TWO-DIMENSIONAL LOCAL TERNARY PATTERNS USING SYNCHRONIZED IMAGES FOR OUTDOOR PLACE CATEGORIZATION

Hojung Jung, Ryo Kurazume, Yumi Iwashita*

Kyushu University, Graduate Faculty of
Information Science and Electrical Engineering
Fukuoka 819-0395, Japan

Oscar Martinez Mozos

University of Lincoln
School of Computer Science
Lincoln, LN6 7TS, United Kingdom

ABSTRACT

We present a novel approach for outdoor place categorization using synchronized texture and depth images obtained using a laser scanner. Categorizing outdoor places according to type is useful for autonomous driving or service robots, which work adaptively according to the surrounding conditions. However, place categorization is not straight forward due to the wide variety of environments and sensor performance limitations. In the present paper, we introduce a two-dimensional local ternary pattern (2D-LTP) descriptor using a pair of synchronized texture and depth images. The proposed 2D-LTP describes the local co-occurrence of a synchronized and complementary image pair with ternary patterns. In the present study, we construct histograms of a 2D-LTP as a feature of an outdoor place and apply singular value decomposition (SVD) to deal with the high dimensionality of the place. The novel descriptor, i.e., the 2D-LTP, exhibits a higher categorization performance than conventional image descriptors with outdoor place experiments.

***Index Terms*—** Two-dimensional Local Ternary Pattern (2D-LTP), Place categorization, Laser scanner, Reflectance image, Depth image

1. INTRODUCTION

Place categorization is an important capability for robots that allows them to identify their current type of location. This type of information can greatly improve communication between robots and humans[1, 2] and allows robots to make decisions with context-based understanding when completing high-level tasks[3]. Moreover, if a robot has the ability to categorize places according to type, then the robot will be able to properly execute a task even in unfamiliar surroundings. Furthermore, the scope of the present research can be extended to autonomous vehicles, so that decisions for safety driving can be made based on the environmental conditions.

Place classification has been investigated using various approaches, such as scene recognition, topological place

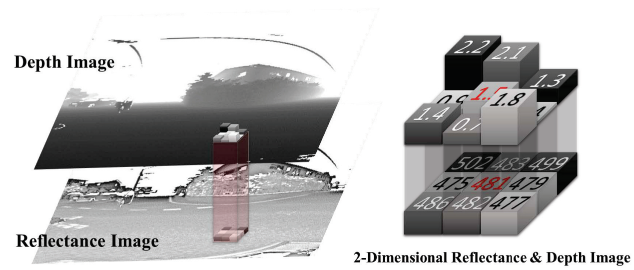


Fig. 1. A pair of synchronized reflectance and depth images obtained using a laser scanner

recognition, and place categorization. Using visual information, a place type can provide general information regarding location. A natural scene is recognized as a semantic label of an image by extracting features, such as GIST[4], SIFT[5], SURF[6] and PLISS[7]. Scene recognition is applied to representative scene images from websites or image libraries[4, 8]. The problem of outdoor place categorization has been addressed by an unclear categorization of places and a limited performance of a camera. Previous research on scene recognition assumed that RGB image data is obtained under proper illumination conditions. However, in real outdoor environments, images are taken under various conditions, such as bright sunlight, darkness, and sudden illumination changes. Therefore, robustly categorizing outdoor environments is an open problem.

In the present paper, we proposes a novel feature descriptor for place categorization using synchronized texture and range images obtained using a laser scanner. In the proposed approach, we extract two-dimensional local ternary patterns (2D-LTP) from two complementary and synchronized image models. The spatial synchronized texture and range image models can be provided by a laser scanner as reflectance and depth point information. The proposed novel descriptor can describe local patterns in texture and range images simultaneously and can obtain more detailed information than the local binary pattern (LBP), which is a state-of-the-art descriptor.

In order to describe the synchronized place images, we

*hojung@irvs.ait.kyushu-u.ac.jp

978-1-4799-5751-4/14/$31.00 ©2014 IEEE

create histograms of the LTP values. These histograms represent global feature vectors for synchronized texture and range images. Finally, we reduce the numbers of dimensions of the images by singular value decomposition (SVD). Moreover, we apply the average intensity value of neighborhood pixels, called the NI-LBP[9], to the proposed method. The integrated reflectance and depth NI-LBP, the 2D-LTP descriptor, provided the best classification results with very-low-dimensional feature vectors.

2. RELATED RESEARCH

The application of place categorization to robotics has gained interest in recent years. Several approaches have been investigated in an attempt to solve the classification problem.

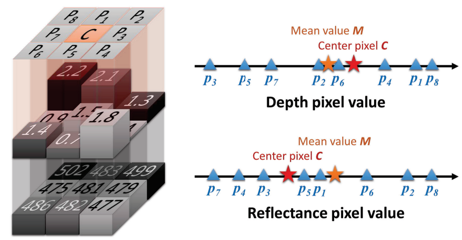
Place categorization has been investigated using different types of sensors, such as vision cameras[7, 10, 11, 12, 13], laser scanners[14, 15, 16], and RGB-D sensors[17, 18]. Recently, researchers have started to use multi-image models, such as gray-scale, reflectance, and depth images, obtained using RGB-D sensors[19], cameras, and laser scanners[20]. However, previous multi-image models simply concatenate individual models. The proposed method simultaneously describes a pair of synchronized images using a single laser scanner.

There are several approaches to indoor place categorization that use different types of feature descriptors, such as GIST[4], SIFT[5], and CENTRIST[12]. Originally, the CENTRIST descriptor was introduced for indoor environments using gray-scale images[10] but has since been extended to the HSV color space[21]. Finally, the LBP has been used to categorize indoor and outdoor places[13, 19, 20].

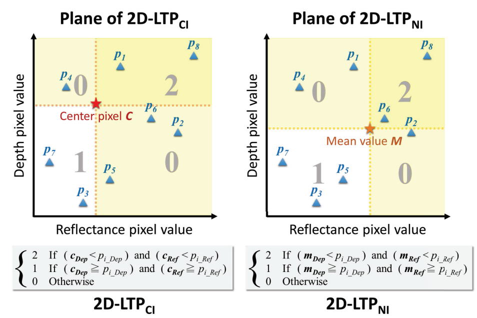
In the present paper, we adopt the LBP philosophy for describing local image pixels, but we describe the complementary texture and depth images simultaneously by assigning a ternary pattern. An LTP descriptor using a threshold parameter with a single texture image was investigated in a previous study[22]. However, the 2D-LTP proposed in this paper, describing a pair of synchronized images, is completely a different descriptor except using a concept of ternary pattern. The primary advantage of the proposed method is that rich local patterns of features can be obtained in a simple manner using a pair of synchronized images. Moreover, the performance of the 2D-LTP descriptor is verified to be highly reliable through outdoor experiments using a laser scanner.

3. FEATURE EXTRACTION

In the present paper, we use a laser scanner to generate reflectance and depth images from the same laser pulse. The reflectance and depth images as shown in Fig. 1 are fundamentally synchronized with respect to individual pixels. The depth image can describe the structure of the target and the reflectance image indicates the texture of the target. Therefore, these images are synchronized, but complementary.



(a) Local binary patterns for reflectance and depth images



(b) Two-dimensional local ternary pattern plane

Fig. 2. Overview of 2D Local Ternary Patterns

3.1. 2D Local Ternary Patterns

We propose the two-dimensional local ternary pattern (2D-LTP), a novel descriptor, for a pair of synchronized images. The 2D-LTP operator, which is a local transformation, indicates the relationships between the values of pixels and their neighboring pixels.

Let $I_{Ref}(i), I_{Dep}(i)$ be the value of pixel $i = (x, y)$ in a pair of reflectance and depth images I_{Ref}, I_{Dep}, and let $\mathcal{N}_{p_{Ref}}(i), \mathcal{N}_{p_{Dep}}(i)$ define the pixels in its P-neighborhood. Then, the $\mathcal{LTP}_{CI}$ operator compares the individual reflectance and depth pixel values, $I_{Ref}(i)$ and $I_{Dep}(i)$, respectively, with the corresponding values ($I_{Ref}(j)$ and $I_{Dep}(j)$) for every pixel in its neighborhood, $j \in \mathcal{N}_{p_{Ref}}(i), \mathcal{N}_{p_{Dep}}(i)$. On the other hand, the $\mathcal{LTP}_{NI}$ operator compares the mean values in neighborhood with the corresponding.

Figure 2(a) shows an image divided in a conventional manner into reflectance and depth data using a local binary pattern (LBP). As shown in Fig. 2(b), we can create a 2D-LTP plane. Intuitively, we set the value of every neighboring pixel j to 2 if $x_{Ref} \geq 0$ and $x_{Dep} \geq 0$, or to 1 if $x_{Ref} < 0$ and $x_{Dep} < 0$. However, since a pair of synchronized images is used, we assign as another value 0 for conflicting conditions of x_{Ref} and x_{Dep}. The conflicting condition contains ($x_{Ref} \geq 0$ and $x_{Dep} < 0$) and ($x_{Ref} < 0$ and $x_{Dep} \geq 0$). If we assign the condition individual, we can define it as local quaternary pattern (LQP).

The obtained ternary values are concatenated clockwise and are transformed into a corresponding decimal value d in the range of $[0, \cdots, 6561]$. This decimal value is then assigned to pixel i in the resulting transformed image I_{LTP}. Formally, we have

$$I_{LTP}(i) = \sum_{j=1}^{P} s(I(j) - I(i))3^{j-1}, \forall j \in N_p(i), \quad (1)$$

$$s(x) = \begin{cases} 2 & \text{if } x_{Ref} \geq 0 \text{ and } x_{Dep} \geq 0 \\ 1 & \text{if } x_{Ref} < 0 \text{ and } x_{Dep} < 0 \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

where P indicates the P-neighboring pixels of i. In the present paper, we adopted $P = 8$. An example of a local ternary transformation is shown in Fig. 2.

Algorithm 1: Two-dimensional LTP

Input:
– Reflectance and depth images
– Number of bins for 2D-LTP

Output:
– Array of two-dimensional ternary pattern vectors

Definition:
– A local pixel structure contains:
1) Center pixel i
2) Neighborhood pixel j
3) Average of neighborhood pixels j_{avg}

Algorithm:
1) create local pixel structures for each images
2) **for** *each pixel* i **do**
    A) create the 2D-LTP plane
    B) **for** *each neighborhood pixel* $\mathcal{N}_p(i)$ **do**
        a) define the reference; center pixel for $\mathcal{LTP}_{CI}$, mean value for $\mathcal{LTP}_{NI}$
        b) assign the 2D-LTP neighborhood value
        c) calculate the ternary pattern value
3) create a histogram of 2D-LTP images: H_{global}
4) reduce the dimensionality using the SVD: H_{svd}

3.2. Global Feature Model

As in the conventional method, we use the histogram of a LBP [19]. The histogram can represent a global feature for each image. For reflectance and depth images, the number of 2D-LTP feature dimensions is $3^8 = 6,561$ bins. In addition, we define any out-of-range and non-returnable value obtained using the laser scanner as a Not-A-Number (NAN) value. Therefore, we assigned an additional NAN value to the histogram of bins, but not as 2D-LTP feature values. In summary, the total number of 2D-LTP histogram bins is 6,562.

$$H_{global} = \{H_{2D-LTP}, NAN\} \quad (3)$$

Furthermore, we reduce the global feature model H_{global} into H_{svd} by applying singular value decomposition (SVD) to deal with its high-dimensional features. The final dimension of the 2D-LTP is smaller than that of the conventional concatenating LBP feature[20].

4. CLASSIFICATION

In the present paper, we used a supervised learning approach based on support vector machines (SVMs). The feature vectors of multiple image models are used as labeled inputs. In the present study, we apply a one-against-one approach, in which a SVM is learned for each pair of categories. In our experiments, we use the implementation given by the LIBSVM library[23]. The input feature vectors are first normalized in the range of 0 to 1. Moreover, the parameters C and γ are selected by a grid search using cross-validation. The ranges of C and γ are $C \in [2^0, ..., 2^{18}]$ and $\gamma \in [2^{-20}, ..., 2^0]$.

5. EXPERIMENT

In the proposed approach, we use a single laser scanner, which provides range data by measuring the round-trip time of a laser pulse reflected by an object. In addition to range data, the laser scanner can measure the strength of the reflected laser pulse, i.e., the reflectivity. The laser scanner used in the present study is a SICK LMS151 laser with a maximum range of 50 meters and an angular resolution of 0.25 degrees. In the configuration of the present study, we rotate the laser around the vertical axis in order to obtain a complete panoramic range image, as shown in Fig. 3. A complete panoramic scan has a resolution of 3,753 $\times$ 760 points.

In order to evaluate the performance, we have created a dataset of four different outdoor categories: forest, residential, parking, and urban in Fig. 4. For each of these four categories, we made panoramic 3D laser scans covering 360 degrees with the laser positioned 95 cm above the ground. In capturing the panoramic scans, we situated the laser scanner in five to seven different locations for each type of place scanned. For example, we placed the laser scanner on a straight road, at a corner, and at an intersection. Some examples of panoramic scans are shown in Fig. 5. Table 1 shows that each category contains 35 pairs of images corresponding to different locations that belong to one specific category, and 143 pairs are obtained in total.

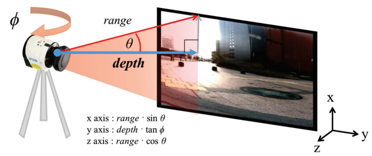


Fig. 3. Depth image extraction from a laser scanner

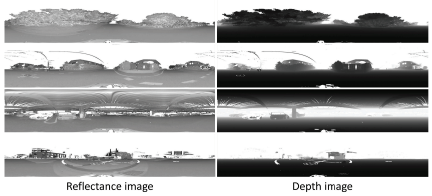


Fig. 4. Four types of outdoor places: forest, residential area, parking lot, and urban area from top to bottom (left column is reflectance images and right column is depth images)

6. RESULTS

Table 2 shows the correct classification rates (CCRs) for the conventional LBP, the LBP with the uniformity criterion ($U = 4$) [19], the proposed method ($2D\text{-}LTP_{CI}$ and $2D\text{-}LTP_{NI}$), and the $2D\text{-}LQP$ using depth and reflectance images. The proposed $2D\text{-}LTP_{NI}$ provided a better CCRs than the $2D\text{-}LQP_{NI}$, which assigned four different sections individually. The first 25 eigenvalues of the $2D\text{-}LTP$ and $2D\text{-}LQP$ account for 97% of the variance. We verified that the proposed $2D\text{-}LTP_{NI}$ image has a better categorization performance by reducing the number of dimensions through singular value decomposition. Table 3 lists the values for $2D\text{-}LTP_{CI}$ and $2D\text{-}LTP_{NI}$ for various dimensionalities. This table indicates that a final dimension of 25 provides the best categorization results when using $2D\text{-}LTP_{NI}$.

7. CONCLUSION

In the present paper, we presented the 2D-LTP, a novel feature descriptor for categorizing outdoor environments using a laser scanner. The results of the experiments indicate that the classification performance of a 2D-LTP image model exceeds that of a LBP image model. Moreover, even in the case of using only 25 feature dimensions, we were able to classify different types of outdoor places with high accuracy. The 2D-LTP descriptor is robust and exhibits the best categorization performance for outdoor environments.

Table 1. Dataset of outdoor places containing 143 pairs of reflectance and depth images obtained using a laser scanner

Category	Number of images by location							Total
Forest	4	2	3	6	7	6	8	36
Residential	5	5	4	4	13	0	0	31
Parking	6	6	8	8	4	0	0	32
Urban	5	5	5	7	8	6	8	44
Total number of place images								143

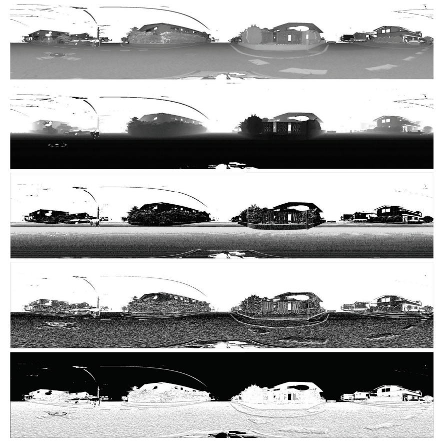


Fig. 5. Residential area image: reflectance, depth, LBP_{Ref}, LBP_{Dep}, $2D\text{-}LTP_{NI}$ (in order)

Table 2. CCR of Reflectance and Depth Images [%]

Feature	Pattern	Number of dimensions	Accuracy
LBP[20]	Binary	514	79.42
LBP_{u4}[20]	Binary	398	84.09
$2D\text{-}LTP_{CI}$	Ternary	25	81.93
$2D\text{-}LTP_{NI}$	Ternary	25	**93.49**
$2D\text{-}LQP_{CI}$	Quaternary	25	86.62
$2D\text{-}LQP_{NI}$	Quaternary	25	91.15

8. ACKNOWLEDGEMENT

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

9. REFERENCES

[1] Hendrik Zender, O Martínez Mozos, Patric Jensfelt, G-JM Kruijff, and Wolfram Burgard, "Conceptual spatial representations for indoor mobile robots," *Robotics and Autonomous Systems*, vol. 56, no. 6, pp. 493–502, 2008.

Table 3. Comparison of CCR for $2D\text{-}LTP$[%]

Number of dimensions	25	50	75	100
$2D\text{-}LTP_{CI}$	81.93	81.16	83.80	86.65
$2D\text{-}LTP_{NI}$	**93.49**	83.92	84.65	87.49
$2D\text{-}LQP_{CI}$	85.48	85.05	87.32	89.76
$2D\text{-}LQP_{NI}$	91.15	83.30	86.70	84.46

[2] Andrzej Pronobis and Patric Jensfelt, "Large-scale semantic mapping and reasoning with heterogeneous modalities," in *Robotics and Automation (ICRA), 2012 IEEE International Conference on*. IEEE, 2012, pp. 3515–3522.

[3] Cyrill Stachniss, Óscar Martínez Mozos, and Wolfram Burgard, "Efficient exploration of unknown indoor environments using a team of mobile robots," *Annals of Mathematics and Artificial Intelligence*, vol. 52, no. 2-4, pp. 205–227, 2008.

[4] Aude Oliva and Antonio Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[5] Stephen Se, David Lowe, and Jim Little, "Vision-based mobile robot localization and mapping using scale-invariant features," in *Robotics and Automation, 2001. Proceedings 2001 ICRA. IEEE International Conference on*. IEEE, 2001, vol. 2, pp. 2051–2058.

[6] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool, "Surf: Speeded up robust features," in *Computer Vision–ECCV 2006*, pp. 404–417. Springer, 2006.

[7] Ananth Ranganathan, "Pliss: Detecting and labeling places using online change-point detection.," in *Robotics: Science and Systems*, 2010.

[8] Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. IEEE, 2006, vol. 2, pp. 2169–2178.

[9] Li Liu, Lingjun Zhao, Yunli Long, Gangyao Kuang, and Paul Fieguth, "Extended local binary patterns for texture classification," *Image and Vision Computing*, vol. 30, no. 2, pp. 86–99, 2012.

[10] Jianxin Wu, *Visual place categorization*, Ph.D. thesis, Georgia Institute of Technology, 2009.

[11] Andrzej Pronobis, O Martínez Mozos, Barbara Caputo, and Patric Jensfelt, "Multi-modal semantic place classification," *The International Journal of Robotics Research*, vol. 29, no. 2-3, pp. 298–320, 2010.

[12] Jianxin Wu and Jim M Rehg, "Centrist: A visual descriptor for scene categorization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1489–1501, 2011.

[13] Ehsan Fazl-Ersi and John 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.

[14] O Martınez Mozos, Cyrill Stachniss, and Wolfram Burgard, "Supervised learning of places from range data using adaboost," in *Robotics and Automation, 2005. ICRA 2005. Proceedings of the 2005 IEEE International Conference on*. IEEE, 2005, pp. 1730–1735.

[15] Stephen Friedman, Hanna Pasula, and Dieter Fox, "Voronoi random fields: Extracting topological structure of indoor environments via place labeling.," in *IJCAI*, 2007, vol. 7, pp. 2109–2114.

[16] Lei Shi, Sarath Kodagoda, and Gamini Dissanayake, "Laser range data based semantic labeling of places," in *Intelligent Robots and Systems (IROS), 2010 IEEE/RSJ International Conference on*. IEEE, 2010, pp. 5941–5946.

[17] Lei Shi, Sarath Kodagoda, and Ravindra Ranasinghe, "Fast indoor scene classification using 3d point clouds," in *Proc. Australasian Conference on Robotics and Automation, Melbourne, Australia (December 2011)*, 2011.

[18] E Fernández-Moral, W Mayol-Cuevas, V Arévalo, and J González-Jiménez, "Fast place recognition with plane-based maps," in *Robotics and Automation (ICRA), to appear in 2013 IEEE International Conference on*, 2013, pp. 5210–5215.

[19] Oscar Martinez Mozos, Hitoshi Mizutani, Ryo Kurazume, and Tsutomu Hasegawa, "Categorization of indoor places using the kinect sensor," *Sensors*, vol. 12, no. 5, pp. 6695–6711, 2012.

[20] Oscar Martinez Mozos, Hitoshi Mizutani, Hojung Jung, Ryo Kurazume, and Tsutomu Hasegawa, "Categorization of indoor places by combining local binary pattern histograms of range and reflectance data from laser range finders," *Advanced Robotics*, vol. 27, no. 18, 2013.

[21] Wei-Ta Chu and Chih-Hao Chen, "Color centrist: a color descriptor for scene categorization," in *Proceedings of the 2nd ACM International Conference on Multimedia Retrieval*. ACM, 2012, p. 33.

[22] Xiaoyang Tan and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[23] Chih-Chung Chang and Chih-Jen Lin, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, pp. 27, 2011.