

# Multi-modal Panoramic 3D Outdoor Datasets for Place Categorization

Hojung Jung<sup>1</sup>, Yuki Oto<sup>1</sup>, Oscar M. Mozos<sup>2</sup>, Yumi Iwashita<sup>3</sup> and Ryo Kurazume<sup>4</sup>

**Abstract**—We present two multi-modal panoramic 3D outdoor (MPO) datasets for semantic place categorization with six categories: forest, coast, residential area, urban area and indoor/outdoor parking lot. The first dataset consists of 650 static panoramic scans of dense (9,000,000 points) 3D color and reflectance point clouds obtained using a FARO laser scanner with synchronized color images. The second dataset consists of 34,200 real-time panoramic scans of sparse (70,000 points) 3D reflectance point clouds obtained using a Velodyne laser scanner while driving a car. The datasets were obtained in the city of Fukuoka, Japan and are publicly available in [1], [2]. In addition, we compare several approaches for semantic place categorization with best results of 96.42% (dense) and 89.67% (sparse).

## I. INTRODUCTION

Understanding the surrounding environment is an important capability for autonomous robots and vehicles that allows them to identify their current type of location. This information greatly improves communication between robots and humans [3], [4] and it allows autonomous robots to make decisions with context-based understanding when completing high-level tasks [5]. Moreover, if a robot has the ability to categorize places according to their type, then it will be able to properly execute a task even in unfamiliar surroundings. In addition, autonomous vehicles can make decisions based on the environmental conditions.

This paper focuses on semantic categorization of places in outdoor scenarios where a mobile robot or vehicle should determine the type of place where it is located. Examples of outdoor places are shown in Figure 1. Place categorization using 2D images has been achieved in high-level image understanding. 2D images can be captured by a camera or be collected by search engines, therefore, 2D image understanding and scene recognition are actively researched and evaluated by using several datasets, such as ImageNET [6], SUN database [7], or Places2 [8]. The ImageNET and Sun datasets consist of a large collection of images with variability in visual appearance, which can be searched using category terms. The Places2 dataset includes images from

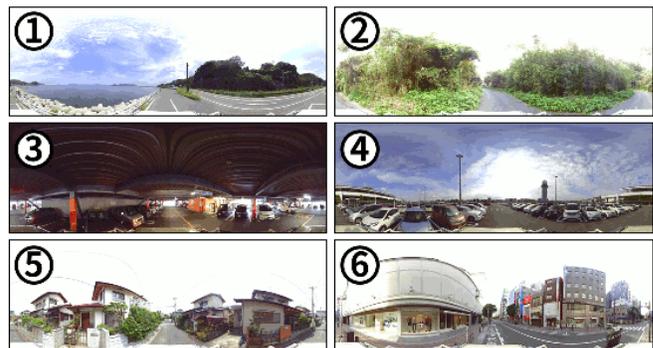
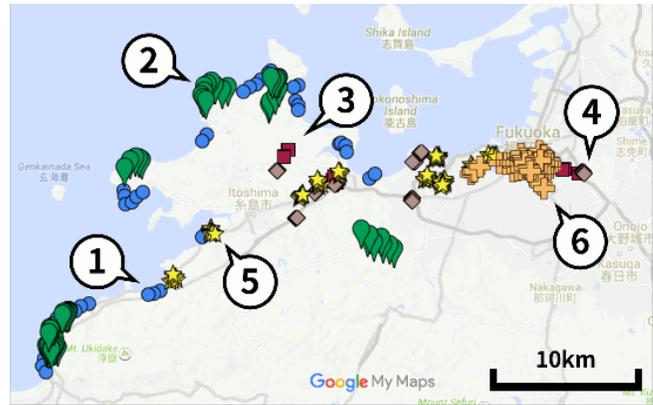


Fig. 1. **Top image:** Map indicating the location where dense panoramic images were obtained. **Bottom image:** Example panoramic images for forest (1), coast (2), indoor parking lot (3), outdoor parking lot (4), residential area (5), and urban area (6).

different scene categories that are used to train convolutional neural networks (CNN) [9]. New images can be classified using the provided CNN models. However, those datasets do not provide 3D information.

To extend the scope of 2D images, 3D place categorization has been studied. In indoor environments, several researchers used low cost RGB-D sensors to provide sequences of depth and color point clouds. For example, the NYU-Depth V2 dataset [10] contains segmented 3D images of indoor environments with per-frame accelerometer data. The 3DSUN RGB-D dataset [11] is an annotated benchmark scene for 3D scene/place categorization and reconstruction. Finally, the Kyushu University Kinect Place Recognition Database [12] contains RGB-D sequences of indoor places divided into six categories, i.e. corridor, kitchen, lab, office, study room and toilet. In addition, laser scans have been used to increase the range and resolution of panoramic point clouds in indoor place datasets, and to add reflectance information like in the Kyushu University Indoor Semantic Place Dataset [13].

Outdoor 3D datasets for place categorization are less common since they need more expensive sensors and ve-

<sup>1</sup>Hojung Jung and Yuki Oto are with Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan {hojung, y.oto}@irvs.ait.kyushu-u.ac.jp

<sup>2</sup>Oscar Martinez Mozos is with Technical University of Cartagena (UPCT), Spain oscar.mozos@upct.com

<sup>3</sup>Yumi Iwashita is with Jet Propulsion Laboratory, 4800, Oak Grove Drive Pasadena, 91109 CA, USA. This research was performed when she was at Kyushu University, and this paper does not contain any results from the researches at Jet Propulsion Laboratory. Yumi.Iwashita@jpl.nasa.gov

<sup>4</sup>Ryo Kurazume is with Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka 819-0395, Japan kurazume@ait.kyushu-u.ac.jp

hicles. The majority of available 3D datasets are used for localization and mapping [14], [15] but not for high level semantic place categorization and therefore they do not include place labels. Only the KITTI [16] dataset includes four categories in the labeled scans: city, residential, road and campus. However, this dataset is used as benchmark for other purposes such as optical flow, visual odometry, 3D object detection or 3D tracking. In comparison to the KITTI dataset our dense MPO datasets include higher resolution panoramic 3D point clouds. In addition all our panoramic instances include a place label from our six categories.

In this paper, we present two Multi-modal Panoramic 3D Outdoor (MPO) datasets for outdoor semantic place categorization in static and dynamic environments for mobile robots and autonomous vehicles. Our datasets are recorded by multi-modalities sensors, and provided as multi-resolution point clouds. Each dataset consists of six place categories including ‘forest’, ‘coast’, ‘indoor parking lot’, ‘outdoor parking lot’, ‘residential area’ and ‘urban area’.

The dense MPO dataset consists of static panoramic scans of dense 3D color and reflectance point clouds as shown in Figure 1, while the sparse MPO dataset consists of real-time panoramic scans of sparse 3D reflectance point clouds recorded while driving a car. The datasets were obtained in the city of Fukuoka, Japan, and are publicly available in [1], [2]. The main motivation for creating two different sensor modalities datasets (dense and sparse) for the same outdoor places is to study different sensor performances, and thus present the researchers with different options when selecting their sensors. In this paper we use time-of-flight laser scanners which have the advantage of their robustness under various environmental conditions, such as bright sunlight, darkness, and sudden illumination changes.

In addition, we applied several approaches for place categorization to the MPO datasets using different global descriptors and we obtain best categorization results of 96.42% (dense) and 89.67% (sparse) using local binary patterns [13].

## II. MPO DATASETS

We present two Multi-modal Panoramic 3D Outdoor (MPO) Datasets for semantic place categorization in static and dynamic environments for mobile robots and autonomous vehicles. The first dataset contains dense panoramic point clouds obtained using a FARO sensor. The second dataset contains sparse point clouds acquired using a Velodyne sensor.

### A. DENSE MPO DATASET

The dense MPO dataset consists of 650 static panoramic scans of high resolution dense 3D point clouds obtained using a FARO Focus3D laser scanner. Each panoramic scan is composed of three synchronized sensor modalities: depth, reflectance, and RGB color.

The FARO Focus3D sensor has a maximum range of 150 meters and a field of view of  $360^\circ$  horizontally and  $\times 300^\circ$  vertically (Figure 2). Our dataset contains 3D panoramic point clouds of  $5140 \times 1757$  pixels (total of 9,030,980

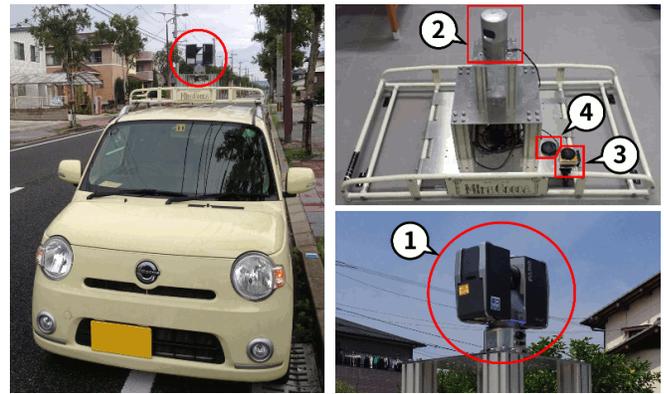


Fig. 2. **Left:** Vehicle setup for data acquisition. **Right Top:** Velodyne HDL-32E laser scanner (2), Kodak PIXPRO SP360 camera (3), and GARMIN GPS 18x LVC (4) for the sparse MPO dataset. **Right Bottom:** FARO Focus3D sensor (1) for dense MPO dataset.

pixels) with  $0.07^\circ$  horizontal and  $0.17^\circ$  vertical angular resolution. This sensor is additionally equipped with a rotating camera which is calibrated with the laser sensor.

The FARO sensor was installed on top of a vehicle at a height of 1.8 meters as shown in Figure 2. Each scan was obtained in a static way with the vehicle stopped. The measurement time needed to obtain one synchronized depth and color panoramic scan was 3 minutes. Approximately, the depth and reflectance images takes 1 minute, and color images take 2 minutes. For each scan we moved the car 10-100 meters, stopped and took the full panoramic scan and color images. We repeated this process several times in each area. Point clouds and color images were synchronized off-line using the SCENE software provided by FARO, which processes and manages scanned data easily and efficiently by recognizing primitive objects as well as scan registration and positioning.

3D panoramic scans were label with the corresponding category (‘forest’, ‘coast’, ‘indoor parking lot’, ‘outdoor parking lot’, ‘residential area’ or ‘urban area’.) The full dataset contains 650 panoramic scans distributed among the six categories according to Table I. For each category we recorded seven sets of scans corresponding to different physical places inside the same category. Example 3D panoramic scans are shown in Figure 3 in their three sensor modalities: depth, reflectance, and RGB color. The full dense MPO dataset is publicly available in [1].

Each panoramic scan in the dense MPO dataset is stored in a PTX file, which contains a header and one line for each point in the format ‘‘X, Y, Z, intensity, R, G, B.’’ Full details on this files can be obtained from the dataset’s website [1].

### B. SPARSE MPO DATASET

The sparse MPO dataset contains a total of 34,200 sparse 3D panoramic scans obtained using a Velodyne HDL-32E laser scanner sensor. This sensor contains a vertical field of view of  $41.3^\circ$  ranging from  $10.67^\circ$  to  $-30.67^\circ$  and a horizontal field of view of  $360^\circ$ . Each panoramic 3D panoramic point cloud has size of  $32 \times 2166$  with horizontal

TABLE I  
DISTRIBUTION OF 3D PANORAMIC SCANS BY CATEGORY IN THE DENSE MPO DATASET

Category	Number of scans by location							Total
	Set1	Set2	Set3	Set4	Set5	Set6	Set7	
Coast	14	14	16	12	17	14	16	103
Forest	16	16	17	18	16	16	17	116
Indoor parking	16	16	13	15	17	13	13	105
Outdoor parking	15	17	16	15	15	14	16	108
Residential area	14	16	14	15	16	15	16	106
Urban area	16	17	16	16	15	16	16	112
Total number of panoramic scans								650

TABLE II  
DISTRIBUTION OF 3D PANORAMIC SCANS BY CATEGORY IN THE SPARSE MPO DATASET

Category	Number of scans by location										Total
	Set1	Set2	Set3	Set4	Set5	Set6	Set7	Set8	Set9	Set10	
	Set6	Set7	Set8	Set9	Set10						
Coast	511	254	571	221	314	376	872	506	386	287	4298
	376	872	506	386	287						
Forest	440	824	980	707	730	720	439	311	797	531	6479
	720	439	311	797	531						
Indoor parking lot	520	357	274	873	583	343	466	592	344	428	4780
	343	466	592	344	428						
Outdoor parking lot	874	579	388	370	477	536	581	563	460	617	5445
	536	581	563	460	617						
Residential area	674	787	667	724	563	973	717	720	977	662	7464
	973	717	720	977	662						
Urban area	490	572	587	487	410	566	712	565	606	739	5734
	566	712	565	606	739						
Total scans	3509	3373	3467	3382	3077	3514	3787	3257	3570	3264	34200
	3514	3787	3257	3570	3264						

angular resolution of  $0.17^\circ$  and vertical angle resolution of  $1.33^\circ$ . Scans were acquired at 2 Hz of sampling frequency.

In addition the dataset contains GPS information obtained by a Garmin GPS 18x LVC, which stored in NMEA 0183 format at a frequency of 2 Hz.

Additionally, the dataset provides panoramic color images captured by Kodak PIXPRO SP360 camera a reference of scanning environment. Each color image covers  $360^\circ$  and has a resolution of 16.36 Megapixels. Images were acquired 6~7 Hz.

The sensors were located on a vehicle as shown in Fig. 2. The sparse data was acquired while driving a vehicle at 30~50 kph through different areas of Fukuoka city as shown in Fig. 1. The maximum displacement error of a scan is 1.4m at the maximum velocity of 50kph with 10Hz of scanning speed for a single scan.

The dataset is divided into the same six place categories as the dense MPO dataset, i.e. ‘forest’, ‘coast’, ‘indoor parking lot’, ‘outdoor parking lot’, ‘residential area’ or ‘urban area’. For each category we recorded seven trajectories corresponding to different physical places inside the same category. Table II shows the distribution of the sparse MPO dataset. The full sparse MPO dataset is available at [2].

### III. PLACE CATEGORIZATION IN MPO DATASETS

In this paper we present several approaches for semantic place categorization and compare their effectiveness in our two MPO datasets.

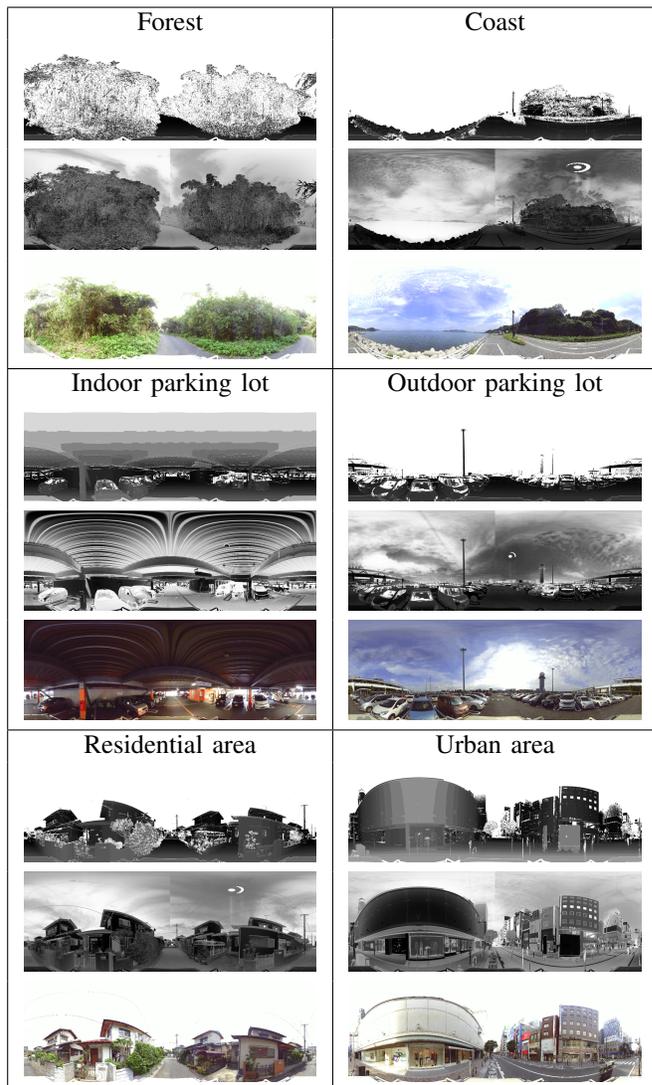


Fig. 3. **Dense MPO Dataset:** examples of high-resolution range (top), reflectance (middle) and color (bottom) panoramic images for six outdoor place categories: forest, coast, indoor/outdoor parking lot, residential and urban area. In range images, brighter colors indicate closer distances and in reflectance images, darker colors indicate higher intensity.

#### A. Feature Descriptors

In this work we compared three feature descriptors that can be applied to 3D and multimodal sensor cues: local binary patterns (LBP) [17], spin images [18], and textons [19].

1) *Local Binary Pattern:* LBP [17] is a visual transformation applied to grayscale images that describes the relationship between the values of each pixel and their neighbors. Formally, given a pixel  $i = (x_i, y_i)$  in an image  $I$ , and its P-neighborhood  $\mathcal{N}_P(i)$ , the new decimal value for pixel  $i$  in the transformed image  $I_{LBP}$  is given by:

$$I_{LBP}(i) = \sum_{k=0}^{P-1} LBP(I(i) - I(j_k))2^k, \forall j_k \in \mathcal{N}_P(i), \quad (1)$$

where  $\{j_0, j_1, \dots, j_{P-1}\}$  are the pixels in the P-neighborhood  $\mathcal{N}_P(i)$  and  $LBP(\cdot)$  is a binary operator over the difference  $z$  of two pixels' values defined as:

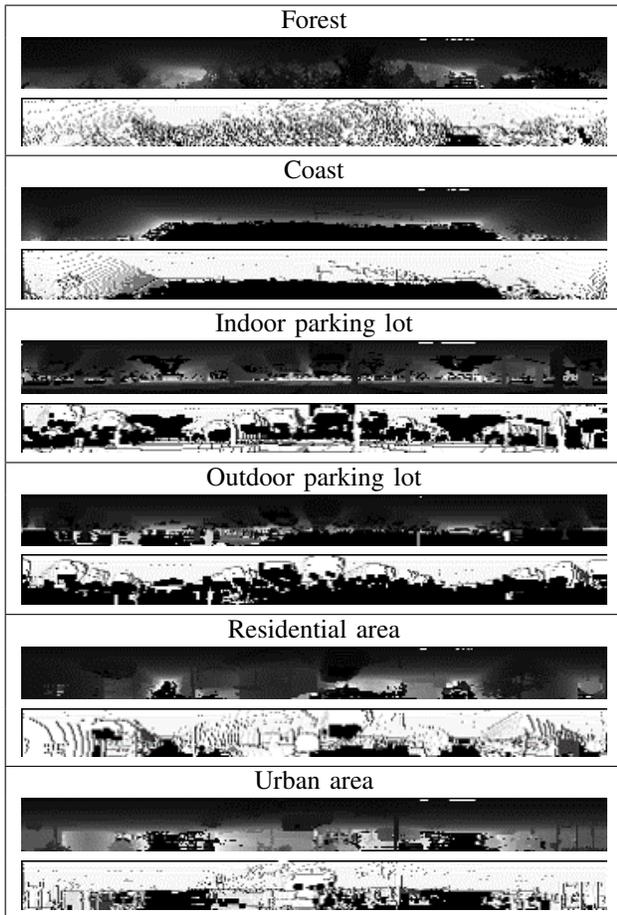


Fig. 4. **Sparse MPO Dataset:** examples of range (top) and reflectance (bottom) panoramic images for six outdoor place categories: forest, coast, indoor/outdoor parking lot, residential and urban area. In range images, brighter colors indicate closer distances and in reflectance images, darker colors indicate higher intensity.

$$LBP(z) = \begin{cases} 1 & \text{if } z \geq 0 \\ 0 & \text{if } z < 0 \end{cases} \quad (2)$$

Each new pixel value in the transformed image image  $I_{LBP}$  is a decimal values  $d$  in the range  $[0, \dots, 255]$ . In a final step, the transformed image  $I_{LBP}$  is represented by a histogram  $h_{LBP}$  of length  $l$  in which each bin  $h_{LBP}(l)$  indicates the frequency of appearance of the decimal value  $d$  as:

$$h_{LBP}(l) = \sum_i \mathcal{I}(I_{LBP}(i) = d), \quad (3)$$

where  $\mathcal{I}(\cdot)$  denotes the indicator function which returns 1 if its argument is true, and 0 otherwise. In our case, the LBP values are restricted to the 8-neighborhood, and thus the dimension of the final histogram is 256. LBPs have been successfully applied to categorization of indoor and outdoor categorization in single and multiple sensor modalities [12], [13], [20].

2) *Spin image:* The spin image [18] is one of popular technique for surface matching and 3D object recognition. The conventional spin images encode the global properties of any surface in an object-oriented coordinate system. In this

paper we applied spin image in a scanner-oriented view point in cylindrical coordinate system rather than in an object-oriented view point using 3D surface of object. A entire 3D point cloud can be represented as the scanner-oriented image with the cylindrical coordinate system when the position of laser scanner is  $(0,0,0)$  for each scan data. To achieve this we use the tangent plane through oriented perpendicularly to  $n$  and the line through parallel to [18].

3) *Texton:* Texton [19] is a filter-based technique based on Leung-Malik filter banks and the maximum response filter bank contains filters at multiple orientations and scales. In standard texton, the maximum response filter is applied among several Gaussian and Laplacian of Gaussian filters with the same scale, but different orientations are selected. In our case, we adopted the maximum response filter among filters with different scales but same orientation in order to evaluate the directions of edges and not the size.

### B. Classification Method

We use support vector machines (SVM) [21], [22] with a radial basis function (RBF) kernel for the final categorization of places. Multi-class classification is performed by a “one-against-one” approach [23]. In our experiments, we use the LIBSVM library [24]. Following the method reported in [25], the parameters  $C$  and  $\gamma$  are selected by a grid search using cross-validation. The ranges of  $C$  and  $\gamma$  are  $C \in [2^{-1}, \dots, 2^{20}]$  and  $\gamma \in [2^{-20}, \dots, 2^0]$  in the grid search.

For trajectories in the sparse MPO dataset we additionally apply a classic linear-time majority vote technique [26] for consecutive frames as follows. Given a time  $t$ , we define  $M$  previous consecutive frames  $\{C_t, C_{t-1}, \dots, C_{t-M}\}$ . Finally a majority vote is obtained from the classification results of the  $M$  as:

$$C_t^{Final} = \text{Majority vote}(C_t, C_{t-1}, \dots, C_{t-M}). \quad (4)$$

where  $C_t$  indicates the classification of frame  $t$  using SVM.

## IV. EXPERIMENTAL EVALUATION

We applied the previous approaches for place categorization to our dense and sparse MPO datasets respectively.

### A. DENSE MPO DATASET

We applied the previous feature descriptors and a SVM classifier to the dense MPO dataset using different modalities independently. The classification results are shown in Table III. Best classification results are obtained using LPBs. These results are in accordance with our previous work [13]. In particular, reflectance data provide the highest correct classification ratio. Detailed classification results using LBP and reflectance are shown in Table IV for all six place categories.

In addition, we present results combining laser and reflectance modalities as in [12]. Combination of modalities is achieved by concatenating the feature vectors of each modality into a single vector, and then applying SVMs [12]. This multi-modal categorization gets a CCR of  $95.67 \pm 3.69\%$ .

TABLE IV  
CONFUSION MATRIX FOR REFLECTANCE LBP IMAGES FOR THE DENSE MPO DATASET (CCR%)

	Coast	Forest	Indoor parking	Outdoor parking	Residential area	Urban area
Coast	<b>93.45 ± 17.39</b>	6.54 ± 17.39	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Forest	8.95 ± 8.27	<b>91.04 ± 8.27</b>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Indoor parking	0.00 ± 0.00	0.00 ± 0.00	<b>100.00 ± 0.00</b>	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00
Outdoor parking	1.91 ± 2.92	0.00 ± 0.00	0.00 ± 0.00	<b>98.08 ± 2.92</b>	0.00 ± 0.00	0.00 ± 0.00
Residential area	1.33 ± 2.68	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	<b>97.32 ± 4.43</b>	1.33 ± 4.00
Urban area	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	1.87 ± 4.00	<b>98.12 ± 4.00</b>

TABLE V  
CONFUSION MATRIX (CCR %) OF LBP IMAGES WITH MAJORITY VOTE FOR THE SPARSE MPO DATASET.

	Coast	Forest	Indoor parking	Outdoor parking	Residential area	Urban area
Coast	<b>82.99 ± 4.19</b>	5.20 ± 0.54	0.00 ± 0.00	7.15 ± 2.20	1.50 ± 0.05	3.16 ± 0.56
Forest	8.64 ± 1.12	<b>90.87 ± 1.20</b>	0.00 ± 0.00	0.49 ± 0.01	0.00 ± 0.00	0.00 ± 0.00
Indoor parking	0.00 ± 0.00	3.31 ± 0.01	<b>93.95 ± 0.65</b>	2.16 ± 0.56	0.59 ± 0.02	0.00 ± 0.00
Outdoor parking	0.96 ± 0.06	0.00 ± 0.00	5.25 ± 0.12	<b>89.16 ± 1.31</b>	2.44 ± 0.26	2.19 ± 0.26
Residential area	0.45 ± 0.01	0.00 ± 0.00	0.00 ± 0.00	2.70 ± 0.01	<b>93.23 ± 0.20</b>	3.63 ± 0.20
Urban area	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	3.84 ± 0.76	8.76 ± 0.79	<b>87.39 ± 0.20</b>

TABLE III  
CORRECT CLASSIFICATION RATIO (CCR %) RESULTS FOR THE DENSE MPO DATASET.

Descriptor	Range	Reflectance	Greyscale
Spin image [18]	89.43 ± 0.00	-	-
LBP [13]	94.35 ± 2.67	<b>96.42 ± 2.68</b>	93.86 ± 3.85
Texton [19]	89.04 ± 5.58	73.59 ± 13.73	81.81 ± 10.67

TABLE VI  
CORRECT CLASSIFICATION RATIO (CCR %) FOR THE SPARSE MPO DATASET

	Single frame	Majority vote
Spin image [18]	79.23 ± 4.51	88.34 ± 0.12
LBP [13]	83.98 ± 4.59	<b>89.67 ± 0.21</b>

## B. SPARSE MPO DATASET

We present categorization results for the sparse MPO dataset using range data by comparing single observations VS a majority vote over the last  $M$  frames. Table VI shows the corresponding CCRs. For the majority vote we selected  $M = 40$  since it provided the best categorization results according to Fig. 5. According to this results the majority vote approach significantly improves the outdoor categorization results. Finally we present detailed categorization results for the six place categories using LBP and majority vote in Table V.

## V. CONCLUSIONS

We presented two multi-modal panoramic 3D outdoor (MPO) datasets for semantic place categorization with six categories: forest, coast, residential area, urban area and indoor/outdoor parking lot. While the dense MPO dataset contains static high density multi-modal panoramic 3D point clouds, the sparse dataset contains sparse panoramic point clouds captured by a moving vehicle. Both MPO datasets are publicly available on the Internet under [1], [2]. In addition we have presented preliminary categorization results using different approaches. The datasets together with preliminary

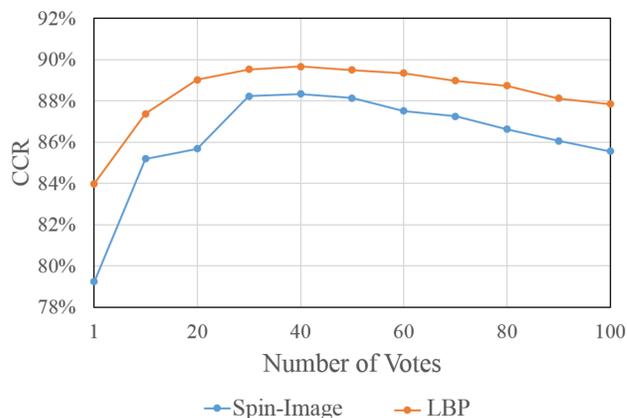


Fig. 5. Majority vote categorization for different consecutive frames  $M$ .

categorization results are thought to be used as benchmark for researches interested in outdoor semantic place categorization.

## ACKNOWLEDGMENT

The present study was supported in part by a Grant-in-Aid for Scientific Research (A) (26249029).

## REFERENCES

- [1] "Dense Multi-modal Panoramic 3D Outdoor Dataset for Place Categorization." <http://robotics.ait.kyushu-u.ac.jp/~kurazume/research-e.php?content=db#d07>.
- [2] "Sparse Multi-modal Panoramic 3D Outdoor Dataset for Place Categorization." <http://robotics.ait.kyushu-u.ac.jp/~kurazume/research-e.php?content=db#d08>.
- [3] H. Zender, O. M. Mozos, P. Jensfelt, G.-J. M. Kruijff, and W. Burgard, "Conceptual spatial representations for indoor mobile robots," *Robotics and Autonomous Systems*, vol. 56, pp. 493–502, June 2008.
- [4] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proc. of the 2012 IEEE International Conference on Robotics and Automation (ICRA)*, (Saint Paul, MN, USA), May 2012.
- [5] C. Stachniss, O. M. Mozos, and W. Burgard, "Efficient exploration of unknown indoor environments using a team of mobile robots," *Annals of Mathematics and Artificial Intelligence*, vol. 52, pp. 205–227, April 2008.

- [6] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Miami, FL, USA), pp. 248–255, June 2009.
- [7] J. Xiao, J. Hays, K. Ehinger, A. Oliva, A. Torralba, et al., "Sun database: Large-scale scene recognition from abbey to zoo," in *IEEE conference on Computer vision and pattern recognition (CVPR)*, (San Francisco, CA), pp. 3485–3492, June 2010.
- [8] B. Zhou, A. Khosla, A. Lapedriza, A. Torralba, and A. Oliva, "Places: An image database for deep scene understanding," *Arxiv*, 2016.
- [9] Y. Bengio, *Learning deep architectures for AI*. Foundations and trends in Machine Learning, now, 2009.
- [10] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, "Indoor Segmentation and Support Inference from RGBD Images," in *European Conference of Computer Vision (ECCV)*, (Florence, Italy), pp. 746–760, October 2012.
- [11] S. Song, S. P. Lichtenberg, and J. Xiao, "SUN RGB-D: A RGB-D scene understanding benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Boston, Massachusetts), pp. 567–576, June 2015.
- [12] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, pp. 6695–6711, May 2012.
- [13] O. M. Mozos, H. Mizutani, H. Jung, R. Kurazume, and T. Hasegawa, "Categorization of indoor places by combining local binary pattern histograms of range and reflectance data from laser range finders," *Advanced Robotics*, vol. 27, pp. 1455–1464, October 2013.
- [14] J.-L. Blanco-Claraco, F.-A. Moreno-Dueñas, and J. Gonzalez-Jimenez, "The malaga urban dataset: High-rate stereo and lidar in a realistic urban scenario," *The International Journal of Robotics Research*, vol. 33, pp. 207–214, 2014.
- [15] G. Pandey, J. R. McBride, and R. M. Eustice, "Ford campus vision and lidar data set," *The International Journal of Robotics Research*, vol. 30, pp. 1543–1552, 2011.
- [16] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Providence, Rhode Island), pp. 3354–3361, June 2012.
- [17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 971–987, jul 2002.
- [18] A. E. Johnson and M. Hebert, "Surface matching for object recognition in complex three-dimensional scenes," *Image and Vision Computing*, vol. 16, no. 9, pp. 635–651, 1998.
- [19] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision*, vol. 43, pp. 29–44, June 2001.
- [20] H. Jung, O. M. Mozos, Y. Iwashita, and R. Kurazume, "Local n-ary patterns: a local multi-modal descriptor for place categorization," *Advanced Robotics*, pp. 1–14, 2016.
- [21] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [22] C. M. Bishop, *Pattern Recognition and Machine Learning*. Springer, 2006.
- [23] S. Knerr, L. Personnaz, and G. Dreyfus, "Single-layer learning revisited: a stepwise procedure for building and training a neural network," in *Neurocomputing: Algorithms, Architectures and Applications* (J. Fogelman, ed.), Springer-Verlag, 1990.
- [24] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011.
- [25] C.-W. Hsu, C.-C. Chang, and C.-J. Lin, "A practical guide to support vector classification," 2010.
- [26] R. S. Boyer and J. S. Moore, *MJRTYa fast majority vote algorithm*. Springer, 1991.