Performance of Global Descriptors for Velodyne-based Urban Object Recognition

Tongtong Chen[1] and Bin Dai[1] and Daxue Liu[1] and Jinze Song[2]

***Abstract*— Object Recognition is an essential component for Autonomous Land Vehicle (ALV) navigation in urban environments. This paper presents a thorough evaluation of the performance of some state of the art global descriptors on the public Sydney Urban Objects Dataset[1], which was collected in the Central Business District of Sydney. These descriptors are Bounding Box descriptor, Histogram of Local Point Level descriptor, Hierarchy descriptor, and Spin Image (SI). We also propose a novel Global Fourier Histogram (GFH) descriptor. Experimental results on the public data set show that GFH descriptor turns out to be one of the best global descriptors for the object recognition in urban environments, and the results on the data collected by our own ALV in urban environments also demonstrate its usefulness.**

I. INTRODUCTION

Navigation in an unknown urban environment is a challenging task for an ALV. It is a hot research topic in the areas related to perception, path planning, and control. The research outcomes of this topic have many applications. For instance, an ALV can not only be used in military, where tasks are too risky for soldiers, it can also help decrease fatalities caused by traffic accidents. Since there are so many traffic participants and public facilities in urban environments, object recognition, which is a high-level task in scene understanding, is very essential to ALVs both for local path planning and position estimation, especially in the dynamic environments. For example, local path planning can incorporate prior knowledge of typical behavior of a certain known dynamic object, i.e. pedestrian, vehicle, cyclist et al. Fig. 1 shows a typical dynamic urban scene.

We are particularly interested in a object-wise recognition in this paper. For every scan (the set of points collected during the Velodyne LIDAR spins once), we assume that the ground segmentation and object cluster steps ([1], [2], [3], [4], [5], [6]) have already been performed, and the outputs of the two steps are different objects. In the fact that many global descriptors have been proposed for urban object recognition, there is a lack of a thorough comparison of these proposed global descriptors. How to select a better global descriptor for the objects is the first task for anyone who want to recognize the urban objects using Velodyne LIDAR. In this paper, five global descriptors (Bounding Box descriptor, Histogram of Local Point Level descriptor, Hierarchy descriptor, SI, GFH descriptor) are computed for every object, respectively. Eventually, the recognition accuracy, which is obtained with the trained Support Vector Machine (SVM) classifiers, is used to evaluate the performance of these different global descriptors.

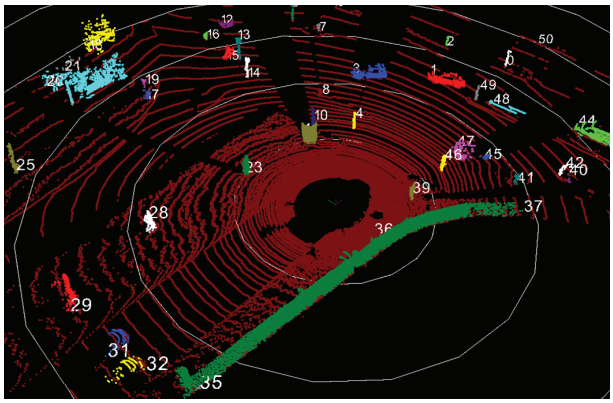

Fig. 1. A typical segmentation result of one scan collected by Velodyne LIDAR (LIght Detection And Ranging) in an urban environment. Different colors mean different objects

The main contributions of this paper are two-fold. The most significant one is that a novel GFH descriptor is proposed, it can obtain the best performance among all the global descriptors involved in this paper on the public Sydney Urban Objects Dataset. Also, we analyze the influence of some parameters of GFH descriptor to the recognition accuracy. The second contribution is that this paper presents a thorough evaluation of the performance of five global descriptors. From these experimental results, one could know which global descriptor is suitable to classify a certain class in urban environments according to their own requirements.

The rest of the paper is structured as follows. In the next section, an outline of the related work is described. Section 3 describes some state of the art global descriptors and the proposed GFH descriptor. Some experimental results are shown in Section 4, the performance of the GFH descriptor outperforms that of all the state of the art global descriptors on the public Sydney Urban Objects Dataset. We also give some recognition results on the data collected by our own ALV in urban environments. Finally, conclusion and an outline of the future work are given in Section 5.

II. RELATED WORK

Much research has been carried out over the last few years concerning object recognition, and various descriptors have

*This work is supported by National Nature Science Foundations under Grant 61075043 and 61375050.

[1]Tongtong Chen, Bin Dai, and Daxue Liu are with College of Mechatronic Engineering and Automation, National University of Defense Technology, Changsha 410073, Hunan, China bindai.cs@gmail.com

[2]Jinze Song is with the Graduate School, National University of Defense Technology, Changsha 410073, Hunan, China nwsac97@gmail.com

[1]www.acfr.usyd.edu.au/SydneyUrbanObjectsDataset.shtml

be proposed recently. Generally, these descriptors can be roughly divided into two groups: the local descriptors for every point of the object [7], [8], [9], [10], [11], [12], [13] and the global descriptors for the whole object [3], [5], [14], [15], [16].

Johnson et al. [7] proposed SI to describe the distribution of points around an oriented point. The support region is a cylinder and the cylindrical coordinates of the points in the support region are projected into a 2D histogram while neglecting the polar angle. Frome et al. [8] introduced 3D Shape Context (SC) and Harmonic SC to describe the distribution of the points in the spheral support region of a basic point over the De Espona 3D model library. The support region is divided into bins in the elevation and azimuth dimensions evenly, and along the radial dimension logarithmically. For the sake of removing the degree of freedom along azimuth angle of 3D SC, Tombari et al. [10] proposed Unique 3D SC on the Stanford dataset which employs a unique, unambiguous local Reference Frame (RF). Tombari et al.[9] also proposed SHOT. The difference to Unique 3D SC is that a local histogram is constructed in every bin. Chen et al. [11] proposed Intrinsic Shape Signature to character the spheral support region of a basic point over the Princeton Shape Benchmark. The intrinsic RF at a basic point is defined by the eigen analysis of the scatter matrix of the points within a support radius. A discrete spherical grid, computed from a base octahedron recursively, is used to partition the spherical angular space into unevenly distributed cells. Zhang et al. [12] introduced Local Surface Patch, which includes a 2D histogram and surface type, for 3D object classification. Rusu et al. [13] proposed a novel local descriptor Point Feature Histogram (PFH) to characterize the local geometry of 3D points. Four geometric features proposed in [17] are computed for every pair of points in the $k-$neighborhood of the basic point to construct a 16-bin histogram for the corresponding basic point.

All the above local descriptors only describe the 3D geometric character for the basic points. For the sake of improving the performance of recognition, Conditional Random Field (CRF) [18], [19] and Markov Random Field (MRF) [20], [21], [22] are often used as classifiers.

The global descriptors are used to describe the whole objects segmented in advance. Golovinskiy et al. [3] used the number of the points, average height, estimated volume, standard deviations in the two principle horizonal directions, standard deviation in height, and a SI for the center of the object as the global descriptors to conduct urban object recognition experiment. Douillard [5] proposed a serial of binary descriptors and Hierarchy descriptor (without any object recognition experiments) in the appendix of his phd thesis. Wang et al. [14] concatenated a SI computed at the centroid, shape distributions, shape factors, and Bounding Box as the global descriptors for the dynamic urban object recognition. Himmelsbach et al. [15] described some object level descriptors (maximum object intensity, mean object intensity, object intensity variance, object volume) and Histogram of Point Level Features for the urban object recognition. Rusu et al. [16] presented the Viewpoint Feature Histogram (VFH) descriptor, that describes both the geometry and viewpoint, for fast 3D kitchenware object recognition.

All the above papers only employed these global descriptors to classified the urban or kitchenware objects, there is no comparison of the performance of these global descriptors, which is done in this paper. Though there have been several studies [8], [18], [23], [24] on the evaluation of the performance of local descriptors for object recognition already, still to our best knowledge, this paper is the first one which presents a thorough comparison of the performance of the global descriptors in the context of urban object recognition with Velodyne LIDAR.

III. GLOBAL DESCRIPTORS

This section presents some state of the art global descriptors and the proposed GFH descriptor for every object instance, and their performance will be compared on the public data set in the next section.

A. Bounding Box Descriptor

The three dimensions of the Bounding Box descriptor [14] (also known as Object Volume [15]) along the PCA directions, which represent three dimensional extend of the point cloud of the object, are amongst the simplest global descriptors. In urban environments of this paper, the first two dimensions contain the length and width along the directions of eigenvectors, which are corresponding to the biggest and smallest eigenvalues of the point clouds in x and y, respectively. The z dimension is treated separately to capture the height of object, as we believe that most of the objects in urban environments are normal to the ground surface.

B. Histogram of Local Point Level Descriptor

The Histogram of Local Point Level descriptor [15] is based on the local point statistic features [25] for every point of the object. Lalonder et al.[25] calculated local point statistic feature at every point to represent its three saliences (point-ness, curve-ness, surface-ness) by performing eigenvalue decomposition of the covariance matrix of the 3D coordinates of its neighboring points. For a point p, given a set of N 3D neighboring points $P = \{p_1, p_2, \ldots, p_N\}$, the covariance matrix is defined as (1):

$$K = \frac{1}{N} \sum_{i=1}^{N} \big(p_i - \bar{p}\big)\big(p_i - \bar{p}\big)^T \tag{1}$$

where $\bar{p} = \frac{1}{N} \sum_{i=1}^{N} p_i$. The eigenvalues of matrix K are in the decreasing order $\lambda_0 \geq \lambda_1 \geq \lambda_2$. The symbols $L_1 = \lambda_0, L_2 = \lambda_0 - \lambda_1, L_3 = \lambda_1 - \lambda_2$ represent the point-ness, curve-ness and surface-ness of point p, respectively.

In order to transfer these local descriptors for each point of the object to a global one, Himmelsbach et al.[15] introduced three histograms, each consisting of 4 bins spaced over the range from 0 to 1, equally, for the three saliences of points, respectively. As there is no upper bound to any of the three

saliences, these eigenvalues $\lambda_0, \lambda_1, \lambda_2$ are firstly normalized by $\sum_{i=0}^{2} \lambda_i$.

C. Hierarchy Descriptor

The Hierarchy descriptor (also known as Ki description [5]) is a context-dependent model of a 3D object which describes the typical vertical extrusion observed in urban environments. It assumes that different classes of objects in urban environments often have distinctive vertical profiles. In our implementation, the z dimension of each object, ranging from 0 to 5 meters, is discretized into levels, and the resolution is set to 0.2 meter. For each vertical level, a 2D horizonal rectangle is fitted to the points falling in the corresponding level [5]. The Hierarchy descriptor is a data structure which stores four numbers (the length, width, area of the fitting rectangle and the ratio of the points in the corresponding level) for each level of the object. It gives rise to a 100-dimensional feature vector. In terms of rotation, the Hierarchy descriptor only achieves independence around the vertical axis. Even though, It will be sufficient to an ALV running in urban environments, since the objects sitting on the ground rarely rotate around other axes excepted the vertical axis.

D. Spin Image as Global Descriptor

The SI is a local descriptor which is used for object retrieval, matching, recognition in 3D scenes. It was initially proposed by Johnson et al. [7], and commonly-used in robotic community [3], [26], [27] recently. Given a oriented point p (surface vertex) with associative direction n (surface normal), the SI of point p is calculated by spinning a grid around the reference axis n, and the cells of the grid accumulate the neighboring points $q \in N_p^{\delta}$. An entry of SI $\in \mathbb{R}^{b \times b}$ with indexes (i, j) is represented by the non-negative perpendicular distance α to the line through the normal n and the signed perpendicular distance β to the tangent plan P defined by the point p and normal n. A spin map S_O, which is a projection function of a 3D neighboring point q of the oriented point p to (α, β), is defined as (2):

$$\begin{aligned} S_O &= \mathbb{R}^3 \rightarrow \mathbb{R}^2 \\ (\alpha, \beta) &= \left(\sqrt{\| q-p \|^2 - (n \cdot (q-p)^2)}, n \cdot (q-p)\right) \end{aligned}, \tag{2}$$

The indexes (i, j) in the SI are calculated by (3) [18],

$$i = \left[\frac{\beta + \delta}{2\rho}\right], \quad j = \left[\frac{\alpha}{\rho}\right] \tag{3}$$

where $[.]$ is the rounding function, δ is the radius of the support, $\rho = \frac{\delta}{b}$ is the grid resolution of SI, and b is the width of SI(the height is the same as the width).

In this paper, we follow Douillard's saying (from [28]) "the large image size provides an almost global descriptor ". In our implementation, only one SI is calculated for every object. The oriented point p is chosen as the center of the object with the associated direction n perpendicular to the ground.

E. Global Fourier Histogram Descriptor

The GFH descriptor is motivated by the drawback of SI and the fact that objects sitting on the ground in urban environments rarely rotate around other axes except the vertical axis. For the sake of removing the degree of freedom along azimuth angle, SI omits the cylindrical angular coordinate at the expense of decreasing the descriptiveness while spinning the grid. In such a case, we propose GFH descriptor, that can not only exploit the cylindrical angular coordinate, but also achieve independence with respect to rotation around the vertical axis.

The GFH descriptor is created for an oriented point which defines an object-centered cylindrical coordinate system, and the way to calculate the radial and elevation coordinates is the same as SI. In the urban object recognition, the oriented point is often chosen as the center point $o = [o_x, o_y, o_z]^T$ of the object, and the associative direction n is the direction of the normal vector. The cylindrical support region, whose axis is perpendicular to the ground ($z = [0, 0, 1]^T$), is partitioned into I, J, K bins by equally spaced boundaries in the elevation, azimuth and radial dimensions, as shown in Fig. 2. The GFH descriptor is calculated by counting the points that fall within each bins within the support region, and forming a 3D histogram.

Dividing the cylindrical support region in the way above can improve the descriptiveness. Meanwhile, it can also bring in a degree of freedom along the azimuth dimension. For the sake of achieving independence with respect to rotation around the vertical axis, we introduce a global reference frame [18] and Fourier transform. The global reference frame is a rotation invariant reference frame which is constructed using the global $z-$axis and the associative direction n of the oriented point o. The global reference frame $R_{global} \in \mathbb{R}^{4 \times 4}$ [18] is defined as (4),

$$R_{global} = \begin{bmatrix} \frac{(n \times z) \times z}{\|(n \times z) \times z\|} & \frac{n \times z}{\|n \times z\|} & \frac{z}{\|z\|} & o \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{4}$$

where $z = [0, 0, 1]^T$, $n \in \mathbb{R}^{3 \times 1}$, $o = [o_x, o_y, o_z]^T$.

For the sake of improving the robust of the GFH descriptor, the 3D histogram is analyzed using 1D Fast Fourier Transform (FFT) in the azimuth dimension. A key property of FFT is the rotation in the azimuth dimension results in a phase shift in the frequency domain, and the amplitudes

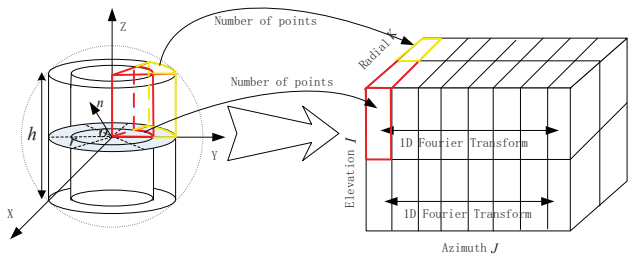

Fig. 2. Subdivisions of GFH descriptor ($I = 2, J = 8, K = 2$), n is the associate direction of the oriented point o, r is the support radius

of transformation are invariant to rotation in the azimuth dimension (5). The GFH descriptor is a $I \times J \times K$ matrix of amplitudes of the transformation, which is invariant to the rotation in the azimuth dimension. Thus, the degree of freedom in the azimuth dimension is removed.

$$|(F\{f(x - x_0)\})| = |F\{f(x)\}|, \tag{5}$$

where $F\{.\}$ represents the Fourier transform, $|.|$ represents the amplitudes of the transformation, and $f(.)$ represents the one dimensional discrete function.

IV. EXPERIMENTAL EVALUATION AND ANALYSIS

The GFH descriptor is evaluated quantitatively using the public Sydney Urban Objects Dataset and is compared against other global descriptors involved in this paper. Meanwhile, in order to present the performance of GFH descriptor intuitively, some qualitative experiments are conducted in the City of Changsha with our own ALV, which is equipped with a Velodyne HDL-64E S2, Three SICK LMS 291, and a NovAtel SPAN-CPT GPS-aided inertial navigation system (INS). Our ALV is shown in Fig. 3.

A. Training SVM Classifier with RBF Kernel

In this paper, the SVM classifier (one-vs-one) with RBF kernel is employed to evaluate the performance of these global descriptors on the public Sydney Urban Objects Dataset. There are two parameters which can not be obtained in advance for a RBF kernel: C, γ. For choosing the best parameters, we use a grid-search on them with cross-validation [29]. We choose the ones that get the best cross-validation accuracy as the best parameters. Meanwhile, in order to prevent the over-fitting problem, four-cross-validation procedure is used in this paper. The data set is divide into 4 subsets. Sequentially, three subsets are used to train the classifier, and the other one is tested with the trained classifier. In this way, each sample in the data set is tested once, and the accuracy is the ratio of the samples that are classified correctly.

Fig. 3. Our ALV used for the experiments in urban environments. It is equipped with Three SICK LMS 291, two color cameras, a Velodyne HDL-64E S2, a RIEGL laser and a NovAtel SPAN-CPT GPS-aided inertial navigation system

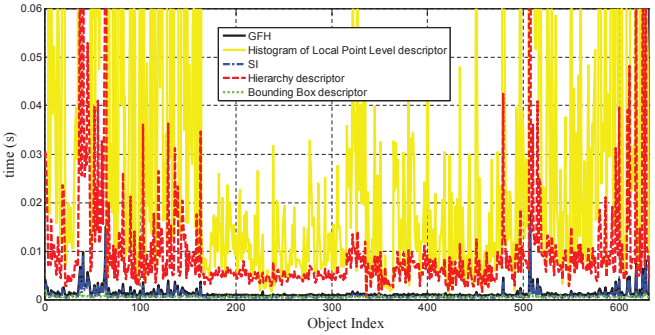

Fig. 4. Computational time of each descriptor over the public Sydney Urban Objects Dataset (in MATLAB), the sizes of Bounding box descriptor, Hierarchy descriptor, Histogram of Local Point Level descriptor, SI, and GFH descriptor are 4, 100, 12, 144 and 864, respectively

B. Quantitative Experiment

The Sydney Urban Objects Dataset contains of 631 urban objects which were collected in the Central Business District of Sydney. These objects were labeled into 26 categories, such as *cyclist, pedestrian, car, traffic light, traffic sign, tree* and so on. In this paper, the categories that have few samples are combined into a category named *others*, and we classify all the 631 urban objects into 15 classes (*4wd, building, bus, car, pedestrian, pillar, pole, traffic lights, traffic sign, tree, truck, trunk, ute, van* and *others*).

1) Computational Time of Each Descriptor: As these descriptors are applied to the ALV, the computational time of these descriptors is our focus. Fig. 4 shows the computational time of each descriptor over the 631 objects with MATLAB. One can find that the computational time of Histogram of Local Point Level descriptor, with an average of 0.052 ms per object, is the slowest, because local point statistic feature is calculated for every point of the object. The fastest is Bounding Box descriptor, whose computational time is 0.0006 ms. The computational time of GFH descriptor ($12 \times 6 \times 12$), with an average of 0.0018 ms per object, is twice as much as that of SI (12×12). Though it is slower that SI, we believe the GFH descriptor can achieve real-time performance for ALV[2].

2) Recognition Accuracy: Compared with SI, the azimuth dimension of GFH descriptor is divided into J bins to improve the descriptiveness. For the sake of demonstrating it, the influence of the number of azimuth division of GFH descriptor is analyzed while the support radius is fixed to 2 meters and the number of radial division[3] is fixed to 6, 8, 10, 12 and 14, respectively. Fig. 5 shows the recognition results of SI and GFH descriptor with different azimuth divisions. One can find that the performance of GFH descriptor outperforms that of SI in every radial division K, and the variational trend of results is almost the same (the

[2]In our implementation for ALV, a grid map typically covers an area of 65 by 30 meters. According to our statistics, there are around 50 objects in such a grid map

[3]In this paper, the radial, elevation divisions of GFH descriptor are equivalent to the width and height of SI, respectively. They are equal in number

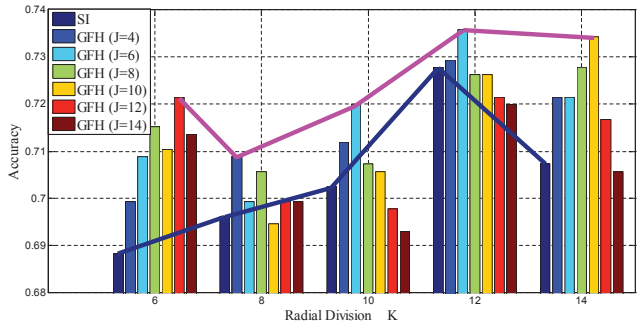

Fig. 5. The recognition results using SI and GFH descriptor when azimuth angle is divided into different bins, the blue and pink lines represent the best results of SI and GFH descriptor in different bins

pink and blue lines). Meanwhile, the influence of the number of azimuth division is also very severe for the recognition accuracy. Given the support radius, elevation division and radial division, we should choose suitable azimuth division of GHF descriptor to get the best result. From Fig. 5, it can be noticed that the best parameters for the GFH descriptor is $I = 12, J = 6, K = 12$ over the public Sydney Urban Objects Dataset.

The class-wise accuracy reflects the efficiency of different global descriptors. Fig. 6 shows the confusion matrix of object recognition with the GFH descriptor and Fig. 7 shows the confusion matrices with Bounding Box descriptor, Hierarchy descriptor, Histogram of Local Point Level descriptor, and SI, respectively. From the two figures, one can find that the GFH descriptor gets the best performance with an accuracy of 0.7358, which is followed by SI, whose accuracy is 0.7278. The performance of Histogram of Local Point Level descriptor is the worst (0.5158).

Generally, the classes *car* and *pedestrian* can be well distinguished from the other classes, while *pillar, pole, traffic light, traffic sign* and *trunk* turn to be more difficult for all

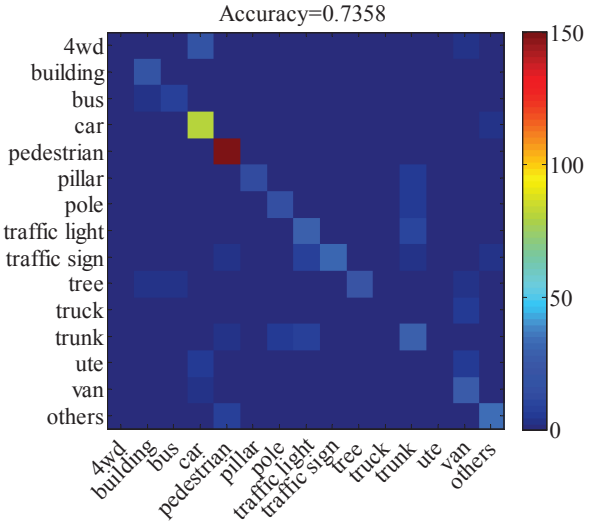

Fig. 6. The confusion matrix of object recognition with the GFH descriptor ($I = 12, J = 6, K = 12$)

these global descriptors. Especially, the class *trunk* is often confused with the classes *pillar, pole, traffic light, traffic sign, trees* and vice versa. Likewise, the class *4wd* is mostly misclassified as *car*; and the classes *building, bus* can often be confused, that is just because they have similar shapes. Though these global descriptors have something in common, they get different performance for each class too. Table I shows the F-Measure [5] of each class while using different global descriptors, and the red number is the best result for each class among these global descriptors. From Table I, one can find that GFH descriptor, SI and Hierarchy descriptor are the best ones among all the global descriptors. Specifically, one can obtain the best results for the classes *bus, pedestrian, pole traffic light, others* with SI, the best ones for *building, pillar, traffic sign, tree, van* with GFH descriptor, and the best for *4wd, car, trunk, ute* with Hierarchy descriptor. In addition, Bounding Box descriptor turns out to be the best choice for the recognition of truck.

C. Qualitative Experiment

In order to evaluate the usefulness of GFH descriptor to our own ALV, this descriptor is testified on the Velodyne scans collected by our own ALV in the City of Changsha, China. For every scan, we implement the Gaussian-Process-based ground segmentation algorithm in [30] to remove the ground points, the left obstacle points are clustered into many objects with the Radially Bounded Nearest Neighbor (RBNN) graph in [6]. The GFH descriptor is calculated for every object of each scan. All the labeled samples of the public Sydney Urban Objects Dataset are employed to train the SVM classifier ($C = 8.000, g = 0.500$). We use point cloud library (PCL) [31] to show the results of recognition.

Fig. 8 shows the performance of the GFH descriptor in two urban environments. The different colors and numbers represent different classes, for instance, the green and num-

TABLE I

F-MEASURE OF EACH CLASS WITH DIFFERENT GLOBAL DESCRIPTORS. BBD, HD, HLPLD ARE SHORT FOR BOUNDING BOX DESCRIPTOR, HIERARCHY DESCRIPTOR, HISTOGRAM OF LOCAL POINT LEVEL DESCRIPTOR, RESPECTIVELY. $b = 12, I = 12, J = 6, K = 12$, THE RED NUMBER REPRESENTS THE BEST RESULT FOR EACH CLASS

	F-Measure				
	BBD	HD	HLPLD	SI	GFH
4wd	0.0769	**0.3265**	0	0.2308	0
building	0.5128	0.6977	0.4444	0.7317	**0.8085**
bus	0.5	0.2222	0	**0.5806**	0.5455
car	0.7805	**0.8229**	0.4984	0.8159	0.8223
pedestrian	0.9359	0.9419	0.7660	**0.9484**	0.9434
pillar	0	0.4211	0	0.5263	**0.6842**
pole	0.5532	0.3684	0.5	**0.7727**	0.7273
traffic light	0.5	0.5306	0.4651	**0.6522**	0.6316
traffic sign	0.5833	0.7045	0.6598	0.7368	**0.7442**
tree	0.5797	0.6452	0.5152	0.6471	**0.7333**
truck	**0.5**	0.32	0	0.4211	0.25
trunk	0.2353	**0.5323**	0.4286	0.4727	0.5225
ute	0.2609	**0.4286**	0	0.1111	0.1905
van	0.3824	0.4179	0	0.5610	**0.5909**
others	0.6957	0.7174	0.3478	**0.8090**	0.7816

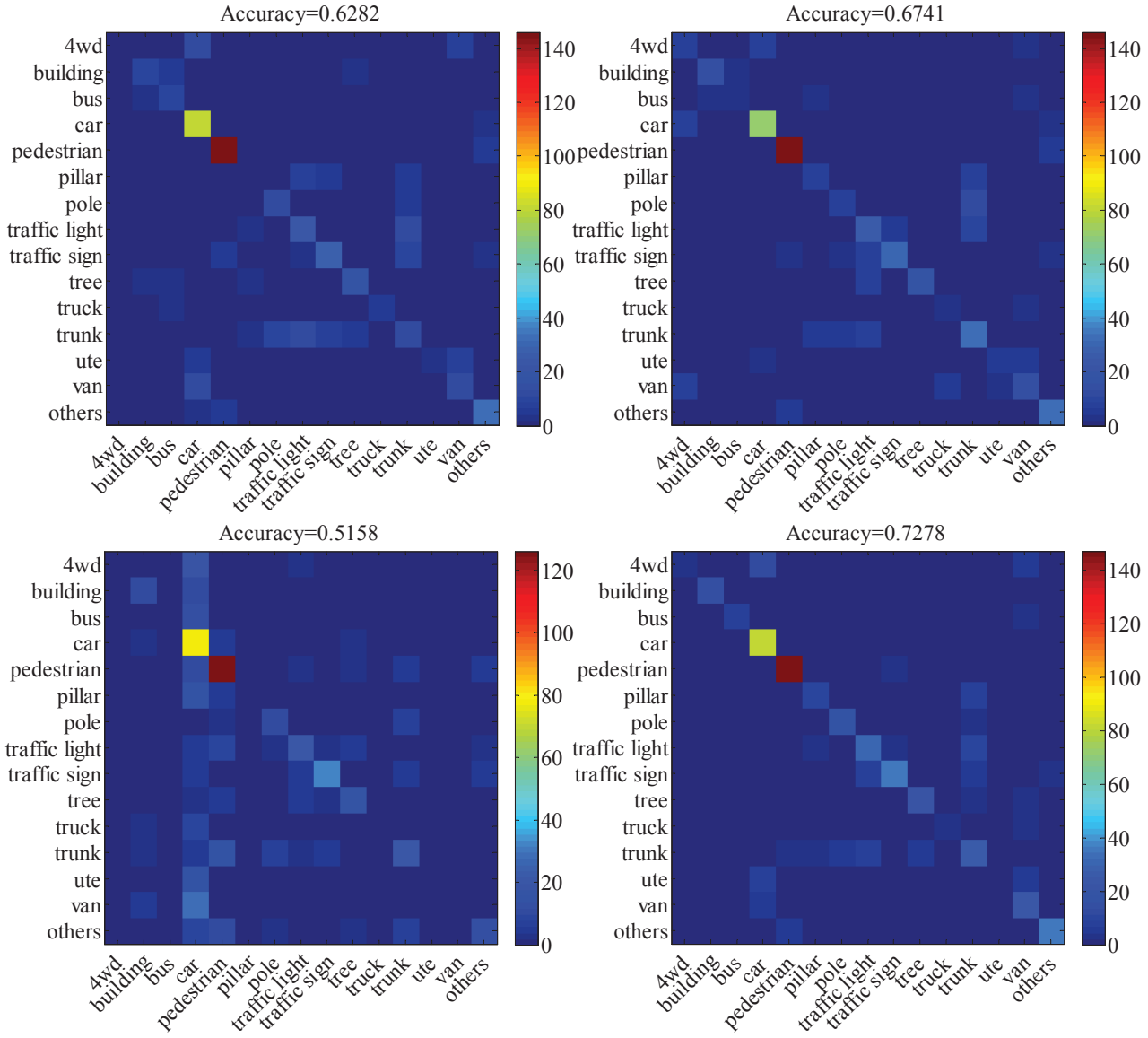

Fig. 7. The confusion matrices. The left top, left bottom, right top, and right bottom images are the recognition results of Bounding Box descriptor, Histogram of Local Point Level descriptor, Hierarchy descriptor, SI ($b = 12$), respectively

ber 4 represent the class *car* and the blue and number 10 represent the class *tree* et al. As one can see, the GFH descriptor can recognize almost all the objects correctly, such as pedestrian (cyan), car (green), pole (white), tree (blue) and so on. As two pedestrians (the red rectangles in the bottom image) walk together, the RBNN graph segments them into one object. Thus, they were classified as the class *others* because of the lack of the training samples. There are also some recognition errors in urban environments with the GFH descriptor. For example, in the green polygon of the top image, the building is misclassified as the class *bus*; and a real bus is misclassified as the class *building* in the blue rectangle of the bottom image. There are two reasons: (1) they have similar shape, as both of them have a large surface; (2) only the GFH descriptor at the center is calculated for the object. There are a few points in the cylindrical support region with 2m radius in this paper. Thus, the descriptiveness decreases for the classes *bus, building*.

V. CONCLUSIONS AND FUTURE WORKS

The performance of the global descriptors will dramatically influence the results of the recognition approaches. In this paper, we evaluate several state of the art global descriptors for the object recognition on public Sydney Urban Objects Dataset. We also proposed a novel GFH descriptor, which could get the best recognition results on that data set, and the performance on the data collected by our own ALV also demonstrates its usefulness.

In the future, we will employ Bayesian optimization to gain better parameters for RBF kernel, and will also compare different classifiers such as nearest-neighbor, Gaussian-Process, and random forest to present the descriptiveness of the GFH descriptor.

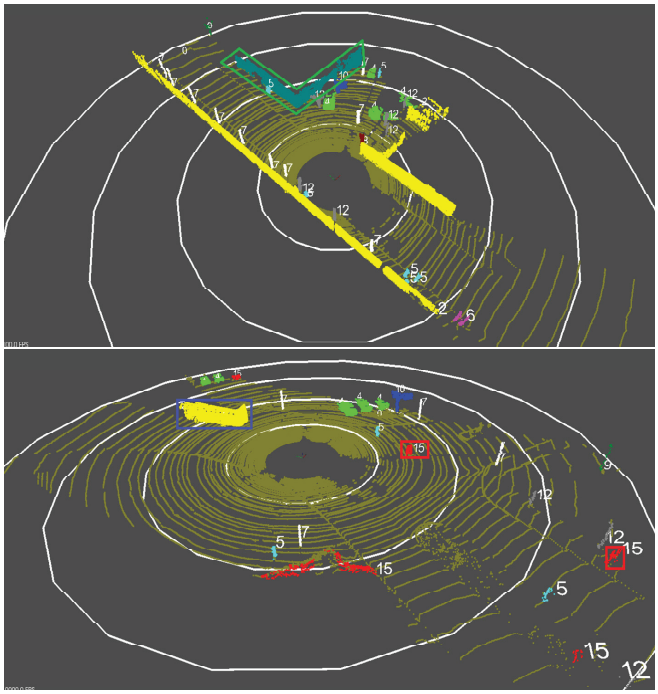

Fig. 8. The performance of the GFH descriptor. The numbers $0-15$ represent the classes *ground, 4wd, building, bus, car, pedestrian, pillar, pole, traffic light, traffic sign, tree, truck, trunk, vte, van* and *others*, respectively. The radius of the four white circles are $10, 20, 30, 40$ meters, respectively

ACKNOWLEDGMENT

Thanks Alastair Quadors at Australian Center for Field Robotics (ACFR) who released some experimental data.

REFERENCES

[1] M. Himmelsbach, F. Hundelshausen, and H. J. Wuensche, "Fast segmentation of 3d point clouds for ground vehicles," in *IEEE Intelligent Vehicles Symposium*, 2010, pp. 560–565.

[2] F. Moosmann, O. Pink, and C. Stiller, "Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion," in *IEEE Intelligent Vehicles Symposium*, 2009, pp. 215–220.

[3] A. Golovinskiy, V. G. Kim, and T. Funkhouser, "Shaped-based recognition of 3d point clouds in urban environments," in *Internation Conference on Computer Vision*, 2009, pp. 2154–2161.

[4] A. Das, J. Servos, and S. L. Waslander, "3d scan registration using the normal distributions transform with ground segmentation and point cloud clustering," in *Internation Conference on Robotics and Automation*, 2013, pp. 2207–2212.

[5] B. Douillard, "Laser and vision based classification in urban environments," Ph.D. dissertation, The University of Sydney, 2009.

[6] K. Klasing, D. Wollherr, and M. Buss, "A clustering method for efficient segmentation of 3d laser data," in *Internation Conference and Robotics and Automation*, 2008, pp. 4043–4048.

[7] A. E. Johnson and M. Hebert, "Using spin images for efficient object recognition in cluttered 3d scenes," *Pattern Analysis and Machine Intelligece, IEEE Transactions on*, vol. 21, no. 5, pp. 433–449, 1999.

[8] A. Frome, D. Huber, R. Kolluri, T. Bulow, and J. Mailk, "Recognizing objects in range data using regional point descriptors," in *European Conference on Computer Vision*, 2004, pp. 224–237.

[9] F. Tombari, S. Salti, and L. D. Stefano, "Unique signature of histograms for local surface descriptors," in *European Conference on Computer Vision*, 2010, pp. 356–369.

[10] F. Tombari, S. Salti, S. Salti, and L. D. Stefano, "Unique shape context for 3d data description," in *Proceedings of ACM workshop on 3D object retrieval*, 2010, pp. 57–62.

[11] H. Chen and B. Bhanu, "3d free-form object recognition in range images using local surface patch," *Pattern Recognition Letters*, vol. 28, no. 10, pp. 1252–1262, 2007.

[12] Y. Zhang, "Intrisic shape signature: A shape descriptor for 3d object recognition," in *International Conference on Computer vision workshops*, 2009, pp. 689–696.

[13] R. B. Rusu, Z. C. Marton, N. Blodow, and M. beetz, "Persistent point feature histograms for 3d point clouds," in *International Conference on Intelligent Autonomous Systems*, 2011, pp. 119–128.

[14] D. Z. Wang, I. Posner, and P. Newman, "What could move? finding cars, pedestrains and bicyclist in 3d laser data," in *Internation Conference on Robotics and Automation*, 2012, pp. 4038–4044.

[15] M. Himmelsbach, A. Muller, T. Luttle, and H. J. Wunsche, "Lidar-based 3d object perception," *Automatic Control*, vol. 50, no. 4, pp. 511–515, 2009.

[16] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *International Conference on Intelligent Robots and Systems*, 2010, pp. 2156–2162.

[17] E. Wahl, U. Hillenbrand, and G. Hirzinger, "Surflet-pair-relation histograms: A statistical 3d-shape representation for rapid classification," in *Fourth International Conference on 3-D Digital Imaging and Modeling*, 2003, pp. 474–481.

[18] J. Behley, V. Steinhage, and A. B. Cremers, "Performance of histogram descriptors for the classification of 3d laser range data in urban environments," in *Internation Conference on Robotics and Automation*, 2012, pp. 4391–4398.

[19] D. Munoz, N. vandapel, and M. Hebert, "Directional associative markov network for 3d point cloud classification," in *Interantional symposium on 3D Data Processing, Visualization and Transmission*, June 2008.

[20] F. Tombari and L. D. Stefano, "3d data segmentation by local classification and markov random field," in *Internation Conference on 3D Imageing, Modeling, Processing, Visualization and Transmission*, 2011, pp. 212–219.

[21] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Ng, "Discriminative learning of markov random fileds for segmentation of 3d scan data," in *International Conference on Computer Vision and Pattern Recognition*, 2005, pp. 169–176.

[22] D. Munoz, N. Vandapel, and M. Hebert, "Onboard contextual classification of 3d point clouds with learned high-order markov random fields," in *Internation Conference on Robotics and Automation*, 2009, pp. 4273–4280.

[23] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligece*, vol. 27, no. 10, pp. 1615–1630, 2005.

[24] R. B. Rusu, Z. C. Matron, N. Bildow, and M. Beetz, "Learning informative point classes for the acquisition of object model maps," in *International Conference on Control, Automation, Robotics and Vision*, 2008, pp. 643–650.

[25] J. F. Lalonde, N. vandapel, D. Huber, and M. Hebert, "Natural terrian classifiation using three-dimensional ladar data for ground robot mobility," *Journal of field Robotics*, vol. 23, no. 10, pp. 839–861, 2006.

[26] B. Douillard, J. Underwood, V. Vlaskine, A. Quadros, and S. Singh, "A pipeline for the segmentation and classification of 3d point clouds," in *International Symposium on Experimental Robotics (ISER)*, 2010, pp. 585–600.

[27] X. Xiong, D. Munoz, J. A. Bagnell, and M. Hebert, "3d scenes analysis via sequenced predictions over points and regions," in *Internation Conference and Robotics and Automation*, 2011, pp. 2609–2016.

[28] B. Douillard, A. Quadros, P. Morton, J. Underwod, and M. D. Deuge, "A 3d classifier trained without field samples," *Automatic Control*, vol. 50, no. 4, pp. 511–515, 2012.

[29] C. W. Hsu, C. C. Chang, and C. J. Lin, *A practical Guide to Support Vector Classification*, 2013.

[30] T. Chen, B. Dai, R. Wang, and D. Liu, "Gaussian-process-based real-time ground segmentation for autonomous land vehicle," *Journal of Intelligent and Robotic systems*, 2013, Accept.

[31] R. Rusu and S. Cousins, "3d is here: point cloud library (pcl)," in *International Conference on Robotics and Automation (ICRA)*, 2011, pp. 1–4.