applied sciences

MDPI

Article

Visual Place Recognition of Robots via Global Features of Scan-Context Descriptors with Dictionary-Based Coding

Minying Ye * and Kanji Tanaka

Human and Artificial Intelligent System Course, Graduate School of Engineering, The University of Fukui, Fukui 910-8507, Japan; tnkknj@u-fukui.ac.jp
* Correspondence: yymm2280@g.u-fukui.ac.jp

Abstract: Self-localization is a crucial requirement for visual robot place recognition. Particularly, the 3D point cloud obtained from 3D laser rangefinders (LRF) is applied to it. The critical part is the efficiency and accuracy of place recognition of visual robots based on the 3D point cloud. The current solution is converting the 3D point clouds to 2D images, and then processing these with a convolutional neural network (CNN) classification. Although the popular scan-context descriptor obtained from the 3D data can retain parts of the 3D point cloud characteristics, its accuracy is slightly low. This is because the scan-context image under the adjacent label inclines to be confusing. This study reclassifies the image according to the CNN global features through image feature extraction. In addition, the dictionary-based coding is leveraged to construct the retrieval dataset. The experiment was conducted on the North-Campus-Long-Term (NCLT) dataset under four-seasons conditions. The results show that the proposed method is superior compared to the other methods without real-time Global Positioning System (GPS) information.

Keywords: visual place recognition; 3D point clouds; scan context; CNN classification; feature extraction

Citation: Ye, M.; Tanaka, K. Visual Place Recognition of Robots via Global Features of Scan-Context Descriptors with Dictionary-Based Coding. *Appl. Sci.* **2023**, *13*, 9040. https://doi.org/10.3390/app13159040

Academic Editors: Dimitris Mourtzis and Yutaka Ishibashi

Received: 26 April 2023
Revised: 28 July 2023
Accepted: 31 July 2023
Published: 7 August 2023

1. Introduction

Visual place recognition in robots can be quickly and accurately realized by a pre-built environment map. Many researchers have studied this topic in-depth. Some of their research is based on image sensors, rich texture and object information obtained through analysis [1–4]. In recent years, the research object has been changed from a two-dimensional image to three-dimensional Lidar data. Three-dimensional Lidar technology provides a new means of collecting information. Through the movement of the Lidar scanner, large-scale and high-precision 3D data in the target area can be collected directly. Furthermore, this technology has the advantages of low measurement dependence, a high degree of automation and is little influenced by weather conditions. It can also fully reflect the characteristics of the survey place [5–8].

Many researchers have analyzed 3D Lidar for place recognition [9–12]. They mainly focused on how to use the image sensor method on 3D Lidar. The original scan context [13] converted from the 3D Lidar data has achieved successful place recognition. The scan context converts the 3D Lidar data into a compact descriptor. The height information of the 3D point clouds is very characteristic and is affected much less by environmental changes. It exhibits good performance in robot self-localization. In addition, the scan context was optimized in a paper from 2019 [14]. It has been made more suitable for inputting into convolutional neural network (CNN) training. Therefore, we chose this descriptor as the input for our system. Although the researchers considered the rotation of the 3D Lidar in their article, they just rotated the training 3D point clouds by 180 degrees to increase the training samples. However, it was observed that there were very different scan-context descriptors within the same region, and very similar scan context ones in different regions.

The above problems are because the conventional methods divide the original descriptors based on the geographical location. There are bound to be some limitations and irrationality for these 2D descriptors compared with the 3D Lidar [15,16]. Although the scan context descriptor extracts and retains important features of the 3D point cloud, it is also influenced by non-important objects and the distance between the vehicles at two sampling moments, and so causes some perceptual confusion. Therefore, it is difficult to predict the set to which the data collected at the boundaries of each region belongs. In addition, some very similar descriptors appear in non-adjacent areas which can cause perceptual aliasing and misrecognition.

To solve this problem, in this study, we proposed a Lidar place recognition system, called the Based on Global features-system (BGF-system), which is shown in Figure 1. We continue to use the improved version of the scan context and leverage it into our system. Our idea is to introduce a reclassification system based on global features. This method eliminates the influence of similar descriptors in different regions. After the global feature is extracted, it reclassifies the descriptor. Furthermore, for practical applications, we encode the region ID under the same feature set through dictionary-based coding [17,18]. The result is a single compact region ID word, in contrast to the multi-word representation of the method named BoW [19,20]. Thus, the performance improves compared to the improved scan context method. As the retrieval dataset has been transformed into text information, our retrieval system is lightweight and its speed is greatly enhanced under similar conditions. The contribution of our study is as follows:

- The accuracy of Lidar place recognition has been improved according to the global features of the descriptor. This can weaken the influence of different regions but similar descriptors.
- The retrieval dataset has been transformed into text information by dictionary-based coding. We simplify the final retrieval dataset, which can significantly improve the retrieval speed.
- Four-seasons datasets (North Campus Long-Term (NCLT) datasets [21–23]) were evaluated. The proposed method was successful for place recognition using the four-seasons dataset.

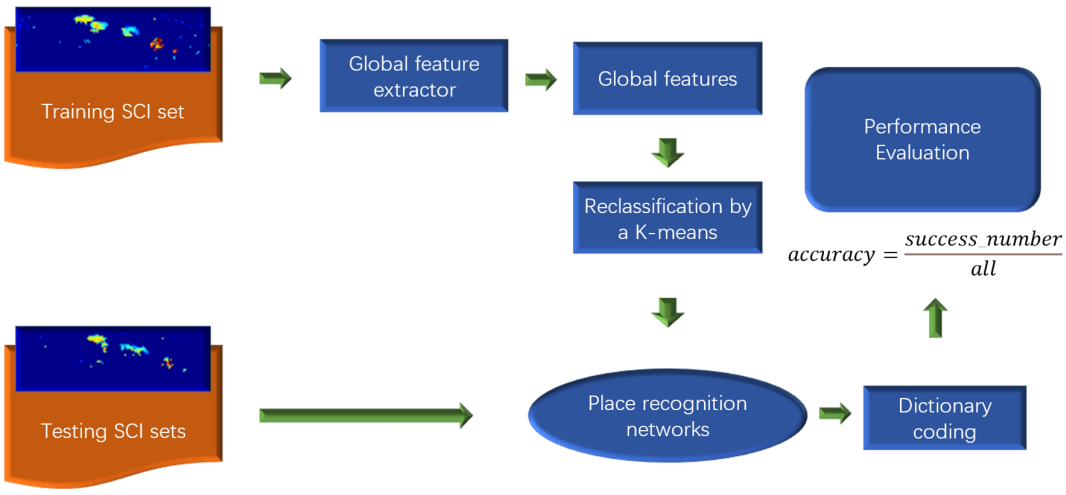

Figure 1. The pipeline of the place recognition method using the proposed global features (BGF) based on the scan-context image (SCI).

2. Related Works

The important issue of visual robot self-localization based on 3D point cloud data has been exploited by many researchers. In reference [24], the authors suggested a novel global descriptor, M2DP. They projected a 3D point cloud to multiple 2D planes and obtained the density features. They used these singular vectors of features as the descriptor. Yara Ali

alnaggar et al. [25] proposed a novel multi-projection fusion system, which uses spherical and aerial projection, and results in fusion using a soft voting mechanism to segment point cloud semantics. It has high throughput and can achieve improved segmentation results compared to the single projection method. In reference [26], Jacek Komorowski proposed a new 3D structure descriptor, named minkloc3d, based on a sparse voxelized representation and 3D Feature Pyramid Network [27]. Many experiments show that this is superior to the existing location recognition methods based on the point cloud. These methods have two advantages. First, the sparse convolution structure generates information-rich local features which construct global point cloud descriptors. Second, the improved training process can train large databases and the distinguishability and generalization ability of the obtained descriptors are positively affected. Lun Luo et al. [28] introduced a new descriptor called bird's-eye view feature transform (BVFT). They leveraged the BVFT descriptor to recognize the place and RANSAC to predict the poses of the 3D point cloud. They were inspired by the scan context and other papers. They used the height information of the 3D point cloud as a key point to form the BV image. The SIFT feature was extracted for pose estimation. However, the time complexity of the project was quite high. Vinicio Rosas-Cervantes et al. [29] proposed an automatic localization system for a mobile robot for collecting large-scale site uneven surface maps. They proposed a three-dimensional localization algorithm for mobile robots in non-uniform and unstructured environments by combining the point cloud and Monte Carlo algorithm. The processing of the point cloud includes the 3D point cloud obtained by projection and a 2D feature is generated from the 3D point cloud. These features are used to rebuild the robot environment map. For real-time place recognition, the robot system uses an occupancy grid map and two-dimensional features as inputs, combined with the Monte Carlo algorithm. Xuyou Li et al. [30] proposed a 3D-point cloud segmentation method, which divided the earth point cloud, then clustered it on the ground, combining it with the 2D image of the location for Iterative Closest Point (ICP) matching preprocessing. Rafael Barea et al. [31] proposed a localization method combining camera images and 3D Lidar data by using the Karlsruhe Institute of Technology and Toyota Technological Institute (KITTI) database. The 2D images were segmented by ERFNet, which was a semantic segmentation CNN, then the 3D point clouds were analyzed to obtain the box. Finally, the localization of the aerial view was obtained. Giseop Kim et al. [14] proposed the scan context method, by which a 3D point cloud is converted into a 2D image and maintains certain point cloud features. It effectively transforms the 3D problems into 2D ones. Xuecheng Xu et al. [32] proposed the Differentiable Scan Context with Orientation (DiSCO) system. The relative direction and position of the candidates can be estimated simultaneously. While this considers the direction problem, the problem of the scan-context descriptor is still not solved because some features of the 3D point cloud are still missing when the 3D Lidar data is converted to 2D descriptors.

Although many studies have focused on 3D Lidar data visual place recognition, there are still some valuable unknown areas which are not covered. The scan-context descriptors largely retain the characteristics of the 3D Lidar scanning results, but when dividing the regions, many descriptors at different regions have similar features, which can have a great impact on the results [16]. As we showed in Part 4.3 of the evaluation results section. Our study resolves this by bringing a reclassification using image global features.

3. Approach

In this chapter, we introduce a Robot Place Recognition system. This system primarily uses the extracted global features and re-classification of image descriptors to improve the results because it realizes the place recognition of the robot based on a 2D image descriptor. Therefore, we first introduce the conversion part of the image descriptor. Next, we introduce the re-classification based on the global features of image descriptors. The localization results are obtained by analyzing the data of the four seasons for the NCLT dataset. All the processes are shown in Figure 1.

3.1. Generation of Descriptors

To ensure that the images input into the neural network can be used to obtain good training and prediction results, we have been inspired by the recent development of the scan-context descriptor in article [14], which can ensure that some features of the 3D point cloud can be retained. Compared to the old version of the scan context [13], this descriptor is improved and is more suited for CNN input. The scan context converts the 3D point cloud data from a Cartesian coordinate system into a polar coordinate system. The highest value of z-coordinate of each angle is kept and plotted into the descriptors. This method divides a 3D point cloud into azimuthal and radial bins. N_s is the number of sectors and N_r is the number of rings. We set the maximum sensing range of the Lidar sensor as $L_{\max}$. In this study, $N_s = 120$, $N_r = 40$ and $L_{\max} = 80$. Thus, we obtained the max z value at each bin. The bin encoding formula is $B_{ij} = \max_{p \in B_{ij}} z(p)$, where z is the z coordinate value of a point p. Finally, a scan-context descriptor I is shown as Equation (1).

$$I = (a_{ij}) \in \mathbb{R}^{N_r \times N_s}, a_{ij} = \mathrm{B}_{ij} \tag{1}$$

To facilitate the evaluation of the accuracy of robot place recognition, we divide the robot's operation range into small areas of 10 m × 10 m and 3D point clouds are allocated to different area sets according to the GPS information. Comparing the training dataset and testing dataset, we may find some area sets that do not appear in the training dataset. These are defined as invisible places while the others are visible places. The image size is 40 × 120 and is saved as a jet image. This method is convenient for the subsequent steps, which are feature extraction and result evaluation.

3.2. Global Feature Extraction by CNN and Reclassification

To generate global features from the image descriptors, we convert image data into a digital feature matrix through a neural network. The VGG19 [33–35] is a Convolutional Neural Network (CNN) which is a scientific research tool developed by the Oxford University in 2015. The VGG19 is famous because it has strong feature extraction ability. In our system, the neural network used for feature extraction is VGG19.

As shown in Figure 2, there are a total of 19 layers in VGG19, including 16 layers in the convolution layer and the last three layers in the full connection layer. The convolutional layers in the middle part are connected to a max-pool layer. Finally, the last layer is the soft-max layer. Then, the global features of the image descriptors are extracted at the second full connection layer of the VGG19. Therefore, the feature is a vector of size 1 × 4096. In this experiment, we do not need to train the model. We use the VGG19 model directly and the VGG19 uses the pre-trained weights on the dataset ImageNet [36,37]. The input image size is 224 × 224. Therefore, to get a better performance, we set the input image size to 224 × 224. At this point, we do not need to input the label information for the training model. Therefore, we only use the scan-context descriptor as VGG19 input, and the feature is output.

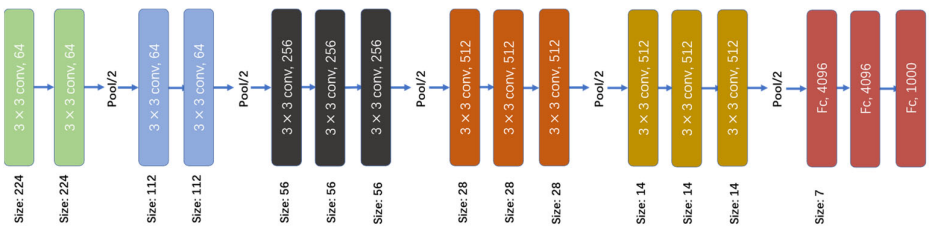

Figure 2. The structure of VGG19.

The next step is re-classification based on the extracted global features. If we directly apply the extracted features to the next step, it will cause unnecessary calculations for time and space. Therefore, we used the following method to reduce feature space dimension.

In our project, we chose random projection to reduce the feature matrix. The dimension reduction matrix is randomly generated to reduce the size of the feature vector from 1×4096 to 1×256. From there, we use the k-means algorithm [38–40] to classify these scan-context descriptors through the dimension-reduced feature vectors. We set category K to 2000, which is the number of clusters obtained by K-means. If K is set very low, this method's advantage will not be obvious. The data of the training dataset is re-distributed to 2000 clusters. Next, we use this result as the input to a neural network for training.

3.3. Deep-Learning Network

Here, the K-classification training dataset is trained to obtain the CNN model, which is then used to obtain label information prediction based on the test datasets (the four-seasons dataset). According to the flow chart in Figure 1, the initial modeling of the training database is calculated to be established by the CNN. Then, the VGG-16 [41–43] network based on transfer learning is used to train the model. Then, we selected the four-seasons dataset as a test dataset to predict and obtain each test descriptor's K-classification value. Compared to the scan-context paper, the input to the CNN is changed. The CNN input image size was not set to be 40×120, but 128×128, because of the characteristics of the VGG-16 network. The model is trained with a batch size of 8 and the SGD optimizer is used with certain parameters (learning rate = 0.0001, momentum = 0.9). The detailed structural information is shown in Table 1.

Table 1. CNN network.

In	(batch_size, 128, 128, 3)
Conv	VGG-16 network
Conv1	Flatten (input_shape = vgg16.output_shape)
Fc1	ReLU (FC (256, Conv1))
Fc2	Softmax (FC (N, Dropout (Fc1)))
Out	(batch_size, N)

3.4. Place Recognition

In the NCLT database, the path that the robot visits at different times may not be the same. There may be some newly visited locations, so our place recognition effect will inevitably be affected. Therefore, dealing with new unknown locations makes a difference. According to Kim et al. [14], we define these new locations as invisible places and identify those places existing in the dataset as visible places. The main difference is that they use the information entropy algorithm to improve place recognition results. In our study, we try to reduce this step to simplify the complexity of the algorithm and ensure the accuracy of the results.

This method of place recognition has to integrate the division of visible places into the algorithm. The place recognition results are obtained after the joint analysis of the two items. First, the original label information of the image descriptors and the labels obtained by re-classification by K-means are combined to build a document retrieval database. The retrieval speed is accelerated by constructing a dictionary structure. We take the K-means classification category as the key and the original label of the images as the values for retrieval. Second, the retrieval database is searched based on these results. We compare the place labels with the same prediction as that of the CNN, and sort them by the repetition rate of the search results. Top_1 and top_5 can be selected for comparison and analysis. If the results show the correct place number, the place recognition is successful.

4. Experiments

4.1. Benchmark Datasets

In this article, we adopted the NCLT database, which is a Long-Term Vision, and the Lidar dataset collected at the University of Michigan North Campus. Even though the same place is visited again and again, the path of the robot car is different in each session and the driving time of each session is also different from morning to dusk.

We normalized the 3D point cloud into the image with a size of 40 × 120 in polar coordinates. We retained the partial features of the current 3D point cloud via different pixel values at different heights in the scan-context image descriptor.

We divided the map of the robot activities into a small grid of 10 m × 10 m. We extracted the point cloud data at every meter. Then, we pre-processed it and directly converted it into a scan-context image. We chose a database with a relatively large volume for training. Finally, the dataset of the other four seasons was compared and predicted. In Table 2, we summarize the number of visible and invisible places.

Table 2. Description of datasets.

Datasets	Training		Testing			
	15 January 2012		8 January 2012	17 March 2012	4 August 2012	28 September 2012
NCLT	579 places	visible	6171	5449	5464	4626
		invisible	292	428	490	919

4.2. Baseline Method

4.2.1. Scan Context

Each scan context [14] is compared with our system and the results are obtained. In the same method, the retrieval database is constructed by the single-day dataset with simple information data. The other four-seasons databases are selected for comparison to prove the advantages and practicality of our system.

The scan-context image [14] has been partially improved from the original version of the scan-context method [13]. From the method without a basic learning function to the one with a learning function, the concept of information entropy is integrated to optimize the evaluation indicators and obtain significant results.

4.2.2. Pole Extraction

The pole extraction method [44] has been suggested. This method is that robot localization is via a Pole point extracted from 3D point cloud. It can run online without a GPU. The authors validated the method on the NCLT dataset.

4.3. Experimental Results

In this chapter, the effectiveness and stability of our proposed method are proven by in-depth analysis. We display the specific information of the training dataset and test dataset in Table 3. The data of the 3D point clouds were collected at every meter and are divided into visible and invisible parts. After re-classification, we can divide the images into new classifications based on their global features.

Table 3. Result of Accuracy.

NCLT Dataset		8 January 2012	17 March 2012	4 August 2012	28 September 2012
Accuracy	BGF-system (Top_5)	0.9033	0.7962	0.6817	0.7187
	Scan-context (Top_1)	0.6924	0.6206	0.5644	0.5443
	Scan-context (Top_25)	0.8004	0.7507	0.7313	0.7122
	Pole_extraction	0.3725	0.3462	0.2888	0.2477

4.3.1. Comparison of Accuracy

The experimental results are evaluated for accuracy and are used to test and evaluate the four-seasons database selected from the NCLT database. The correct matches are known as true positives and the invisible place number is known as the true negative. The accuracy is the ratio of the number of correctly predicted samples to all the samples [45,46], that is Equation (2). In a perfect system, the *accuracy* would be closer to one.

$$accuracy = \frac{TP + TN}{all} \tag{2}$$

As seen in Table 2, there are invisible parts in our dataset, so we obtained an accurate estimate of this result. The specific method is shown in Table 3. When the visible part and the test data are consistent with the actual place recognition, it is called place recognition success. Unlike the scan-context method, which can extract Top_25, we only used the rank Top_1 of the CNN output. We also used only the CNN result for the search of the retrieval dataset. We can already get a good recognition accuracy when we take Top_5. After comparison, our experimental results show some improvement in place recognition accuracy under the four seasons of the dataset.

From Table 3 and Figure 3, we observe that the accuracy, under Top_5, of our method, is better than that of the scan-context Top_25. Only on 4 August 2012 did the accuracy not surpass that of the original method. Compared with the scan-context Top_1, all the accuracies are better. It can be observed that the dataset for the same season has a higher accuracy. In addition, compared with the pole extraction method, the data results for the four-seasons pole extraction method are lower than ours. In addition, compared with the pole extraction method, our results are significant higher under the four-seasons datasets.

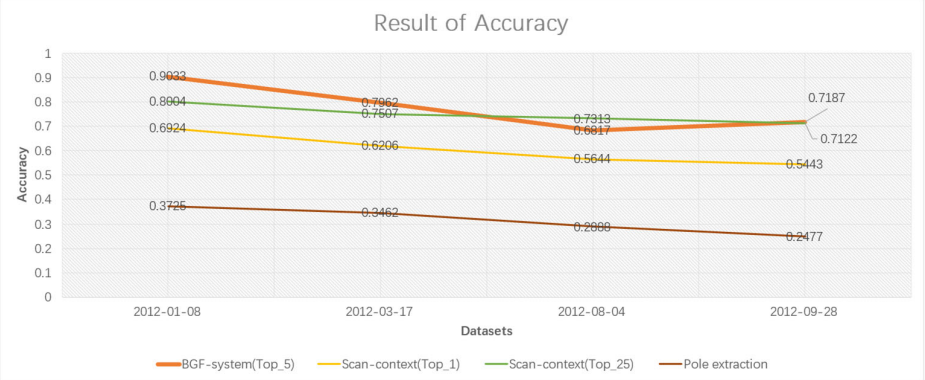

Figure 3. Result of Accuracy.

4.3.2. Global Feature Analysis

In this chapter, we compare the global features which are extracted from image descriptors. Theoretically, the global features in the same place are similar. After re-classification by K-means, in each class, there should be images from the same label. From this point, we can estimate that our idea is right. At the same original label, the scan-context images vary a little. From Figure 4, it is observed that the global features are similar in the same new class after re-classification but are from different labels.

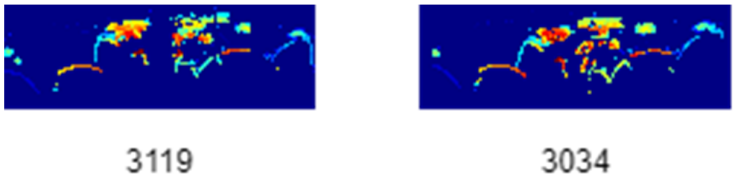

Figure 4. Two images from different labels in the same new class after re-classification.

After reclassification according to the global features, the images with similar global features are aggregated, which centralizes the features in each category to facilitate CNN training and prevents the deviation in training caused by the difference in the number of images in each category.

From Figures 4 and 5, it is observed that in some classes, there may be some images whose original labels are not adjacent. Although, these image data had similar global features, they were not collected from nearby places. Hence, the place recognition accuracy is affected, which needs to be analyzed and studied in future research.

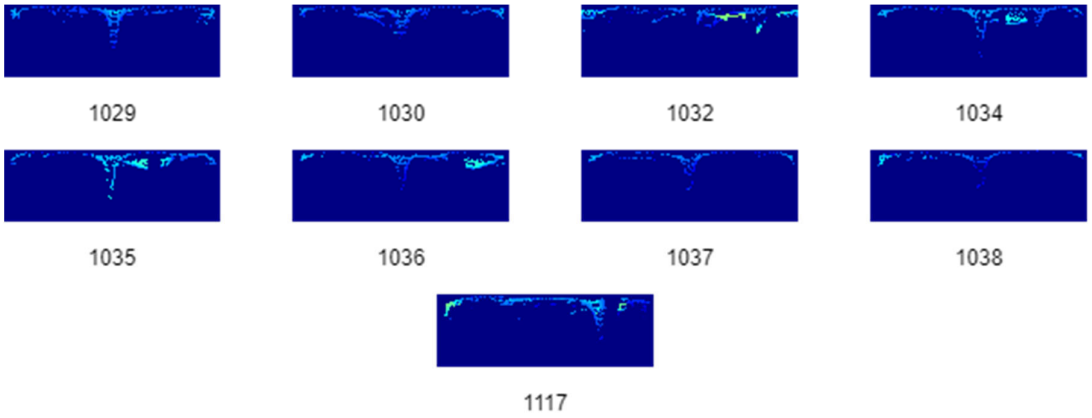

Figure 5. The same classification images from more than two different labels.

4.3.3. Runtime Evaluation

The advantage of our method is that it can be viewed as an extension of the highly efficient BoW framework that enables simplification of the retrieval database and greatly reduces its retrieval time. The generated scan-context part of our programs is written in MATLAB, and the other part is written in Python. In the CNN part, our Python code runs under NVIDIA GTX 960, and the batch size is set to eight.

Table 4 and Figure 6 show the timing performance. As shown, our system is generally faster than the scan context as a whole.

Table 4. Time Performance.

Method	Descriptor Generation (s)	Retrieving (s)	All (s)
BGF-system	0.0434	0.000016	0.04341
Scan-context	0.0434	0.0047	0.0481
Pole_extraction	0.09	0.1	0.19

Figure 6. Time performance.

5. Conclusions

We proposed a method to build a digital retrieval database based on global features reclassification and other technologies derived from the field of document retrieval. Through this method, the visual robot can quickly recognize the robot's place. This system packs all the label information into an inverted index. In the experiment, this method was verified in the NCLT database and has achieved remarkable accuracy and quick place recognition. The system is robust and can be used for long-term place recognition in four seasons. Compared to the improved scan-context method, our method can recognize places more accurately using the NCLT dataset. In future, we again plan to optimize the data processing on the original basis, add new technical and theoretical methods, and realize more functions through the newly studied system.

Author Contributions: Conceptualization, M.Y. and K.T.; methodology, M.Y.; software, M.Y.; validation, M.Y. and K.T.; formal analysis, M.Y.; investigation, M.Y.; resources, K.T. and M.Y.; data curation, M.Y.; writing—original draft preparation, M.Y.; writing—review and editing, M.Y. and K.T.; project administration, M.Y. and K.T.; supervision, K.T.; funding acquisition, M.Y. and K.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been supported in part by JSPS KAKENHI Grant-in-Aid for Scientific Research (C) 17K00361, and (C) 20K12008.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Lin, H.Y.; He, C.H. Mobile Robot Self-Localization Using Omnidirectional Vision with Feature Matching from Real and Virtual Spaces. *Appl. Sci.* **2021**, *11*, 3360. [CrossRef]
2. Jiao, H.; Chen, G. Global self-localization of redundant robots based on visual tracking. *Int. J. Syst. Assur. Eng. Manag.* **2021**, *14*, 529–537. [CrossRef]
3. Liwei, H.; De, X.; Yi, Z. Natural ceiling features based self-localization for indoor mobile robots. *Int. J. Model. Identif. Control* **2010**, *10*, 272–280.
4. Jabnoun, H.; Benzarti, F.; Morain-Nicolier, F.; Amiri, H. Visual substitution system for room labels identification based on text detection and recognition. *Int. J. Intell. Syst. Technol. Appl.* **2018**, *17*, 210–228.
5. Kim, S.; Kim, S.; Lee, D.E. Sustainable application of hybrid point cloud and BIM method for tracking construction progress. *Sustainability* **2020**, *12*, 4106. [CrossRef]
6. Li, N.; Ho, C.P.; Xue, J.; Lim, L.W.; Chen, G.; Fu, Y.H.; Lee, L.Y.T. A Progress Review on Solid—State LiDAR and Nanophotonics—Based LiDAR Sensors. *Laser Photonics Rev.* **2022**, *16*, 2100511. [CrossRef]
7. Xu, X.; Zhang, L.; Yang, J.; Cao, C.; Wang, W.; Ran, Y.; Tan, Z.; Luo, M. A Review of Multi-Sensor Fusion SLAM Systems Based on 3D LIDAR. *Remote Sens.* **2022**, *14*, 2835. [CrossRef]
8. Yan, Z.; Duckett, T.; Bellotto, N. Online learning for human classification in 3D LiDAR-based tracking. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Vancouver, BC, Canada, 24–28 September 2017; pp. 864–871.
9. Bosse, M.; Zlot, R. Place recognition using keypoint voting in large 3D lidar datasets. In Proceedings of the IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 2677–2684.
10. Barros, T.; Garrote, L.; Pereira, R.; Premebida, C.; Nunes, U.J. Attdlnet: Attention-based dl network for 3d lidar place recognition. *arXiv* **2021**, arXiv:2106.09637.
11. Zhou, B.; He, Y.; Huang, W.; Yu, X.; Fang, F.; Li, X. Place recognition and navigation of outdoor mobile robots based on random Forest learning with a 3D LiDAR. *J. Intell. Robot. Syst.* **2022**, *104*, 72. [CrossRef]
12. Vidanapathirana, K.; Moghadam, P.; Harwood, B.; Zhao, M.; Sridharan, S.; Fookes, C. Locus: Lidar-based place recognition using spatiotemporal higher-order pooling. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Xi'an, China, 30 May–5 June 2021; pp. 5075–5081.
13. Kim, G.; Kim, A. Scan Context: Egocentric Spatial Descriptor for Place Recognition Within 3D Point Cloud Map. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Madrid, Spain, 1–5 October 2018; pp. 4802–4809.
14. Kim, G.; Park, B.; Kim, A. 1-Day Learning, 1-Year Localization: Long-Term Lidar Localization Using Scan Context Image. *IEEE Robot. Autom. Lett.* **2019**, *4*, 1948–1955. [CrossRef]

15. Tian, X.; Yi, P.; Zhang, F.; Lei, J.; Hong, Y. STV-SC: Segmentation and Temporal Verification Enhanced Scan Context for Place Recognition in Unstructured Environment. *Sensors* **2022**, *22*, 8604. [CrossRef]
16. Yuan, H.; Zhang, Y.; Fan, S.; Li, X.; Wang, J. Object Scan Context: Object-centric Spatial Descriptor for Place Recognition within 3D Point Cloud Map. *arXiv* **2022**, arXiv:2206.03062.
17. Nam, S.J.; Park, I.C.; Kyung, C.M. Improving dictionary-based code compression in VLIW architectures. *IEICE Trans. Fundam. Electron. Commun. Comput. Sci.* **1999**, *82*, 2318–2324.
18. Sajjad, M.; Mehmood, I.; Baik, S.W. Image super-resolution using sparse coding over redundant dictionary based on effective image representations. *J. Vis. Commun. Image Represent.* **2015**, *26*, 50–65. [CrossRef]
19. Zhao, R.; Mao, K. Fuzzy Bag-of-Words Model for Document Representation. *IEEE Trans. Fuzzy Syst.* **2018**, *26*, 794–804. [CrossRef]
20. Wu, L.; Hoi, S.C.H.; Yu, N. Semantics-Preserving Bag-of-Words Models and Applications. *IEEE Trans. Image Process.* **2010**, *19*, 1908–1920.
21. NCLT Dataset. Available online: http://robots.engin.umich.edu/nclt/ (accessed on 4 April 2022).
22. Carlevaris-Bianco, N.; Ushani, A.K.; Eustice, R.M. University of Michigan North Campus long-term vision and lidar dataset. *Int. J. Robot. Res.* **2015**, *35*, 1023–1035. [CrossRef]
23. Bürki, M.; Schaupp, L.; Dymczyk, M.; Dubé, R.; Cadena, C.; Siegwart, R.; Nieto, J. Vizard: Reliable visual localization for autonomous vehicles in urban outdoor environments. In Proceedings of the IEEE Intelligent Vehicles Symposium (IV), Paris, France, 9–12 June 2019; pp. 1124–1130.
24. He, L.; Wang, X.; Zhang, H. M2DP: A novel 3D point cloud descriptor and its application in loop closure detection. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Daejeon, Republic of Korea, 9–14 October 2016; pp. 231–237.
25. Alnaggar, Y.A.; Afifi, M.; Amer, K.; ElHelw, M. Multi Projection Fusion for Real-time Semantic Segmentation of 3D LiDAR Point Clouds. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 1800–1809.
26. Komorowski, J. Minkloc3d: Point cloud based large-scale place recognition. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 1790–1799.
27. Lin, T.-Y.; Dollar, P.; Girshick, R.; He, K. Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
28. Luo, L.; Cao, S.Y.; Han, B.; Shen, H.L.; Li, J. BVMatch: Lidar-Based Place Recognition Using Bird's-Eye View Images. *IEEE Robot. Autom. Lett.* **2021**, *6*, 6076–6083. [CrossRef]
29. Rosas-Cervantes, V.; Lee, S.-G. 3D Localization of a Mobile Robot by Using Monte Carlo Algorithm and 2D Features of 3D Point Cloud. *Int. J. Control Autom. Syst.* **2020**, *18*, 2955–2965. [CrossRef]
30. Li, X.; Du, S.; Li, G.; Li, H. Integrate point-cloud segmentation with 3D Lidar scan-matching for mobile robot localization and mapping. *Sensors* **2020**, *20*, 237. [CrossRef]
31. Barea, R.; Pérez, C.; Bergasa, L.M.; López-Guillén, E.; Romera, E.; Molinos, E.; Ocana, M.; López, J. Vehicle detection and localization using 3d Lidar point cloud and image semantic segmentation. In Proceedings of the 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; pp. 3481–3486.
32. Xu, X.; Yin, H.; Chen, Z.; Li, Y.; Wang, Y.; Xiong, R. DiSCO: Differentiable Scan Context with Orientation. *IEEE Robot. Autom. Lett.* **2021**, *6*, 2791–2798. [CrossRef]
33. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
34. Gao, Z.; Zhang, Y.; Li, Y. Extracting features from infrared images using convolutional neural networks and transfer learning. *Infrared Phys. Technol.* **2020**, *105*, 103237. [CrossRef]
35. Mateen, M.; Wen, J.; Nasrullah; Song, S.; Huang, Z. Fundus Image Classification Using VGG-19 Architecture with PCA and SVD. *Symmetry* **2018**, *11*, 1. [CrossRef]
36. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Andrej, K.; Aditya, K.; Michael, B.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
37. Beyer, L.; Hénaff, O.J.; Kolesnikov, A.; Zhai, X.; Oord, A.V.D. Are we done with imagenet? *arXiv* **2020**, arXiv:2006.07159.
38. Lloyd, S. Least squares quantization in PCM. *IEEE Trans. Inf. Theory* **1982**, *28*, 129–137. [CrossRef]
39. Karegowda, A.G.; Jayaram, M.A.; Manjunath, A.S. Cascading k-means clustering and k-nearest neighbor classifier for categorization of diabetic patients. *Int. J. Eng. Adv. Technol.* **2012**, *1*, 147–151.
40. Wang, F.; Wang, Q.; Nie, F.; Li, Z.; Yu, W.; Ren, F. A linear multivariate binary decision tree classifier based on K-means splitting. *Pattern Recognit.* **2020**, *107*, 107521. [CrossRef]
41. Chowdary, G.J. Class dependency based learning using Bi-LSTM coupled with the transfer learning of VGG16 for the diagnosis of Tuberculosis from chest x-rays. *arXiv* **2021**, arXiv:2108.04329.
42. Mascarenhas, S.; Agarwal, M. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. In Proceedings of the International Conference on Disruptive Technologies for Multi-Disciplinary Research and Applications (CENTCON), Bengaluru, India, 22–24 December 2021; Volume 1, pp. 96–99.
43. Desai, P.; Pujari, J.; Sujatha, C.; Kamble, A.; Kambli, A. Hybrid Approach for Content-Based Image Retrieval using VGG16 Layered Architecture and SVM: An Application of Deep Learning. *SN Comput. Sci.* **2021**, *2*, 170. [CrossRef]

44. Dong, H.; Chen, X.; Stachniss, C. Online range image-based pole extractor for long-term lidar localization in urban environments. In Proceedings of the European Conference on Mobile Robots (ECMR), Bonn, Germany, 31 August–3 September 2021; pp. 1–6.
45. Lowry, S.; Sünderhauf, N.; Newman, P.; Leonard, J.J.; Cox, D.; Corke, P.; Milford, M.J. Visual Place Recognition: A Survey. *IEEE Trans. Robot.* **2015**, *32*, 1–19. [CrossRef]
46. Stallings, W.M.; Gillmore, G.M. A note on "accuracy" and "precision". *J. Educ. Meas.* **1971**, *8*, 127–129. [CrossRef]