Journal of Signal and Information Processing, **2013, 4, 102-108**
doi:10.4236/jsip.2013.43B018 Published Online August 2013 (http://www.scirp.org/journal/jsip)

Scientific Research

Segmenting Salient Objects in 3D Point Clouds of Indoor Scenes Using Geodesic Distances

Shashank Bhatia, Stephan K. Chalup

School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan 2308, NSW, Australia.
Email: shashank.bhatia@uon.edu.au, stephan.chalup@newcastle.edu.au

Received May, 2013.

ABSTRACT

Visual attention mechanisms allow humans to extract relevant and important information from raw input percepts. Many applications in robotics and computer vision have modeled human visual attention mechanisms using a bottom-up data centric approach. In contrast, recent studies in cognitive science highlight advantages of a top-down approach to the attention mechanisms, especially in applications involving goal-directed search. In this paper, we propose a top-down approach for extracting salient objects/regions of space. The top-down methodology first isolates different objects in an unorganized point cloud, and compares each object for uniqueness. A measure of saliency using the properties of geodesic distance on the object's surface is defined. Our method works on 3D point cloud data, and identifies salient objects of high curvature and unique silhouette. These being the most unique features of a scene, are robust to clutter, occlusions and view point changes. We provide the details of the proposed method and initial experimental results.

Keywords: Saliency Detection; 3D Image Analysis; Image Segmentation

1. Introduction

Traditionally, methods of landmark extraction for the purpose of robot localization have been dependent on the type of environment and nature of landmarks. Such methods follow the standard procedure of sequentially scanning the input percept and aim to match pre-defined patterns for recognition of landmarks. The necessity of pre-defining patterns associated with landmark locations, limits the use of the robot to a specific environment. Furthermore sequential processing of data necessitates high computational power to be ported on a mobile plat- form. This sequential processing of pixels or image windows is in contrast to human visual mechanism. The latter incorporates an attention mechanism that helps humans to focus on the most relevant stimuli based on the task at hand [1-2]. Incorporation of similar strategy in computational vision systems, especially for application of robotics, can have many advantages. Computational Attention (CA), commonly known as "Saliency Detection" or "Interest Point Detection", aims to identify the regions of sensory input that stand out from their neighbors and attract the attention of the subject [3-4]. Due to the convenience that CA offers, it has been adapted in many applications requiring either judicial selection of inputs [5], minimizing computational cost [6], or to achieve invariance to clutter [7].

Computational attention in the 2D image domain has been investigated from past five decades. C. Koch and S. Ullman [8] were the first to provide theoretical foundations of visual attention mechanisms. The authors proposed creation of different conspicuity maps each selecting locations in visual space, which differ from their surroundings in terms of color and orientation. Further, a Winner Take All (WTA) neural network was proposed to combine different conspicuity maps and select the most salient region. Most of the current visual attention systems are based on the implementation of the WTA networks, developed by L. Itti *et al*. [9]. In their implementation, the authors extended the existing theoretical concepts by adding intensity as another feature for computing the conspicuity map. However, it should be noted that most of the existing approaches use the intrinsic properties of an input image. These properties depend on factors like the presence of ambient light, amount of re- flections, visibility of colors and presence of occlusions and are therefore unstable. This factor has motivated re- searchers to utilize 3D depth information (which is inde- pendent of ambient light) in the process of determining salient regions. Furthermore, the availability of the low- cost 3D capturing devices in recent years, has motivated the usage of 3D depth information, especially for systems on mobile robotic agents.

To identify salient regions in a scene, mechanisms that evaluate intrinsic properties of raw data elements to spot

the regions of potential interest are said to follow bottom-up approaches [10]. Such methods explore the neighborhood of each data point present in the input, and assemble points into small clusters having salient characteristics. In this process, the clusters of higher saliency thus obtained, may comprise of arbitrary points that may belong to multiple objects. On the other hand, methods that instead of operating on individual data elements evaluate a collection (with all data elements belonging to same object) are known as top-down approaches. Here, the evaluation is performed after isolation of the points into different objects. Many cognitive science studies and experimental evaluations described in [11] have shown that bottom-up methods are well suited to applications involving explorative tasks, but may not be suitable for goal-directed searching. These studies encourage ideas of taking an "object" as a unit for attention selection and support suitability of top-down approaches for applications involving goal directed search. However, since most of the existing methods of 3D saliency detection are based on bottom-up approaches, their usage has remained limited. Only the most simplistic methods like [12] and [6] have been used in robotic applications.

In this paper, we propose a simple top-down approach to extract salient regions from raw 3D point cloud data. The top-down nature of our approach segments the scene into physically disconnected regions and then compares properties of each region for saliency. We define saliency measures that capture variations in curvature and silhouette (an outline of an object/scene consisting of featureless interior) of the corresponding regions, and compare them with other objects present in the surroundings. We report the initial experiments and results of by testing it in an environment containing objects of different shapes and degrees of curvature. Section 2 of the paper provides a short review of related work, followed by the motivation behind this research. Details of the proposed approach are provided in section 3 and 4. Section 5 describes the measures and initial experiments conducted, and concludes the paper.

2. Related Work

Available methods in majority either cannot handle large size point clouds, or to achieve computational efficiency, reduce the dimensionality of the point cloud. In [13], a multi-scale filtering operator is derived by the convolution of a Gaussian kernel with the operating surface. The operator has the property of being proportional to the curvature of the local area at which it is applied. In effect, it is directly applicable to 3D point clouds and captures the variation in shape of the neighborhood of point of application. However, the need of processing over multiple scales renders its utility limited to very small point clouds. To overcome this disadvantage, J. Stuckler and S. Behnke [7] extended the interest operator to work on depth images. Similar to the historical intensity-driven visual attention algorithms, their approach builds a multi-scale pyramid representation of the depth image to be used with the operator. However, approximation of the depth image limits the factual description of the interest points limited to detection of blobs and corner-like features. It is important to note that in goal-driven applications, salient regions are useful only if they can be recognized as important features. Common to most applications, extracted regions should be invariant to noise, scale, and viewpoint transform. Following the same convention of bottom-up mechanisms provided by [7] and [12] limits the input space to depth images, and pro- vide a simplistic method to extract boundary regions as areas of interest. While this approach is computationally efficient, it does not take account of the 3D shape of the objects in determining the saliency values.

Cole and Harrison were one of the first to incorporate the 3D curvature information directly to identify regions of interest for the application of robot Simultaneous Localization and Mapping (SLAM) [6]. The authors utilized an information-theoretic entropy measure to identify the regions of maximum random curvature. In contrast to [7], authors of [12] used spherical regions to define the scale space. The degree of saliency was based on the entropy of normals in each spherical region and its variation over multiple scales. Using a spherical shape for defining a scale space leads to the selection of points with the highest variation in curvature as the most salient regions. These points in general may correspond to more than one object, and therefore do not provide any recognizable information of the selected region.

Flint et al. [14] defined an interest point to be the one having the largest principal curvature in all three Euclidian axes. The magnitude of all three principal curvatures was estimated by calculating the Hessian matrix convolved with a Gaussian kernel. Finally, the areas having the highest determinant of the Hessian matrix were deemed to be most salient regions. The experimental results revealed that the proposed method extracts all corner and edge points, representing the areas with the highest variations in curvature. Using spatial properties of the cloud data, Akman and Jonker [4] proposed to include the depth as a criterion for saliency. In their approach, two different saliency maps were used in combination to obtain the final saliency map. The first saliency map was calculated using values of curvature. The second map was taken as being inversely proportional to the depth of a region. The farther a region, the lower its saliency would be. In the application of saliency for object classification, Potapova and Zillich [15] proposed the extraction of the orientation of objects in point clouds relative

to the surface on which they are located. This relative orientation was used to create a saliency map, and was combined with the traditional 2D saliency approach to obtain a complete saliency map.

As evident from our review, most of the present research on visual attention mechanism on 3D data has been focused on bottom-up approaches. These approaches assume homogeneity of input data, and the detection of saliency is mostly based on the intrinsic properties. There is no mechanism that suggests the grouping of raw data into identifiable objects/artifacts present in the input space. Some of these methods require multi-scale processing, while others approximate 3D data to depth images. For this research, a particular highlighted drawback is the grouping of multiple objects into a region of interest. This potentially can degrade the efficiency of goal directed search applications. Research in cognitive science has shown the suitability of top-down attention mechanisms in goal directed applications [11]. Further, top-down mechanisms are more computationally efficient, as they do not necessarily process all the data sequentially. In view of these findings, in this paper we propose a top-down approach for salient region detection. Our approach first clusters objects present in an input percept, and then evaluates each separated object for the degree of attention that it may acquire. With the application of our approach, goal directed search applications have the possibility of increasing computational performance, and time efficiency. The major contribution of our approach is its top-down nature of saliency estimation, which considers an object as an elemental unit for attention selection.

3. Planar Region Extraction

Indoor scenes constitute planes as a major part of their point cloud input. The planar part of the 3D point cloud does not contribute to any variations in curvature, and therefore become distractions and increase computational cost. Additionally, removal of planar regions is required for isolation of salient objects from other artifacts present in the scene. Therefore, while seeking to identify regions with higher curvature, we choose to remove these planar regions using the RANSAC method as described in [16].

3.1. Local Surface Normals

The planar regions have low or zero curvature associated with them. Leveraging on these regions can be identified if the surface normal to each point is known. These LSNs are estimated using the method described in [17]. The surface normal to a query point can be estimated using the eigenvalues and eigenvectors of the of the covariance matrix comprising the k-nearest neighbors of the query point (equation 1).

$$C = \frac{1}{k}\sum_{i=1}^{k}\cdot\left(p_i - \overline{p}\right)\cdot\left(p_i - \overline{p}\right)^T \tag{1}$$

$$C\cdot\vec{v}_j = \lambda_j\cdot\vec{v}_j, j \in \{0,1,2\}$$

$$\sigma = \frac{\lambda_0}{\lambda_0 + \lambda_1 + \lambda_2} \tag{2}$$

Here $\overline{p}$ is the 3D centroid of the nearest neighbors. λ_j is the j^{th} eigenvalue, and $\vec{v}_j$ its corresponding eigenvector. The eigenvector $\vec{v}_0$ corresponding to smallest eigenvalue λ_0 is the approximation of the normal at the query point, and the ratio of the eigenvalues (equation 2) provides an estimate of variation in curvature at the query point.

3.2. Iterative RANSAC

Plane extraction using Random Sample Consensus (RANSAC) method described in [16], identifies best fitting planar region in the input point cloud. As a result, only one largest planar region is extracted from the input point cloud, leaving the rest intact. In order to remove subsequent planar regions, we adapted recursive use of RANSAC. The input point cloud is processed multiple times, separating one best planar surface at a time. Recursive RANSAC works by feeding back the residual cloud obtained in previous iteration. The extraction is executed until 95% of the point cloud is processed. As a result, a list of all extracted planar regions is obtained. This list contains all possible planar regions present in the input percept. In addition, the list may also contain parts of planar regions embedded on objects present in the environment. The resultant cloud may thus contain occlusions and holes. These holes also lead to incorrect object based clustering. To overcome this disadvantage, we employ a statistical high pass filter, which removes only significantly large planar regions present in the input cloud. The high pass threshold value is set to $\mu + \sigma$, where μ and σ are the mean and standard deviation of the total number of points contained in all the candidate planes. Finally the candidate planes with number of points above the threshold are removed from the original cloud.

Figure 1 displays raw point cloud data of The Newcastle Robotics Lab, with different objects placed on the ground. The objects comprise of a toy bear, a humanoid robot, a carton box, and a basketball. These objects have different properties of curvature and were chosen to demonstrate the behaviour of the proposed method with different types of objects. The extracted planar region points are marked different shades of grey. Objects that remain after plane extraction are marked in black. The data was collected using the Kinect RGB-D camera [18]. More details of the experimental setup are

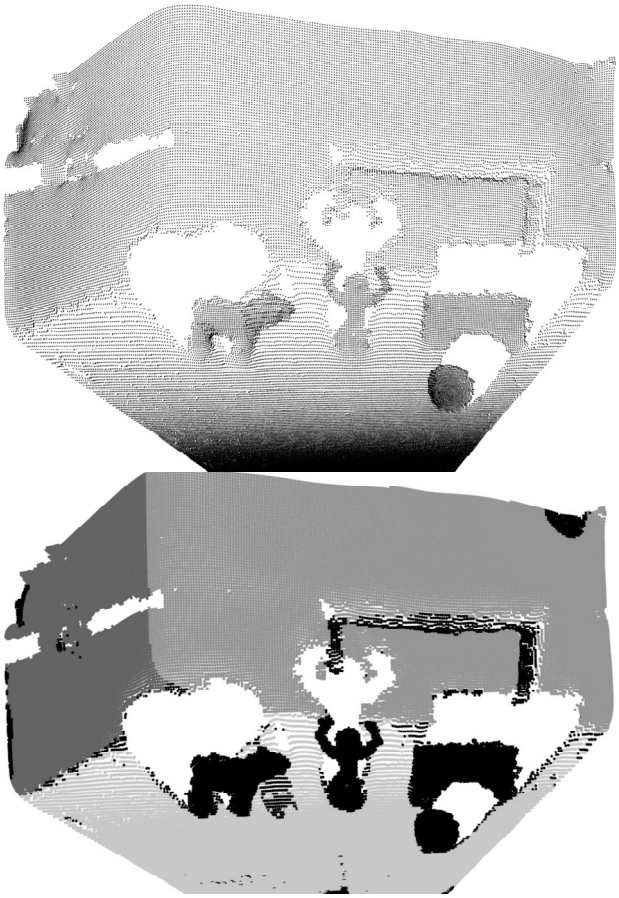

Figure 1. Complete point cloud (top), corresponding detected planar regions marked in different shades of grey (bottom), residual objects after plane extraction marked in red (bottom).

provided in Section 4.

4. Salient Region Extraction

In this section we provide details of how saliency is measured, and relevant areas with high saliency extracted. There are two major phases involved. First, the Euclidean clustering, and second the ranking of the extracted clusters for saliency. Euclidean clustering divides the point cloud into smaller objects/regions to be considered as elements for saliency computation. Finally, the separated objects are evaluated for saliency. The following subsections provide details of each phase.

4.1. Euclidean Clustering

The residual cloud obtained after removal of planar regions contains unlabeled points, some belonging to isolated objects and others being noisy residual of planar extraction. In order to perform a top-down object-level comparison of uniqueness, these objects have to be identified as separate entities. In other words, the point cloud needs to be divided in multiple parts, each containing one isolated object. This is achieved by comparing the Euclidean distance between neighboring points. Clustering is performed using a k-nearest neighbor search [19]. Nearest neighbor search starts by selecting a random point from the residual cloud, and computes the Euclidean distance of the point from its nearest neighbors. Points that fall under a threshold value are labeled to be part of the object. The search stops when no nearest neighbor falls under the threshold distance. At this stage the points found so far are labeled into one group, and the search starts again by removing the object from the cloud, and randomly selecting another un-labeled point. The clustering stops when all points are labeled. For the nearest neighbor search, we make use of a binary KdTree implementation as in [20]. This approach divides the residual cloud into a binary tree structure, enabling easy and fast nearest neighbor searches. The result of the clustering can be seen in **Figure 2**. The distance threshold used here is 0.2 m and as evident, the method identifies four different objects present in the scene. In the figure, each object is presented by different color of the points it contains.

4.2. Saliency Ranking

Multiple point clouds obtained as a result of clustering are evaluated for uniqueness in two aspects: 1) Variance of curvature on the object's surface, 2) Shape of silhouette formed from the object. These properties are captured together in one measure, defined using the difference between geodesic and Euclidean distances between all sets of points in the object point cloud. The geodesic distance between any two points of a cloud is the length of the shortest curve on the surface, connecting these points. Due to the embedding of the curve on the surface, in Euclidean space the geodesic distance between two points having non-zero curvature is always greater than or equal to their Euclidean distance. Additionally, surfaces with high amount of variation in the curvature of their boundary/silhouette may also have their geodesic

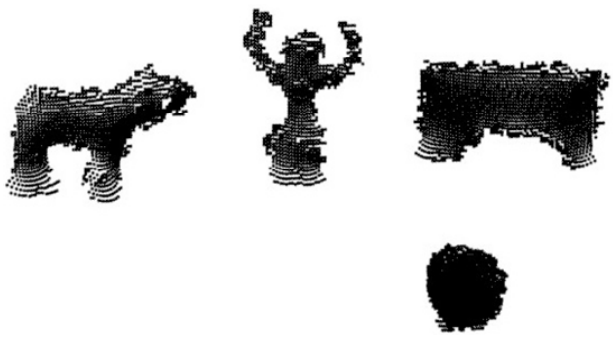

Figure 2. Clusters obtained after removal of planar regions, and performing Euclidean Clustering on the residual cloud.

distance between any two points on their boundary granter than corresponding Euclidean distance. **Figure 3** illustrates these facts by means of two simple examples. First, a half sphere is presented with the geodesic distance (green curve) and Euclidean distance between two points on the surface of the half sphere. The grater value of geodesic distance is evident from the figure. Secondly, a curved silhouette of an arbitrary object is presented. Again, the geodesic distance between two points on the boundary is grater than their corresponding Euclidean distance.

Figure 3 conveys that for any object with curved silhouette and higher curvature on the surface, the values of sum of geodesic distance between all points would be high. More precisely, the difference between geodesic and Euclidean distances between all points of the surface can be used to identify objects with higher curvature and complex shape of silhouette. Exploiting the properties of geodesic distance, we formulate a saliency measure that captures variation in the curvature as well as the silhouette of the object under study. Consider an object point cloud C_k comprising of a total n_k points, for each point p_i in we define the following:

$$G_{ij}^k = \langle p_i, p_j \rangle,\ i \neq j, \forall i, j \in C_k$$

$$E_{ij}^k = \| p_i, p_j \|,\ i \neq j, \forall i, j \in C_k$$

$$P_k = \sum_{ij} \left(G_{ij}^k - E_{ij}^k \right)^2$$

$$V_k = \sum_{ij} \frac{\left(P_k - \overline{P_k} \right)^2}{n_k - 1}$$

$$S_k = \frac{V_k}{\sum_k V_k} \tag{3}$$

G_{ij}^k denotes the geodesic distances from point p_i to p_j in the k^{th} object point cloud. E_{ij}^k denotes corresponding Euclidean distance between the same points.

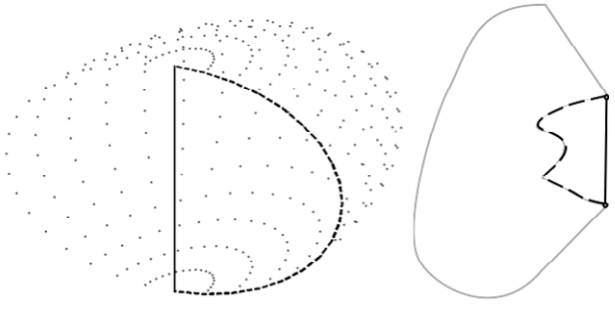

Figure 3. Comparison of geodesic distance (dashed line) and Euclidean distance (solid line) in presence of curvature (left), on a curved silhouette (right). Note that in both cases, the geodesic distances are higher than Euclidean distances.

Since the geodesic distances are always grater or equal to the Euclidean distances for surfaces with higher curvatures, any object having higher variability in the difference between G_{ij}^k and E_{ij}^k will stand out from its surroundings. This variance is captured by V_k and finally, the geodesic saliency S_k is defined as the normalized value of V_k (normalized over all clusters). The graph in **Figure 4** presents the quantity S_k normalized by the size of the cluster. This normalization factor also ensures that the value of saliency does not depend on the size and number of points contained in the point cloud of the object. The graph illustrates the variation in the values of the proposed saliency measure against changes in distance. There are four different objects present in the input cloud namely a toy bear, Nao humanoid robot, a basketball, and a flat box object. It can be noticed that as the distance increases, the value of saliency reduces. This happens due to the addition of noise. Moving away from objects, the curvature is less observable. This particular noise addition is sensor dependent, and current experiments report the results obtained using Microsoft Kinect-RGBD Sensor.

Geodesic distances between all pairs of points in each object point cloud are computed using Floyd-Warshalls algorithm [21]. The point cloud of each object is converted into a fully connected graph, with each point treated as a vertex. The algorithm compares distance-minimizing paths between two vertices in the given connected graph and incrementally improves the estimate of geodesic distance between two vertices iteratively. A more detailed explanation of the algorithm can be found in [21].

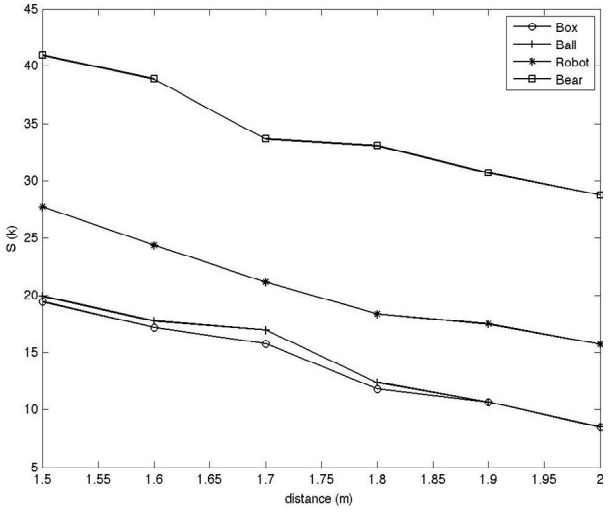

Figure 4. Values of the quantity $G_{ij}^k - E_{ij}^k$, normalized over the size of cloud. Note that as the distance increases, the saliency values of ball and box converge. This is due to increasing noise in the calculation of curvature with increase in distance.

5. Experimental Evaluation

In order to evaluate our approach we captured point clouds from different view-angles in the laboratory. Four objects of varying shapes and curvature were placed on the floor. 3D point clouds were recorded with 3D Time-Of-Flight (TOF), Microsoft Kinect Sensor [18] that was mounted on a tripod. The viewpoint was varied from -25 to 25 degrees, in steps of 5 degree. The distance was varied between 1.5 m to 2 m, in steps of 0.1 m. The input clouds were processed on a Dell workstation equipped with Intel Xeon® 3.40 GHz processor and 16 GB of RAM.

5.1. Performance Measures

We utilized the existing measures of repeatability and overlap, as described in [7] to evaluate our approach. These measures are known to evaluate qualitative and quantitative performance of saliency extraction method. The repeatability of detection of salient regions is defined as the frequency with which the same cluster is ranked with a similar level of saliency. This measures the stability of the approach with variations in distance and viewpoint changes. Overlap rate on the other hand, is calculated by comparing the location of salient regions found with variations in distance and viewpoint. If the salient regions belong to same location, overlap is incremented and vice versa. Location of the salient region was calculated as the centroid of the point cloud cluster. Since the position of the sensor was changed, these cen- troids were transformed from the local frame of reference into the global frame of reference of the environment. Heatmaps are used to graphically represent this perfor- mance measure. The heatmap used consists of cells, with each cell representing the value of repeatability/overlap (scaled between 0 and 1). The rows of the heatmap represent the distance of evaluation, and the columns represent the angle in degrees. All together, presented heatmaps are a visual representation of robustness of proposed method.

5.2. Discussion

Figure 5 shows the resultant heat map reflecting the overlap and repeatability of salient object detection. The values of repeatability and overlap are scaled (between 0 and 1) to provide an accurate account of the performance. The figure demonstrates the repeatability of the toy bear is higher than that of the robot. This is due to complexity of silhouette of the toy bear. Reason behind higher saliency values of the toy bear are depicted in **Figure 3**. Additionally **Figure 4** conforms to results in **Figure 5**, where the toy Bear has attained highest values of saliency. The humanoid robot, having highly curved surface

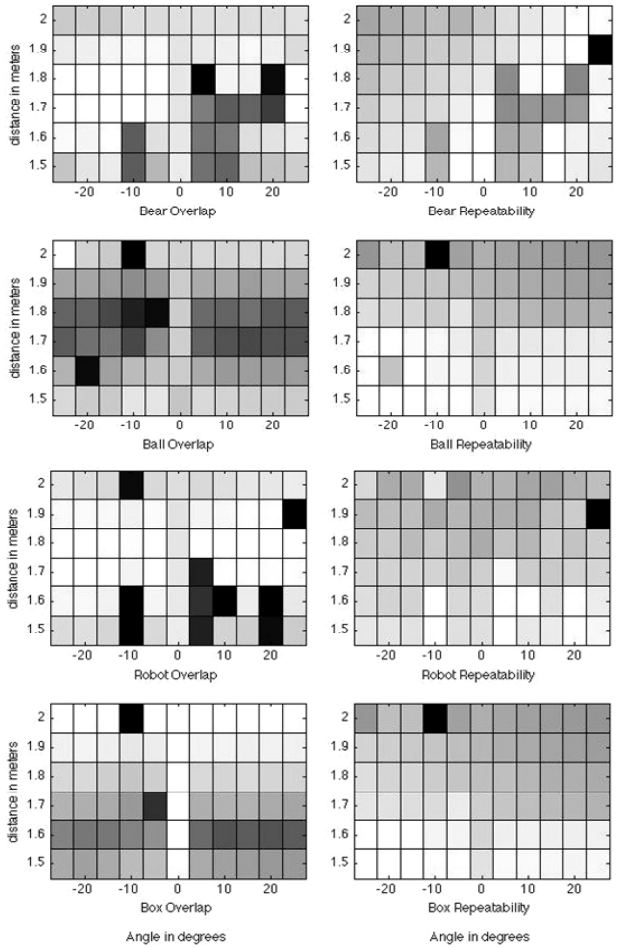

Figure 5. Bear (top), Ball (row 2), Robot (row 3), and Box (bottom) performance (left: overlap and right: repeatability) wrt. view angle and distance. This performance measure was adapted from [7]. We can see that the proposed method is robust to viewpoint and scale changes.

follows next in saliency ranking. Despite being smaller as compared to the flat box, it has higher values of saliency. Finally, the robustness is demonstrated by the high values of repeatability, which in most cases ranges between 0.7 to 1. It should be noted that lower values of overlap and repeatability in case of bear and ball are due to the restricted exposure of the objects with change in angle of the sensor. Moving beyond 10 degree, the bear was not completely visible in the Field of View (FOV) of the sensor. Similarly the ball, that was not visible in the FOV while changing the angle of the sensor below 0 degree. Apart from the missing values, all other observations presented high values of the two measures, which are most desirable characteristics of salient region extraction methods [7].

6. Conclusions

In this paper, we present a top-down approach for ex-

tracting salient regions/objects from indoor environments. Our method segregates significant planar regions, and extracts isolated objects present in the residual point cloud. Each object is then ranked for saliency based on higher curvature complexity of the silhouette. These properties are captured together using the proposed geodesic distance measure (**Figure 4**). The paper has reported initial experiments and demonstrates capacity of the method in identifying objects/regions of higher curvature. Further, testing with variations in viewpoint and distance, reveal stability of proposed saliency criterion.

These initial experiments demonstrate the advantages of adapting top-down clustering for the purpose of saliency ranking. A possible limitation of the method could be identified as lack of using RGB information to support the selection of salient regions, and future developments of this research aim to include variations in color for saliency computation.

REFERENCES

[1] T. N. Vikram, M. Tscherepanow and B. Wrede, "A Saliency Map Based on Sampling an Image Into Random Rectangular Regions of Interest," *Pattern Recognition*, Vol. 45, No. 9, Sep. 2013, pp. 3114-3124. doi:10.1016/j.patcog.2012.02.009

[2] S. Frintrop, E. Rome and H. I. Christensen, "Computational Visual Attention Systems and Their Cognitive Foundations," *ACM Transactions on Applied Perception*, vol. 7, no. 1, Jan. 2010, pp. 1-39. doi:10.1145/1658349.1658355

[3] N. Riche, M. Mancas, B. Gosselin and T. Dutoit, "3D Saliency for Abnormal Motion Selection: The Role of the Depth Map," in J. Crowley, B. Draper, and M. Thonnat, Eds. *Computer Vision Systems*, Springer Berlin/Heidelberg, 2011, pp. 143-152. doi:10.1007/978-3-642-23968-7_15

[4] O. Akman and P. Jonker, "Computing Saliency Map from Spatial Information in Point Cloud Data," *Advanced Concepts for Intelligent Vision Systems*, Vol. 6474, 2010, pp. 290-299. doi:10.1007/978-3-642-17688-3_28

[5] D. Simon, "Fast and Accurate Shape-Based Registration," PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburg, PA, 1996.

[6] D. Cole and A. Harrison, "Using Naturally Salient Regions for SLAM with 3D Laser Data," *In Proceedings of International Conference on Robotics and Automation*, Workshop on SLAM, 2005

[7] J. Stückler and S. Behnke, "Interest Point Detection in Depth Images Through Scale-Space Surface Analysis," *In 2011 IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 3568-3574. doi:10.1109/ICRA.2011.5980474

[8] C. Koch and S. Ullman, "Shifts in Selective Visual Attention: Towards the Underlying Neural Circuitry," *Human Neurobiology*, Springer-Verlag, Vol. 4, No. 4, 1985, pp. 219-227.

[9] L. Itti, C. Koch and E. Niebur, "A Model of Saliency Based Visual Attention for Rapid Scene Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 11, 1998, pp. 1254-1259. doi:10.1109/34.730558

[10] C. E. Connor, H. E. Egeth and S. Yantis, "Visual Attention: Bottom-Up Versus Top-Down," *Current Biology*, Vol. 14, No. 19, 2004, pp. R850-R852. doi:10.1016/j.cub.2004.09.041

[11] M. Begum and F. Karray, "Visual Attention for Robotic Cognition: A Survey," *IEEE Transactions on Autonomous Mental Development*, Vol. 3, No. 1, 2011, pp. 92-105. doi:10.1109/TAMD.2010.2096505

[12] B. Steder and R. Rusu, "NARF: 3D Range Image Features for Object Recognition", *In Workshop on Defining and Solving Realistic Perception Problems in Personal Robotics at the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2010.

[13] R. Unnikrishnan and M. Hebert, "Multi-Scale Interest Regions From Unorganized Point Clouds", *In IEEE Computer Society Conference, Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-8.

[14] A. Flint, A. Dick and A. v. d. Hengel, "Thrift: Local 3D Structure Recognition," *In Society on Digital Image Computing Techniques and Applications, 9th Biennial Conference of the Australian Pattern Recognition*, 2007, pp. 182-188.

[15] E. Potapova, M. Zillich and M. Vincze, "Calculation of Attention Points Using 3D Cues," *35th Annual Workshop of the Austrian Association for Pattern Recognition* (OAGM/AAPR), May 2011.

[16] R. B. Rusu, "Semantic 3d Object Maps for Everyday Manipulation in Human Living Environments," PhD Thesis, Computer Science Department, The University of Rochester, Oct. 2009.

[17] D. Holz, S. Holzer and R. Rusu, "Real-Time Plane Segmentation Using RGB-D Cameras," *In Proceedings of the 15th RoboCup International Symposium*, Isntanbul, Turkey, 2011, pp. 306-317.

[18] Microsoft kinect sensor, http://www.xbox.com/en-au/kinect?xr=shellnav," 2013.

[19] E. Fix and J. L. Hodges, "Discriminatory analysis. Nonparametric discrimination: Consistency properties," *International Statistical Review/Revue Internationale de Statistique*, 1989, pp. 238-247.

[20] R. Rusu and S. Cousins, "3D Is Here: Point Cloud Library (PCL)," *In IEEE International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1-4.

[21] R. W. Floyd, "Algorithm 97: Shortest path", *Communications of the ACM*, Vol. 5, No. 6, Jun. 1962, p. 2. doi:10.1145/367766.368168

