Proceedings of the RAAD 2011
20th International Workshop on Robotics in Alpe-Adria-Danube Region
October 5-7, 2011, Brno, Czech Republic

Fast Range Image Segmentation for a Domestic Service Robot (RAAD 2011)

Peter Einramhof[a], Robert Schwarz[a] and Markus Vincze[a]

[a] *Automation and Control Institute, Vienna University of Technology, Austria*
E-mail: {einramhof, schwarz, vincze}@acin.tuwien.ac.at
URL: www.acin.tuwien.ac.at

Abstract. In this paper we present a fast approach to range image segmentation. The segmentation results are intended to serve as input to the perception system of a domestic service robot. In the first step mixed pixels at depth discontinuities are identified. This is followed by extracting step and roof edges. Planar patches are detected with a focus on horizontal and vertical planar structures. Finally, the range data is labelled with the locally dominant features, which results in a label map. Our approach was tested on real range images recorded in a home-like office environment using a tilting laser range finder.

Keywords. Range Image, Segmentation, Real-Time, Service Robotics

1. Introduction

In "classic" service robotics the focus lies on transporting objects in well-structured and controlled environments like factories, hospitals or offices. In such environments a simple perception system is sufficient to provide data for the necessary tasks of the robot, namely self localisation and obstacle avoidance. The most popular sensor is a 2D laser range finder scanning parallel to the ground. The environment is modelled in a 2D fashion (2D grid or feature maps) since the presence of many unobstructed vertical structures such as walls or the faces of file drawers and closets allow for this approach. Processing the sensor data, for instance 361 range measurements for one 180° scan of the laser range finder, requires only very little computational power to work in real-time, which is a necessity since the robot is moving in a dynamic environment.

In domestic service robotics the demands to the perception system are higher in comparison. Home environments are challenging since they are cluttered and less structured. There are amorphous surfaces like curtains or other home textiles, and protruding surfaces such as table tops. The three-dimensional nature of this environment can no longer be neglected. To be useful for the in general technically non-trained users, the offered services must be more than just safely navigating from A to B. The perception system has to support grasping and object detection, for instance. The latter also contributes to more natural human-robot interaction since it is closer to human perception – the user would rather like to call the robot to "the sofa in the living room" than to (x, y, θ).

When changing over from 2D to 3D data (more exactly: 2.5D data), the demands to the computational part of the perception system rise drastically due to the increased amount and complexity of the data. Nevertheless, the real-time requirement still stands. Since mobile robots run on batteries and a long time of autonomy is desired, the onboard computational power cannot be increased arbitrarily (not to mention cost, size, heat and noise generation). To achieve real-time responsiveness despite this restriction, fast segmentation algorithms are required.

In this work we address range image segmentation under real-time constraints as well as detection and removal of mixed pixels, outliers that occur at discontinuities within the range image. The features extracted in the course of our segmentation approach are intended to be useful for various tasks of indoor service robotics: horizontal planar structures provide information for safe navigation and about support planes, vertical planar structures can be used for map building and self localisation, and finally, step and roof edges define object boundaries and the transition between objects and support plane or between object parts.

The remainder of this paper is structured as follows: Section 2 gives an overview of related work. Section 3 provides the motivation for the features we

have selected as target result and describes our approach in detail. Section 4 provides experimental results on data recorded with a tilting laser range finder. Finally, Section 5 concludes with a summary and an outlook.

2. Related work

When compared to intensity or colour images, which provide information about the surface properties of the objects observed by a sensor, range images encode the three-dimensional structure of the observed scene.

The purpose of range image segmentation is to divide the image into features or regions that are meaningful with respect to a given task. Despite a history of about three decades, there still is no standard approach to range image segmentation. Depending on the task, constraints, assumptions made about the nature of the content of the range image, and whether specific properties of the sensor are incorporated, algorithms range from fast ad hoc solutions to slow(er) sophisticated methods.

Comparison of quality and performance of different segmentation methods is difficult due to the lack of sound experimental evaluation. An exception is the field dealing specifically with the segmentation of objects with planar faces, published in (Hoover et al., 1996) together with experimental data. A detailed overview of literature in that field is available in (Haindl and Zid, 2007). Segmentation methods can be roughly divided into edge-based and region-based approaches.

Edge-based methods are inspired by human vision since humans have the principle that there is an edge or discontinuity of some kind between two separable objects (Zhang and Zhao, 1995; Palmer et al., 1996). Edge pixels have a gradient assigned, that is, magnitude and direction of the greatest local change. (Sappa and Devy, 2001) use an edge-based segmentation technique that consists of two stages: the first stage generates a binary edge map based on scan line approximation that considers only two orthogonal scan line direction. The second stage links the edge points by applying a graph strategy. In (Bellon and Silva, 2002) the authors present range image segmentation based on edge detection techniques with the aim of better preserving the object topology and shape in noisy range images. Their approach avoids fixed thresholds for being useful in unsupervised systems. (Han et al., 2004) propose a jump-diffusion method for segmenting a range image and its associated reflectance image in a Bayesian framework. In (Harati et al., 2007) the authors propose the metric "bearing angle", which the incidence angle between the measurement beam and a surface. By thresholding the bearing angle and its first derivative, step and roof edges are detected. Since their target application is 3D indoor SLAM they are rather interested in the remaining planar patches and thus remove all edges.

Region-based methods group pixels into regions using the criteria of proximity and homogeneity. These methods achieve grouping either by splitting the image into smaller regions (Lee et al., 1998), merging small regions into larger ones (Hoover et al., 1996), or splitting and merging until all criteria are maximally satisfied (Chang and Li, 1994; Jiang and Bunke, 1994; Hijjatoleslami and Kittler, 1998). In more recent work, (Gotardo et al., 2003) present a robust estimator, derived from the RANSAC and MSAC estimators, whose optimization process is accelerated by a genetic algorithm. Their range image segmentation algorithm is based on planar surface extraction in preserving small regions and edge locations when processing noisy images. Similarly, (Wang and Suter, 2004) propose a highly robust estimator (Maximum Density Power Estimator), which applies nonparametric density estimation and density gradient estimation techniques in parametric estimation ("model fitting"). According to the authors it can tolerate more than 85% outliers. (Weingarten et al., 2004) use probabilistic plane fitting to extract large planar surfaces from range images as input to mapping the environment for mobile robotics.

Similarly to (Bellon and Silva, 2002), we also make use of standard image processing as much as possible for edge detection. To increase robustness, various methods are combined in a voting scheme, and also the metric "bearing angle" (Harati et al., 2007) is incorporated.

3. Approach

In this section we discuss what features are extracted in the course of our segmentation approach as well as their relevance in the context of domestic service robotics. This is followed by a detailed discussion of the individual processing steps of our approach.

3.1. Target features

The vertical axis is an important reference (direction of gravity). This information is incorporated into the sensor data and associated processing algorithms via the known geometry and kinematics of the setup consisting of sensor and robot, and optionally via inclinometers.

In (manmade) indoor environments horizontal and vertical planar structures are dominant; together they also define the room structure.

Horizontal planar structures such as the ground, table tops or the seats of chairs play the role of support planes on which the robot moves and on which obstacles or objects of interest are located.

Vertical planar structures stem from walls, the faces of closets or the bodies of objects. Especially walls define the boundaries of the indoor

environment and can serve as features for the robot's self localisation.

Step edges occur at object boundaries, more exactly at the transition between foreground and background. We consider step edges as part of foreground objects.

Roof edges occur at the transition between parts of objects or between object and support plane, and also at top rims and high-curvature surfaces. Roof edges can be concave or convex.

3.2. Mixed pixels

In range images (or equivalently: depth images) from tilting laser scanners, time-of-flight cameras and also stereovision, there are range measurements at depth discontinuities that do not correspond to any physical structure. These mixed pixels have values that lie somewhere between the valid foreground and background range measurements (Fig. 1). They are problematic because they seemingly connect foreground and background points to one contiguous object and thus have to be removed.

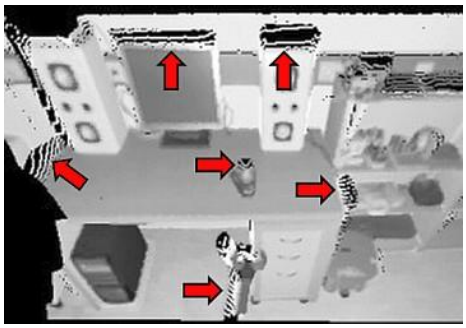


Fig. 1. 3D data of a gripper approaching a table scene. At depth discontinuities mixed pixels occur (marked by arrows)

The first step in detecting mixed pixels is to find depth discontinuities, that is, abrupt changes in the values of neighbouring pixels of the range image. But what extent of change is "abrupt"? Clearly, this depends on the measurement noise of the used sensor. To determine the noise level, we compute the standard deviation of the range values for each pixel within a 3x3 neighbourhood. Fig. 2, right, shows that the standard deviation is high at depth discontinuities – in fact, the result is qualitatively equivalent to the gradient magnitude produced by a Sobel filter.

The standard deviation computed for each pixel votes for a bin of a histogram with a bin width of 1mm. We determine the location of the (first) peak of the histogram and consider it as "sigma" of the noise. Three times this "sigma" serves as threshold to determine depth discontinuities within the standard deviation image. While this threshold works well, there is a problem when neighbouring pixels' beams intersect with planar structures at larger distances and at a flat angle (Fig. 3, left). In such cases the local change in measured range is well above the threshold and would thus register as discontinuity.

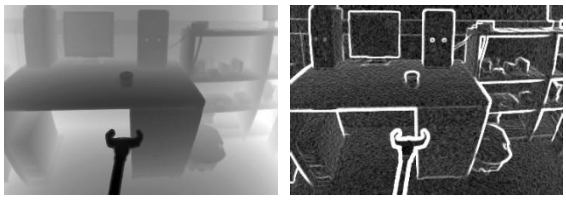


Fig. 2. *Left:* range image. *Right:* standard deviation for each pixel of the range image within a 3x3 neighbourhood. Maximum deviation was clipped to 0.05m for better visibility of low deviation regions

In (Harati et al., 2007) the authors propose the metric "bearing angle" (Fig. 3, right), which the incidence angle between the measurement beam and a surface. At real depth discontinuities this angle (β_i), Eq. (1) and (2), takes values close to 0° or 180°, i.e. the beam is close to parallel to the surface.

$$d_i = \sqrt{r_i^2 + r_{i+1}^2 - r_i r_{i+1} \cos \Phi_i} \qquad (1)$$

$$\beta_i = arccos\left(\frac{r_i - r_{i+1} \cos \Phi_i}{d_i}\right) \qquad (2)$$

Although Harati et al. use solely this metric, it is problematic at short range i.e. objects close to the sensor, like in the case of a table scene. Due to the in general small angular increment Φ_i between two neighbouring measurement beams, r_i and r_{i+1}, and the resulting small lateral distance between them at close range, the bearing angle β_i rather reflects the measurement noise than the geometry of the scanned object in such cases.

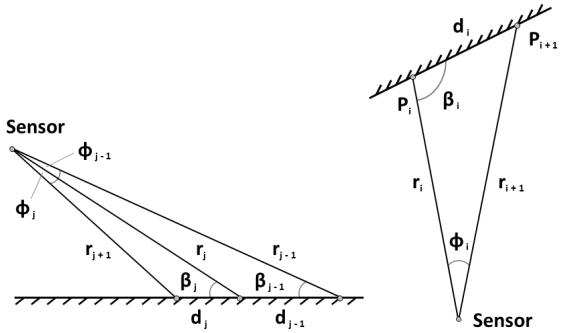


Fig. 3. *Left:* three neighbouring range measurement r_{j-1}, r_j and r_{j+1} of a column of the range image that intersect with a horizontal plane, e.g. the ground at a flat angle (viewed from the side). *Right:* horizontal bearing angle β_i (viewed from above). r_i and r_{i+1} are two neighbouring range measurements of a row of the range image, enclosing an angular increment Φ_i. d_i is the distance between the intersection points P_i and P_{i+1} of the measurement beams r_i and r_{i+1} with the surface.

Since thresholding the standard deviation image yields wrong depth discontinuities at greater distances and thresholding the bearing angle yields wrong depth discontinuities at close distances, we multiply both thresholding results, which leaves only discontinuities where both methods agree (Fig. 4, top left). The threshold value for the bearing angle depends on the sensor; for our setup we experimentally found angles of smaller or equal 5° or greater or equal 175°, respectively, indicating discontinuities.

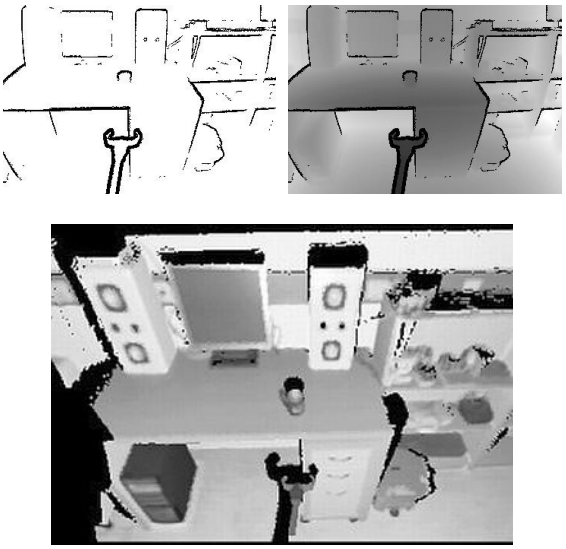


Fig. 4. *Top left:* mask for mixed pixels removal. *Top right:* mixed pixels removed from range image. *Bottom:* 3D data computed from masked range image

3.3. Step edges

In our definition step edges are pixels of the range image at depth discontinuities that (1) are valid pixels and (2) belong to the foreground. Thus, they represent boundaries of foreground objects.

We first smooth the range image using a 7x7 Gauss filter (σ = 1.0). The previously computed mixed pixel mask (Fig. 4, top left) defines "no-go areas" so as not to smooth over depth discontinuities. In the next step, the filtered range image is convoluted with a 3x3 mask that has "-8" as central element and ones as 8-neighbours (basically a Laplace filter). The result is an image that has positive values at the edges of foreground objects (Fig. 5, left). At smooth parts of the range image, there is only a small response due to noise.

Negative values are replaced by zero and a histogram of the positive values is created. Like described in the previous section, the location of the first peak is detected and three times its value is used as threshold. Finally, single-standing pixels that have no further pixel in its 8-neighbourhood are removed since they stem from noise (Fig 5, right).

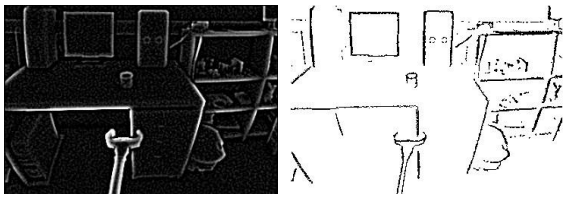


Fig. 5. *Left:* result of Gauss-filtering the range image and subsequently convoluting it with a 3x3 Laplace mask. *Right:* step edges after thresholding the Gauss- and Laplace-filtered range image

3.4. Roof edges

Step edges are added to the mask containing the mixed pixels. Again, this mask defines no-go areas for a second smoothing step. A 3x3 median filter is applied to the previously Gauss-filtered range image. From the smoothed range image, the horizontal and vertical bearing angles are computed. Furthermore, the bearing angles at each pixel are adjusted by the horizontal and vertical deflection angle of the measurement beam so that pixels belonging to planar structures have constant values of the bearing angles (Fig. 6, top row). As suggested by (Harati et al., 2007), each bearing angle image is subjected to edge detection; we use a horizontal 3x3 Sobel mask on the horizontal and a vertical Sobel mask on the vertical bearing angle image. The resulting two edge images are thresholded. We use 15° as minimum local change of the bearing angle. The roof edges from the thresholding results for the horizontal and vertical bearing angle are combined (Fig 6, bottom left).

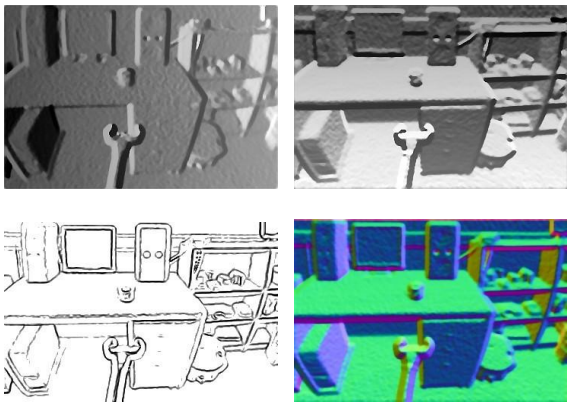


Fig. 6. *Top left:* horizontal bearing angle. *Top right*: vertical bearing angle. *Bottom left*: roof edges. *Bottom right*: surface normals

3.5. Horizontal and vertical planar patches

From the same smoothed range image as the bearing angles were computed, we also calculate the surface normals (Fig. 6, bottom right). First, we compute 3D points from the valid points of range image. The coordinates of each 3D point are stored in individual arrays of the same size as the initial range image and at the same array cell position as its associated range

value. In this way the initial neighbourhood is maintained. For 3x3 pixel patches in regions without mixed pixels or step and roof edges two vectors are computed from the left- and rightmost and the top- and bottom-most 3D point in that patch. Finally, by applying the cross product to both vectors and by normalizing the resulting vector's length, we get the surface normal.

The surface normals are multiplied (dot product) with the unit vector of the vertical axis. As stated earlier, the information about the vertical direction has to be supplied from outside, either from the known geometry of the setup or from inclination sensors. The result of the dot product is thresholded. We allow a deviation of 10° from the vertical axis (horizontal plane) for the surface normals of horizontal (vertical) planar structures.

3.6. Label map

Mixed pixels, step edges, roof edges, vertical and horizontal planar patches have so far been stored in individual maps that have the same size as the initial range image from which they were derived. To each pixel of the range image we assign a label according to the locally dominant feature type. If more than one feature type has activation at a pixel location, a prioritization is applied: Mixed pixels, then step edges, roof edges, vertical and finally horizontal planar patches. The result is a label map (Fig. 8, right column).

4. Results

The following two subsections describe the sensor used for data acquisition and the data itself as well as practical results achieved on that test data.

4.1. Test data

A tilting 2D laser range finder (Fig. 7) built from a SICK LMS 100-10000 scanner and a SCHUNK PW 70 rotary tilt unit was used to capture test data. Each captured frame provides 360x500 range and intensity measurements. With an angular resolution of 0.25° horizontally and 0.125° vertically, the field of view is 90°(H)x62.5°(V).

Fig. 7. Tilting laser range scanner for capturing the test data

The sensor was mounted onto a mobile robot at a height of about 125cm with respect to the ground. The top of the vertical field of view is parallel to the ground plane, its bottom is tilted downwards by 62.5°. This configuration allows scanning table scenes as well as detecting obstacle directly in front of the robot and up to the robot's height.

One 3D scan takes about 20 seconds. In order to simulate a frame rate of about 11Hz, a stop-motion technique was applied. That is, after each scan the robot and dynamic objects in the scene were moved by a small distance or angle according to the simulated speed and frame rate.

The data consists of 14 sequences with a total of 2,136 frames. The recorded sequences address robotic tasks such as obstacle detection, self localisation, object detection, and grasping.

4.2. Experimental results

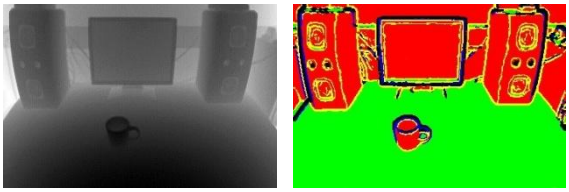
Our approach was implemented in C++ and tested on an Intel Core i5-430M notebook (2.24GHz, 4GB RAM) running 32bit OpenSUSE Linux 11.2. No optimizations such as SSE or multi-threading have been incorporating yet. The total amount of memory allocated for various buffers and lookup tables is slightly less than 3MB.

The segmentation processing chain was applied to the recorded sequences at four different resolutions. Tab. 1 provides the respective average processing times per frame.

Tab. 1. Computation times at different resolutions

	Total point count	Processing time (ms)
360x500	180,000	45.4
360x250	**90,000**	**22.3**
250x160 (bilinear)	40,000	10.8
180x125	22,500	5.7

Fig. 8 shows range images and associated label maps for three tasks a domestic service robots might have: grasping a cup on a table (top row), opening or closing a door (mid row), and detecting the closest obstacles within a relevant height region for obstacle avoidance (bottom row). Fig. 9 shows applications of the extracted features in the context of service robotics.



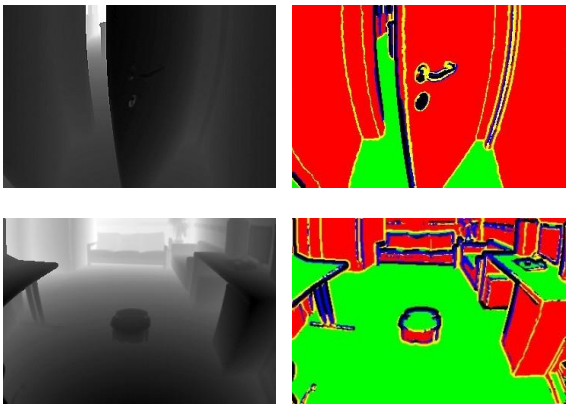


Fig. 8. Range images (left column) and associated label maps for three scenes: objects on a table (top row), door handle (middle row), obstacles on the ground (bottom row). In each label map mixed pixels are black, step and roof edges are blue and yellow, and horizontal and vertical structures are green and red

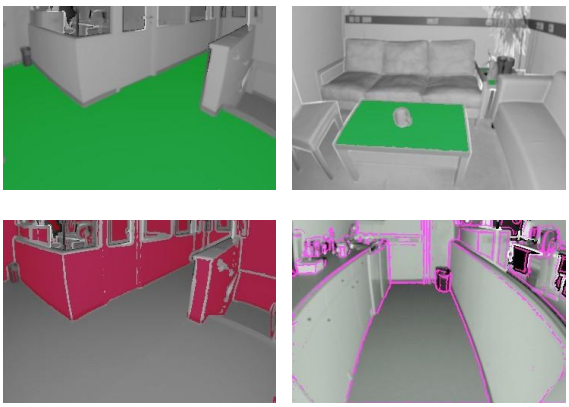


Fig. 9. Examples for applications of the extracted features. Horizontal planar patches for obstacle detection (floor) and a table scene (table top), vertical planar patches (walls) for self localisation and edges reflecting transition between and boundaries of patches

5. Conclusion and Outlook

In this paper we have presented a fast approach to segmentation of dense range images. In the first part of our approach mixed pixels at depth discontinuities in the range images are detected and masked out. The second part extracts step edges, roof edges as well as planar horizontal and vertical structures. We demonstrated that our approach is capable of real-time processing of range images on mainstream notebook hardware.

Although the test data used in this paper stemmed from a tilting laser range finder, the approach can also be applied to range images from other sources such as stereovision or a time-of-flight camera after an adjustment of the algorithms to the specific sensor noise characteristics and geometry.

This work has been done in the context of the EC-funded project TACO[1] (Thielemann et al, 2010) that develops a time-of-flight camera with object detection and foveation. The camera is based on oscillating micro-mirrors that deflect a pulsed laser beam in two dimensions so that a scene can be scanned like with a conventional tilting laser range finder, however, at frame rates comparable to conventional cameras. The pulsed laser beam provides up to one million range measurements per second, so that at 25 frames per second range images at a resolution of 250x160 pixels can be captured. The resolution can be increased or decreased, e.g. 360x250 pixels at around 11 frames per second or 180x125 at 44 frames per second.

The use of micro-mirrors offers flexibility in steering the laser beam across the scene. For example, regions of interest can be scanned at higher spatial resolution or moving objects can be tracked at higher frame rate, which means spatial or temporal foveation. To control the mirror movement and thus the trajectory of the laser beam in response to the contents of the scanned scene requires fast analysis of the range images. Features must be extracted at frame rate and fed into attention algorithms that decide where to foveate in one of the consecutive frames based on a saliency map; the necessary features can be provided by our range image segmentation.

Future work will address improving the quality of feature extraction. Speedups will be incorporated to still achieve real-time performance on computationally weaker embedded processing hardware of the sensor under development.

6. Acknowledgments

The research leading to these results has received funding from the European Union's Seventh Framework Programme (FP7/2007-2013) under grant agreement n°248623.

7. References

Bellon, O. R. P. and L. Silva. 2002. New improvements to range image segmentation by edge detection, *Signal Processing Letters*, IEEE 9(2), 43–45.

Chang, Y. L. and X. Li. 1994. Adaptive image region-growing, *IEEE Transaction on Image Processing*, vol. 3, pp. 868–872.

Gotardo, P.F.U., O.R.P. Bellon and L. Silva. 2003. Range image segmentation by surface extraction using an improved robust estimator, *Computer Vision and Pattern Recognition*, IEEE Computer Society Conference on. Vol. 2. pp. II–33–8.

[1] http://www.taco-project.eu/

Han, F., Z. Tu and S.-C. Zhu. 2004. Range image segmentation by an effective jump-diffusion method, *Pattern Analysis and Machine Intelligence*, IEEE Transactions on 26(9), 1138–1153.

Harati, A., S. Gächter and R. Siegwart. 2007. Fast Range Image Segmentation for Indoor 3D-SLAM, *6th IFAC Symposium on Intelligent Autonomous Vehicles, IAV 2007*, Toulouse, France, September 3–5.

Hijjatoleslami, S. A. and J. Kittler. 1998. Region growing: A new approach, *IEEE Transaction on Image Processing*, vol. 7, pp. 1079–1084.

Hoover, A., G. Jean-Baptiste, X. Y. Jiang, P. J. Flynn, H. Bunke, D. B. Goldof, K. Bowyer, D. W. Eggert, A. Fitzgibbon and R. B. Fisher. 1996. An Experimental Comparison of Range Image Segmentation Algorithms, *IEEE Trans. PAMI*, 18, no.7, pp. 673–689.

Haindl, M. and P. Zid. 2007. Multimodal Range Image Segmentation, *Vision Systems: Segmentation and Pattern Recognition*, Goro Obinata and Ashish Dutta (Ed.), ISBN: 978-3-902613-05-9, InTech.

Jiang, X. and H. Bunke. 1994. Fast segmentation of range images into planar regions by scan line grouping, *Machine Vision and Applications* 7(2), 115–122.

Lee, K. M., P. Meer and R. H. Park. 1998. Robust Adaptive Segmentation of Range Images, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 20, No. 2, pp. 200–205.

Palmer, P. L., H. Dabis and J. Kittler. 1996. A performance measure for boundary detection algorithms, *Computer Vision and Image Understanding*, vol. 63, pp. 476–494.

Sappa, A. D. and M. Devy. 2001. Fast range image segmentation by an edge detection strategy, *3-D Digital Imaging and Modeling*, Third International Conference on. pp. 292–299.

Thielemann, J. T., T. Sandner, S. Schwarzer, U. Cupcic, H. Schumann-Olsen and T. Kirkhus. 2010. TACO: A Three-dimensional Camera with Object Detection and Foveation, *Smarter sensors, easier processing - SAB 2010 workshops*, Paris, France, August 24.

Wang, H. and D. Suter. 2004. MDPE: A very robust estimator for model fitting and range image segmentation, *International Journal of Computer Vision* 59(2), 139–166.

Weingarten, J. W., G. Grüner and R. Siegwart. 2004. Probabilistic plane fitting in 3D and an application to robotic mapping, *IEEE International Conference on Robotics and Automation*, Vol. 1. pp. 927–932.

Zhang, X. and D. Zhao. 1995. Range image segmentation via edges and critical points, *Proc. SPIE*, 2501, no. 3, pp. 1626–1637.