2010 IEEE International Conference on Robotics and Automation
Anchorage Convention District
May 3-8, 2010, Anchorage, Alaska, USA

Scene Understanding in a Large Dynamic Environment through a Laser-based Sensing

Huijing Zhao, Yiming Liu, Xiaolong Zhu, Yipu Zhao, Hongbin Zha

Abstract— **It became a well known technology that a map of complex environment containing low-level geometric primitives (such as laser points) can be generated using a robot with laser scanners. This research is motivated by the need of obtaining semantic knowledge of a large urban outdoor environment after the robot explores and generates a low-level sensing data set. An algorithm is developed with the data represented in a range image, while each pixel can be converted into a 3D coordinate. Using an existing segmentation method that models only geometric homogeneities, the data of a single object of complex geometry, such as people, cars, trees etc., is partitioned into different segments. Such a segmentation result will greatly restrict the capability of object recognition. This research proposes a framework of simultaneous segmentation and classification of range image, where the classification of each segment is conducted based on its geometric properties, and homogeneity of each segment is evaluated conditioned on each object class. Experiments are presented using the data of a large dynamic urban outdoor environment, and performance of the algorithm is evaluated.**

I. Introduction

As the rapid development of sensing and mapping technologies, especially the significant advances in SLAM (Simultaneous Localization And Mapping) using laser scanners (i.e. LiDAR sensors), it became a well known technology that a geometric representation (i.e. map) of an environment can be generated by a robot with multi-modal sensors. Researchers demonstrated successful results even in complex environments, such as urban scenes [6,7,17,24]. However many of the results represent environments directly using the integration of laser points, or low-level geometric primitives such as feature points, planar surfaces and so on. Such a map has limited capacity in representation, as it tells only spatial existence. An operator can easily understand from the data where objects are, and what kinds of objects they are, however a robot can not. In order for a robot to have semantic knowledge of the environment, such as objects, types and their spatial relationships, an automatic technique of converting those low-level map representation into high-level one is important.

A robot vehicle system (POSS-*v*) was developed in our previous research (see Fig.1) [28], where five single-row laser scanners (briefly called "laser scanner") are mounted on a car, for the profiling of the surroundings from different viewpoints and at different directions; a GPS (Global Positioning System)/IMU (Inertial Measurement Unit)-based navigation unit was also used. A localization module was developed by fusing the GPS/IMU navigation unit with a horizontally scanning laser scanner "L1". The problem is formulated as a SLAM (Simultaneous Localization And Mapping) with MODT (Moving Object Detection and Tracking)[27], so that it finds the vehicle pose of both global and local accuracy. It also conducts a 2D mapping and moving object detection/tracking at the same time. With the vehicle pose and sensor geometric parameters, the range data from laser scanners (L2-L5) (see Fig.2) can be geo-referenced into a global coordinate system to provide a three-dimensional representation of the environment (see Fig.3). Such a mapping technology is quite efficient in generating a detailed copy of a dynamic environment at the moment. However, no semantic knowledge is associated to the data at such a low-level representation. This paper focuses on a scene understanding technique, where, given such a low-level representation of the environment, we generate semantic knowledge by segmenting the data into individual objects, meanwhile labeling objects into different classes. Segmentation and classification are formulated in a simultaneous framework. Currently, we consider an off-line procedure after a robot explored and collected low-level data of an unknown environment. In the followings, a literature review is given in section 2, a framework of simultaneous segmentation and classification is addressed in section 3, experimental results are discussed in section 4, followed by conclusions and future studies in section 5.

Laser #	Scan Mode	Major Purpose
L1	Horizontal	SLAM with MODT
L2	Downward	Road geometry
L3	Upward	Objects above the road
L4	Vertical Right	Objects to the right of the car
L5	Vertical Left	Objects to the left of the car

Fig. 1. An intelligent vehicle (POSS-*v) with multiple single-row laser scanners.*

II. Literature Review

Laser scan data represents environmental geometry directly, using the sequence of 2D or 3D laser points. Through a bottom-up procedure, they can be first processed to find

This work is partially supported by the NSFC Grants (No.90920304 and No.60975061)

H.Zhao, Y.Liu, X.Zhu, Y.Zhao and H.Zha are with the State Key Lab of Machine Perception (MOE), Peking University. {zhaohj}@cis.pku.edu.cn

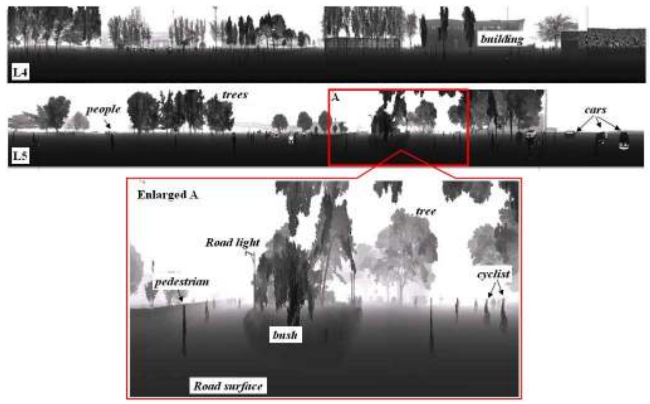

Fig. 2. Range images by laser scanners L4-L5 as well as an enlarged figure reflecting a very dynamic environment. The horizontal axis is the scan line number ($\propto$ time), and the vertical strip represents the range values of each laser scan

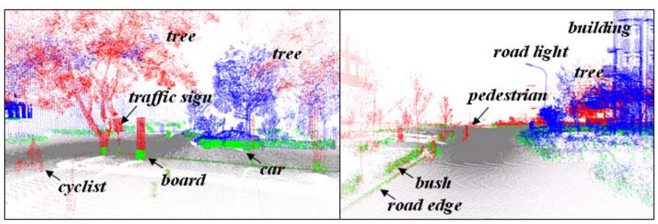

Fig. 3. Geo-referenced laser points, providing a low-level geometric representation of the environment. Colors denote for the data of different laser scanners (L2-Green,L4-Blue,L5-Red). In order for visualization, the laser points on the road are extracted according to their local surface normal and elevation value, and are colored in gray.

data clusters, i.e., laser points in which are most likely to be the measurements of the same objects. They can be then recognized as certain kind of objects, e.g., planar surfaces [9], line feature objects [1], cars [16], natural objects [21] etc. In order to tackle the large number of laser points, [5,18] tessellated the 3D space, and projected laser points to voxels. The sequence of laser points can also be used as a whole, to represent the geometric appearance of local surroundings, a higher level knowledge of the robot's concurrent location, such as a doorway, corridor, room etc. can then be inferred through machine learning techniques[15]. In addition, each individual laser measurement is considered to be dependent on its neighborhood in [23]. Their relationships are modeled using Markov Networks, where labeling of each laser point is influenced by the labeling of others in its local vicinity. [4] also associated image cues with each laser point and varied the probabilistic framework using Conditional Random Field. [20] further extended the method of [23], so that each node in Markov Network corresponds to a data patch, i.e. a superpixel (image patch) with corresponding laser points in successive scans, rather than a laser point. In addition, a recent report can be found in [8], where an air-borne laser scan data is processed to label the small objects, such as post, light, car, etc., in urban environment. The problem is solved by localization, segmentation, representation and classification procedures in a sequential way.

On the other hand, laser scan data can be represented in the form of a range image, where each pixel represents a depth value, and its index corresponds to the sequential order of measurements. Thus, beam origin and angle of each depth value can be retrieved, depth value can then be easily extended to a 3D coordinate. There is a large body of work addressing range image segmentation. A very famous report comparing the major segmentation methods can be found in [11]. Many of the methods are motivated by the needs for recognizing industry parts [13] or registering the data taken at different locations [26]. These works always assume simple or well-defined object geometry. There are still few research works processing range images of real world scenes. [10] considers a real-world indoor and outdoor scene by modeling the man-made objects using planes and conics, modeling free-form objects using splines, and modeling trees using 3D histogram; segmentation and model fitting for each segment is formulated in a data-driven Markov Chain Monte Carlo procedure.

In addition, image segmentation and semantic interpretation have been studied extensively in the field of computer vision. As the data form of a range image is consistent with that of a visual image, many methods developed in the field of computer vision are of great reference for the processing of range image [2,14,19,25].

In this research, we choose the form of range image as an interface for data representation, while, estimations are conducted in both 2D (i.e. the coordinates in the frame of range image, representing a spatial continuity relative to scanning order and view points) and 3D (i.e. the coordinates in a world coordinate system, representing an absolute spatial geometry). As this research is motivated by the need of generating a semantic map of a large urban outdoor environment, we need to consider a scene that contains many kinds of objects, such as buildings, roads, trees, bushes, people, cars, etc., which have different scales in 3D space, with different geometric characteristics (see Fig.4). We refer to the researches in [2,14,19,25] that generate unified frameworks for the segmentation and recognition problems in facing a complex scene. In addition, the robot may explore a large environment, yielding a large among of range data; thus, efficient data processing is of great importance. Our method first computes super-segments using the bottom-up heuristics, such as scan-line segments and edge points at different scales, then merges the super-segments, considering both modeling costs and classification probabilities.

III. A Framework of Simultaneous Segmentation and Classification

A. Geometrical homogeneity vs. Contextual homogeneity

Segmentation is making a partition in data, where in each partition cell (i.e., segment), data has the property of certain homogeneity. Recognition is associating semantic knowledge with each segment, where the problem gets difficult if a segment provides only partial knowledge of the object. Such problems might occur in complex scenes when using the

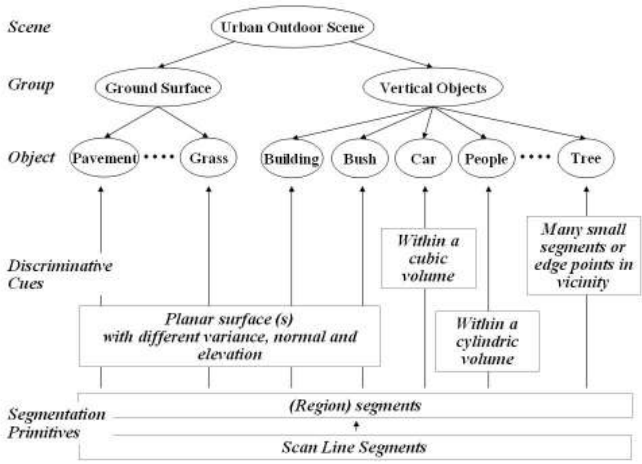

Fig. 4. A hierarchical structure of urban outdoor scene.

geometric primitive such as a planar/ conic/ quadric surface or a spline, to estimate the homogeneity of each segment. For example, a car might contain a number of surfaces; a plane-based segmentation might divide the data of a single car into a number of surfaces. The same problem occurs with people and other complex-shaped objects, which makes recognition an extremely difficult problem.

In this research, we specify homogeneity that data in each segment should belong to the same object(i.e., contextual homogeneity). However, it is difficult to define a unified contextual homogeneity for all kinds of objects, as they have different geometric properties. We consider a normal urban scene containing objects such as roads, buildings, trees, cars, people etc.. The objects can be broadly divided into two groups, i.e., ground surface and vertical objects. Some discriminative characteristics are also summarized in Fig.4. We consider that large scale objects, such as building surfaces, roads and bushes can be represented using one or a group of large planar surfaces with different surface normal, variance and elevation cues. The segment of a car or a person might be composed of a number of small surfaces, while basically, their data in 3D space can be restricted within a limited cubic or cylindric volume, respectively. It is difficult to find a geometric model to describe tree data, however normally, there will be many small segments or edge points in their neighborhood that are defined in 2D range image.

B. Implementation details

As a single-row laser scanner measures the environment in a mode of scan-line by scan-line. A scan-line measurement on a planar object, such as the surface of a building, road etc., can be modeled using a straight line segment. The measurement on a free form object, such as a tree, will form many small line segments or irregular points. It is a straightforward consideration to use scan line segments as the segmentation primitives, instead of laser points, so as to improve computation efficiency [12]. Here, we refer to [22] and their source codes to extract straight line segments from each scan line.

We explore a framework that first split the range image into super segments, then merge the segments belonging to the same objects. In split procedure, a region growing is first conducted to merge the scan line segments according to their planarity; a region growing is then conducted to merge the irregular points and isolated scan line segments according to their spatial connectivity. There is a requirement to the result: each super segment could be a part of an object, but it should not be a mixture of different objects, so that strict criteria are used in the above region growings. In merge procedure, a key problem is to define a model that evaluates contextual homogeneity of the data segments of multi-class objects. In this research, the super segments of the same object are merged into one unit, meanwhile classified into different classes. They are formulated in a simultaneous framework as discussed below.

C. Probabilistic model

Let L denotes the set of object classes, i.e., $L = \{l|building, road, tree, person, car, ...\}$, and $y = l \in L$ denotes the label of a segment. If we know the object type y, we can previously train a model. With the model and data cue I, we can estimate the likelihood of a segment s by $P(s|y, I)$. If we already know a segment s, we can estimate the probability of object type $P(y|s, I)$ using a previously trained classifier. However, if we do not have any a prior knowledge of the object type and data partition, we need to guess these simultaneously, i.e., $P(s, y|I)$. In this research, we generate a unified framework to model the simultaneous segmentation and classification problems.

Let s_i and s_j be a pair of neighboring segments with the label of y_i and y_j, respectively. Let s_{i+j} denote the union of segments s_i and s_j, and y_{i+j} for its label. The probability for s_i merge with s_j, i.e. $P(s_{i+j}|I)$, can be estimated as follows.

$$P(s_{i+j}|I) = \sum_{l \in L} P(y_i = l|I) P(y_j = l|I) P(s_{i+j}|y_{i+j} = l, I) \quad (1)$$

here, $P(s_{i+j}|y_{i+j} = l, I)$ estimates the likelihood, where given the knowledge of object type $y_{i+j} = l$ and data cues I, the probability that s_i and s_j be the measurements to a single object, i.e. s_i and s_j be merged to s_{i+j}. $P(y_i = l|I)$ and $P(y_j = l|I)$ evaluate the probabilities of s_i and s_j be labeled to l, respectively, given the data cues I. As a segment can be represented as a composite of scan lines (omitting the irregular points that do not belong to any line segments) or a cloud of laser points, we define the estimation of $P(y_i = l|I)$ as follows.

$$P(y_i = l|I) = \frac{1}{Z} P(Y_i = l|I) \prod_k P(y_i^{(k)} = l|I) \quad (2)$$

where, $y_i^{(k)}$ is the label of the kth line segment, Y_i is the label of the laser point cloud. $P(y_i^{(k)} = l|I)$ evaluates the probability of each line segment ($yk_i = y_i^{(k)}$) be labeled to l, based on the cues extracted from each individual line segment, such as length, height, direction of the line segment, and scattering property of laser points on line segment. $P(Y_i = l|I)$ evaluates the probability of laser point cloud Y_i be labeled to l, based on the properties of the entire cloud, such as spatial layouts, size, shape etc.. Z is a normalization factor that is calculated as follows.

$$Z = \sum_{l \in L} P(Y_i = l|I) \cdot \prod_k P(y_i^{(k)} = l|I) \qquad (3)$$

So that in calculating the probability for merging segment s_i and s_j, we need to find analytic solutions for the following estimations.

- $P(yk|I)$: the probability of classifying a scan line segment yk, given data cues I.
- $P(Y|I)$: the probability of classifying a scan line segment union Y, given data cues I.
- $P(s|y, I)$: the likelihood that given object label y and data cues I, range segment s be a measurement to a single object.

We train the first two classifiers using machine learning techniques, and define the third one based on the a prior knowledge of a certain kind of object. We detail the methods below.

IV. Training and Definition to Classifiers

A. Training samples

An interactive tool is developed to generate training data samples. A range image is first processed to find local surface normal for each laser point, edge point that has discontinuous change in range value, scan line segments, super segments, etc.. These results are compared by an operator to discriminate the boundaries of individual objects. For each object, the operator will manually draw boundary and assign a label. The software will output the scan line segments that are inside the boundary, as well as the label, for training classifier $P(yk|I)$. The software will also output all laser points within the boundary, as well as the label, for training classifier $P(Y|I)$. The manually labeled data are used as true values for evaluation too.

B. Classification of a scan line segment

Given a scan line segment with a set of data cues $I = d_1, d_2, ...$, we want to estimate the probability of the line segment be labeled to $yk = l$, i.v., $P(yk = l|d_1, d_2, ...)$. It can be further extended according to Bays' rule.

$$P(yk = l|d_1, d_2, ...) = P(yk = l) \prod_i P(d_i|yk = l) \qquad (4)$$

where, $P(yk = l)$ is a prior knowledge of a scan line segment be labeled to $yk = l$. It is initialized as a equal distribution, and updated after each iteration according to the percentages of data labeling. $P(d_i|yk = l)$ is the likelihood measure (denoted by $\lambda_l^{yk}(d_i)$), when given a label $yk = l$,

TABLE I

Data cues of a scan line segment

Feature	Definition
d_1	length of the scan line segment
d_2	maximal height value
d_3	minimal height value
d_4	Z factor of the directional vector
d_5	mean of line regression
d_6	variance of line regression

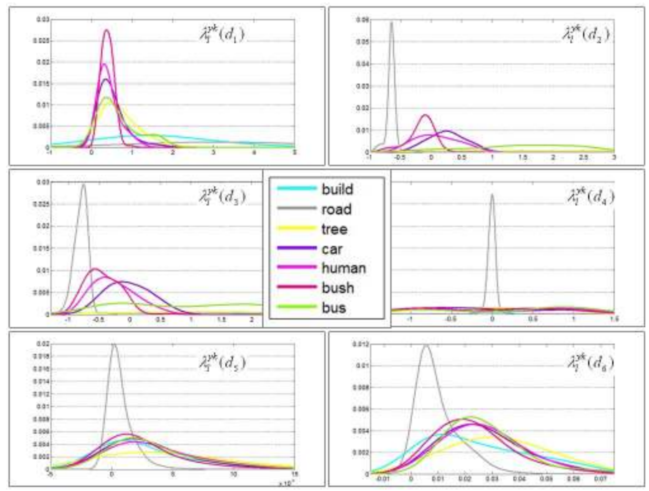

Fig. 5. The likelihood measures for the classification of line segments.

the probability d_i of the scan line segment be observed. The data cues extracted from each scan line segment are listed in Table 1.

For any pair of data cue d_i and object label yk, a likelihood measure $\lambda_l^{yk}(d_i)$ is trained using a set of manually labeled scan line segments. A histogram is first generated on data samples, then Gaussian fittings are conducted on each distinctive picks, followed by a normalization so that integration of the graph be 1. The likelihood measures trained in this research are shown in Fig.5.

C. Classification of a cloud of laser points

Data cues extracted from a cloud of laser points are listed in Table 2. They are different with those in scan line classification, as the classification of a laser point cloud does not evaluate the properties of each individual laser point, but treats them as a whole. The likelihood measures $\lambda_l^Y(d_i)$ of a certain pair of data cue and object label are generated and demonstrated in Fig.6. However, through experimental study, we found that the results of classifying a cloud of laser points using the method described in the previous section is less good than that of a scan line segment. One reason for this might be the limited and unbalanced number of sample data (see Table 3). In classifying a cloud of laser points, we finally use an off-the-shelf method, LIBSVM [3].

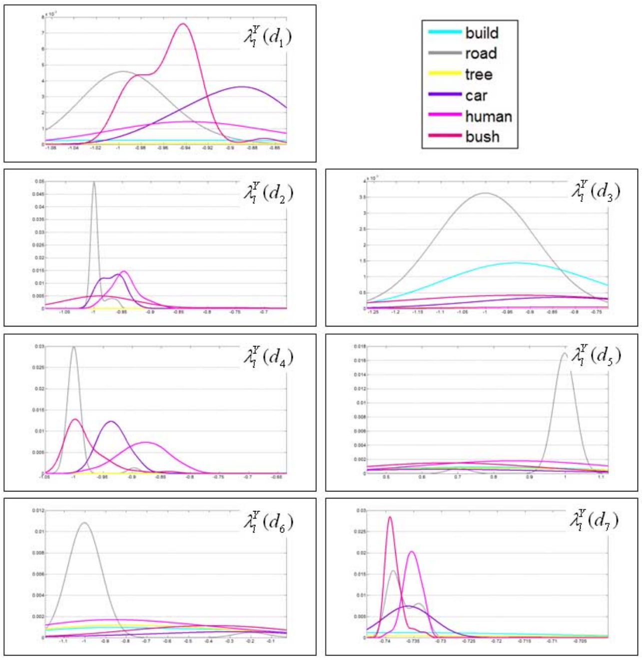

Fig. 6. Likelihood measures for classifying a cloud of laser points.

TABLE II

DATA CUES EXTRACTED FROM A CLOUD OF LASER POINTS

Feature	Definition
d_1	minimal height value
d_2	maximal height value
d_3	ratio of boundary point number vs total point number
d_4	mean of height values
d_5	variance of a histogram distribution on normal vectors
d_6	major picks of a histogram distribution on normal vectors
d_7	ratio of width vs. length

D. Likelihood evaluation of a range segment with known object class

Estimation of $P(s|y, I)$, i.v., given object label y and data cues I, the probability that a range segment s be a measurement to a single object, is different with the previously discussed classifiers. Here, we know what kind of object it is, so that an evaluation based on a prior knowledge to the object model is required. We give the following definitions based on experiences.

- $y = building\ or\ road$:

Normally only a partial surface of building or road are measured, which could be modeled using a planar surface, with a certain volume (e.g. $\varepsilon = \pm 20cm$) representing the errors in laser range measurements and modeling generalization. Let S denotes the total number of laser points in the range segment, N for the number of laser points within the volume, and α be the angle between the surface normal and a vertical normal vector $(0, 0, 1)^T$. We define

$$P(s|y = building, I) \propto \frac{N * sin\alpha}{S} \tag{5}$$

TABLE III

THE NUMBER OF SAMPLE DATA IN TRAINING CLASSIFIERS

Class	Scan line segments	Clouds of laser points
Building	9394	96
Road	10714	23
Tree	4122	148
Car	6080	41
People	394	120
Bush	1176	39
Bus	253	1
Total	32133	468

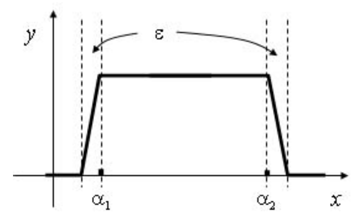

Fig. 7. A step function for likelihood evaluation

$$P(s|y = road, I) \propto \frac{N * cos\alpha}{S} \tag{6}$$

- $y = car$:

Normally data of a car can be restricted within a cube, so that we define τ_W, τ_L and τ_H according to the largest width, length and height of a normal car. We use a cubic model to fit on the laser points of the segment and obtain a width (W_f), a length (L_f), and a height (H_f). A step function is defined (see Fig.7)

$$y = f(x, \alpha_1, \alpha_2), \text{ and} \tag{7}$$

$$\int_{max(\alpha_1-\varepsilon,0)}^{\alpha_2+\varepsilon} ydx = 1 \tag{8}$$

We define

$$P(s|y = car, I) \propto f(W_f, 0, \tau_W) * f(L_f, 0, \tau_L) * f(H_f, 0, \tau_H) \tag{9}$$

- $y = bus$: similar with the evaluation of a car.
- $y = person$:

Normally data of a person can be restricted within a cylinder, so that we define a radius (τ_r) and a height threshold (τ_h) according to the size of a normal person. We use a cylindrical model to fit on the laser points of the segment and obtain a radius (r_f) and a height (h_f).We define

$$P(s|y = person, I) \propto \frac{\tau_r}{max(\tau_r, r_f)} * \frac{\tau_h}{max(\tau_h, h_f)} \tag{10}$$

- $y = tree$:

Normally the range segment of a tree consists of many small line segments and edge points. Let S denote the total

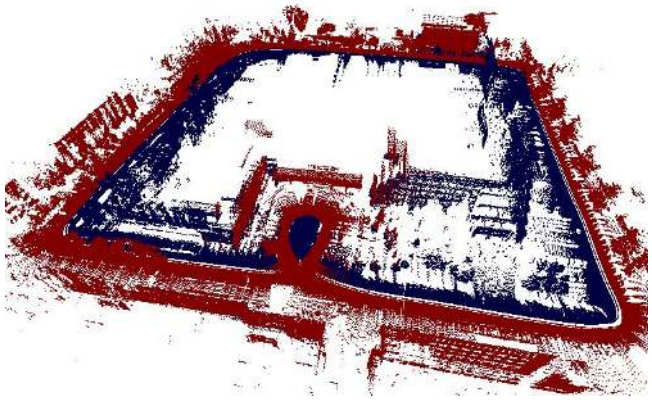

Fig. 8. A 3D view of integrated laser points measured by the laser scanners L4 and L5.

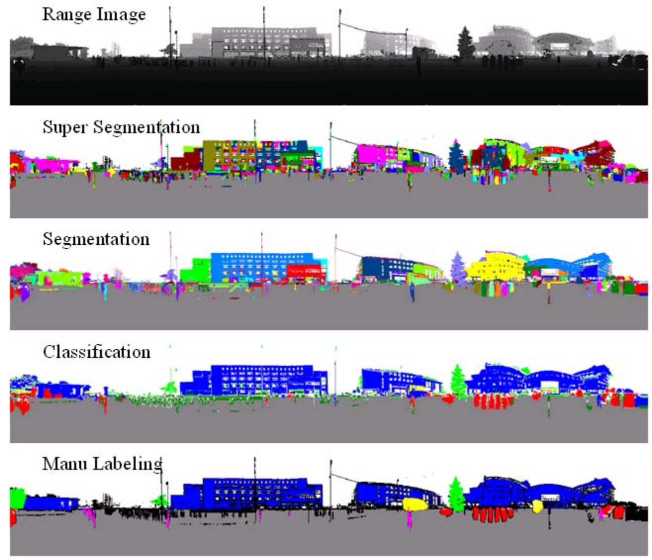

Fig. 9. A comparison of segmentation and classification results.

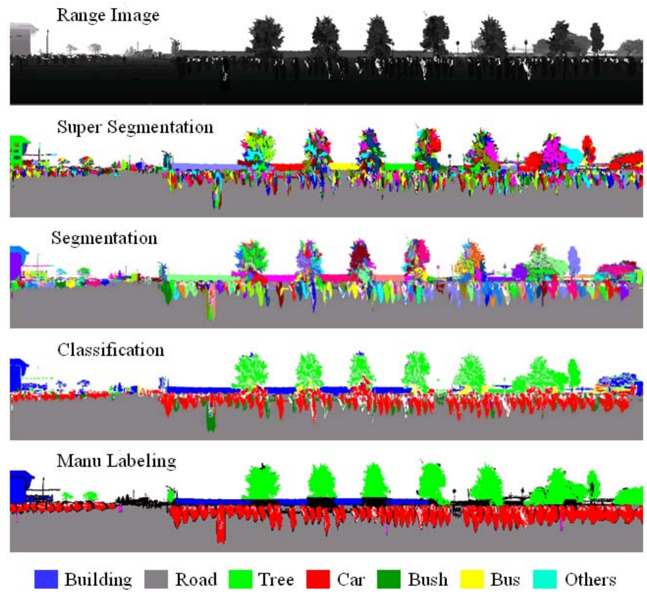

Fig. 10. A comparison of segmentation and classification results.

number of laser points of the range segment, E for the number of laser points on scan line segments. We define

$$P(s|y = tree, I) \propto \frac{E}{S} \tag{11}$$

- $y = bush$:

We also use a cubic model to fit on the bush data. The method is similar to that used for a car or bus.

V. Experimental Results

We present results of an experiment that were taken placed in the campus of Peking Univ. The vehicle run a course about 1km around a campus building. Streams of range images can be found in Fig.2, Fig.9 and Fig.10, a 3D view of the integrated laser points measured by the laser scanners L4 and L5 (refer to sensor layout in Fig.1) is demonstrated in Fig.8. Colors denote for different sensor data. However, no semantic knowledge is associated to the data at such a low-level map representation. We manually labeled the data on range images as shown in Fig.9 and Fig.10 using the method of extracting training samples. The labeled data of laser scanner L4 are used in training classifiers, while those of L5 are used in examining the automated processing results.

Using the method developed in this research, range image is partitioned into segments, meanwhile, labels representing object types are associated to each individual segment. Two sets of results are presented in Fig.9 and Fig.10. In order for comparison, each contains a view of range image, the super segments after split procedure, a segmentation and a classification result after merge procedure, and a manually labeled result as the ground truth. On the other hand, a view of laser points colored on object types is demonstrated in Fig.11, compared with a view of manually labeled result in Fig.12. Definition to color legend is consistent with those in Fig.9 and Fig.10, where blue for building, gray for road, green for tree, red for car, dark green for bush, yellow for bus, water blue. In addition, the black point in manually labeled result means that the data is not labeled, which happens in confusing zone that difficult for an operator to interpret; the white point in classification result means that the data is not classified, which happens for irregular or edge point. From the results, it can be found that the algorithm has good performance for buildings, trees and cars. However, some of the pedestrians are mis-classified into bushes, still some part of tree are mis-classified into buildings. The algorithms are to be improved through future works.

VI. Conclusion and Future Studies

This research is motivated by the need for obtaining semantic knowledge of a large urban outdoor environment after a robot explores and generates a map consisting of a low-level sensing dataset. An algorithm is developed in the representation of range image, while data are processed in both 2D and 3D coordinates. A framework of simultaneous segmentation and classification is developed in this research, where classification of each segment is conducted based on its geometric properties, while homogeneity in each segment is evaluated conditioned on object class. We presented experimental results using the data of a large dynamic urban outdoor environment, and evaluated the performance of the algorithm. All the laser scan data, training samples and processing results in this research will be opened at http://poss.pku.edu.cn.

Future studies will be addressed in extending object types in classification, and improving robustness in the processing of small scale objects.

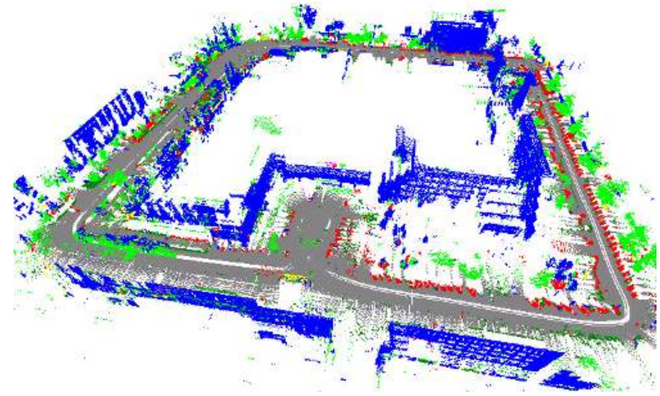

Fig. 11. The annotated laser points on automated segmentation and classification result.

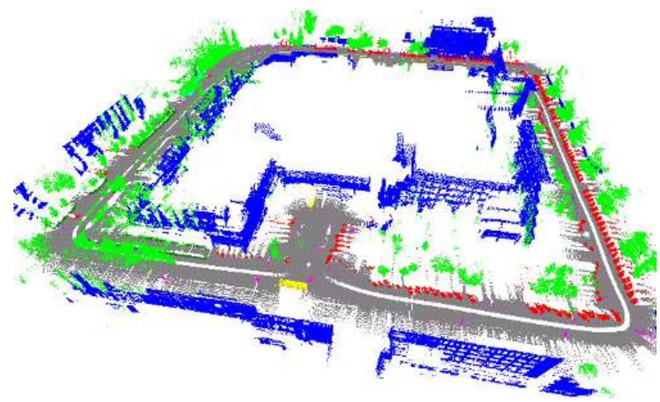

Fig. 12. The annotated laser points on manually labeled result.

REFERENCES

[1] Althaus, P., Christensen, H., "Behavior coordination in structured environments" *Advanced Robotics*, 17(7):657-674, 2003.

[2] Borenstein, E., Ullman, S., "Combined top-down/bottom-up segmentation" *IEEE Trans. Pattern Analysis and Machine Intelligence*, 30(12):2109-2125, 2008.

[3] Chang, C.C., Lin, C.J., "LIBSVM : a library for support vector machines", 2001. Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm

[4] Douillard, B., Fox, D., Ramos, F.T., "Laser and vision based outdoor object mapping" *Proc. Robotics: Science and Systems*, 2008.

[5] Douillard, B., "Vision and laser based classification in urban environments" *PhD thesis, Univ. of Sydney*, 2009.

[6] Fruh, C., Zakhor, A., "3D model generation for cities using aerial photographs and ground level laser scans" *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2001.

[7] Georgiev, A., Allen, P. "Localization methods for a mobile robot in urban environments" *IEEE Trans. Robotics*, 20(5):851-864, 2004.

[8] Golovinskiy, A. Kim, V.G., Funkhouser, T., "Shape-based Recognition of 3D Point Clouds in Urban Environments" *IEEE Int. Conf. on Computer Vision*, 2009.

[9] Hahnel, D., Burgard, W., Thrun, S., "Learning compact 3d models of indoor and outdoor environments with a mobile robot" *Robotics and Autonomous Systems*, 44:15-27, 2003.

[10] Han, F., Tu, Z., Zhu, S.C, "Range image segmentation by an effective jump-diffusion method" *IEEE Trans. Pattern Analysis and Machine Intelligence*, 26(9):1138-1153, 2004.

[11] Hoover, A., Jean-Baptiste, G., Jiang, X., Flynn, P.J., Bunke, H., Goldgof, D., Bowyer, K., "A comparison of range image segmentation algorithms" *IEEE Trans. Pattern Analysis and Machine Intelligence*, 18(7):673-689, 1996.

[12] Jiang, X., Bunke, H., "Fast segmentation of range images into planar regions by scan line grouping" *Machine Vision and Application*, 7:115-122, 1994.

[13] Katsoulas, D., Bastidas, C.C., Kosmopoulos, D., "Superquadic segmentation in range image via fusion of region and boundary information" *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(5):781-795, 2008.

[14] Malisiewicz, T., Efros, A.A., "Recognition by Association via Learning Per-exemplar Distances" *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2008.

[15] Martinez-Mozos, O., Stachniss, C., Burgard, W. "Supervised Learning of Places from Range Data using AdaBoost" *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 1742-1747, 2005.

[16] Mendes, A., Nunes, U., "Situation-based multi-target detection and tracking with laserscanner in outdoor semi-structured environment" *Proc. Conf. on Intelligent Robots and Systems*, 2004.

[17] Nuchter, A., Surmann, H., Hertzberg, J., "Planning robot motion for 3d digitization of indoor environment" *Proc. 11th Int. Conf. on Advanced Robotics (ICAR)*, 2003.

[18] Nuchter, A., Lingemann, K., Hertzberg, J., Surmann, H., "6D SLAM - 3D mapping outdoor environments" *J. Field Robotics*, 24(8-9):699-722, 2009.

[19] Porway, J., Wange, K., Zhu, S.C., "A hierarchical and contextual model for aerial image understanding" *Proc. Computer Vision and Pattern Recognition (CVPR)*, 2008.

[20] Posner, I., Cummins, M., Newman, P., "A generative framework for fast urban labeling using spatial and temporal context" *Autonomous Robots*, 26(2-3):153-170. April 2009.

[21] Posner, I., Cummins, M., Newman, P., "Online generation of scene descriptions in urban environments" *Robotics and Autonomous Systems*, 56(11):901 - 914. 2008.

[22] Rosin, P.L., West, G.A.W., "Nonparametric segmentation of curves into various representations" *IEEE Trans. Pattern Analysis and Machine Intelligence*, 17:1140-1153, 1995. Software available at http://users.cs.cf.ac.uk/Paul.Rosin

[23] Triebel, R., Kersting, K., Burgard, W., "Robust 3D scan point classification using associative Markov Networks" *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2006.

[24] Thrun, S., Burgard, W., Fox, D., "A real-time algorithm for mobile robot mapping with applications to multi-robot and 3D mapping" *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2000.

[25] Tu, Z., Chen, X., Yuille, A., Zhu, S.C., "Image parsing: unifying segmentation, detection and recognition" *Int. J. Computer Vision*, 63(2):113-140, 2005.

[26] Weingarten, J., Siegwart, R., "EKF-based 3D SLAM for structured environment" *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, 2089-2094, 2005.

[27] Zhao, H., Chiba, M., Shibasaki, R., Shao, X., Cui, J., Zha, H., "SLAM in a Dyanmic Large Outdoor Environment using a Laser Scanner" *IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2008.

[28] Zhao, H., Xiong, L., Jiao, Z., Cui, J., Zha, H., "Sensor alignment towards an omni-directional measurement using an intelligent vehicle" *IEEE Intelligent Vehicle Symposium*, 292-298, 2009.