman et al. [6] computes a visual saliency score based on the dissimilarity of an image patch to its surroundings. The dissimilarity of a pair of patches is determined by comparing their color histograms. Patches which are highly-dissimilar to their surrounding patches receive a high saliency score. This method demonstrates good performance and it serves as the basis for our depth-based saliency approach. Inspired by the findings of Theeuwes *et al*. [23], we extend the dissimilarity framework to model the joint interaction between layout and shape features obtained from the depth channel, and the color information.

Hou and Zhang present a saliency method based on spectral analysis [10]. The spectral residual is introduced as the frequency domain equivalent of the spatial saliency map. Thresholding the resulting saliency map produces a proto-object segmentation of the scene. More recently, Cheng et al. extend [6] to base the saliency computation on superpixels instead of simple image patches [4]. They also introduce an alternative object segmentation approach in which a coarse threshold-based segmentation is used to seed an iterative grab cut. In contrast, we present a segmentation method which integrates saliency measurements directly within an MRF model at the superpixel level.

2.2 Saliency Computation with Depth

While [15] mentions disparity as a potential cue in saliency computation, very little work has been done on integrating depth information into a saliency model. Maki et al. use depth in an attentional framework to help select the nearest out of a number of moving targets in a tracking application [17, 18]. This is a task-dependent use of depth information and not a bottom up integration of depth features in the computation of saliency. Ouerhani and Hügli extend the approach of Itti et al. from [12] by adding a conspicuity map built directly from the depth map [20]. This approach treats depth as just another channel, along with color and other cues. We demonstrate that better performance can be obtained by explicitly constructing 3D layout and shape features from the depth measurements.

Lang et al. present a depth prior for saliency learned from human gaze information [16]. This saliency prior produces a saliency map that is then either directly added or multiplied by the saliency results of other methods. A limit to this approach comes from decoupling the depth features from other saliency cues (i.e. color, intensity, gradient, etc) as complicated joint interactions are not modeled in the resulting saliency scores.

2.3 Object Segmentation Using Depth

There have been many previous works which exploit simple depth-based cues to segment generic objects for robot manipulation applications. A representative example is [21], which fits a plane to a tabletop and then uses connected components on the non-tabletop points to segment individual objects. These existing methods have two main limitations: they assume that all salient objects in the scene are associated with a planar support surface, and they require that the supporting plane be visible. Such methods have been shown to break down in highly cluttered scenes where not enough of the table surface is visible [22]. Depth-based saliency followed by object segmentation provides an alternative approach to these methods which can still make use of planar cues.

Much work has been done on performing object detection and recognition with depth information. Representative work can be found in [1, 8, 9]. Recently an object detection dataset of registered color and depth information was released [13]. This dataset provides only bounding box annotation for a fixed set of object classes. In contrast our dataset provides segmentation masks of all objects in the scene.

3 Method

In this section we explain our approach for incorporating depth into the computation of saliency. We explain our method for depth feature extraction and introduce a saliency model that combines depth and color features. We then present a novel method for saliency-based segmentation of generic objects.

3.1 Depth Model for Saliency

We base our approach to saliency on the computation of dissimilarity between a given pixel and all other locations in the image. This is done by constructing feature vectors at all image locations and designing a distance measure for comparing feature vectors. We then compute the saliency at a given pixel location from a ranking of the distances to all other image sites, based on the comparison of feature vectors.

It follows that we must perform two main tasks in order to incorporate depth into saliency: (1) design depth-based features which can be extracted at each image site, and (2) determine how to measure the dissimilarity between two feature vectors at multiple scales. A standard approach to these tasks is to treat depth as simply another image channel and use existing methods for feature extraction and distance measurement. Our hypothesis is that depth is a fundamentally-different source of information, and the correct exploitation of depth will necessarily change the structure of the saliency computation. Psychophysical experiments in [23] support this belief, which demonstrate that depth and color cues interact jointly, and in a complementary manner, to determine saliency.

We exploit two distinct, complementary cues from the depth channel which inform our saliency computation. First, we exploit the fact that depth encodes object shape. Depth images tend to be locally smooth at almost all points, except for the boundaries between objects. The smooth variation in depth corresponds to the shape of scene elements. We incorporate estimates of the surface normals at each image pixel into our feature vector in order to identify regions which are salient as a function of their shape or 3D orientation.

In addition to shape, the depth image also encodes information about the scene layout. In particular, we can use depth to organize the scene into planes. The findings of [23] suggest that humans use coplanarity to guide their assessment of saliency. We fit planes to the 3D points of the depth image, allowing us to associate each pixel with the dominant plane that contains it. In computing the feature distance between two image locations, we penalize points which lie on different depth planes. This has the effect of incorporating layout into our saliency computation. We describe this model in more detail in Section 3.2.

3.2 Computational Model

We represent each pixel by the patch surrounding it (we use 5×5 patches). Let p_i be the patch representation of pixel i. The saliency value of each pixel i depends upon the similarity between the colors and surface normals in p_i and those of every other patch extracted from the image. Let $d_{col}(p_i, p_j)$ and $d_{norm}(p_i, p_j)$ respectively be the Euclidean distance between vectorized patches of colors (in CIE $L*a*b$ space) and surface normals of p_i and p_j, normalized to the range $[0,1]$. Let $d_{pos}(p_i, p_j)$ be the 3D Euclidean distance between pixel i and

pixel j in the scene. Let $d_{plane}(p_i, p_j)$ be a binary measure of whether pixel i and j belong to the same plane. We define a dissimilarity measure between patches as:

$$d(p_i, p_j) = \frac{d_{col}(p_i, p_j) + \alpha \cdot d_{norm}(p_i, p_j) + \beta \cdot d_{plane}(p_i, p_j)}{1 + \gamma \cdot d_{pos}(p_i, p_j)} \tag{1}$$

Here α, β, and γ are weights which determine the influence of the surface normal, plane, and euclidean distances with respect to distance in color space. We subsequently use the distance defined by Equation 1 to compare pixel p_i with all other pixels in the image.

To suppress the influence of the many image patches which are distant in feature space, only the top K most similar patches determine the saliency value at each pixel. By limiting computation to include only the top K patches, one can interpret $d_{plane}(p_i, p_j)$ in equation 1 as prioritizing patches from the same plane for use in computing the saliency score. The saliency value of pixel i is defined as:

$$S_i = 1 - exp\{-\frac{1}{K}\sum_{k=1}^{K} d(p_i, p_k)\} \tag{2}$$

Patches are compared both with other patches at the same image scale as well as with patches extracted from the image at other scales (e.g. at 50% and 25% of the current scale). Comparison across multiple scales is helpful in suppressing the effects of large uniform regions, such as the image background, which are more likely to have similar patches at different scales. In addition, we compute a saliency map, S_i, at M different scales (we use 100%, 80% and 50%) and average the saliency score obtained from each different scale, giving a final saliency score for each pixel of $\bar{S}_i = \frac{1}{M}\sum_{m=1}^{M} S_i^m$.

Figure 2 illustrates the impact of the normal and plane features on the determination of saliency. In order to highlight these effects, we compute differences between the saliency maps produced by different feature combinations. These differences are visualized in Figure 3. In Figure 3(a) we see that the addition of the depth plane feature substantially increases the saliency of the light switch in the upper right corner of the image. This can be attributed to the fact that the light switch is located on a different plane from the other objects with similar color (the door and window blind). From Figure 3(b), we see that the use of surface normals increases the saliency score at many parts of the image, particularly around the door area. We attribute this to the greater variation in surface normal directions on the surface of the door area, including the door and the blinds, in comparison to the side wall. Figure 3(c) demonstrates that the combination of depth features significantly amplifies the saliency of the door and the light switch, which are the key elements in this scene.

(a) Original color image (b) Color Alone (c) Color & Planes (d) Color & Normals (e) Color, Normals, & Planes

Figure 2: Saliency results computed using different features.

3.3 Segmentation from Saliency

Given a saliency map we wish to determine a pixel level segmentation mask of the salient objects in the scene. This is commonly done by performing a simple thresholding operation on the saliency map [4, 6, 10, 11]. This simple procedure produces very rough boundaries

(a) (Color & planes) minus (color alone) (b) (Color & normals) minus (color alone) (c) (Color, normals, & planes) minus (color alone)

Figure 3: Saliency map differences for the plane and surface normal features.

and is difficult to tune from image to image. Instead, we wish to accumulate and propagate saliency information across similar pixels to give a more refined result. To this end we construct a Markov Random Field (MRF) model of the image connecting extracted superpixels. We wish to label each superpixel as either object or background. In order to do this we apply a unary potential for each superpixel which gives high cost to labeling a superpixel as background if its average saliency score is high. Similarly we penalize labeling superpixels with low average saliency as objects. We propagate information between neighboring superpixels by applying a smoothness term which penalizes regions of similar color distribution for having different labels. However, we allow the data term to dominate and do not force like labeled regions to have similar color distributions.

Our MRF connects neighboring superpixels and produces the following energy for which we wish to find the minimum cost labeling $E = \sum_{i \in I} g(l_i) + \sum_{i,j \in N} f(l_i, l_j)$. Here l_i describes the ith superpixel, $f(l_i, l_j)$ defines the smoothness term between neighboring superpixels, and $g(l_i)$ is the data potential which penalizes high saliency regions when given a background label. We define this data cost as

$$g(l_i) = \begin{cases} \bar{S}(l_i) \cdot |l_i| & \text{if } l_i \text{ has background label} \\ \max(\theta - \bar{S}(l_i), 0) \cdot |l_i| & \text{if } l_i \text{ has object label} \end{cases}$$

where $\bar{S}(l_i)$ is the average of all pixel saliency values in l_i, $|l_i|$ is the number of pixels in l_i, and θ is a threshold which scales how salient a region must be in order to receive no cost when assigned an object label.

We define the smoothness cost between superpixel i and j as

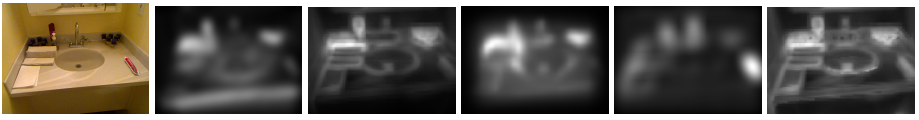
$$f(l_i, l_j) = \begin{cases} \bar{L} \cdot hist(l_i, l_j) & \text{if } label(l_i) \neq label(l_j) \\ 0 & \text{otherwise} \end{cases}$$

where $\bar{L}$ is the average superpixel size in the image and $hist(l_i, l_j)$ is the histogram intersection between the color histograms of l_i and l_j. Note that we normalize the color histograms to sum to one for each superpixel in order to control for different superpixel sizes. The superpixel size term is necessary to weight the smoothness term to be of similar magnitude to the data cost term.

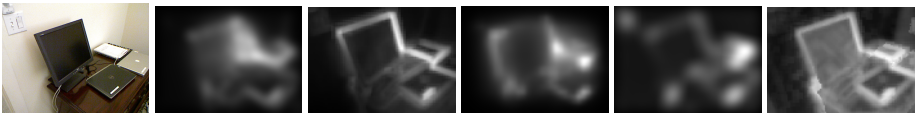
We extract superpixels from the color image using the method presented in [19]. We perform the optimization using a binary graph cut as described in [3]. Our experiments demonstrate the added benefit of our MRF approach over simple thresholding (cf. Section 4.2).

3.4 Surface Normal Computation and Plane Estimation

The depth images contain missing data due to limitations of the sensor. To overcome this issue we apply mode filters to the image as a smoothing operation. These filters produce values to fill in the missing data, while preserving the known depth values at all other pixel locations. Additionally, we crop the depth image to remove areas at the border where no depth



(a) Color image (b) [20] (c) [6] (d) [7] (e) [11] (f) Ours



(g) Color image (h) [20] (i) [6] (j) [7] (k) [11] (l) Ours

Figure 4: Results comparing various saliency methods.

data is present. We compute surface normals at all pixel locations following the "PlanePCA" procedure explained in [14]. This method makes a locally planar assumption of the 3×3 patch centered at a given pixel and computes the normal of this plane as the surface normal to the point. This produces fairly stable results, but has issues at depth discontinuities occurring at the boundaries between objects. We compute dominant planes in the depth image using RANSAC plane estimation [5]. Once the best sampled plane is chosen, we refine the plane estimate with a linear least squares fit using all support points in order to reduce noise effects from the depth data. After finding the first plane, we remove all support points and extract the most dominant plane in the remaining set. We continue this iterative process of adding additional planes until no plane with at least $20,000$ support pixels can be found. In our experiments, this process results in at most 6 planes for a 640×480 image.

4 Results

We collected a new dataset of 80 color and depth images using a mobile-manipulator robot in a real-world home environment. We created pixel-level segmentations of the objects in all of the images in the dataset. We present qualitative comparisons of our saliency method to four previous approaches. We provide quantitative evaluation of our saliency maps in the context of generic object segmentation.

4.1 Qualitative Results

Figure 4 provides a qualitative comparison of our saliency method to the previous methods of center surround with depth information [20], context-aware saliency [6], graph-based visual saliency [7] and image signature [11] approaches. In all examples, our depth features produce much better results than the previous depth based saliency work of [20]. Additionally we produce results which outperform the state of the art color saliency methods of [6, 7, 11].

Examining the color-based methods, we see they all produce fairly different results. The image signature method stands out for its much sparser result. When depth information is incorporated in the center surround method as an additional channel, it tends to over-smooth the result. For example, it fails to highlight the tissue paper and toothpaste in Figure 4(b). In contrast, our method is largely succesful in highlighting as salient those regions corresponding to objects in the scene (see Figure 4(b) to 4(f)).

We can see the effects of the different saliency methods by examining the resulting segmentations they produce. In Figure 5 we compare our segmentation results to those produced using the saliency measure of Goferman et al. [6] which, based on our observations, tends to produce the best results in comparison to the other baselines. Our saliency method labels

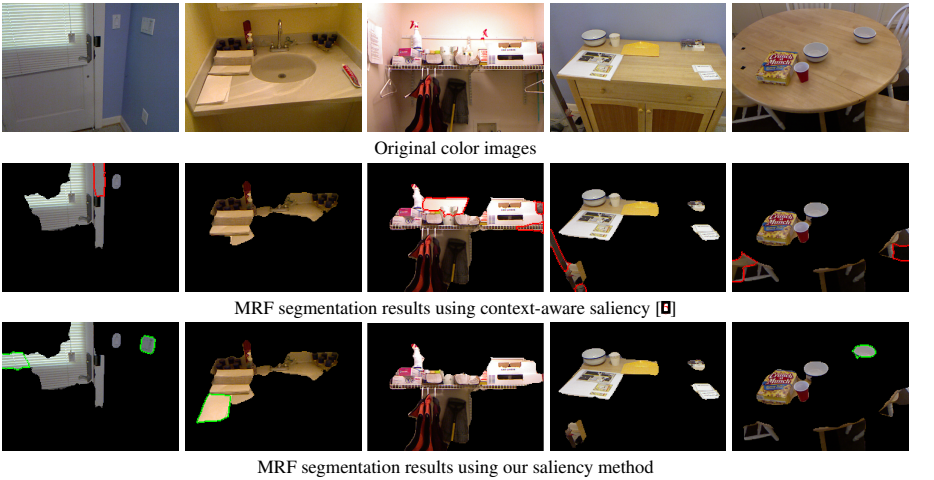


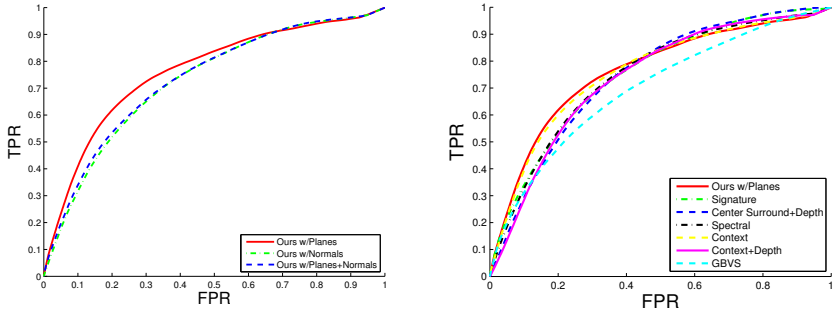
Figure 5: Comparison of segmentation results for the images presented in the first row. The second row are the segmentation results produced by running our MRF framework using the context-aware saliency scores [6]. The third row shows segmentation results for our best performing saliency method. Regions labeled foreground by the color based context-aware saliency and not by our method are outlined in red. We outline in green those regions our method detected as foreground which were not found by the color only method.

more of the objects in the scene as objects while capturing less of the background across these images. For example, our method detects the second set of light switches in the first column, while the method without depth cues does not. In the utility closet scene (third column), our segmentation results in a tighter fit to the objects of interest, and captures less of the background wall. Importantly, our method detects the second bowl in the scene of the final column, while the color-based method does not. In the second column, our method successfully labels the low-profile paper towels on the counter as foreground. This is a particularly difficult task, as the object depth is quite similar to the counter beneath it. The same can be said for the papers on the table in the fourth column. Taken collectively, these results demonstrate that the incorporation of depth cues for shape and layout into saliency can produce substantial benefits.

4.2 Quantitative Segmentation Results

We now present a quantitative evaluation of the benefit of depth information in saliency computation. Figure 6 shows ROC curves for all competing methods (image signature [11], center surround with depth information [20], spectral residual [10], context-aware saliency [6], depth-extended context-aware saliency, and graph-based visual saliency [7]). In these experiments segmentation is performed by a direct thresholding of the saliency map. Figure 6(a) summarizes results of three variations of our method. We see that the plane based method gives the best overall performance. In Figure 6(b), we compare the plane based method from Figure 6(a) to previously published approaches as well as the depth-extended context-aware saliency method.

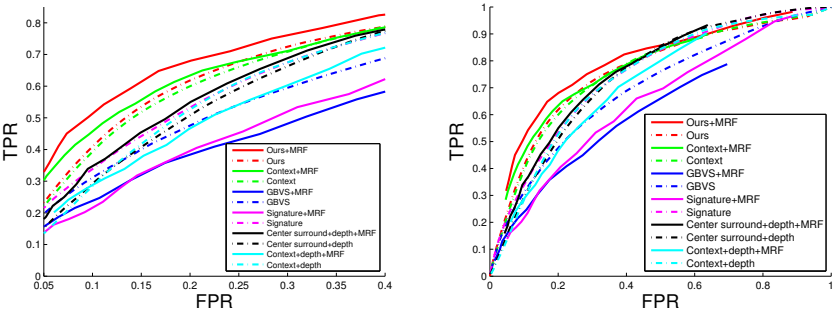
We can draw several insights from Figure 6. Raw depth data does not help much, since local changes are mostly smooth. This is true for both the center surround and context-aware saliency methods when depth is added into the model. Normal estimation does increase



(a) ROC curve for our methods (b) ROC curve for all methods

Figure 6: ROC curves for object segmentation. False Positive Rate vs True Positive Rate

segmentation performance, since curvature varies much faster than absolute depth in the scene. However the use of plane information increases performance the most, a finding reflective of the human attentional study in [23]. Adding surface normal information to the plane based saliency decreases the performance compared to using planes alone, while improving slightly above surface normals alone. We believe this is caused by the relatively high noise present in normal estimation. The RANSAC based plane estimation method is capable of suppressing this noise by finding dominant planes, while the raw normal estimates distract from the plane estimates when the two are combined.



(a) Zoomed in results of segmentation with the MRF model (b) Full results for MRF segmentation

Figure 7: ROC of object segmentation results with MRF. False Positive Rate vs True Positive Rate.

Figure 7 shows segmentation performance of our proposed MRF model. This MRF model improves the performance of our method, as well as those of [6] and [20]. However, our method benefits more by this MRF model and outperforms all others by a larger margin than before. Note that incorporating the depth map as just another image channel does not improve the saliency result and may even decrease the performance of the methods (see center surround+depth and context+depth in Fig. 7). This lends more support to our approach of using higher level depth features in performing unsupervised object segmentation.

5 Conclusion

We have presented a novel approach to saliency computation using shape and layout features derived from depth measurements. Specifically, we propose dissimilarity features based on

modeling the joint interaction between layout and shape features obtained from the depth channel, and the color information. Moreover we demonstrate that simply treating depth as an additional channel produces little improvement over purely color based methods. We compare our approach to standard techniques and show improved performance both qualitatively and quantitatively in the context of segmentation. We presented a novel MRF based object segmentation method based on saliency computation.

6 Acknowledgments

This work was supported in part by NSF Award 0916687.

References

[1] L. Bo, X. Ren, and D. Fox. Depth kernel descriptors for object recognition. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS)*, September 2011.

[2] Ali Borji and Laurent Itti. Sate-of-the-art in visual attention modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(1):185–207, January 2010.

[3] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23 (11):1222–1239, 2001.

[4] Ming-Ming Cheng, Guo-Xin Zhang, Niloy J. Mitra, Xiaolei Huang, and Shi-Min Hu. Global contrast based salient region detection. In *IEEE Conference On Computer Vision and Pattern Recognition (CVPR)*, 2011.

[5] Martin A. Fischler and Robert C. Bolles. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24:381–395, June 1981.

[6] Stas Goferman, Lihi Zelnik-Manor, and Ayellet Tal. Context-aware saliency detection. In *IEEE Conference On Computer Vision and Pattern Recognition (CVPR)*, 2010.

[7] Jonathan Harel, Christof Koch, and Pietro Perona. Graph-based visual saliency. In *Neural Information Processing Systems*, 2006.

[8] Scott Helmer and David G. Lowe. Using stereo for object recognition. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3121–3127, 2010.

[9] Stefan Hinterstoisser, Stefan Holzer, Cedric Cagniart, Slobodan Ilic, Kurt Konolige, Nassir Navab, and Vincent Lepetit. Multimodal templates for real-time detection of texture-less objects in heavily cluttered scenes. In *IEEE International Conference on Computer Vision (ICCV)*, 2011.

[10] Xiaodi Hou and Liqing Zhang. Saliency detection: A spectral residual approach. In *IEEE Conference On Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.

[11] Xiaodi Hou, Jonathan Harel, and Christof Koch. Image signature: Highlighting sparse salient regions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2012.

[12] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 20(11):1254–1259, 1998.

[13] Allison Janoch, Sergey Karayev, Yangqing Jia, Jonathan T. Barron, Mario Fritz, Kate Saenko, and Trevor Darrell. A category-level 3-d object dataset: Putting the kinect to work. In *ICCV Workshop on Consumer Depth Cameras for Computer Vision*, 2011.

[14] Klaas Klasing, Daniel Althoff, Dirk Wollherr, and Martin Buss. Comparison of surface normal estimation methods for range sensing applications. In *Proc. of the IEEE International Conference on Robotics and Automation (ICRA)*, pages 3206–3211, 2009.

[15] Cristof Koch and Shimon Ullman. Shifts in selective visual attention: towards the underlying neural circuitry. *Human Neurobiology*, 1985.

[16] Congyan Lang, Tam V. Nguyen, Harish Katti, Karthik Yadati, Mohan Kankanhalli, and Shuicheng Yan. Depth Matters: Influence of Depth Cues on Visual Saliency. In *Proceedings of the European Conference on Computer Vision*, 2012.

[17] Atsuto Maki, Jan-Olof Eklundh, and Peter Nordlund. A computational model of depth-based attention. In *International Conference on Pattern Recognition*, 1996.

[18] Atsuto Maki, Peter Nordlund, and Jan-Olof Eklundh. Attentional scene segmentation: Integrating depth and motion from phase. *Computer Vision and Image Understanding*, 2000.

[19] Greg Mori. Guiding model search using segmentation. In *International Conference on Computer Vision*, 2005.

[20] Nabil Ouerhani and Heinz Hüglie. Computing visual attention from scene depth. In *International Conference on Pattern Recognition*, 2000.

[21] Radu Bogdan Rusu, Nico Blodow, Zoltan Csaba Marton, and Michael Beetz. Close-range scene segmentation and reconstruction of 3d point cloud maps for mobile manipulation in human environments. In *Proc. of the IEEE/RSJ International Conference on Intelligent Robotics and Systems (IROS)*, St. Louis, MO, USA, 10/2009 2009.

[22] Martin J. Schuster, Jason Okerman, Hai Nguyen, James M. Rehg, and Charles C. Kemp. Perceiving Clutter and Surfaces for Object Placement in Indoor Environments. In *IEEE/RAS International Conference on Humanoid Robots (Humanoids)*, 2010.

[23] Jan Theeuwes, Paul Atchley, and Arthur F. Kramer. Attentional control within 3-d space. *Journal of Experimental Psychology: Human Perception and Performance*, 24 (5):1476–1485, 1998.

[24] Dirk Walther and Christof Koch. Modeling attention to salient proto-objects. *Neural Networks*, 19(9):1395–1407, 2006.