The 2009 IEEE/RSJ International Conference on
Intelligent Robots and Systems
October 11-15, 2009 St. Louis, USA

Extraction of Planar Features from Swissranger SR-3000 Range Images by a Clustering Method Using Normalized Cuts

GuruPrasad M. Hegde and Cang Ye, Senior Member, IEEE

Abstract—**This paper describes a new approach to extract planar features from 3D range data captured by a range imaging sensor—the SwissRanger SR-3000. The focus of this work is to segment vertical and horizontal planes from range images of indoor environments. The method first enhances a range image by using the surface normal information. It then partitions the Normal Enhanced Range Images (NERI) into a number of segments using the Normalized-Cuts (N-Cuts) algorithm. A least-square plane is fit to each segment and the fitting error is used to determine if the segment is planar or not. From the resulting planar segments, each vertical or horizontal segment is labeled based on the normal of its least-square plane. A pair of vertical or horizontal segments is merged if they are neighbors. Through this region growing process, the vertical and horizontal planes are extracted from the range data. The proposed method has a myriad of applications in navigating mobile robots in indoor environments.**

I. Introduction

INDOOR environments commonly consists of regular structures such as, stairways, hallways, and doorways, etc.. To operate efficiently in such conditions it is import for a mobile robot to identify these structures and deal with them. For instance, a tracked robot may use the information to traverse steps and stairways. Extracting and recognizing these structures is also useful in building a symbolic map of the environment. A robot may use these structures as landmarks for navigation and localize itself. Since these indoor structures are often constituted by planar surfaces, efficient planar feature extraction becomes an essential capability for the robot. Pattern recognition algorithm built on the plane extraction method may allow the robot to group the planar surfaces into structures and identify them based on their geometric constituent. For instance, a stairway can be characterized by an occurrence of alternate horizontal (treads) and vertical (raisers) planes and a floor can be characterized by a large horizontal plane. Also, extracting planar segments is an important problem in range data processing, and it serves a number of purposes. First, the information about where two planes intersect in 3D space can be used to extract prominent linear features such as corners in a room. These features are important for registering multiple scans of range data or to register range data with 2D images.

Researchers have addressed the problem of planar feature extraction from range data either in the original input domain (3D point cloud) [1,2,3,4] or by representing the range data as an image [5,6]. Venable and Uijit de Haag [3] propose a so-called histogramming method for planar surface extraction for the SR-3000. The method first divides a range image into a number of sub-images equally. It then fits a least-square plane to the set of 3D data points belonging to each sub-image, and the plane with the smallest fitting error is chosen as a candidate planar feature. A histogram of the distances d from the rest of the data points to the candidate plane is computed. Data points that are closely located at the surrounding of d=0 and d=D in the histogram are classify as points in the candidate plane and points in a parallel plane, respectively. The advantage of the method is its real-time performance. The limitation is that it can not be applied to a scenario where planes have multiple orientations (e.g., perpendicular planes). In addition, the set of data points in a sub-image with the minimum fitting error does not necessarily form a planar surface. Stamos and Allen's method [4] identifies planar structures from 3D range data of a precision laser scanner by dividing the data into $k \times k$ patches and merging the planar patches based on plane-fitting statistics. A patch is classified as a locally planar patch if the plane-fitting error is below a threshold. Two locally planar patches are considered to be in a same planar surface if they have similar orientations and are close in 3D space. The plane-fitting based classification method is sensitive to the threshold value. In addition it is not easy to determine an appropriate patch size—a trade-off between computational cost and the granularity of data segmentation.

In this work we use the SwissRanger SR-3000 imaging sensor [7,8] as we are investigating the plane extraction problem for possible application of navigating a small mobile robot in indoor environments. In this case the SR-3000 is advantageous over a LADAR. It has a much higher data throughput—25344 points per frame and up to 50 frames per second, and is much smaller in size ($50 \times 48 \times 65$ mm^3). Also the SR-3000 works well in featureless environments which is a big advantage over a stereovision system. However, the SR-3000's sensing technology is nascent and its range data has relatively large measurement errors (much bigger than that of a LADAR [9]) due to random noise (e.g., thermal noise, photon short noise) and environmental factors (e.g., surface reflectivity). Previous research efforts [10,11] have demonstrated a proper calibration process may reduce the errors in the SR-3000's range data to certain extent. However, it can not eliminate the errors induced by random noise. In [12] the authors of this paper developed a Singular Value Decomposition (SVD) filter to deal with the noise in the Normal Enhanced Range Image (NERI) of the SR-3000. The SVD filter demonstrates some success in smoothing the surface. However, there is still certain amount of corruption in the

Manuscript received March 1, 2009. This work was supported in part by NASA and the Arkansas Space Grant Consortium under grants UALR18800; by the NASA EPSCoR RID Award, and a matching fund from the Arkansas Science and Technology Authority.

C. Ye is with the Department of Applied Science, University of Arkansas at Little Rock, Little Rock, AR 72204, USA (phone: 501-683-7284; fax: 501-569-8020; e-mail: cxye@ualr.edu).

G. Hegde is with the same department (e-mail: gmhegde@ualr.edu).

NERI. In such a case, a pixel-by-pixel region growing method can not perform segmentation very well as it is susceptible to disturbance in local features. It is required to use a global criterion to segment a NERI. The criterion must take into account both dissimilarity between the segments as well as the total similarity within the segments (i.e., among image pixels).

In this paper we present a new range image segmentation method based on the Normalized Cuts (NC) method [13]. The NC method was originally proposed for the segmentation of intensity images and it uses the total dissimilarity between groups and the total similarity within the groups to partition an image. It may result in inappropriate grouping of pixels in case that an object does not have distinctive dissimilarity from the background. This problem may be alleviated in segmenting a range image since additional metrics, such as the surface normal of the least-square plane to the data points of a segment and the fitting error, may be used to evaluate the correctness of the segmentation.

The reminder of this paper is organized as follows: In the following section we briefly describe briefly the NC method. In section III, we explain our proposed method for extracting planar features from range data. In Section IV, we present experimental results followed by section V where we discuss a recursive method to extract planar pixels from misclassified clusters. The paper is concluded in section VI where we discuss some directions for our future work.

II. Image Segmentation Using Normalized Cuts

A. Image segmentation as a graph partitioning problem

Image segmentation can be modeled as graph partitioning problem. An image is represented as a weighted undirected graph $G=(V,E)$ wherein each pixel is considered as a node V_i and an edge E_{ij} is formed between each pair of nodes. The weight for each edge is recorded in a Pixel Similarity Matrix (PSM) calculated as a function of similarity between each pair of nodes. In partitioning an image into various disjoint sets of pixels or segments $V_1, V_2, V_3, ..., V_m$, the goal is to maximize the similarity of nodes in a subset V_i and minimize the similarity across different sets V_j. For the NC algorithm the optimal bipartition of a graph into two sub-graphs *A* and *B* is the one that minimizes the Ncut value given by:

$$Ncut(A,B)=\frac{cut(A,B)}{assoc(A,V)}+\frac{cut(A,B)}{assoc(B,V)}, \quad (1)$$

where $cut(A,B)=\sum_{(u\in A, v\in B)} w(u,v)$ is the dissimilarity between *A* and *B*, and $w(i,j)$ is the weight calculated as a function of the similarity between nodes i and j. $assoc(A,V)$ is the total connection from nodes in A to all nodes in *V*. $assoc(B,V)$ is defined similarly. From (1) we can see that a high similarity among nodes in *A* and a low similarity across different sets *A* and *B* can be maintained by the minimization process. Given a partition of nodes that separates a graph V into two sets A and B, let x be an $N=|V|$ dimensional indicator vector, $x_i=1$ if node i is in A and -1, otherwise. Let $d_i=\sum_j w(i,j)$ be the total connection from node *i* to all other nodes. With the above definition, $Ncut(A,B)$ in (1) can be calculated. According to [13] an approximate discrete solution to minimize $Ncut(A,B)$ can be obtained by solving the following equation:

$$\min_x Ncut(x)=\min_y \frac{y^T(D-W)y}{y^T Dy}, \quad (2)$$

where $D=diag(d_1,d_2,...,d_n)$, $d_i=\sum_j w(i,j)$, $W=[w_{ij}]$, and $y=(1+x)-(1-x)\sum_{x_i>0} d_i \Big/ \sum_{x_i<0} d_i$. If $y\in\Re$ ($\Re$ is a set of real numbers), then (2) can be minimized by solving the following generalized Eigen value system:

$$(D-W)y=\lambda Dy \quad (3)$$

B. Grouping Algorithm

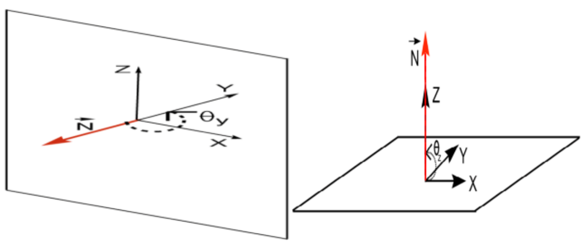
The grouping of pixels in an image ***I*** consists of the following steps:

a) Consider image ***I*** as an undirected graph $G=(V,E)$ and construct a PSM. As stated before, each element of the PSM is the weight of edge $w(i,j)$ and is calculated by $w(i,j)=\exp\frac{-\|F(i)-F(j)\|_2^2}{\sigma_I^2} * \exp\frac{-\|X_{(i)}-X_{(j)}\|_2^2}{\sigma_X^2}$ if $\|X(i)-X(j)\|_2<r$ and $w(i,j)=0$, otherwise. Here, $X(i)$ is the spatial location of node i, $F(i)=I(i)$ is the brightness value of pixel i. It is noted that $w(i,j)=0$ for any pair of nodes i,j that is greater than r pixels apart. The reason for calculating $w(i,j)$ in such a manner is substantiated by the following argument: any two pixels that have similar brightness value and are spatially nearer belong to the same object more possibly than two pixels with different brightness values and are distant from each other.
b) Solve (3) for the Eigenvectors with the smallest Eigen values.
c) Use the Eigen vector with the second smallest Eigen value to bipartition the image by finding the splitting points such that its *Ncut* value is minimized.
d) Recursively re-partition the segments (go to step a)
e) Exit if Ncut value for every segment is over some specified threshold.

III. The Proposed Method

In this work we adopt the method in [14] for range image enhancement. The authors of this paper demonstrate in [12] that the use of surface normal in the SR-3000's range images make the surfaces and edges of an object more distinct. We first construct a tri-band color image where each pixel's RGB values represent the *x*, *y* components of its surface normal and its depth information, respectively. The tri-band image is then converted to a gray image which we call a Normal-Enhanced Range Image (NERI). The proposed segmentation method is divided into three steps. First, the NC algorithm is applied to the NERI and partitions the NERI into a number of segments. A pre-specified number is needed in our current implementation. Second, the least-square plane to the data points in each segment is computed and the plane-fitting statistics are used to label the segments as planar and non-planar. Third, adjacent planar surfaces with the same orientation are merged.

To simplify the description we only consider the extraction of vertical and horizontal planes in 3D space. As shown in Fig. 1 a vertical plane is defined as the one whose normal direction is along the $-Y$ axis (Fig. 1a) and a horizontal plane is the one whose normal direction is along the Z axis (Fig. 1b).



(a) (b)

Fig.1 Diagram of vertical and horizontal planes

For the N segments resulted from the NC method, we need to identify those that best describe either vertical or horizontal planes. To do this we perform a Least-Square Plane (LSP) fit to the data points associated with each of the N segment and calculate the normal direction and the fitting error. Let the normal to the LSP be denoted by $\vec{N} = (n_x, n_y, n_z)$ and the residual of the fit, also known as Plane Fit Error (PFE), is computed by $\Delta = \sqrt{\sum_{k=1}^{P} d_k \Big/ P}$, where P denotes the number of pixels in the segment and d_k is the distance between the k^{th} data point (x_k, y_k, z_k) and the LSP. The LSP is found by minimizing Δ. The minimization can be obtained by the Singular Value Decomposition (SVD) method. First the following matrix is constructed using the data points of the segment:

$$\mathbf{M} = \begin{bmatrix} x_1 - x_0 & y_1 - y_0 & z_1 - z_0 \\ x_2 - x_0 & y_2 - y_0 & z_2 - z_0 \\ . & . & . \\ . & . & . \\ x_p - x_0 & y_p - y_0 & z_p - z_0 \end{bmatrix} \times \begin{bmatrix} x_1 - x_0 & x_2 - x_0 & . & . & x_p - x_0 \\ y_1 - y_0 & y_2 - y_0 & . & . & y_p - y_0 \\ z_1 - z_0 & z_2 - z_0 & . & . & z_p - z_0 \end{bmatrix}$$

where $(x_0, y_0, z_0) = (\frac{1}{P}\sum_{k=1}^{P} x_k, \frac{1}{P}\sum_{k=1}^{P} y_k, \frac{1}{P}\sum_{k=1}^{P} z_k)$ is the centroid of the data points. Then the Eigen values $\lambda_1, \lambda_2, ..., \lambda_p$ of $\mathbf{M}$ and their corresponding eigenvectors are computed. It can be proven that $\vec{N}$ equates the Eigen vector corresponding to the minimum Eigen value $\lambda_{\min} = \min(\lambda_1, \lambda_2, ..., \lambda_p)$ and Δ equates to $\sqrt{\lambda_{\min} \Big/ P}$. The deviation of the normal direction from Y and Z axes are computed by $\theta_y = \cos^{-1}(-n_y)$ and $\theta_z = \cos^{-1}(n_z)$, respectively. The value of the PFE determines whether the segment forms a planar surface while the values of θ_y and θ_z determine if the plane is vertical or a horizontal. To be specific the data points in a segment whose PFE is sufficiently small form a planar surface; and the planar segment is vertical (horizontal) if the value of θ_y (θ_z) is sufficiently close to 0º.

Extracting an entire plane from the scene involves merging of two or more planar segments. In our work, merging is performed only if two segments are neighboring. Figure 2 shows three typical configurations of two close-by segments. The common boundary between the two segments is highlighted in green. Any two segments are considered as neighbors if there exists at least two consecutive common points (i.e., they belong to both segments and are continuous in space) on their boundaries. Thus the segments in Fig. 2a do not qualify as neighbors since they have a single common point whereas the two segments in Fig. 2b or Fig. 2c are considered as neighbors.

(a) (b) (c)

Fig.2 Definition of neighboring segments: for simplicity the segments are drawn as rectangles. (a) two non-neighboring segments, (b) and (c) two neighboring segments.

Our proposed method for extracting vertical and horizontal planes in a range image is as follows:

1) Construct the NERI and apply the NC algorithm to partition the NERI into N segments c_i for $i = 1, ..., N$.
2) The planar segments c_i for $i = 1, ..., N$ that satisfy $\Delta_i < \delta$ are selected to form a set of planar segments $S = \{s_1, s_2, s_3, ..., s_n\}$, i.e., a segment with a PFE smaller than δ is taken as a planar surface. Here $n \le N$ and δ is a suitable threshold.
3) Each $s_j \in S$ is then labeled as vertical or horizontal based on the normal direction of its LSP
4) A pair of vertical or horizontal segments is merged if they are neighbors.
5) Terminate the process when all the neighbors are merged.

IV. Experiments And Results

We have validated our planar surface extraction method through experiments in various indoor environments that contain most commonly occurring structures. As mentioned before our current method requires a pre-specified number of clusters *N*. In all our experiments we use *N*=100 that is bigger than the actual number of planar segments each NERI contains. The reason for choosing such a big N is to ensure a correct segmentation. This can be demonstrated by the example in Fig. 3. The test scene is shown in Fig. 3a and its NERI is depicted in Fig. 3c. Apparently the NERI has 5 segments that are hand-labeled and shown in Fig. 3c. We now apply the NC algorithm to Fig. 3c using the actual number of segments (N = 5). The result is shown in Fig. 3d. We can observe in Fig. 3d that there is a misclassification in cluster 5 that contains two regions with different brightness. Each of the two regions represents a planar surface that is perpendicular to one another in 3D space. To avoid such a misclassification scenario we need to assign N a number that is much bigger than the actual number of segments in a scene.

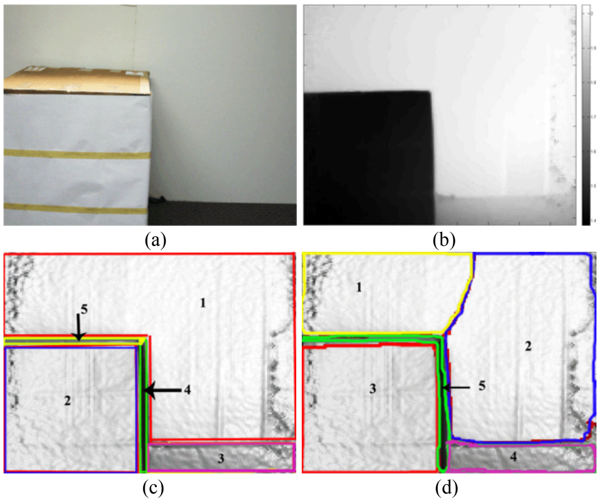


Fig. 3: Misclassification of the NC method using the exact number of segments of a scene: (a) Actual Scene, (b) Raw range image of (a), (c) NERI with labeled segments, (d) Segmentation results of the NC over the NERI.

In all our segmentation results hereafter, we label a vertical and a horizontal plane in blue and green, respectively. In the first experiment we consider an indoor scene with a stairway as depicted in Fig. 4a. The raw range image of the SR-3000 is shown in Fig. 4b and the NERI representation of the range data in Fig. 4c. The NC algorithm partitions the NERI into 100 segments as shown in Fig. 4d. Fig. 4e displays selected planar segments. A token in blue indicates that the corresponding segment belongs to a vertical plane and a token in green means that the segment lies on a horizontal plane. Finally the extracted planes are shown in Fig. 4f. We can see that the majority of the planar surfaces are correctly extracted. We can also observe in Fig. 4f that region P (marked in red) is under-extracted because its adjacent regions (circled in yellow) are misclassified. As we can see from Fig. 4e that pixels inside region A or B are classified as in the same segment by the NC algorithm. However, each of them contains data points on different planar surfaces that are perpendicular to one another in the 3D space. They are considered as non-planar segments due to their large PFE's. It should be noted that the misclassification will not have big impact in recognizing the stairway and guiding the robot. This is because there are enough number of treads and risers indentified and the misclassification occurred at a location far away from the robot. We can also see that there are minor misclassifications at the right or left end of some extracted planar segments. This suggest for future research effort that go beyond the scope of this paper. Fig. 4g renders the segmentation result in a 3D point cloud where the unclassified points are represented in black.

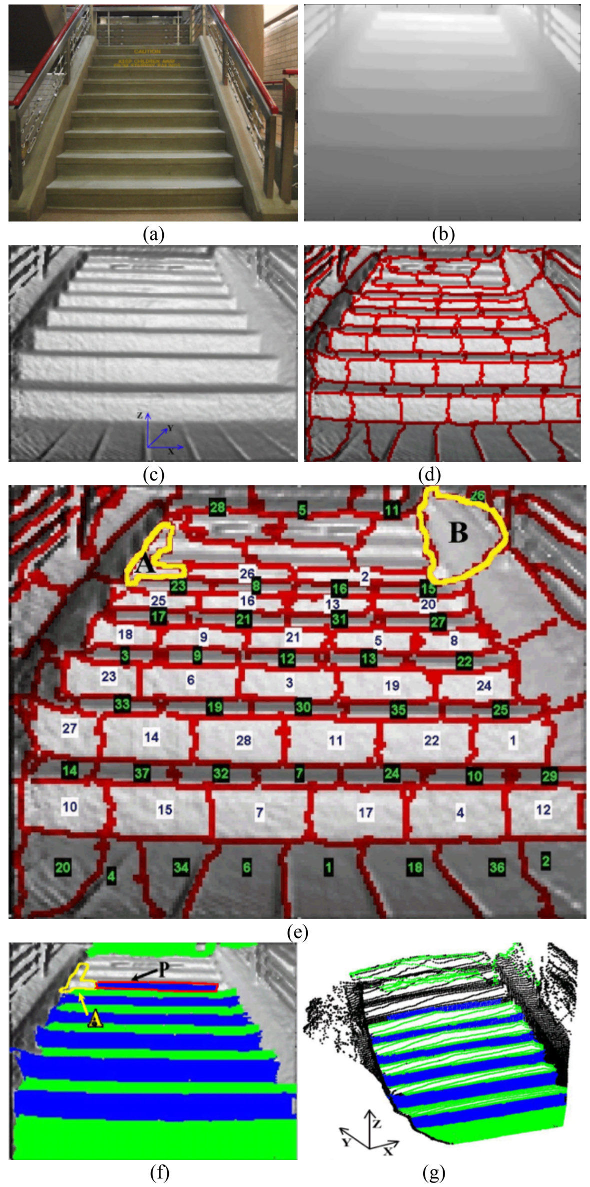


Fig. 4 Segmentation of the scene with a stairway: (a) Actual Scene, (b) Raw range image, (c) NERI of (b), (d) Initial clustering results from the NC, (e) Labeling of vertical and horizontal planar segments, (f) Segmentation results after merging the homogeneous segments in (e), (g) Segmented data shown in a 3D point cloud.

The 2nd experiment shows the plane extraction of a hallway. The result is depicted in Fig. 5. We can see that the floor, door and the wall regions have been properly extracted. Fig. 6 displays the result of our experiment in the lobby in the ETAS building at the University of Arkansas at Little Rock. The result demonstrates a satisfactory planar feature extraction of our proposed method.

V. Discussion

We have seen in the previous section that some segments (mainly in the boundary regions) fail to qualify as planar ones due to misclassification in the initial clustering phase. In this section we put forward a recursive approach that may identify planar pixels from the non-planar segment and hence can extract a plane to its entirety.

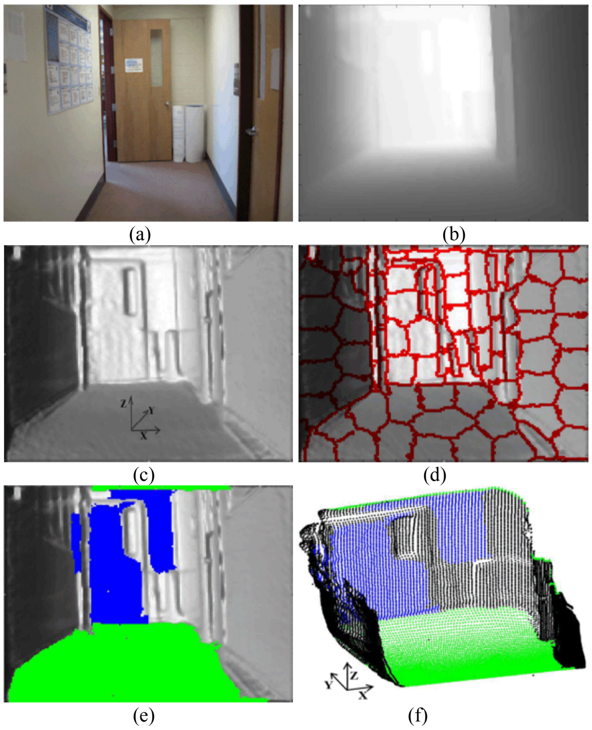


Fig. 5 Segmentation of the scene with a hallway:
(a) Actual Scene, (b) Range image from SR-3000, (c) NERI of (b), (d) Initial clustering results from NCCT, (e) Segmentation results after merging clusters, (g) Extracted points in 3D.

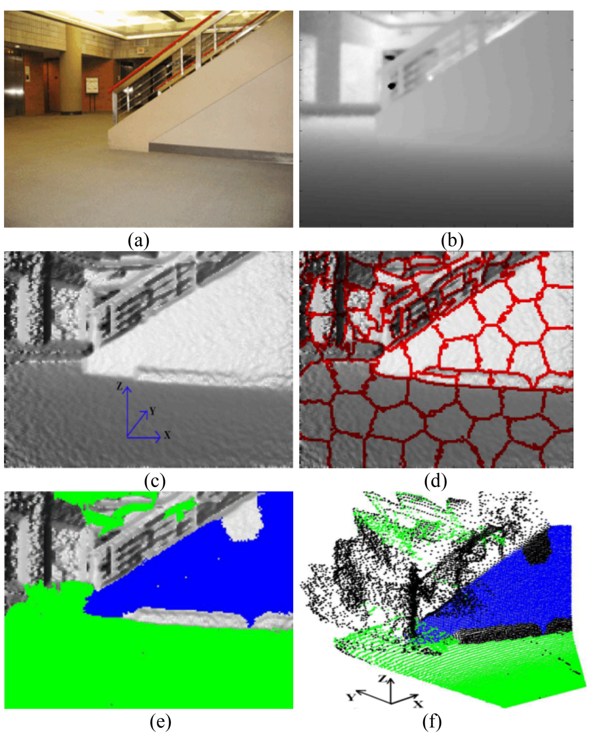


Fig. 6 Segmentation results of the scene with a Lobby:
(a) Actual Scene, (b) Range image from SR-3000, (c) NERI of (b), (d) Initial clustering results from NCCT, (e) Segmentation results after merging clusters, (f) Extracted points in 3D.

Consider Fig. 7a which is a magnified view of region A in Fig.4f. Our objective is to extract the part of A that belongs to the adjacent vertical plane P (Fig. 4f). To achieve this we apply our proposed method recursively as follows:

a) The NC algorithm is applied to A with N=2. This breaks A into two sub-segments
b) For each sub-segment compute the PFE and normal of its LSP.
c) A sub-segment is then merged with P or discarded based on the criterion we set forth in Section IV.
d) If none of the sub-segments is merged with P, the NC is again applied to A with N=N+1, i. e., repeat the process with N=N+1.

In this example, we did not make a merger with N=2. But with N=3 we obtained three sub-segments (C, D, E in Fig. 7a). Through steps b) and c), a larger segment (Fig. 7b) is formed with C joining P. From this we can see that a finer segmentation can be achieved by further splitting the non-planar segments that are adjacent to the extracted segments (blue and green segments in this case).

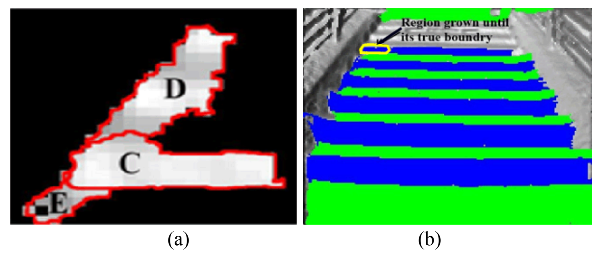


Fig. 7: (a) Enlarged view of a misclassified cluster, (b) Segmentation result after extracting planar region from (a).

Currently, the proposed method is implemented in Matlab using the Normalized Cuts library [15]. As a consequence it is not real-time. Efforts are being made to translate the code entirely into C/C++. We expect that this will significantly reduce the runtime.

VI. Conclusion And Future Work

We have presented a method that may reliably extract the vertical and horizontal planes from the range images captured by the SwissRanger SR-3000. We use a split and merge approach to achieve this. In the proposed method, surface normal information is used to convert a range image into a NERI for image enhancement. To deal with surface normal errors caused by noise in range data we apply the NC algorithm over a NERI to get homogenous segments. They are merged to form larger segments (horizontal and vertical planes) based on the LSP fitting data statistics. Our method works efficiently without a prior knowledge of the number of vertical or horizontal planes in a scene. We also have also demonstrated that under-extraction of planes due to misclassification of non-planar segments can be resolved by further splitting the related segment and apply our method to the sub-segments. We have validated the method's efficacy by real experiments in various indoor environments. Although the method is intended for the segmentation of flat surfaces, it can be adapted to handle non-flat cases as well. One possible approach is to assume that a non-flat surface comprises a number of planar segments with small changes in their orientations (normals). The merging of neighboring segments then takes place based on the rate of change of their normal directions.

It should be noted that we use a gray image for the NERI in order to use the Normalized Cuts library. The mapping of a tri-band image pixel to a NERI pixel is multiple-to-one. If two non-parallel planes happen to have such orientations that their pixels have the same brightness in theirs NERI representations, the distinctness of the planes in the NERI will be very small. This may potentially add difficulty to the N-Cuts method. Fortunately, in this case the intersection of the planes (a straight line segment) has a different normal direction from both planes. This will likely help the N-Cuts method locate correct planar segments. In case that the N-Cuts method fails, we will have to use the tri-band color image representation for NERIs and develop new N-Cuts method to segment the color NERIs. We will carry out more experiment to test this point in our future work. Another direction for future research is to develop method that may find the optimal number of clusters for the N-Cuts method. This might improve the execution time of our method in both splitting and merging phases.

The method proposed in this paper can be employed by a mobile robot for autonomous navigation in indoor environments.

REFERENCES

[1] R. Unnikrishnan and M. Hebert, "Robust extraction of multiple structures from non-uniformly sampled data," in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2003, pp. 1322-1329.

[2] R. Triebel, W. Burgard, and F. Dellaert, "Using hierarchical EM to extract planes from 3d range scans," in *Proc. IEEE International Conference on Robotics and Automation*, 2005, pp. 4437-4442.

[3] V. Don and H. Maarten Uijt de, "Near real-time extraction of planar features from 3d flash-ladar video frames," in *Proc. SPIE,* vol. 6977 of Optical Pattern Recognition, pp. 69770N-69770N-12, 2008.

[4] I. Stamos and P. E. Allen, "3-d model construction using range and image data," in *Proc. IEEE International Conference on Computer Vision and Pattern Recognition,* 2000, pp. 531-536.

[5] A. Bab-Hadiashar and N. Gheissari, "Range image segmentation using surface selection criterion," *IEEE Transactions on Image Processing,* vol. 15, no. 7, pp. 2006-2018, 2006.

[6] A. D. Sappa, "Automatic extraction of planar projections from panoramic range images," in *Proc. 2nd International Symposium on 3D Data Processing, Visualization and Transmission,* 2004, pp. 231-234.

[7] T. Oggier, *et al.*, "An all-solid-state optical range camera for 3D real-time imaging with sub-centimeter depth resolution," in *Proc. SPIE,* 2003, vol. 5249, pp. 534-545.

[8] T. Oggier, B. Büttgen,F. Lustenberger, "SwissRanger SR3000 and first experiences based on miniaturized 3D-TOF Cameras," Swiss Center for Electronics and Microtechnology (CSEM) Technical Report, 2005.

[9] C. Ye and J. Borenstein, "Characterization of a 2-D laser scanner for mobile robot obstacle negotiation," in *Proc. IEEE International Conference on Robotics and Automation*, 2002, pp. 2512-2518.

[10] S. A. Guomundsson, H. Aanaes, and R. Larsen, "Environmental effects on measurement uncertainties of time-of-flight cameras," in *International Symposium on Signals, Circuits and Systems,* 2007, pp. 1-4.

[11] K. Young Min, D. Chan, C. Theobalt, and S. Thrun, "Design and calibration of a multi-view TOF sensor fusion system," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2008, pp. 1-7.

[12] G. M. Hegde and C.Ye, "Swissranger sr-3000 range images enhancement by a singular value decomposition filter," in *Proc. IEEE International Conference on Information and Automation,* 2008, pp. 1626-1631.

[13] S. Jianbo and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 8, pp. 888-905, 2000.

[14] K. Pulli, "Vision methods for an autonomous machine based on range imaging," Master's Thesis, University of Oulu.

[15] http://www.cis.upenn.edu/~jshi/software/