2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 14-18, 2024. Abu Dhabi, UAE

SDFT: Structural Discrete Fourier Transform for Place Recognition and Traversability Analysis

Ayumi Umemura[1,2*], Ken Sakurada[2], Masaki Onishi[2] and Kazuya Yoshida[1]

***Abstract*— The ability to associate the current location with previously visited places is an essential aspect of autonomous ground robots. Unstructured environments such as planetary surfaces pose a significant challenge for robots because their terrain is less distinctive. Meanwhile, traversability must be analyzed simultaneously for safe navigation. In the past, place recognition research has rarely considered traversability analysis despite its significance. This is because the structural information of terrains becomes quickly implicit during the encoding process. This paper provides a method that explicitly addresses both problems: place recognition and traversability analysis. It proposes a discrete Fourier transform (DFT) to represent the frequency components embedded in ground curvature, which underlies both concepts. Our place recognition function demonstrates excellent performance in extensive experiments using challenging planetary & urban datasets while estimating traversability that other approaches find difficult to handle.**

I. INTRODUCTION

Autonomous robots in unknown environments require simultaneous localization and mapping (SLAM). The key feature of these systems is their ability to recognize previously visited places, enabling them to correct the map distorted by estimation errors. However, unstructured environments such as planetary surfaces pose a significant challenge for visual-based approaches [1]–[3] owing to their visual ambiguity. Meanwhile, traversability analysis is essential for safe navigation in uneven terrains. Despite the importance of traversability analysis, our literature survey revealed no suitable methods capable of addressing both issues.

For visually ambiguous scenes, point clouds' structures have been exploited for place recognition [4]–[6]; these methods offer high accuracy in localization. However, they are sensitive to noises and less informative in unstructured environments. In addition, their implicit nature does not allow robots to interpret incoming hazards.

We propose Structural Discrete Fourier Transform (SDFT), which realizes stable place recognition and traversability analysis by leveraging frequency information embedded in the local point cloud as illustrated in Fig. 1. The aggregation of frequencies can represent gentle ground curvature and urban scenes in a more informative way than approaches describing salient structures [7], [8]. Moreover,

This work is supported by JSPS KAKENHI under Grant No. 23KJ0170.
[1]A. Umemura and K. Yoshida are with the Space Robotics Lab. (SRL) in the Department of Aerospace Engineering, Graduate School of Engineering, Tohoku University, Sendai 980-8579, Japan umemura.ayumi.t6@dc.tohoku.ac.jp
[2]A. Umemura, K. Sakurada, and M. Onishi are with The National Institute of Advanced Industrial Science and Technology (AIST), Tokyo 761-0395, Japan k.sakurada@aist.go.jp
The corresponding author is Ayumi Umemura.

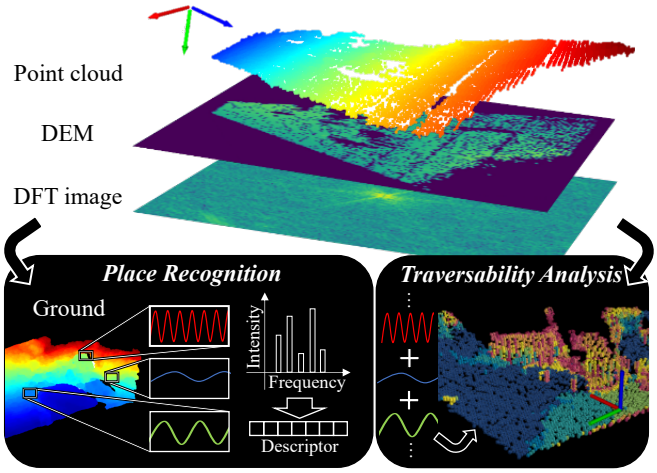


Fig. 1: Our approach is designed to simultaneously solve place recognition and traversability analysis in challenging unstructured terrains for mobile robots based on a frequency analysis of the environment.

Fourier analysis of a point cloud seamlessly addresses both tasks simultaneously, as the understanding of frequency components is a common principle in these tasks. The contributions of this study are featured as follows:

- Provision of a novel structure-based descriptor that exploits underlying frequency components from a 3D point cloud for high-entropy representation.
- Proposal of traversability analysis, which is concurrently accomplished during descriptor creation.
- Experimental validations of our approach with real datasets. Compared to existing methods, the proposed method shows excellent performance w. r. t. accuracy and robustness against noises. Meanwhile, it pinpoints traversable locations with degrees of danger risks that were entirely overlooked by other approaches.

II. RELATED WORK

The approaches that have been used in place recognition studies can be categorized into visual and structural methods. Visual approaches focus on visual similarities between the query and target image. Visual similarities are represented as a cooccurrence of local descriptors, such as ORB [9] and SIFT [10], which use feature aggregation schemes such as bag-of-visual words. However, their performance is sensitive to extreme lighting conditions and viewpoint changes. Moreover, as visually ambiguous scenes, such as planetary sites, degrade their performance, other approaches should be explored for stable detection.

Structure-aware approaches present strong robustness to the aforementioned perceptual challenges. They can be separated into local, global, and learning-based descriptors.

Local descriptors, such as SHOT [11], spin image [12], and shape context [13], begin with key point detection, separate points into several bins, and encode their bin patterns into a histogram. Most of these methods, however, usually suffer from noises included in the point clouds. This is because these approaches compute normal vectors per point to summarize the place. Le Gentil et al. [14] addressed noisy point clouds using Gaussian process regression [15], which gives dense and precise 3D representation. However, as the process requires considerable computation for online processing, it is unsuitable for extreme environment exploration where vehicles have scarce computing resources.

Global descriptors, which exploit the point cloud's appearance, can cope better with noisy point clouds. Using LiDAR, Scan Context [16], and their variations [17], [18] encodes the surrounding structure with high-entropy descriptors by separating clouds into circular bins along the distance. Konrad et al. proposed DELIGHT [19], which encodes spatial information with intensity information originating from LiDAR. A robot can use rich intensity information to localize the map without prior information. Fresco [20], whose key idea was closest to ours, aimed to aggregate frequency components to represent the place. However, these approaches rely on the nature of LiDAR and assume that robots can recognize their omnidirectional environments. Because of their high energy consumption, 3D LiDAR sensors have not been deployed in autonomous robot missions in planetary environments. Therefore, this study considers stereo cameras as a primary sensor and leverages them to represent structural information.

Approaches based on deep learning have received considerable attention in the field of place recognition. Several deep-learned approaches [21]–[23] have been extensively investigated in localization problems. Although deep-learning techniques may significantly improve performance, they present high computational complexity in cases where GPUs are unavailable.

He et al. proposed M2DP [24], which projects the entire point cloud onto multiple 2D planes and generates low-dimensional descriptors. M2DP achieved higher accuracy and strong robustness against decreases in resolution and noises. Recent approaches, such as those in [25] and [26], use distinctive diagrams from key points or landmarks. The current state-of-the-art approach, STD [8], which summarizes places as a group of triangles, connecting local key points from the point cloud, follows this trend.

However, none of the above approaches address traversability analysis because the structure information used for the analysis becomes implicit during encoding. Recent traversability representations rely on deep learning with navigation experiences [27], [28]. Instead, this study proposes a novel 3D descriptor for informative place description and traversability analysis based on frequency properties.

III. METHODOLOGY

Assuming the robot uses a stereo camera to reconstruct point clouds, we propose DFT-based structural encoding (SDFT). We will briefly explain some basic concepts in DFT. Then, we shall provide the details of SDFT.

A. 2D Discrete Fourier Transform

A discrete Fourier transform converts a discrete signal from its original domain to frequency components. Therefore, we can represent any signal defined in 2D space as a weighted summation of frequencies.

Assume that there is a 2D function $f(x, y)$ where x and y are spatial information. $f(x, y)$, whose dimension is $N \times N$, can be converted to a frequency representation $F(u, v)$, where u and v are the frequencies in the horizontal and vertical directions.

$$F(u,v) = \sum_{x=0}^{N-1} \sum_{y=0}^{N-1} f(x,y) \exp\left(\frac{-2\pi j(ux+vy)}{N}\right) \quad (1)$$

The dimension of $F(u, v)$ becomes the same as $f(x, y)$.

Equation 1 is utilized for place recognition. As planetary terrain does not show salient structures, approaches that exploit saliency detections fail to generate informational representations. Conversely, gentle ground curvatures can be regarded as an aggregation of physical waves. Based on this insight, SDFT is constructed using Fourier analysis.

An inverse discrete Fourier transform can be described as follows for any function defined in 2D space:

$$f(x,y) = \frac{1}{N^2} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} F(u,v) \exp\left(\frac{2\pi j(ux+vy)}{N}\right) \quad (2)$$

Equation 2 is exploited for traversability analysis. Applying this to a limited range of $F(u, v)$, specific frequency components can be isolated and extracted.

As described above, decomposing the ground structure into frequency components provides seamless solutions for place recognition and traversability analysis.

B. Algorithm Overview

Given a point cloud, we first project it onto a single plane to represent it as a digital elevation map (DEM). Next, DFT is applied to the DEM to extract underlying frequency information. This DFT output is utilized in place recognition and traversability analysis, as shown in Fig. 2 (a).

We must form sufficiently robust descriptors against significant viewpoint changes for place recognition. Therefore, we sample values from the DFT output through our max-pooling sampling to gain robustness. By sampling a signature, we can represent the point cloud by a data matrix A. We then match two point clouds by comparing their data matrices. For compact representation, we employ SVD for dimensionality reduction.

Traversability is influenced by various factors, including the kinematic model of the unmanned ground vehicle (UGV) and the physical properties of the environment [29]. This

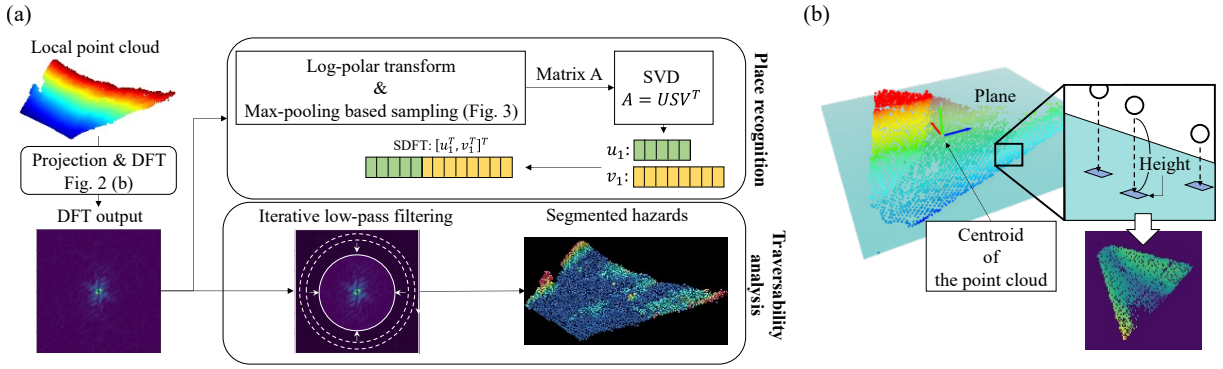


Fig. 2: **(a):** Overview of the proposed pipeline consisting of place recognition and traversability analysis. **(b):** Overview of DEM creation. PCA defines the local reference frame. The plane passes through the centroid of the point cloud.

study, however, focuses on analyzing the frequency characteristics of the terrain. Specifically, it posits that terrains with finer undulations pose a greater risk to traversal. Hazardous regions associated with their degree of travel risk must be located for safe robot navigation. We adopt iterative low-pass filters with different cut-off frequencies to extract portions of the point cloud associated with each frequency.

Details regarding place recognition are described in the next two subsections, while traversability analysis is described in the last section.

C. Preprocessing of a Point Cloud

To enhance the robustness of the descriptor against changes in the camera's pose, the reference frame should be defined based on the structural properties of the local point cloud. To realize this, we first place the reference frame's origin onto the centroid of the point cloud. We then perform PCA on the points to find the two underlying dominant directions in the point cloud and its normal vector. The first and second PCs define the reference frame's x and y-axis. As the normal vector is settled, the descriptor becomes invariant w. r. t. the elevation of the sub-map's origin.

However, as PCA is inherently affected by narrow-angle point cloud shapes, achieving correct associations under significant yaw angle changes is difficult. As an additional process, our max-pooling sampling strategy is proposed as explained in Section III-D.

D. SDFT Creation for Place Recognition

Next, we generate a DEM by projecting local point cloud P to a plane as illustrated in Fig. 2 (b). We first define the plane X passing through the origin with the normal vector m derived in Section III-C. For each point p in P and its corresponding point's location u on X, $u = p - (p^T m)m$. Each pixel value is the distance from the point to the plane.

The pixel map is now integrated with DFT to obtain $F(u, v)$. We denote the size of the pixel map as $M \times N$ and let $X(a, b)$ be the value at the (a, b)-th pixel of X.

$$F(u,v) = \sum_{a=0}^{M-1} \sum_{b=0}^{N-1} X(a,b) \exp\left(-2\pi j \left(\frac{u}{M}a + \frac{v}{N}b\right)\right) \quad (3)$$

For place recognition, we take the logarithm of all absolute values in $F(u, v)$.

Let a polar coordinate system define $F(u, v)$. We define r as its polar radius and θ as its angle. Suppose two DEMs, I_A and I_B, differ by a rotation δ_θ, i.e., $I_A(r, \theta) = I_B(r, \theta + \delta_\theta)$. We have

$$I_A^F(r,\theta) = I_B^F(r,\theta+\delta_\theta) \quad (4)$$

where I^F is a DFT output. This means rotation in the spatial domain causes the same rotation in the frequency domain. Considering Equation 4, our descriptor is sensitive to rotation along the normal vector.

To mitigate this rotation effect, we divide DFT output into polar bins and index them according to their radius and angle. Next, we take the maximum values in each bin to gain max-pooling effects, alleviating the rotation shifts. To deal with opposite revisitation, bins for max-pooling are organized as illustrated in Fig. 3 (a). This is because their frequency responses are inverted at that moment.

The DFT output is unrolled in the log-polar coordinate system to arrange kernel windows for pooling. Fig. 3 (b) illustrates the bin and kernel window organization. Let the unrolled DFT output be Z, and the result from max-pooling be A. We define $Z(a, b)$ as the value of the (a, b)-th element in Z $(0 \leqq a < M, 0 \leqq b < N)$. We denote $R_{(i,j)}$ as a kernel window whose output results in the (i, j)-th element of A. Given Z, each element of A can be formulated:

$$B_{(k,j)} := \left\{ Z(r,j) \middle| \frac{M\theta}{2\pi}k \leqq r < \frac{M\theta}{2\pi}(k+1) \right\}$$

$$R_{(i,j)} := \begin{cases} \left\{ B_{(k,j)} \middle| k = \left\{0, \frac{\pi}{\theta}, \frac{\pi}{\theta} - 1, \frac{2\pi}{\theta} - 1\right\}\right\} & \text{if } i = 0 \\ \left\{ B_{(k,j)} \middle| k = \left\{i, i + \frac{\pi}{\theta}\right\}\right\} & \text{if } i \neq 0 \end{cases}$$

$$A_{(i,j)} = \max_{(a,b)\in R_{(i,j)}} Z(a,b) \quad (5)$$

Finally, to reduce the dimension, SVD is applied to A. The first left and right singular vectors are concatenated as the final form of SDFT. Given SDFT, we coarsely search for loop closure candidates based on the L_2 proximity metric.

E. Traversability Analysis

The traversability analysis process aims to extract the regions of the point cloud forming each frequency that is regarded as a degree of traversability in this study.

To realize this, we iteratively adopt low-pass filters, decreasing their cut-off frequency. As the filtering makes the

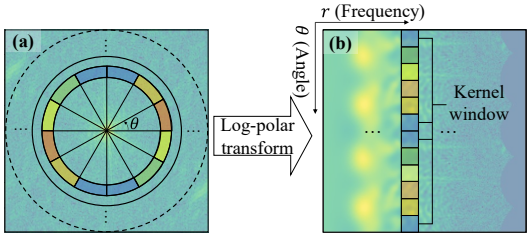


Fig. 3: Partitions where max-pooling is applied. Bins of the same color are combined to form a kernel window that is applied to max-pooling.

point cloud smoother, points constituting the target frequency are highlighted by selecting points distant from the smoothened cloud. The flow of the algorithm is summarized in Algorithm 1. We employ the following different elements:

- $GetSmoothenedCloud$: Applying a low-pass filter to the DFT output, point sets that belong to frequencies lower than F_i are obtained.
- $GetHigherFreq$: Comparing with the original point cloud, a set of points P_i is extracted, which consists of higher frequencies than F_i.
- $GetComponent$: The lower cut-off frequency C_w is applied with the filter, the more points are obtained from $GetHigherFreq$. This function highlights points that become P_i for the first time when C_w decreases to F_i.

Applying a hard threshold in the frequency domain results in a ringing effect in the spatial structure. Employing a filter with a smooth cut-off effectively prevents the occurrence of this distortion phenomenon. Hence, in this study, we utilized a Butterworth filter [30], which is a signal-processing filter. The equation can be expressed as follows:

Algorithm 1 Traversability analysis algorithm

Input: Point cloud P, DFT output of the DEM F_{dem}, frequencies F to be associated with P. F is organized in descending order

Output: Point cloud clusters X constituting each frequency components

$i \leftarrow 0$
$X \leftarrow \{\}$
for all $element \leftarrow F$ **do**
  $s \leftarrow GetSmoothenedCloud(element, F_{dem})$
  $p \leftarrow GetHigherFreq(P, s)$
  if $i > 0$ **then**
    $f \leftarrow GetComponent(p_{prev}, p)$
  else
    $f \leftarrow p$
  end if
  $X.insert(f)$
  $p_{prev} \leftarrow p$
  $i \leftarrow i + 1$
end for
return X

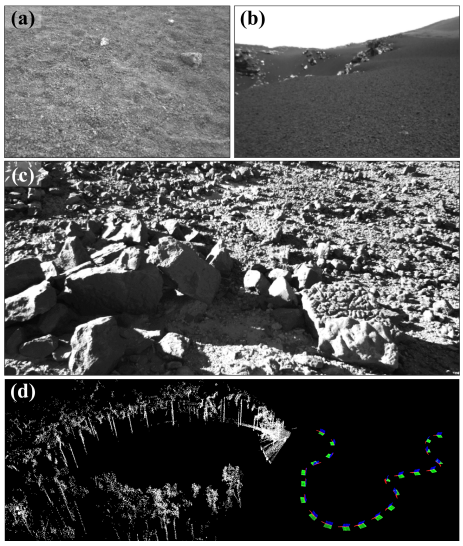


Fig. 4: (a) LRNT sequence. An unstructured and visually ambiguous terrain. (b) S3LI sequence. Dome ground curvature can be seen. (c) MADMAX sequence. More rocky terrain. (d) Livox sequence. A park area filled with trees can be seen in the point cloud.

$$|T(d)| = \frac{1}{\sqrt{1 - \left(\frac{d}{d_c}\right)^{2n}}} \tag{6}$$

T is the magnitude response, d is the distance from the DC component in the frequency domain, and d_c is the cut-off frequency. n is the order of the Butterworth filter, which determines how quickly the function decreases. $n = 4$ was used in our implementation.

IV. EXPERIMENTAL EVALUATION

This section evaluates our method against existing structure-based descriptors on public datasets. We compared SDFT with three descriptors: SHOT, M2DP, and the state-of-the-art STD descriptor. SDFT was implemented in Python while SHOT was coded in C++, M2DP was written in Matlab and C, and STD was coded in C++. Only the official open-sourced codes were employed. Our platform for experiments was a laptop equipped with an Intel Core i7 5GHz.

The proposed method was evaluated using camera sequences in MADMAX [31], LRNT [32], and S3LI [33]. We used the MADMAX-1, 2[1], LRNT-1, 2[2], and S3LI-1, 2[3] sequences. Sequences have few distinctive features as they were recorded to simulate exploration on planetary surfaces. Moreover, an urban dataset, Livox sequence[4] [8], was used to evaluate STD in Section IV-A. The datasets are detailed in Table I. Examples of each scene are listed in Fig. 4.

In the next subsection, STD is separately evaluated from other approaches because different experimental setups were

[1]MADMAX Path E & G sequence
[2]LRNT Path 1 & 2 sequence
[3]S3LI Crater_inout & Loops sequence
[4]Livox Avia park 2 sequence

TABLE I: Properties of the sequences used in the evaluations.

	MADMAX		S3LI		LRNT		Livox
Sequence name	E	G	Crater_inout	Loops	1	2	Avia park 2
Total length (m)	906	394	1338	587	834	1097	802
Number of keyframes	2202	1169	1122	484	2298	2673	538
Number of revisited frames	1787	1096	501	260	1887	1277	538
Route dir. on revisit	Both	Both	Same	Same	Both	Reversed	Both

TABLE II: Results for place recognition with the state-of-the-art key point-based STD descriptor.

	STD	SDFT
Dataset	Top 10 accuracy(Stability)	Top 10 accuracy
Livox	0.586 (100)	**0.877**
MADMAX-1	0.247 (75.5)	**0.870**
MADMAX-2	– (15.5)	**0.979**

required for fair comparison. A comparison with SHOT and M2DP is conducted in the following subsections.

A. Comparison with State-of-the-art Approach

Recent approaches focus on diagram constellations consisting of key points and demonstrate excellent performances in structured environments [8], [34]. However, unstructured environments lack structures that help form stable patterns.

Dataset: To demonstrate this, STD was evaluated using MADMAX and Livox datasets. For every dataset, the detection was regarded as a true positive if the ground truth pose distance between the queries and matched frames was less than 15m. To avoid neighbor detections, 30 frames before/after the current frame were ignored from the evaluation.

Experimental Settings: For fair comparisons, we used STD without geometry checks. STD required processing a sub-map comprising ten frames for Livox and 20 frames for MADMAX dataset. Considering this, the same sub-maps were used for SDFT.

As STD relies on unique triangles consisting of key points, it fails to generate descriptors for some frames if a reasonable number of key points are not detected from the scene. In the remainder of this section, the word "stability" will be referred to as the ratio of frames in which STD succeeded in generating descriptors.

STD's official parameters were used for Livox dataset. Meanwhile, we carefully tuned its parameters to maximize their stability and accuracy for planetary scenes. Testing SDFT, we settled the dimension of the DEM and the angular value to define max-pooling: the width and height of the DEM were 60 m, and the angular value was set to 30°.

Evaluation Metric: Top 10 accuracy was adopted. In contrast to SDFT, which refers to L2 norms between the query description and all in the database, STD selects fewer candidates because of screening. Considering this, accuracy for STD was computed from the top candidates, up to a maximum of ten outputs. As the final output is chosen from these via the following geometry checks in STD, this process is applicable for estimating STD's performance.

Results: The results are listed in Table II. Our approach showed better accuracy with Livox dataset. In their implementation, STD picks the top 50 candidates and determines the best one after geometry checks, which results in higher performance. We concluded that geometry checks play an important role in gaining STD's performance.

It shows that STD's performance significantly declined in unstructured environments. In MADMAX-2, stability was even lower compared to that in MADMAX-1. The reason was that the smaller size of point clouds in MADMAX-2 prevented sufficient key point detections. As STD descriptors were too unstable in MADMAX-2, we were unable to obtain any possible candidates. Meanwhile, SDFT generated more informative descriptions without detections of salient structures, as indicated in favorable accuracies.

B. Comparison with Existing Global Descriptors

Dataset: We employed MADMAX, LRNT, and S3LI datasets. A total of 30 frames before/after the current frame was ignored from the evaluation to avoid neighbor detections.

MADMAX sequences have 6 degrees of freedom ground truth, while others only have position data. For MADMAX dataset, two locations were considered to be true positive if their distance was less than 5m and the orientation change θ_{diff} was in a reasonable bound ($0 \leqq |\theta_{diff}| \leqq 30$ or $150 \leqq |\theta_{diff}| \leqq 180$). Meanwhile, true positive detections were decided based on positions in LRNT and S3LI sequences.

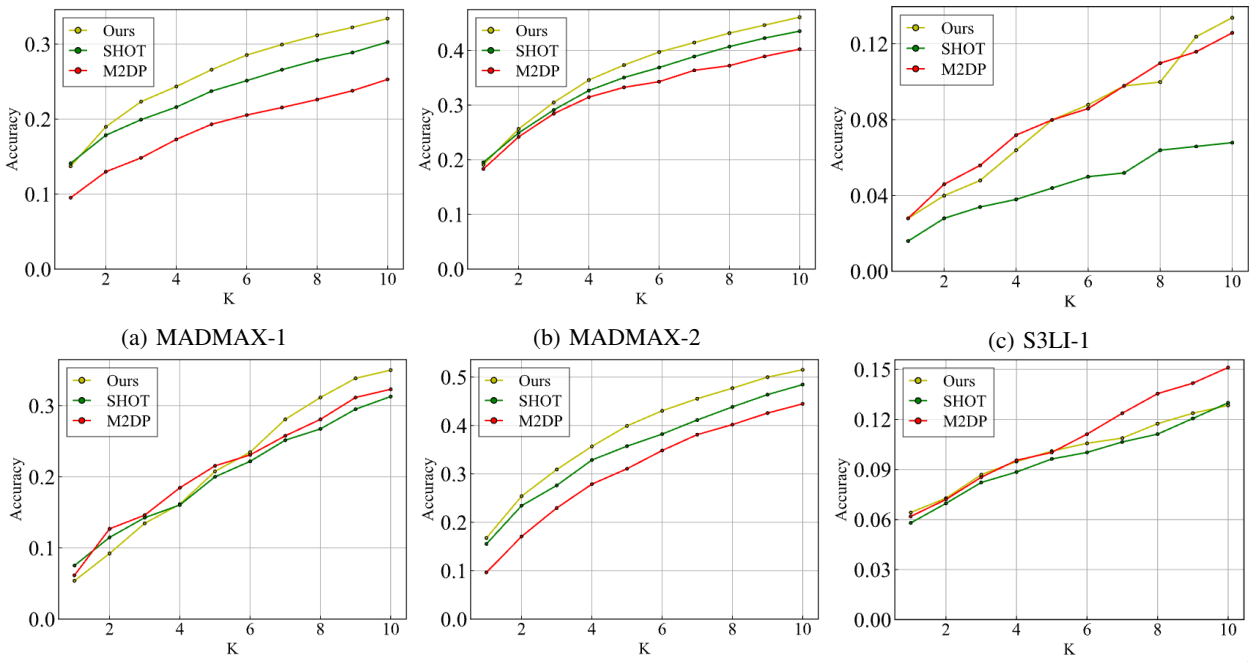
Experimental Settings: As SHOT is a surface normal-based approach, it requires a setting for the normal calculations. Our experiments estimated the normal based on their nearest ten points. SHOT also needs a radius parameter that defines the area taken into account. To regard SHOT as a global descriptor, this parameter was configured to encompass the same spatial range as SDFT, i.e., 15m×15m. For M2DP, the default parameters were used. As SDFT parameters, the width and height of the DEM were 15m, and the angular value was set to 30° for max-pooling.

Evaluation Metric: Three different experiments were used to evaluate the performance of the approaches. Except for the third experiment, the resolution of the point clouds was set to 0.2m.

In the first experiment, SDFT was compared against existing methods, SHOT and M2DP, regarding Top K accuracy ($1 \leqq K \leqq 10$), using all datasets. Fig. 5 illustrates the results.

The second experiment focused on robustness against random noise on point clouds. In this experiment, we added uniform noise ranging from 0 to x_{noise}(m) to each point location in MADMAX sequences, where $x_{noise} = [0, 0.01, 0.02, 0.05, 0.1, 0.15, 0.2, 0.3]$. We iterated this evaluation ten times to increase repeatability and computed the average/std of each Top 10 accuracy. The results are summarized in Fig. 6.

The third experiment investigated the robustness w. r. t. different down sample sizes. The grid size was set as



(a) MADMAX-1 (b) MADMAX-2 (c) S3LI-1

(d) S3LI-2 (e) LRNT-1 (f) LRNT-2

Fig. 5: Top K accuracies on MADMAX, S3LI, and LRNT datasets.

$[0.05, 0.075, 0.1, 0.2, 0.3, 0.5, 0.7]$ in meters. All approaches were applied on the down-sampled clouds. A comparison of the Top 10 accuracies is presented in Fig. 7.

Results: From Figs. 5, 6, 7, we can observe following:

- As shown in Fig. 5, SDFT generally showed better performance than other methods. SHOT and M2DP sometimes presented competing results, as seen in Figs. 5 (b) and (f). However, their performance decreased overall, while SDFT's performance was superior and stable. This means the performance of SHOT and M2DP was highly dependent on point cloud properties and may resulted in fewer unique descriptors.
- From Fig. 6, SDFT showed no noticeable accuracy degradation with $x_{noise} < 0.15$ while other approaches presented some decrease at that stage. As SHOT relies on normal vectors on the surface, the noises caused the most significant decline. M2DP showed a higher durability with $x_{noise} > 0.15$ than SDFT. Although SDFT presented some degradation with significant errors, its performance was still better than M2DP.
- According to Fig. 7, our approach was sensitive to point cloud density, while other methods showed more gentle degradation. Conversely, our approach showed significant improvement if the point cloud was dense. This result was natural, considering the sampling theorem. Therefore, it is crucial to select an appropriate grid size depending on the property of the environment.

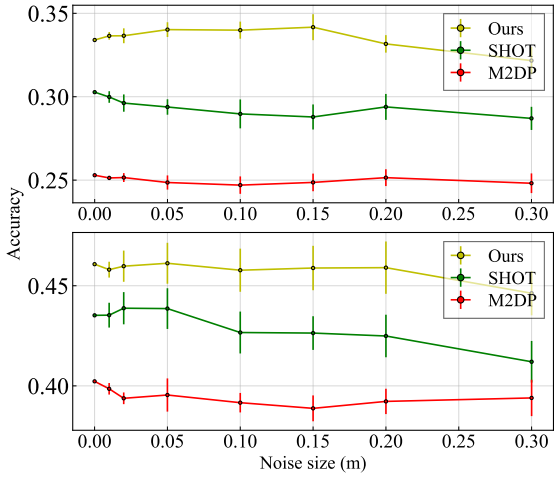


Fig. 6: Top 10 accuracies with different noise amounts. **Top:** MADMAX-1. **Bottom:** MADMAX-2.

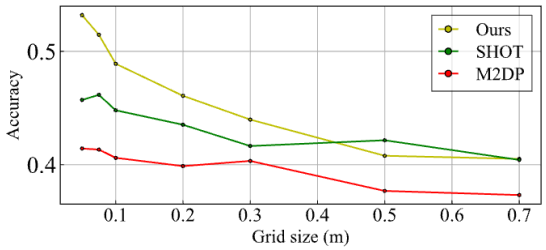


Fig. 7: Top 10 accuracies with different down-sampling sizes in MADMAX-2.

C. Computational Complexity

The average computation time was evaluated in MADMAX-1 while changing point cloud density. The timing for matching was measured in Python. The results are presented in Table III.

In these experiments, SDFT took longer than M2DP for descriptor creation because Matlab and C, which are generally faster than Python we use in SDFT, are employed in M2DP. When a common code was employed for point

TABLE III: Time costs of calculating descriptors and searching candidates in MADMAX-1 sequence in milliseconds.

	Descriptor calculation			Candidate searching		
Grid size(m)	0.05	0.1	0.2	0.05	0.1	0.2
SDFT	198	68.4	12.5	86.3	82.7	12.8
SHOT	477	452	276	82.6	83.9	21.4
M2DP	29.5	11.8	6.63	79.1	77.6	14.3

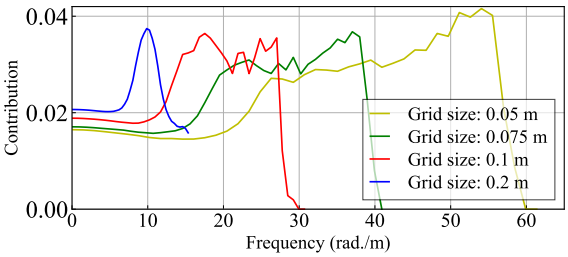


Fig. 8: SDFT's contribution ratio along frequency from MADMAX-2 sequence while changing point cloud density.

cloud matching, the search time depended on the number of dimensions. SDFT took longer than M2DP as the point cloud density grows but in a reasonable bound.

D. Result Interpretation

Before SDFT is formed via SVD, the descriptor's dimensions along columns can be associated with the frequency degree as shown in Fig. 3 (b). As SVD operates as a linear mapping operation, analysis of the final SDFT can be conducted before SVD is employed.

To interpret SDFT's performance, we computed each frequency's contribution to accuracy, while changing the point cloud density. The grid size was set as [0.05, 0.075, 0.1, 0.2] in meters. Contributions were estimated by computing normalized variances along each dimension after gathering all descriptors in MADMAX-2 as shown in Fig. 8.

The result indicates that higher frequencies were no longer interpreted as input point clouds became sparse. As the result in Fig. 7 suggests that denser point clouds improved SDFT's accuracy, we concluded that the range of interpretable frequencies was vital for its performance.

E. Combination with Visual-based Approach

We evaluated DBoW3[5] as a visual-based approach using MADMAX and LRNT sequences. Furthermore, a coarse-to-fine approach was adopted to determine the candidates from SDFT and DBoW3. We first selected the top 50 candidates with SDFT, and the final top 10 candidates were determined based on their appearance. The result is shown in Fig. 9.

Fig. 9 presents that the superiority of an approach depended on the properties of the scene. DBoW3 was more suitable if visual features were sufficiently rich or the point cloud was too small to provide enriched structures, as shown in MADMAX-2 or LRNT-2.

Fusing different domains may help to mitigate their incompleteness. We found that the coarse-to-fine approach mentioned above provided an excellent balance w.r.t. their accuracies. Therefore, the combination of SDFT and DBoW3 provides more stable place recognition regardless of the scene's property.

[5]https://github.com/rmsalinas/DBow3

F. Qualitative Evaluation of Traversability Analysis

We tested our method's traversability analysis in MADMAX-1, which includes hazardous regions. As shown in Fig. 10 (b), the robot recognized hazardous areas with continuous degrees of danger. We built a cost map from these masks that safely navigated the robots to the desired place. As a demonstration, we applied the A* algorithm [35] on the cost map, as shown in Fig. 10 (c).

V. CONCLUSION

This paper presents a discrete Fourier transform-based structural representation method, SDFT, for conducting place recognition and traversability analysis seamlessly. We demonstrate the high performance of the proposed method in unstructured and urban environments, where it exhibits superiority over baseline methods w. r. t. accuracy and robustness against noises. These results indicate that wave decomposition more informatively represents ground structures. Furthermore, unlike existing descriptors, the proposed method simultaneously analyzes traversability based on the frequency components embedded in the ground.

In the future, we will investigate the applicability of sparse point clouds generated from monocular SLAM or insufficient SGM matching in featureless areas.

REFERENCES

[1] D. Gálvez-López and J. D. Tardós, "Bags of binary words for fast place recognition in image sequences," *IEEE Transactions on Robotics*, vol. 28, no. 5, pp. 1188–1197, 2012.

[2] R. Arandjelovic, P. Gronát, A. Torii, T. Pajdla, and J. Sivic, "Netvlad: Cnn architecture for weakly supervised place recognition." in *CVPR*. IEEE Computer Society, 2016, pp. 5297–5307.

[3] C. Toft, W. Maddern, A. Torii, L. Hammarstrand, E. Stenborg, D. Safari, M. Okutomi, M. Pollefeys, J. Sivic, T. Pajdla, F. Kahl, and T. Sattler, "Long-term visual localization revisited," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 4, pp. 2074–2088, 2022.

[4] W. Wohlkinger and M. Vincze, "Ensemble of shape functions for 3d object classification," in *IEEE International Conference on Robotics and Biomimetics*, 2011, pp. 2987–2992.

[5] R. B. Rusu, G. Bradski, R. Thibaux, and J. Hsu, "Fast 3d recognition and pose using the viewpoint feature histogram," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2010, pp. 2155–2162.

[6] N. Muhammad and S. Lacroix, "Loop closure detection using small-sized signatures from 3d lidar data," in *IEEE International Symposium on Safety, Security, and Rescue Robotics*, 2011, pp. 333–338.

[7] L. Luo, S.-Y. Cao, B. Han, H.-L. Shen, and J. Li, "Bvmatch: Lidar-based place recognition using bird's-eye view images," *IEEE Robotics and Automation Letters*, vol. 6, no. 3, pp. 6076–6083, 2021.

[8] C. Yuan, J. Lin, Z. Zou, X. Hong, and F. Zhang, "Std: Stable triangle descriptor for 3d place recognition," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2023, pp. 1897–1903.

[9] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "Orb: An efficient alternative to sift or surf," in *International Conference on Computer Vision*, 2011, pp. 2564–2571.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, p. 91–110, 2004.

[11] S. Salti, F. Tombari, and L. Di Stefano, "SHOT: Unique signatures of histograms for surface and texture description," *Computer Vision and Image Understanding*, vol. 125, pp. 251–264, 2014.

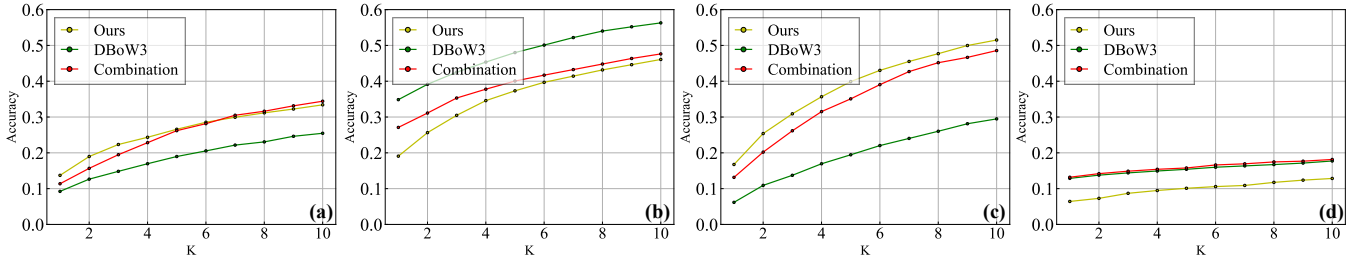


Fig. 9: Top K accuracies for different methodologies: DBoW3, SDFT, and their combination. **(a):** MADMAX-1. **(b):** MADMAX-2. **(c):** LRNT-1. **(d):** LRNT-2.

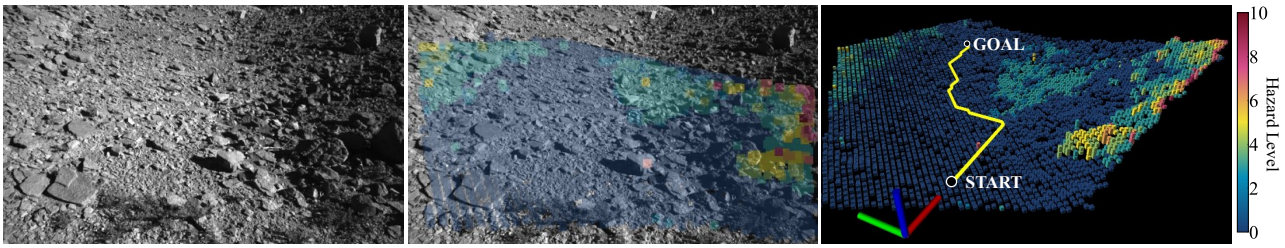


Fig. 10: Examples of traversability analysis. The left shows the appearance of the target terrain. The middle shows overlaid masks, which indicate degrees of danger. The right shows an example of path planning that enables safe navigation based on segmented regions.

[12] A. Johnson, "Spin-images: A representation for 3-d surface matching," Ph.D. dissertation, Carnegie Mellon University, Pittsburgh, PA, 1997.

[13] S. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509–522, 2002.

[14] C. Le Gentil, M. Vayugundla, R. Giubilato, W. Stürzl, T. Vidal-Calleja, and R. Triebel, "Gaussian process gradient maps for loop-closure detection in unstructured planetary environments," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 1895–1902.

[15] E. Schulz, M. Speekenbrink, and A. Krause, "A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions," *Journal of Mathematical Psychology*, vol. 85, pp. 1–16, 2018.

[16] G. Kim and A. Kim, "Scan context: Egocentric spatial descriptor for place recognition within 3d point cloud map," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018, pp. 4802–4809.

[17] G. Kim, S. Choi, and A. Kim, "Scan context++: Structural place recognition robust to rotation and lateral variations in urban environments," *IEEE Transactions on Robotics*, vol. 38, pp. 1856–1874, 2021.

[18] Y. Wang, Z. Sun, C.-Z. Xu, S. E. Sarma, J. Yang, and H. Kong, "Lidar iris for loop-closure detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2020, pp. 5769–5775.

[19] K. P. Cop, P. V. K. Borges, and R. Dubé, "Delight: An efficient descriptor for global localisation using lidar intensities," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3653–3660.

[20] Y. H. Fan, X. Du, and J. Shen, "Fresco: Frequency-domain scan context for lidar-based place recognition with translation and rotation invariance," *17th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, pp. 576–583, 2022.

[21] J. Ma, J. Zhang, J. Xu, R. Ai, W. Gu, and X. Chen, "Overlaptransformer: An efficient and yaw-angle-invariant transformer network for lidar-based place recognition," *IEEE Robotics and Automation Letters*, vol. 7, no. 3, pp. 6958–6965, 2022.

[22] K. Vidanapathirana, M. Ramezani, P. Moghadam, S. Sridharan, and C. Fookes, "Logg3d-net: Locally guided global descriptor learning for 3d place recognition," in *International Conference on Robotics and Automation (ICRA)*, 2022, pp. 2215–2221.

[23] K. Vidanapathirana, P. Moghadam, B. Harwood, M. Zhao, S. Sridharan, and C. Fookes, "Locus: Lidar-based place recognition using spatiotemporal higher-order pooling," in *IEEE International Conference on Robotics and Automation (ICRA)*, 2021.

[24] L. He, X. Wang, and H. Zhang, "M2dp: A novel 3d point cloud descriptor and its application in loop closure detection," in *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2016, pp. 231–237.

[25] G. V. Nardari, A. Cohen, S. W. Chen, X. Liu, V. Arcot, R. A. F. Romero, and V. Kumar, "Place recognition in forests with urquhart tessellations," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 279–286, 2021.

[26] B. Jiang, Y. Zhu, and M. Liu, "A triangle feature based map-to-map matching and loop closure for 2d graph slam," in *IEEE International Conference on Robotics and Biomimetics (ROBIO)*, 2019, pp. 2719–2725.

[27] S. Xie, R. Song, Y. Zhao, X. Huang, Y. Li, and W. Zhang, "Circular accessible depth: A robust traversability representation for ugv navigation," *IEEE Transactions on Robotics*, vol. 39, no. 6, p. 4875–4891, 2023.

[28] I. Cho and W. Chung, "Learning self-supervised traversability with navigation experiences of mobile robots: A risk-aware self-training approach," *IEEE Robotics and Automation Letters*, vol. 9, no. 5, pp. 4122–4129, 2024.

[29] P. Krüsi, P. Furgale, M. Bosse, and R. Siegwart, "Driving on point clouds: Motion planning, trajectory optimization, and terrain assessment in generic nonplanar environments." *Journal of Field Robotics*, vol. 34, no. 5, pp. 940–984, 2017.

[30] S. Butterworth, "On the Theory of Filter Amplifiers," *Experimental Wireless & the Wireless Engineer*, vol. 7, pp. 536–541, 1930.

[31] L. Meyer, M. Smíšek, A. F. Villacampa, L. O. Maza, D. A. Medina, M. J. Schuster, F. Steidle, M. Vayugundla, M. G. Müller, B. Rebele, A. Wedler, and R. Triebel, "The madmax data set for visual-inertial rover navigation on mars," *Journal of Field Robotics*, vol. 38, pp. 833 – 853, 2021.

[32] M. Vayugundla, F. Steidle, M. Smisek, M. Schuster, K. Bussmann, and A. Wedler, "Datasets of long range navigation experiments in a moon analogue environment on mount etna," in *ISR 2018; 50th International Symposium on Robotics*, 2018, pp. 1–7.

[33] R. Giubilato, W. Stürzl, A. Wedler, and R. Triebel, "Challenges of slam in extremely unstructured environments: the dlr planetary stereo, solid-state lidar, inertial dataset," *IEEE Robotics and Automation Letters*, pp. 1–8, 2022.

[34] J. Jiang, J. Wang, P. Wang, B. Peng, and Z. Chen, "Lipmatch : Lidar point cloud plane based loop-closure," *IEEE Robotics and Automation Letters*, vol. 5, pp. 6861–6868, 2020.

[35] E. Galceran and M. Carreras, "A survey on coverage path planning for robotics," *Robotics and Autonomous Systems*, vol. 61, no. 12, pp. 1258–1276, 2013.