

Indoor Place Categorization using Co-Occurrences of LBPs in Gray and Depth Images from RGB-D Sensors

Hojung Jung*, Oscar Martinez Mozos[†], Yumi Iwashita[‡], and Ryo Kurazume[‡]

*Graduate School of Information Science and Electrical Engineering

[‡]Graduate Faculty of Information Science and Electrical Engineering

Kyushu University, 744 Motoooka, Nishi-ku, Fukuoka 819-0395, Japan

Email: hojung@irvs.ait.kyushu-u.ac.jp, yumi@ait.kyushu-u.ac.jp, kurazume@ait.kyushu-u.ac.jp

[†]School of Computer Science, University of Lincoln, Brayford Pool, Lincoln, LN6 7TS, United Kingdom

Email: omozos@lincoln.ac.uk

Abstract—Indoor place categorization is an important capability for service robots working and interacting in human environments. This paper presents a new place categorization method which uses information about the spatial correlation between the different image modalities provided by RGB-D sensors. Our approach applies co-occurrence histograms of local binary patterns (LBPs) from gray and depth images that correspond to the same indoor scene. The resulting histograms are used as feature vectors in a supervised classifier. Our experimental results show the effectiveness of our method to categorize indoor places using RGB-D cameras.

I. INTRODUCTION

The capability to distinguish and categorize the different places that compose an environment is fundamental for service robots working and interacting with humans indoors. Moreover, this ability is indispensable for robots providing different services to people in their daily life. If a robot is able to recognize the category of its surroundings such as the offices, kitchens or laboratories, then it can reduce the ambiguity of the human instructions according to the places, and it can greatly improve its human-robot communication capabilities [1], [2].

The problem of place categorization has been addressed by several researchers using different type of sensors like for example laser scans [3], [4], [5], or vision cameras [6], [7], [8], [9]. Recently, RGB-D sensors are getting popular in robotics due to its low cost and the multi-modal nature of the provided information: RGB and depth data. Therefore, it becomes of great interest to analyze the capabilities of RGB-D sensors for the previous task.

This paper proposes a new method for categorizing indoor places using RGB-D sensors. Our approach is based on the idea of spatial correlations between the gray and depth image modalities that are provided by RGB-D sensors. Whenever gray and depth images are obtained at the same time for the same scene, these images must be spatially correlated. In order to describe this correlation we create 2D co-occurrence histograms of local binary pattern (LBP) values corresponding to the gray and depth images representing the same scene. This histogram is converted into a one-dimension vector by concatenating its rows, and its dimension is finally reduced using singular value decomposition (SVD). We define this new multi-modal transformation as *Co-LBP* (*Co*-occurrence of *LBP*), because it contains the co-occurrence nature of the multi-modal LBP images. An example of this process is shown

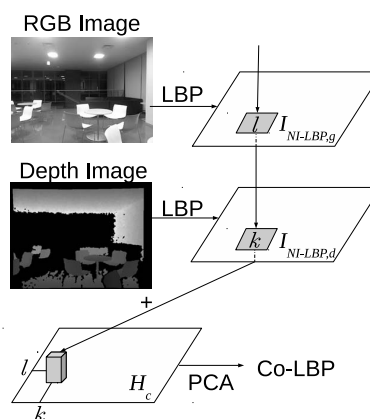


Fig. 1. Process for obtaining the Co-LBP descriptor for the gray and depth images corresponding to an indoor place.

in Fig.1. The resulting Co-LBP descriptors are then used as input to a support vector machine (SVM) for the final classification. As far as we know this is the first work applying co-occurrence histograms of gray and depth modalities to the task of place categorization.

An additional contribution of this paper is the application of a variant of LBP, called NI-LBP [10], for place categorization. In this work, we modify the original NI-LBP descriptor by reducing its dimensionality using a uniformity threshold. The resulting descriptor is then used to calculate the previous Co-LBP descriptor for gray and depth images. The combination of the modified NI-LBP descriptor and the 2D co-occurrence histogram provides the best classification results for very low dimensional feature vectors.

We finally apply our approach to a dataset which contains synchronized pairs of gray and depth images corresponding to five different indoor place categories: offices, corridors, kitchens, laboratories, and study rooms. The results of our experiments show that we obtain high categorization results when applying our approach to the task of indoor place categorization.

II. RELATED WORK

The problem of indoor place categorization applied to robotics has increased its interest in the last years. Several

approaches have been applied to solve this problem by using different sensor modalities.

In [3], 2D laser scans are used as observations in order to obtain the type of the place the robot is located at. In this work several geometrical features are calculated from each scan, and they are used as input for a boosting classifier. This system is extended with the detection of objects in camera images in [11]. In [4], the laser scans are used to classify place nodes inside a Voronoi Random Field. Alternatively, the work in [5] classifies single laser beams instead of complete scans by applying logistic regression.

Whenever service robots are equipped with camera sensors, they can make use of visual techniques for indoor place categorization. The PLISS system from [6] applies spatial pyramids to images using SIFT features for place categorization. In addition, transitions between places are detected as change-points in image sequences. In [7], a multi-modal approach combines different features extracted from indoor images and 2D laser scans for place recognition. Moreover, the method introduced in [8] applies the CENTRIST descriptor to indoor images. The CENTRIST descriptor is extended to HSV color space in [12]. Finally, the approach in [9] uses uniform LPBs to categorize indoor and outdoor places.

Recently, RGB-D sensors are becoming popular in robotics for its low cost and its capability to provide RGB and depth images simultaneously. Therefore, some works have used this sensor for indoor place categorization. The work in [13] uses the point cloud obtained by several Kinect cameras covering 360 degrees around the robot. In addition, the approach in [14] applies the SURE descriptor to point clouds for place recognition. Finally, in [15] places are recognized by extracting planes in 3D. The previous works do not use the RGB data provided by the sensor and only depth data is processed. In comparison, in this paper we exploit gray and depth data simultaneously.

The closest work to our approach is [16], in which LPBs are applied to gray and depth images independently, and a final feature vector is obtained by simple concatenation of both descriptors. Our work differs from [16] in that we apply a different variant of LBP, i.e. NI-LBP. In addition, we use co-occurrence histograms to combine the modalities instead of using simple concatenation. In this way, we exploit the spatial correlation of LBP descriptors between gray and depth images.

III. LOCAL BINARY PATTERNS

The local binary pattern (LBP) operator [17] is a local transformation applied to gray scale images that encodes the relations between the value of every pixel with respect its neighborhood.

Let $I(i)$ be the value of pixel $i = (x, y)$ in image I , and let $\mathcal{N}_P(i)$ define the pixels in its P -neighborhood, then the LBP operator compares the pixel value $I(i)$ with the value $I(j)$ of every pixel in its neighborhood $j \in \mathcal{N}_P(i)$. The new value for every neighboring pixel j is set to 1 if $I(j) > I(i)$, and to 0 otherwise. The obtained binary values are concatenated clockwise and transformed into the corresponding decimal value. This decimal value is then assigned to pixel i in the

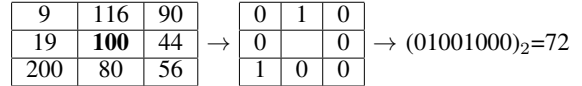


Fig. 2. Local binary transformation of the center pixel (marked in bold) using a 8-neighborhood.

resulting transformed image I_{LBP} . Formally

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

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where P indicates the P -neighboring pixels of i . In our case we use $P = 8$. An example of a local binary transformation is shown in Fig. 2.

The NI-LBP operator [10] is a modification of the previous LBP in which the neighboring pixels are compared with a threshold

$$I_{\text{NI-LBP}}(i) = \sum_{j=0}^{P-1} s(I(j) - \theta(i))2^j, \forall j \in \mathcal{N}_P(i), \quad (2)$$

$$s(x) = \begin{cases} 1 & \text{if } x \geq 0 \\ 0 & \text{otherwise,} \end{cases}$$

where

$$\theta(i) = \frac{1}{P} \sum_{j=0}^{P-1} I(j), \forall j \in \mathcal{N}_P(i). \quad (3)$$

The transformed NI-LBP image $I_{\text{NI-LBP}}$ is then represented by a histogram h of length L in which each bin $h(l)$ indicates the frequency of appearance of the decimal value l as

$$h(l) = \sum_i \mathcal{I}(I_{\text{NI-LBP}}(i) = l), \quad (4)$$

where \mathcal{I} denotes the indicator function which returns 1 if its argument is true, and 0 otherwise. Since the NI-LBP values are restricted to the 8-neighborhood, the dimension of the final histogram is $L = 256$.

The dimension of h can be further reduced by selecting a subset of its bins. The criterion to select these bins is based on the uniformity value U [17], which counts the number of transitions between 0s and 1s (or vice versa) in the binary string representing the decimal value l as

$$U(l) = \sum_{j=0}^{P-1} |b_j - b_{\text{mod}(j+1, P)}|, \quad (5)$$

where $(b_0 \dots b_{P-1})_2$ is the binary representation of the decimal value l , with $b_j \in \{0, 1\}$. As an example $U(72) = 4$ as can be corroborated in Fig. 2.

Finally, a reduced histogram h^u can be obtained by keeping the bins whose uniformity value is lower than or equal to a certain value u as

$$h^u = \{h(l) \mid U(l) \leq u\}. \quad (6)$$

IV. CO-OCCURRENCE OF MULTI-MODAL IMAGES

Whenever we obtain gray and depth images at the same time for the same scene using an RGB-D sensor, then these images must be spatially correlated. Therefore, the LBP values of the corresponding transformed images should be considered to be correlated too.

Let I_g and I_d be the original gray and depth images respectively, then after applying the NI-LBP transformation from Sect. III to each image, we obtain the corresponding histograms h_g and h_d . A simple method to create a feature vector $h_{g,d}$ representing both modalities simultaneously would be to concatenate the previous histograms, i.e. $h_c = \{h_g, h_d\}$. This approach considers the gray and depth images simultaneously, but the spatial relationship of LBP values between the modalities is ignored.

To take into account the spatial correlation of LBPs from the two modalities we create a 2D co-occurrence histogram H_c of dimension $L \times K$ from the LBP values obtained after applying the LBP transformation to the gray and depth images. Each bin in the co-occurrence histogram $H_c(l, k)$ contains the frequency of appearance of a pair of LBP values (l, k) . Formally

$$H_c(l, k) = \sum_{\forall i} \begin{cases} 1 & \text{if } I_{\text{LBP},g}(i) = l \text{ and } I_{\text{LBP},d}(i) = k \\ 0 & \text{otherwise,} \end{cases} \quad (7)$$

where $I_{\text{LBP},g}$ and $I_{\text{LBP},d}$ are the LBP transformations of the gray and depth images respectively. A graphical representation for the creation of H_c is shown in Fig. 1.

In a further step we re-arrange the values of the vector H_c into a 1D vector h_c by concatenating the different rows one by one. Finally, we reduce the dimension of h_c by applying singular value decomposition (SVD), and we obtain \tilde{h}_c . We define this transformation as Co-LBP because the final vector \tilde{h}_c contains the co-occurrence nature of the multi-modal LBP images.

V. CLASSIFICATION USING SVMs

For the final categorization of places we use a supervised approach based on support vector machines (SVMs) [18] with radial basis function (RBF) kernels. For the multi-class case we apply a *one-against-one* approach [19]. As input for the SVM we use the Co-LBP histograms \tilde{h}_c .

We use the LIBSVM library [20] for our experiments. The optimal SVM parameters C and γ are selected by a grid search using cross-validation with values ranging in $C \in [2^{-5}, \dots, 2^{15}]$ and $\gamma \in [2^{-12}, \dots, 2^3]$.

VI. EXPERIMENTS

In order to evaluate the performance of our proposed Co-LBP descriptor, we conducted several experiments using the indoor dataset from [16]. This dataset contains synchronized RGB and depth images of five different categories of indoor places: corridors, kitchens, labs, study rooms, and offices. The images were acquired on a mobile platform equipped with a Kinect camera. Each category contains pairs of images corresponding to different rooms that belong to one specific category. In particular, the category ‘‘corridor’’ contains four

TABLE I. THE NUMBER OF RGB AND DEPTH IMAGES FOR INDOOR SCENE CATEGORIZATION (1228 PAIRS IN TOTAL)

Category	Place	Pairs of RGB and depth images
Corridor	Corridor 1	68
	Corridor 2	42
	Corridor 3	70
	Corridor 4	99
	Total	279
Kitchen	Kitchen 1	73
	Kitchen 2	65
	Kitchen 3	53
	Total	191
Laboratory	Laboratory 1	99
	Laboratory 2	99
	Laboratory 3	81
	Laboratory 4	78
	Total	357
Study room	Study Room 1	71
	Study Room 2	70
	Study Room 3	49
	Study Room 4	62
	Total	252
Office	Office 1	57
	Office 2	45
	Office 3	47
	Total	149

different corridors, the category ‘‘kitchen’’ contains three different kitchens, the category ‘‘laboratory’’ contains five different laboratories, the category ‘‘study room’’ contains four different study rooms, and the category ‘‘office’’ contains three different offices. In our experiments, we use a subset of this dataset containing a total of 1228 pairs of gray and depth images. A summary of our subset is shown in Table I. Some example images are shown in Fig 6.

To evaluate the categorization performance we created ten pairs of training and test sets following the approach from [16]. In each pair of sets, a place appearing in the training set does not appear in the corresponding test set. In this way, we want to analyze the ability of our method to categorize previously unseen places. After applying our approach to each pair of training and test sets, we calculate the average correct categorization rate (CCR) for the ten experiments. In all experiments we used a SVM for the categorization of the Co-LBP descriptors as described in Sect.V.

A. Categorization using Concatenation of LBP Histograms

In the first experiments, we study the performance of the NI-LBP descriptor in comparison with different LBP versions for the task of place categorization. In these experiments, we combine the gray and depth modality by simple concatenation of their LBP histograms.

Table II compares the different correct categorization rates (CCR) using different variations of the LBP operator. The table also shows the final dimension of each descriptor. In this table, LBP refers to the standard LBP transformation presented in [17]. In addition, LBP^{u2} refers to *uniform* LBP [17], which is an LBP descriptor using uniformity value $U = 2$. Moreover, LBP^{u4} indicates an LBP with $U = 4$. The LBP^{u4} descriptor performed the best in the multi-modal categorization experiments of [21]. The NI-LBP descriptor is the one presented in [10]. Finally, we show results using our suggested NI-LBP u4 , which is an NI-LBP descriptor using uniformity value $U = 4$. The NI-LBP u4 descriptor provides the best categorization results in our dataset.

TABLE II. PLACE CATEGORIZATION USING SIMPLE CONCATENATION OF GRAY AND DEPTH LBP DESCRIPTORS

LBP type	Dim	CCR (%)	Feature type	Dim	CCR (%)
LBP [17]	512	88.25	LM filter [22]	512	73.07
LBP ^{u2} [17]	118	85.64	MR filter [23]	512	76.74
LBP ^{u4} [21]	398	89.14	SIFT [24]	200	86.86
NI-LBP [10]	512	90.02	SIFT [24]	400	86.75
NI-LBP ^{u4}	398	91.35	SURF [25]	200	64.16
			SURF [25]	400	89.51

TABLE III. COMPARISON RESULTS FOR SINGLE AND MULTIPLE MODALITIES USING NI-LBP^{u4}

LBP type	Dimension	CCR (%)
NI-LBP ^{u4} (<i>gray</i>)	196	73.38
NI-LBP ^{u4} (<i>depth</i>)	196	86.35
NI-LBP ^{u4} (<i>gray and depth</i>)	398	91.35

In addition, we tested popular texture analysis techniques named "Texton"[22], [23], "SIFT" [24] and "SURF" [25]. The Leung-Malik (LM) filter bank and the maximum response (MR) filter bank were evaluated. In standard texton, the maximum response filter among several Gaussian and LOG (Laplacian of Gaussian) filters with same scale but different orientations is selected. However, we adopted the maximum response filter among filters with different scales but same orientation so as to evaluate the directions of edges, not the size of them. Moreover, SIFT and SURF features with Bag-of-Features technique were evaluated, too. Each pixel of image could transform into SIFT and SURF descriptors by using dense detector, and additionally we used Bag-of-Features technique for selecting clustered features. Table II shows the results of classification performance using Texton, SIFT and SURF. It is clear that the filter bank techniques and dense feature with Bag-of-words are not suitable for this problem comparing with LBP, since they mainly focus on the edges in the images.

In a final experiment, we show the contribution of using multi-modal information in comparison with single modalities when applying the NI-LBP^{u4} descriptor for place categorization. Table III shows the CCRs for single and multi-modal information. It is clear that the recognition performance is improved by utilizing gray and depth images simultaneously when compared with the case in which gray and depth images are utilized individually.

B. Categorization using Co-LBP

In the following experiments we show the performance of our proposed Co-LBP descriptor for the task of place categorization. In this experiments we apply co-occurrences histograms of NI-LBP^{u4} descriptors which are further reduced using SVD as explained in Sect. IV. In this section we will refer to this descriptor as Co-NI-LBP^{u4}.

The Co-NI-LBP^{u4} descriptor is the result of the application of SVD to the co-occurrence histogram h_c . In order to know how the dimension reduction affects the classification performance, we apply SVD to h_c using different final dimensions, and then we calculate the corresponding CCRs.

The results of this experiment are shown in Table IV. From this table we can conclude that a final dimension of

TABLE IV. COMPARISON OF CCRs FOR CO-LBP AND NI-LBP^{u4} (SVD)

Dimension	CCR (%)	
	Co-NI-LBP ^{u4}	NI-LBP ^{u4} (SVD)
5	58.79	69.05
10	83.25	84.90
25	89.39	82.31
50	81.50	78.35
100	76.40	71.86
200	67.44	66.19
300	64.05	61.77

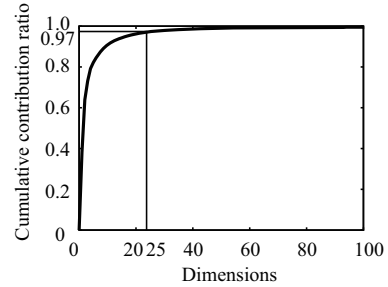


Fig. 3. Cumulative contribution ratio of SVD applied to Co-NI-LBP^{u4}.

25 provides the best categorization results when using Co-NI-LBP^{u4}. In addition, Fig. 3 shows the cumulative contribution ratio according to the different SVD dimensions. We can see in this plot that the first 25 eigenvalues account for 97% of the variance in the data. Calculation time of SVD is 959.8 [sec.] using Matlab (Xeon E5, 2.80GHz) in case that the dimension of the correlation matrix is 39601 for Co-NI-LBP^{u4}.

In addition, Table IV presents a comparison with the NI-LBP^{u4} multi-modal descriptor, which is obtained by simple concatenation of gray and depth descriptors. The NI-LBP^{u4} descriptors are also reduced using SVD. The results of Table IV are plotted in Fig. 4 for a better visualization. From the previous results we can conclude that Co-NI-LBP^{u4} offers better performance for similar dimensions in almost all cases. Moreover, the best global categorization result CCR= 89.39 is obtained when using Co-NI-LBP^{u4} with SVD dimension 25. This result is similar to the CCR obtained using the original NI-LBP^{u4} without SVD reduction (see Table II), however the dimension of our Co-NI-LBP^{u4} descriptor (25) is 93.71% smaller than the dimension of NI-LBP^{u4} (398). We think this dimension reduction is an important characteristic of our proposed method, and it can be a major advantage in large scale databases for indoor places.

In the next experiment, we compare the performance of the Co-LBP approach when applied to different variations of LBPs. The results of this experiment are shown in Table V. In this table, Co-LBP refers to co-occurrences of LBP, Co-LBP^{u2} to co-occurrences of LBP^{u2}, and Co-LBP^{u4} to co-occurrences of LBP^{u4}. Finally, Co-NI-LBP and Co-NI-LBP^{u4} refer to co-occurrences of NI-LBP and NI-LBP^{u4} respectively. In all cases the dimension of the final descriptor after applying SVD is 25. This dimension is selected by taking into account the results of Table IV. The results in Table V confirm that the applications of co-occurrence histograms of NI-LBP^{u4} descriptors provides the best results for multi-modal place categorization.

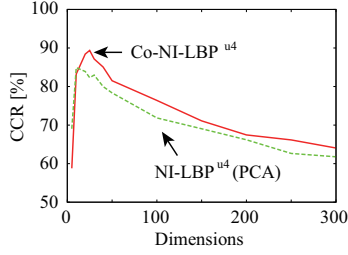


Fig. 4. Comparison of Co-NI-LBP^{u4} and NI-LBP^{u4} (SVD)

Moreover, a global comparison is shown in Fig 5. Here we present the categorization results of different LPB variations with their corresponding dimensions. According to this plot, our proposed Co-NI-LBP^{u4} provides the best trade-off between low dimension and final categorization results.

Table VI presents the confusion matrix after applying our proposed Co-NI-LBP^{u4} to the different categories in the dataset. We can see in this table that, although all categories obtain high rates, some places are still misclassified like for example study rooms and laboratories or kitchens. We think this is due to the similarities between places. For example, we find many tables and chairs in both laboratories and study rooms, and this makes the categorization difficult in some specific viewpoints.

Finally, Table VII shows the CCR by combining NI-LBP^{u4} descriptor and Co-NI-LBP^{u4} descriptor which reduces the dimension to 25 by PCA. A final feature vector is obtained by simple concatenation of both descriptors. It is clearly shown that the descriptor which combines NI-LBP and proposed Co-LBP outperforms the conventional LBP.

TABLE V. CORRECT CLASSIFICATION RATIO FOR VARIOUS CO-LBPS

LBP type	Dimension	CCR (%)
Co-LBP	25	83.78
Co-LBP ^{u2}	25	81.71
Co-LBP ^{u4}	25	83.27
Co-NI-LBP	25	89.17
Co-NI-LBP ^{u4}	25	89.39

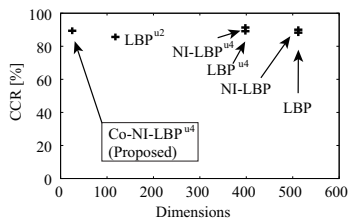


Fig. 5. CCRs of various LBPs

VII. CONCLUSIONS

This paper presented a place categorization method using RGB-D sensors. Our approach is based on co-occurrence of LBP transformations of gray-scale and depth images. The proposed Co-LBP descriptor shows a high classification performance for indoor scene categorization even if the dimension

TABLE VI. CONFUSION MATRIX FOR CO-NI-LBP^{u4} (%)

	Corridor	Kitchen	Lab.	Study room	Office
Corridor	96.06	0	1.27	2.54	0.13
Kitchen	2.05	79.35	1.88	16.55	0.17
Lab.	0	0	98.27	1.50	0.23
Study room	0.59	3.71	14.99	79.82	0.89
Office	0	2.51	4.63	4.83	88.03

TABLE VII. CORRECT CLASSIFICATION RATIO BY COMBINING NI-LBP^{u4} AND CO-NI-LBP^{u4}

LBP type	Dimension	CCR (%)
LBP ^{u4} [21]	398	89.14
NI-LBP ^{u4} & Co-NI-LBP ^{u4}	423	91.70

of the feature vectors is considerably smaller than the conventional LBPs. We think this is an important advantage of our approach that can be exploited in large-scale databases of indoor places.

ACKNOWLEDGMENT

This work was partially supported by Japan Society for the Promotion of Science (JSPS), Grant-in-Aid for Scientific Research (A) (26249029).

REFERENCES

- [1] 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 (RAS)*, vol. 56, no. 6, pp. 493–502, June 2008.
- [2] A. Pronobis and P. Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, Saint Paul, MN, USA, May 2012. [Online]. Available: <http://www.pronobis.pro/publications/pronobis2012icra>
- [3] O. M. Mozos, C. Stachniss, and W. Burgard, "Supervised learning of places from range data using AdaBoost," in *IEEE International Conference on Robotics and Automation (ICRA)*, Barcelona, Spain, 2005.
- [4] S. Friedman, H. Pasula, and D. Fox, "Voronoi random fields: Extracting the topological structure of indoor environments via place labeling," in *In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India, 2007.
- [5] L. Shi, S. Kodagoda, and G. Dissanayake, "Laser range data based semantic labeling of places," in *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, Taipei, Taiwan, October 2010.
- [6] A. Ranganathan, "PLISS: Detecting and Labeling Places Using Online Change-Point Detection," in *Proceedings of Robotics: Science and Systems*, Zaragoza, Spain, June 2010.
- [7] A. Pronobis, O. M. Mozos, B. Caputo, and P. Jensfelt, "Multi-modal semantic place classification," *International Journal of Robotics Research*, vol. 29, no. 2–3, pp. 298–320, February–March 2010.
- [8] J. Wu and J. M. Rehg, "Centrist: A visual descriptor for scene categorization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 8, pp. 1489–1501, 2011.
- [9] E. Fazl-Ersi and J. K. Tsotsos, "Histogram of oriented uniform patterns for robust place recognition and categorization," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 468–483, 2012.
- [10] L. Liu, L. Zhao, Y. Long, G. Kuang, and P. Fieguth, "Extended local binary patterns for texture classification," *Image and Vision Computing*, vol. 30, no. 2, pp. 86 – 99, 2012. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0262885612000066>
- [11] A. Rottmann, O. M. Mozos, C. Stachniss, and W. Burgard, "Semantic place classification of indoor environments with mobile robots using boosting," in *National Conference on Artificial Intelligence (AAAI)*, Pittsburgh, PA, USA, 2005.

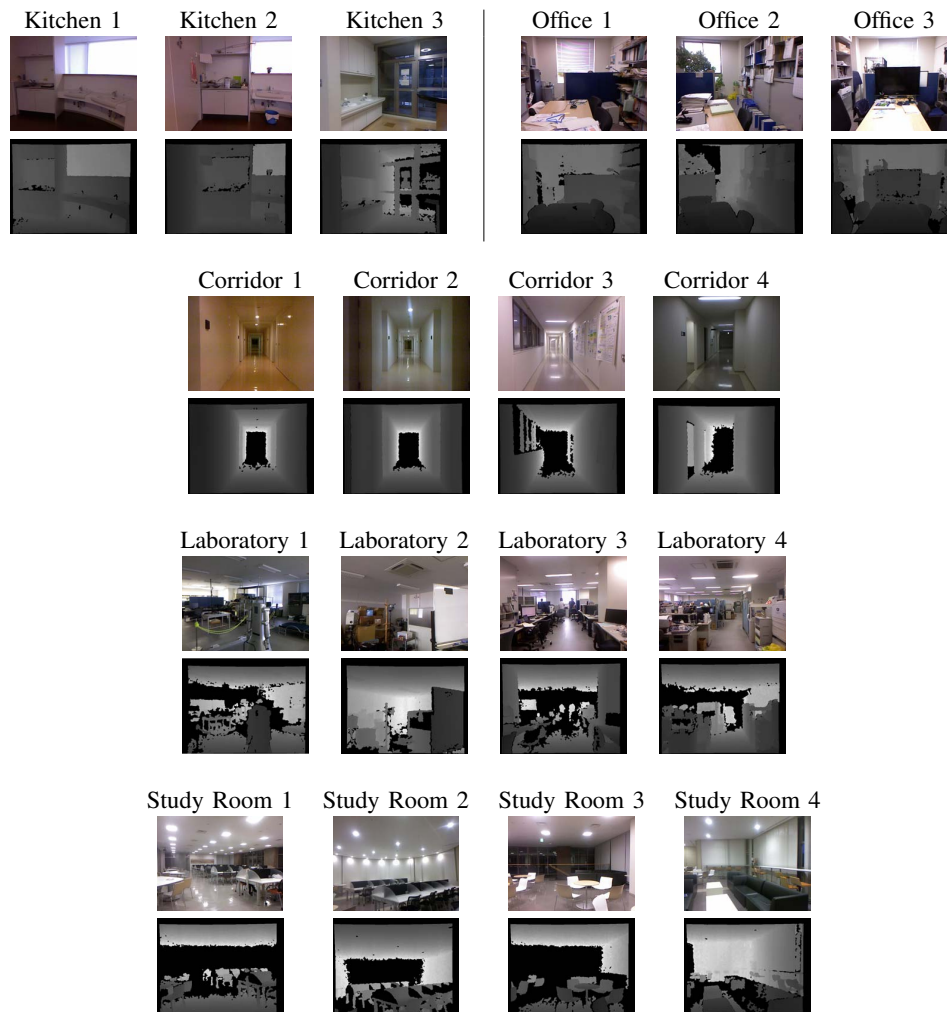


Fig. 6. Examples of synchronized RGB and depth images for different places and categories from our indoor dataset.

- [12] W.-T. Chu and C.-H. Chen, "Color centrist: a color descriptor for scene categorization," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (ICMR)*, Hong Kong, 2012.
- [13] L. Shi, S. Kodagoda, and R. Ranasinghe, "Fast indoor scene classification using 3d point clouds," in *Australasian Conference on Robotics and Automation (ACRA)*, Melbourne, Australia, December 2011.
- [14] T. Fiolka, J. Stueckler, D. A. Klein, D. Schulz, and S. Behnke, "Place recognition using surface entropy features," in *ICRA Workshop on Semantic Perception, Mapping, and Exploration*, Saint Paul, MN, USA, May 2012.
- [15] E. Fernandez-Moral, W. Mayol-Cuevas, V. Arevalo, and J. Gonzalez-Jimenez, "Fast place recognition with plane-based maps," in *IEEE International Conference on Robotics and Automation (ICRA)*, Karlsruhe, Germany, May 2013.
- [16] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, no. 5, pp. 6695–6711, May 2012.
- [17] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.
- [18] C. Cortes and V. Vapnik, "Support-vector network," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [19] 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.
- [20] 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, software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
- [21] O. M. Mozos, H. Mizutani, R. Kurazume, and T. Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, no. 6, pp. 6695–6711, 2012.
- [22] T. Leung and J. Malik, "Representing and recognizing the visual appearance of materials using three-dimensional textons," *Int. J. Comput. Vision*, vol. 43, no. 1, pp. 29–44, Jun. 2001. [Online]. Available: <http://dx.doi.org/10.1023/A:1011126920638>
- [23] M. Varma and A. Zisserman, "A statistical approach to texture classification from single images," *International Journal of Computer Vision*, vol. 62, no. 1-2, pp. 61–81, 2005.
- [24] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [25] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (surf)," *Computer vision and image understanding*, vol. 110, no. 3, pp. 346–359, 2008.