Adaptive Edge Features Estimation for Humanoid Robot Visual Perception

Tianwei Zhang[*] and Yoshihiko Nakamura[*]

[*] Department of Mechino and informatics, The University of Tokyo, Japan

Abstract Lidar scanner is a kind of sensor widely used in robotics visual perception, which provides accurate range data. Scan line grouping is an extremely fast plane segmentation and edge estimation method tailored for 2D Lidar scanners, which obtain 3D environment models by assembling 2D scan lines. In this paper, we propose an adaptive scan line split algorithm to overcome density shift problem of large scale scenes so that the scan line grouping method output more accurate plane segments and edges. The simulated experimental results indicate the proposed method is robust and promising in humanoid robotics applications.

1 Instruction

3D visual perception is an important technology for a number of applications related to robotic perception, wild environment mapping, automatic driving, cultural heritage modelling and architecture. In robotics, it is important for robots to be able to autonomously navigate and localize itself in both known or unknown environments. To fulfil navigation, localization, environment mapping and path planning tasks, various algorithms for different sensors are well studied in recent years (Nguyen et al., 2007).

Compared to other kinds of sensors, Laser scanner has many advantages, such as long range, more accurate range measurement, less noise, high angle resolution and high sampling frequency. However, the time cost of forming 3D point clouds constrains the application of 2D laser scanner in robotics. To obtain 3D space models, 2D laser scan lines need to be assembled due to their rolling angles, this phase may spend several seconds. Therefore, (Jiang and Bunke, 1994) and (Gutmann et al., 2008) applied Ramer-Douglas-Peucker algorithm (RDP) (Douglas and Peucker, 1973) before scan line grouping phase. That is a way to save whole processing time by partly moving point clouds processing to pre-processing.

Different from wheel robot, a humanoid robot can step on or over obstacles, but fall down easily with bad contact conditions. So that the shape and texture of ground and obstacles' surfaces is the main task for visual perception. Many works have been done in plane segments estimation (Okada et al., 2001), surface normals estimation(Holzer et al., 2012), (Bormann et al., 2015) and edge estimation(Asatani et al., 2011). Aiming at real-time humanoid robot continuous locomotion, we apply a modified RDP method before organising point clouds data (PCD) from 2D scan lines, adaptively extract edge point from scan lines and the sensed data is saved as straight line segments and endpoints. After assembling 3D PCD, a region growing method is employed to estimate plane segments by using the straight line segments. the contribution of this paper is that we propose an adaptive RDP method to split polyline segments which improved the accuracy of edge point estimation in both close and far areas.

The rest of the paper is organised as follows. Section 2 introduces related work in visual perception. PCD acquiring and edge point estimation are described in Section 3, experimental results are talked in Section 3 and this paper is concluded in Section 5.

2 Related Works

Platform estimation and surface reconstruction with depth sensors (Marton et al., 2009), (Steinbrucker et al., 2013) are most widely studied families of methods for unknown environment perception. Point normal is one of the essential features of 3D point clouds. Rusu *et al.* introduced a normal estimation framework in (Rusu, 2010), in which point normal is computed by analysing the eigenvectors and eigenvalues of a covariance matrix created by its k-nearest neighbours. Point clouds are captured from a single viewpoint, the orientation of normals are flipped toward viewpoint at the end. This framework has the limitation that the edge and corner points' normals lose their sharp features. Moreover, in (Ioannou et al., 2012), these boundary regions are estimated basing on the difference of estimated normals with different neighbour sizes. Recently, another normal estimation method designed for organized PCD, which supposed the structure of PCD data is fully known, in (Holzer et al., 2012) achieved real-time performance by building integral images of input PCD.

In (Okada et al., 2001), Okada *et al.* used a randomized 3D Hough transform to estimate platforms for stair climbing. Another work of Rusu (Bogdan Rusu et al., 2009) used Random sample consensus (RANSAC) to generate polygons upon point clouds which are represented by small volumes, called cells or voxels. RANSAC is fast but tends to combine small local platforms into big slopes, especially in large clutter scenes. In recent works, Papon *et al.* (Papon et al., 2013) extends an over-segmentation approach to real-time stereo data processing, it cluster voxels with same features, such as voxels' position, colour and normal features, to "supervoxels". Supervoxels decrease PCD size complexity, real-time plane segmentation is achieved in (Papon et al., 2013) by using supervoxels as input. However, this method is designed for dense stereo data. In large scale scenes, the algorithm suffers from slow space dividing.

Jiang and Bunke proposed a plane segment estimation method in (Jiang and Bunke, 1994), which grows platforms from straight lines segments. These straight line segments are extracted as soon as the scan lines are acquired with original RDP algorithm. This plane estimation method is extremely fast since most input points are processed only in line segmentation phase. The sensed data is stored as straight line segments rather than single sampling points. (Gutmann et al., 2008) developed the former scan line grouping method by adding a polyline splitting strategy, which is comparing point numbers rather than compute the distance of all the point set. They also build a height map for humanoid robot walking planning basing on this planar segmentation method. However, the RDP polyline splitting method may result in wrong splitting as the distance grows further. In this paper, we proposed an adaptive threshold setting strategy to deal with the wrong polyline splitting when the sensing distance ranges.

3 Adaptive Straight Line Split

In this paper, the environment information is obtained from a Hokuyo UTM-30LX-EW 2D laser scanner. Comparing to high-frequency RGB-D stereo sensors, Microsoft Kinect, laser scanner has the superiorities of a larger field of view, further detectable range and suitability to bad lighting conditions. The chosen Hokuyo Lidar has 270° field of view, comparing with 57.8° of Kinect. Therefore, the robot can see more features without turning the neck, which is important for real-time mapping and perception. Moreover, depth cameras suffer from their baseline problem. Kinect works from about 0.6 m to 4 m, while, Hokuyo UTM30 works in 0.1 m to 60 m with $\pm 1\%$ error. The baseline limitation makes a big problem for the humanoid robot. If the robot cannot see the staircase or obstacles in front of its foot, the stepping has to be finished without visual guidance. In our case, Hokuyo scanner can help the robot see the stair close to its standing feet accurately.

Hokuyo UTM-30LX-EW returns 1081 points each scan (0.25° resolution) every 25 ms, to achieve dense 3D point clouds, the row data need are assembled according to the tilting angle of the scanner. This assembling

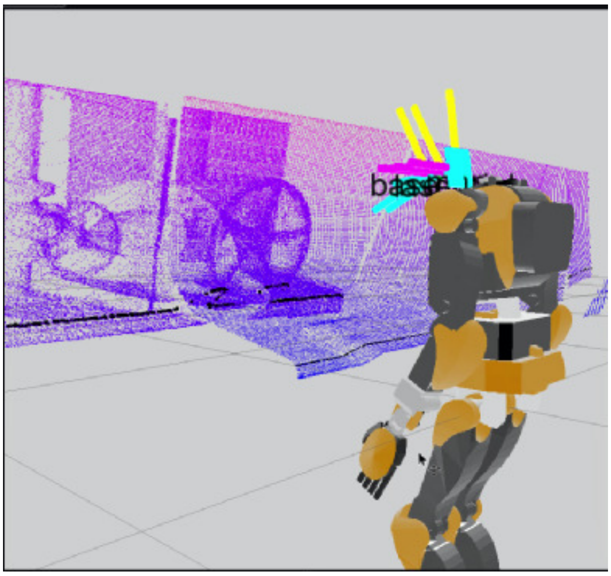
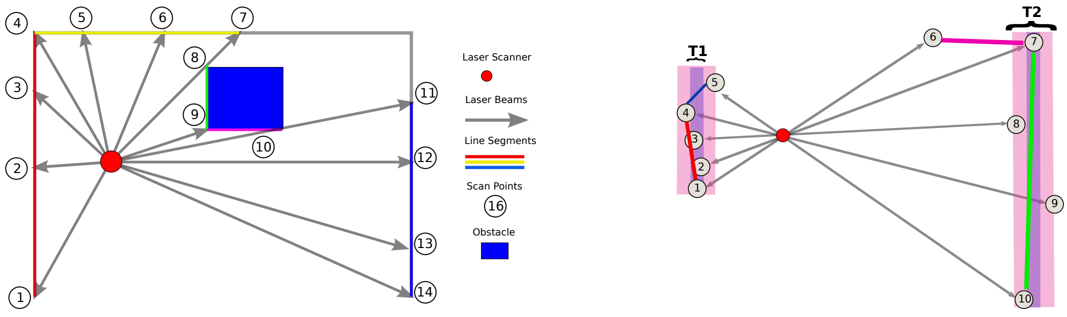


Figure 1. A Hokuyo scanner actuated by a dynamixel MX-64 servo mounted on HRP4's neck. The 3D purple PCD is assembled from 2D scan lines (one shown in black color) according to its pitching angle accurately recorded by the actuator (shown as blue square).

processing limits the frame frequency. In (Osswald et al., 2011), the assembled 3D PCD is noisy with an error up to 5 cm, and the author points out that the error mainly comes from the estimation of the scanner's pose from the joint angle when robot tilts its head. To overcome this problem, our scanner is actuated by dynamixel MX-64 servo actuator which is fixed on the robot neck's joint. As shown in Fig. 1, the robot body keeps static when the scanner is tilting, and the tilting angle with respect to actuator's turning axis is recorded accurately by dynamixel MX-64. Accurate 3D scan can be achieved by assembling scan lines with their turning angle.

One example of the 2D scan output is shown in Figure 2. The RDP method is applied to split the polyline to straight line segments. The split method is shown in Algorithm 1, which processes the scan points in a recursive way: the distance from the query point to the line segment vector (between the start and end points) are computed and compared with a threshold d_T, if it is bigger than the threshold, this line segment will be split on this query point, and the resulted two line segments will be added



(a) (b)

Figure 2. Hokuyo Lidar scanner 2D scan line (a), Output of DP split, the 14 scan points are segmented to 5 line segments shown in different colors: ①④, ④⑦, ⑧⑨, ⑨⑩, ⑪⑭. In (b), T_1 is the split method threshold on the area close to sensor, while T_2 is the threshold on the area far from sensor. ⑧ and ⑨ will be split if we use same threshold.

to the input of RDP.

The problem of the RDP is the setting of the threshold value. In the case of short range stereo vision, this threshold can be set to a fixed value according to the sensor's noise situation. However, in the case of Lidar scanner, the depth date ranges from 0.1 to 30 m, a fix parameter setting may lead to wrong split and wrong segmentation.

To deal with the large range difference of Lidar scanner raw data, an adaptive threshold set as:

$$d_T = \zeta * R \tag{1}$$

in which, R is the depth data set of the query point, ζ is a value according to the sensor property.

4 Experimental results

In the experiments, see Figure 3. a white rectangle board is put in front of the scanner, 1 m, 2m respectively. and (c) and (f) are the wall located around 10 m away. In the upper scenes, the threshold d_T is fixed to 0.03 m, in the lower scenes, d_T is adaptive set as Equation 1, and ζ is set to 0.01. The results show that in short distance scenes split results are similar, however, when the scan distance ascends to 10 m, adaptive threshold result in much

Algorithm 1 RDP split

Input: point cloud data set $L : (L_1, L_2, ...L_n)$, split distance threshold d_T t, distance function D_n, gravity vector $\vec{g}$

Output: edge points P

for each $L_i \in L$ **do**
    for each point $p_i \in L_i$ **do**
        $p_s \leftarrow$ FIRST POINT OF L_i
        $p_e \leftarrow$ END POINT OF L_i
        $d_i \leftarrow$ COMPUTE_DISTANCEp_i *to* $(\vec{p_s, p_e}))$
        if $d_i > d_T$ **then**
            $\mathbb{P} \leftarrow p_i$
            $L \leftarrow push\ back(\vec{p_s, p_i}) and (\vec{p_i, p_e})$
        end if
    end for
end for
return P

fewer edge points on the flat wall surface, therefore, adaptive threshold setting make RDP method robust to distance changing.

5 Conclusion and Future works

In this paper we propose an adaptive threshold setting strategy for scan line grouping plane finding method, and achieved better edge feature estimation performance even when the scan distance range from 0.1 m to 30 m. These experiences suggest that the accurate edge point and straight line segments can be used for efficient plane segmentation and further humanoid applications, like footstep planning and state estimation. In future work, we will develop the method for point clouds registration and robot state estimation applications using straight line vector feature.

Bibliography

Minami Asatani, Shigeki Sugimoto, and Masatoshi Okutomi. Real-time step edge estimation using stereo images for biped robot. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4463–4468. IEEE, 2011.

Radu Bogdan Rusu, Aravind Sundaresan, Benoit Morisset, Kris Hauser, Motilal Agrawal, Jean-Claude Latombe, and Michael Beetz. Leaving flatland: Efficient real-time three-dimensional perception and motion

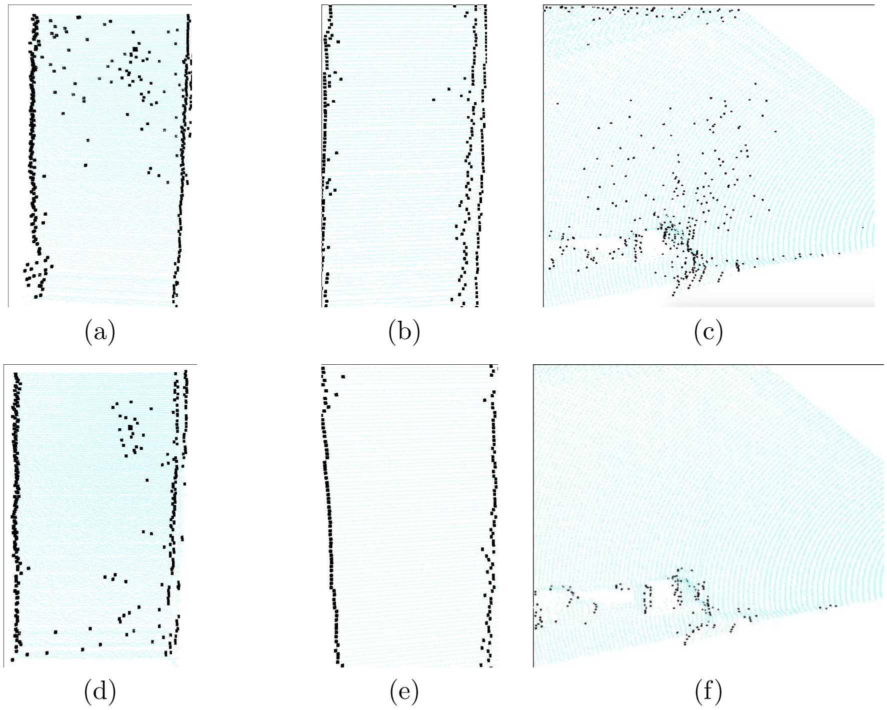


Figure 3. Extracted edge points (a) 1 m distance, $d_T = 0.03$ m (b) 2 m distance, $d_T = 0.03$ m (c) 10 m distance, $d_T = 0.03$ m (d) 1 m distance, adaptive d_T (e) 2 m distance, adaptive d_T (f) 10 m distance, adaptive d_T

planning. *Journal of Field Robotics*, 26(10):841–862, 2009.

Richard Bormann, Joshua Hampp, Martin Hagele, and Markus Vincze. Fast and accurate normal estimation by efficient 3d edge detection. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 3930–3937. IEEE, 2015.

David H Douglas and Thomas K Peucker. Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. *Cartographica: The International Journal for Geographic Information and Geovisualization*, 10(2):112–122, 1973.

Jens-Steffen Gutmann, Masaki Fukuchi, and Masahiro Fujita. 3d perception and environment map generation for humanoid robot navigation. *The International Journal of Robotics Research*, 27(10):1117–1134, 2008.

Stefan Holzer, Radu Bogdan Rusu, M. Dixon, Suat Gedikli, and Nassir Navab. Adaptive neighborhood selection for real-time surface normal estimation from organized point cloud data using integral images. In *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012, Vilamoura, Algarve, Portugal, October 7-12, 2012*, pages 2684–2689. IEEE, 2012.

Yani Ioannou, Babak Taati, Robin Harrap, and Michael Greenspan. Difference of normals as a multi-scale operator in unorganized point clouds. In *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, pages 501–508. IEEE, 2012.

Xiaoyi Jiang and Horst Bunke. Fast segmentation of range images into planar regions by scan line grouping. *Machine vision and applications*, 7(2):115–122, 1994.

Zoltan Csaba Marton, Radu Bogdan Rusu, and Michael Beetz. On fast surface reconstruction methods for large and noisy point clouds. In *Robotics and Automation, 2009. ICRA'09. IEEE International Conference on*, pages 3218–3223. IEEE, 2009.

Viet Nguyen, Stefan Gächter, Agostino Martinelli, Nicola Tomatis, and Roland Siegwart. A comparison of line extraction algorithms using 2d range data for indoor mobile robotics. *Autonomous Robots*, 23(2):97–111, 2007.

Kei Okada, Satoshi Kagami, Masayuki Inaba, and Hirochika Inoue. Plane segment finder: algorithm, implementation and applications. In *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*, volume 2, pages 2120–2125. IEEE, 2001.

Stefan Osswald, J-S Gutmann, Armin Hornung, and Maren Bennewitz. From 3d point clouds to climbing stairs: A comparison of plane segmentation approaches for humanoids. In *Humanoid Robots (Humanoids), 2011 11th IEEE-RAS International Conference on*, pages 93–98. IEEE, 2011.

Jeremie Papon, Alexey Abramov, Markus Schoeler, and Florentin Worgotter. Voxel cloud connectivity segmentation-supervoxels for point clouds. In *Computer Vision and Pattern Recognition (CVPR), 2013 IEEE Conference on*, pages 2027–2034. IEEE, 2013.

Radu Bogdan Rusu. Semantic 3d object maps for everyday manipulation in human living environments. *KI-Künstliche Intelligenz*, 24(4):345–348, 2010.

Frank Steinbrucker, Christian Kerl, and Daniel Cremers. Large-scale multi-resolution surface reconstruction from rgb-d sequences. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3264–3271. IEEE, 2013.