

# Topological Mapping with Visual Place Recognition using Panorama Shift of 360-degree Images and Spatial Consistency Check

Hiroki Tsukoshi

Graduate School of Science and  
Technology, Meiji University,  
Kanagawa, Japan

Yoshitaka Hara

Future Robotics Technology Center  
(fuRo), Chiba Institute of Technology,  
Chiba, Japan  
0000-0002-8667-1061

Yoji Kuroda

School of Science and  
Technology, Meiji University,  
Kanagawa, Japan  
ORCID

**Abstract**—This paper proposed a method for topological mapping with visual place recognition using panoramic image shift and spatial consistency check. In our previous work, We have been working on constructing topological maps that align with the phase structure of the real environment. However, due to the nature of conventional algorithms, loop detection may fail when revisited locations intersect or are in reverse direction along the robot's travel path during data acquisition. This method involves shifting panoramic image sequences acquired from a spherical camera according to a specific criterion, then extracting features from the shifted images. Next, nodes and arcs are created based on the visual similarity of the extracted feature vectors. By detecting loops while considering spatial consistency for these nodes and arcs, a topological map is constructed that aligns the real environment with the topological structure. Based on experimental results in both indoor and outdoor environments, the proposed method constructs a topological map with correct phase structure. This is achieved by improving visual place recognition accuracy independent of the driving path through panoramic image shift, and by suppressing false loop detections via spatial consistency check.

**Index Terms**—Topological Map, Visual Place Recognition, Panoramic Image

## I. 緒言

移動ロボットの自律走行に用いる地図には、メトリックマップとトポロジカルマップがある。メトリックマップとは点群地図や占有格子地図のことを指し、環境を座標系上で表現する。一方でトポロジカルマップは、環境の接続構造をノードとアーケで抽象的に表現する。

これまで筆者らは、実環境との位相構造の整合性を持ち、また接続方向の正確性も備えたトポロジカルマップの構築に取り組んできた [1] [2]。前稿 [1] では、全天球カメラで取得した全周画像から、Visual Place Recognition (VPR) 手法である AnyLoc [3] で特徴抽出を行い、画像間の視覚的類似度をもとに適応的なノード作成とループ検出を行った。しかし前稿 [1] ではそのアルゴリズムの特性上、データ取得時のロボットの走行経路において、再訪地点が交差する場合や逆方向の場合では、ループ検出ができない場合があった。

本稿では、前稿 [1] からの変更点として、全天球カメラで取得した各画像を一定の基準でパノラマシフトする。また、

ループ検出アルゴリズムを変更し、空間的一貫性を考慮したループ検出を行う。これにより、再訪地点が交差・逆方向の場合においても、パノラマシフトによる視覚場所認識の精度向上と、空間的一貫性チェックによるループ検出の誤検出抑制を通じて、実環境との位相構造の整合性を持つトポロジカルマップを構築する。

本稿の貢献は、以下の通りである。

- 全天球カメラで取得した全周画像に対して、方位についてのパノラマシフトを行うことで、進行方向が異なる再訪地点での視覚場所認識を可能とする。
- さらに空間的一貫性チェックを用いることで、ループ検出の誤検出を抑制し、実環境との位相構造の整合性を持つトポロジカルマップを構築する。
- 屋内環境と屋外環境での実験により、走行経路が交差する場合や逆方向の場合においても、正しい位相構造のトポロジカルマップを構築できることを示す。

## II. 関連研究

### A. ロボティクスにおけるトポロジカルマップ

ロボティクスにおける地図としての環境情報は、メトリックマップとトポロジカルマップの2種類に大別される。メトリックマップは環境を座標系上で精密に表現するため高精度な自己位置推定が可能であるが、環境の規模が大きくなるほど経路計画などに要する計算コストやメモリ使用量が増大する。また、物体が移動・変化し環境が大きく変化する場合はマップの再構築が必要であり、環境変化に対して頑健とは言えない。一方、トポロジカルマップは環境を特徴的な場所（ノード）とそれらの接続（アーケ）によるグラフ構造で抽象的に表現する手法である。このような表現は、計算効率や環境変化に対する頑健性の面で優れており、経路計画および自律移動において有効であり、人間の空間認知においてもトポロジカル構造に基づく認知地図の存在が示唆され、ロボティクスへの応用が進められている [4] [5] [6]。

Kuipers ら [4] は、認知地図 (Cognitive Map) の概念をロボットの空間表現に取り入れ、環境を複数の抽象レベルで捉える Spatial Semantic Hierarchy (SSH) を提案した。

また Thrun [5] は、グリッドマップの幾何学的な精密さと、トポロジカルマップの計算効率の高さを統合したハイブリッドな地図構築手法を提案した。このメトリック・ト

ポロジカルマップは、詳細な幾何情報と高次の抽象表現を両立させた地図表現を実現している。

Konolige ら [6] もメトリック・トポロジカルマップを用いたナビゲーション手法を提案している。小領域のメトリックな局所地図による局所的なナビゲーションとトポロジカルグラフによる大域的な経路計画を行うことで、詳細な幾何情報に常時依存する必要はないことを示唆している。

### B. Visual Place Recognition

Visual Place Recognition (VPR) は、画像を用いて一度訪れた地点を認識する技術である。初期の VPR は、SIFT や SURF などの local descriptor を bag of visual words [7] や VLAD [8] により固定長ベクトルに変換し、画像検索タスクに拡張する形式が取られていた。その後、NetVLAD [9] や GeM [10]、CosPlace [11]、AnyLoc [3] といった深層学習ベースの手法が登場し、高い性能を発揮している。

通常のカメラの画像では、同一地点であってもカメラの向きが異なることで撮影できる景色が異なり、適切に VPR を行うことできない。そこで、全天球カメラで取得した全周画像を用いて各地点の全ての景色を網羅する画像を取得することで、この問題に対処した [1]。しかし、再訪地点が交差・逆方向の場合においては全周画像を構成する物体の位置関係が異なることで、同一地点であっても VPR による特徴ベクトルに違いが発生し、適切に VPR を行うことができない場合があった。本稿では、パノラマシフトによる全周画像の切れ目の変更を行うことで上記問題の解決を図る。

## III. 視覚場所認識によるトポロジカルマップの構築

### A. 概要

図 1 に、提案手法のパイプラインと提案手法全体の概念図を示す。入力は全天球カメラから取得した全周画像であり、再訪地点が交差・逆方向の場合においても、視覚場所認識を可能にするために用いる。まず、事前に収集した動画データから、画像列を作成する。次に、画像列に対して、パノラマシフトを行う。その後、パノラマシフトされた画像列から特徴ベクトルを抽出し、それらの視覚的類似度を基に、ノードとアーケ作成する。最後に、空間的一貫性チェックを用いたループ検出を行い、トポロジカルマップを構築する。

### B. 全周画像のパノラマシフトを用いた特徴抽出

入力された動画形式の全周画像を一定の時間間隔でサンプリングし、画像列を作成する。次に、すべての画像に対して、一定の基準でパノラマシフトを行う。

図 2 に、パノラマシフトの概念図を示す。画像内の各縦列の特徴量を比較し、最も高い列を画像の左端に来るようシフトする。この処理により、同一地点に異なる方向から再訪した場合でも、画像を構成する物体の位置関係が変わらない画像を取得することができ、視覚的類似度の向上が期待できる。

次に、パノラマシフトされた画像列から特徴ベクトルを抽出する。提案手法では、特徴抽出器として AnyLoc-VLADDINOv2 [3] を用いる。

式(1)に各画像から抽出された特徴ベクトルの集合を示す。

$$V = \{v_1, v_2, \dots, v_n\} \quad (1)$$

ここで、 $n$  は抽出された特徴ベクトルの総数を表す。

### C. ノードとアーケの作成

画像間の視覚的類似度を基に、ノードとアーケを作成する。視覚的類似度の評価には、各特徴ベクトル間のコサイン類似度を用いる。類似度に対して閾値  $node\_sim\_th$  を設定し、ノードの粗密を調整する。具体的には、特徴ベクトル  $v_i$  を基準として、それ以降の特徴ベクトル  $v_j$  ( $j > i$ ) の類似度が  $node\_sim\_th$  を上回った場合はノードは追加されず、下回った場合は  $v_j$  が新たなノードとして追加される。次の回では、基準となる特徴ベクトルを  $v_j$  とし、同様の処理を繰り返すことでロボットの走行経路に沿ったノードとアーケが作成される。式(2)に、作成されたノード列を示す。

$$N = \{n_1, n_2, \dots, n_m\} \quad (2)$$

ここで、 $m$  は作成されたノードの総数を表す。

提案手法では、前稿 [1] よりも  $node\_sim\_th$  を高く設定し、ノード間の距離を短くする。ノード間隔を短くすることで、同一地点を表現するノードを増やすことができ、後述するループ検出でのクエリ画像と類似度の高い画像を増やすことを目的としている。

### D. 空間的一貫性チェックを用いたループ検出

単一画像のコサイン類似度だけでは、類似している画像が取得できる別地点との誤検出が発生しやすい。また前稿 [1] では、そのアルゴリズムの特性上、同一地点を異なる方向から走行する場合はループ検出できない場合がある。上記のようなループの誤検出や未検出により、実環境と位相構造が異なるトポロジカルマップが構築される。

提案手法では、空間的に連続したノードを用いたループ検出を行うことで、位相的に正しいトポロジカルマップの構築を行う。ループ検出の基本的な考え方として、スライディングウィンドウを用いる。ウィンドウには事前に決められた  $window\_size$  個のノードが格納され、それらとクエリノードとの類似度を計算する。ノード列に対してウィンドウを一つずつずらしながら走査し、ループ検出対象のノードを特定する。

クエリノードを  $n_q$ 、それに対し類似度を計算するノードをターゲットノード  $n_t$  とする。まず、ウィンドウ内のノードのうち、 $n_q$  とループ検出に対する閾値  $loop\_sim\_th$  以上の類似度をもつ  $n_t$  をカウントする。次に、一つのウィンドウに対するカウントがカウント用の閾値  $count\_th$  を超えた場合、そのウィンドウをループ検出の候補ウィンドウとする。最後に、候補ウィンドウ内で  $n_q$  と最も類似度の高い  $n_t$  をループ検出の候補ノード  $n_c$  とし、ノードペア  $(n_q, n_c)$  を保存する。この流れをすべてのノードをクエリノードとして行う。

ウィンドウ内には  $n_q$  自身や、 $n_q$  と時間的に連続した近傍のノードが含まれる場合があり、それらがループ検出されることを防ぐ必要がある。そのため、ウィンドウ内に  $n_q$  自身と、 $n_q$  の近傍ノードが含まれる場合は、そのウィンドウでの処理をスキップする。スキップする範囲は、クエリノード  $n_q$  の前後  $neighbor\_node\_num$  分のノードとする。

最後に、保存されたループ検出のノードペアに対して相互チェックを行い、ループ検出を確定する。ここまで処理から、 $n_c$  もまたクエリノードになるため、 $(n_q, n_c)$  と

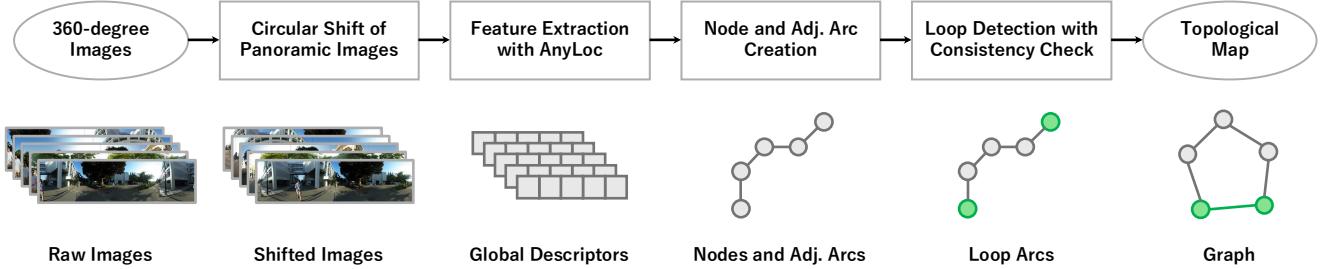


Fig. 1: 提案手法のパイプライン



Fig. 2: パノラマシフトの概念図

`loop_sim_th`よりも高い類似度のノードが、`window_size`に対して `count_th`以上あるウィンドウのうち、最大の類似度のノードがループ検出候補

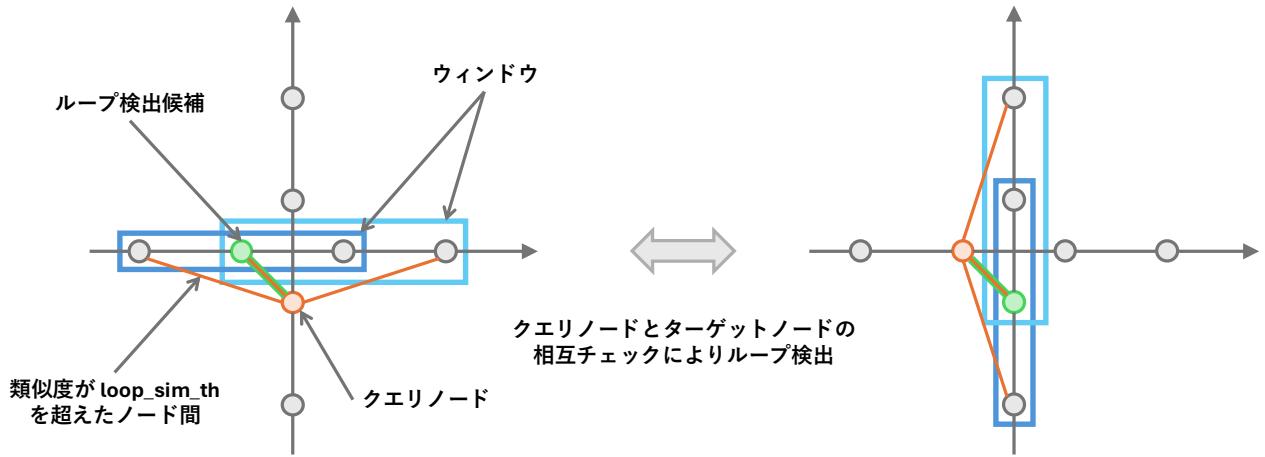


Fig. 3: 空間的一貫性チェックの概念図

$(n_c, n_q)$  の両方が保存されている場合にループ検出を確定し、ノードペア間にアーケを追加する。

これらの処理により、ループ検出の誤検出を抑制しつつ位相的に正しいトポロジカルマップを構築できると考えられる。

#### IV. 実験

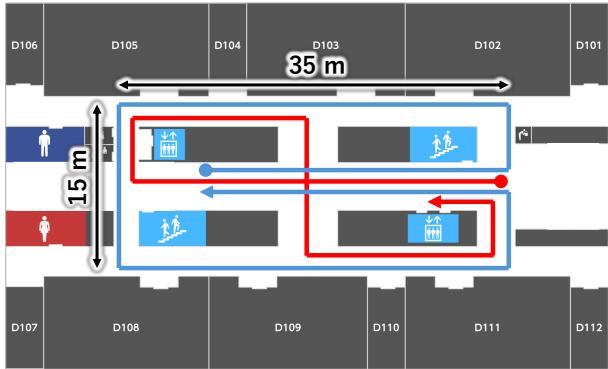
##### A. 実験条件

図 4 に、各実験環境と走行経路を示す。実験環境および走行経路は、屋内交差、屋内逆方向、屋外交差、屋外逆方

向の 4 種類である。図 4(a) の赤色の経路は屋内交差、青色の経路は屋内逆方向である。

WHILL Model CR をベースとした後輪差動駆動の車輪ロボットに RICOH THETA S を搭載し、各環境を走行させ全周画像を取得した。すべてのデータ処理には、CPU: Intel Core i9-14900KF 5.700GHz, RAM: 64GB, GPU: NVIDIA GeForce RTX 4090, VRAM: 24GB を搭載した PC を用いた。

検証用データとして、メトリックな情報も取得した。3D Lidar (Velodyne VLP-32MR), IMU (Xsens MTi-30), 車輪オドメトリを用い、点群位置合わせに基づいてメトリッ



(a) 屋内交差（赤色）・屋内逆方向（青色）



(b) 屋外交差



(c) 屋外逆方向

Fig. 4: 実験環境と走行経路

クなロボット位置を求めた。これらの情報は、トポロジカルマップの構築には一切用いていない。

#### B. 画像のパノラマシフトを用いた特徴抽出によるコサイン類似度の変化

図5に屋外環境で取得した画像列に対するコサイン類似度の変化の例を示す。図5(a)はパノラマシフトなしの場合、(b)はパノラマシフトありの場合である。Query 画像と Target 画像は、図4(b)の走行経路において交差地点で取得した画像であり、同一地点ではあるがロボットの進行方向が異なる際に取得した画像である。

図5(a)では、Query 画像と Target 画像を比較すると、画像を構成する建物や自然物は同じであるが、位置関係が異なる。一方、図5(b)では、パノラマシフトにより画像の切れ目が変更され、人などの動くものを除き、画像を構成する物体の位置関係が同じになっている。コサイン類似度はパノラマシフトなしの場合は 0.360 であるのに対し、パノラマシフトありの場合は 0.641 であり、78.1%の改善が見られる。ただし、コサイン類似度の有効数字は 3 桁として

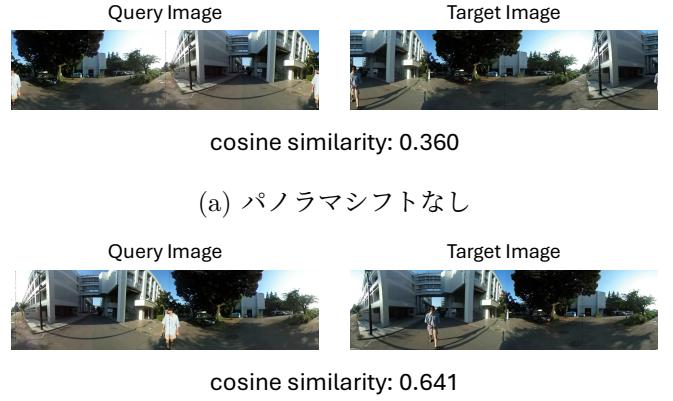


Fig. 5: パノラマシフトによるコサイン類似度の変化

いる。これらの結果から、同一地点を異なる方向から再訪した場合でも、パノラマシフトにより画像の切れ目を変更し、画像内の物体の位置関係を揃えることで、視覚的類似度を向上させることができると考えられる。

#### C. 一貫性チェックを用いたループ検出によるトポロジカルマップの構築

本稿のトポロジカルマップの構築では、動画から画像列を作成する際の一定の時間間隔を 1.0 [s]、ノードの粗密を調整するパラメータ  $node\_sim\_th = 0.80$  で統一した。

図6に各走行経路においてのメトリックマップ上のノードの作成位置を示す。ノードの数や粗密は各画像から抽出される特徴ベクトルのコサイン類似度に依存するため、同じ動画データを用いてもパノラマシフトの有無によりノードの数や位置が異なる。図6には、パノラマシフトを行った後の画像を用いて作成したノードを示す。IV-A 節より、ノードの位置はメトリックな情報から算出している。図6より、 $node\_sim\_th$  を高い値に設定することで、ノード間隔は短くなり、図4に示す走行経路と比較すると、走行経路上のほぼ全ての地点にノードが作成されていることが分かる。

次に、構築したトポロジカルマップと、ループ検出の正誤を評価する。ループ検出の正誤判定には、IV-A 節で述べたメトリックな情報を用いた。本稿では、ループ検出の精度評価のために 3 つの手法でトポロジカルマップを構築、それぞれの結果を比較し、定性・定量評価する。

**単純な閾値**: 通常の全周画像を用いて特徴ベクトルを抽出し、クエリノードとターゲットノードのコサイン類似度が閾値  $loop\_sim\_th$  を上回った場合にループ検出を行う。

**パノラマシフト + 単純な閾値**: パノラマシフトを行った全周画像を用いて特徴ベクトルを抽出し、単純な閾値と同様のループ検出を行う。

**パノラマシフト + 一貫性チェック**: パノラマシフトを行った全周画像を用いて特徴ベクトルを抽出し、空間的一貫性チェックを用いてループ検出を行う。

全ての実験において、 $loop\_sim\_th = 0.55$  に統一した。パノラマシフト + 一貫性チェックにおいては、 $window\_size$

TABLE I: 手法一覧

Method	Simple Th.	P. Shift & Simple Th.	Ours
P. Shift	×	○	○
Sp. Co.	×	×	○

= 5, *count\_th* = 2 に統一し, *neighbor\_node\_num* は走行経路により適応的に変更した。

ループ検出の正誤判定には, IV-A 節で述べたメトリックな情報を用いた。各ノードに対して、ノード間のメトリックな距離が一定の閾値以内であれば同一地点と判定し、閾値以内のノード同士がループ検出された場合は True Positive (TP), 閾値以上のノード同士がループ検出された場合は False Positive (FP) とした。また、閾値以内のノード同士がループ検出されなかった場合は False Negative (FN) とし、それ以外の場合は True Negative (TN) とした。ただし、単純な閾値では、時間的に近いノード同士がループ検出され、アーケが追加されることがある。これはメトリックな距離は閾値の範囲内ではあるものの、トポロジカルマップの位相構造を崩すため、FP として扱う。ループ検出の正誤判定に用いるメトリック距離の閾値は、1.0 [m] とした。

図 7~9 に、各走行経路において構築したトポロジカルマップを示す。以降、屋内環境と屋外環境に分けて、トポロジカルマップの構築結果を比較する。

まず、屋内の走行経路である図 4(a) とその実験結果である図 7, 8 を比較する。単純な閾値とパノラマシフト + 単純な閾値においては、TP なループ検出はあるものの、大きな FP のループが存在し、位相構造の崩れが顕著に見られる。一方で、パノラマシフト + 一貫性チェックでは、他 2 手法で検出された TP を残しつつ、FP を抑制している。

次に、屋外の走行経路である図 4(b), 4(c) とその実験結果である図 9, 10 を比較する。まず屋外交差において、単純な閾値では交差地点をループ検出できていないのに対し、パノラマシフト + 単純な閾値では交差地点をループ検出でき、大きなループを構築できている。これはパノラマシフトによる再訪地点での視覚的類似度の向上により、ループ検出が可能になったと考えられる。さらに、パノラマシフト + 一貫性チェックでは、パノラマシフトによる TP を維持しつつ、FP を抑制し、実環境と同様の位相構造をもつトポロジカルマップが構築されていることが分かる。屋外逆方向においても、パノラマシフトによる TP の増加、一貫性チェックによる FP の抑制が見られる。

これらの結果から提案手法は、パノラマシフトによる再訪地点の視覚的類似度の向上により TP なループ検出を増加させつつ、空間的一貫性チェックにより FP なループ検出を抑制し、実環境との位相構造の整合性をもつトポロジカルマップを構築できることが分かる。

## V. 結言

本稿では、全周画像の一定の基準のパノラマシフトにより、交差や逆方向での再訪においても視覚場所認識を可能とし、また空間的一貫性チェックによりループ検出の誤検出を抑制することで、実環境との位相構造の整合性を持つトポロジカルマップを構築する手法を提案した。

屋内外の 4 種類の走行経路で実験を行い定性評価した。その結果、パノラマシフトにより再訪地点での視覚的類似度

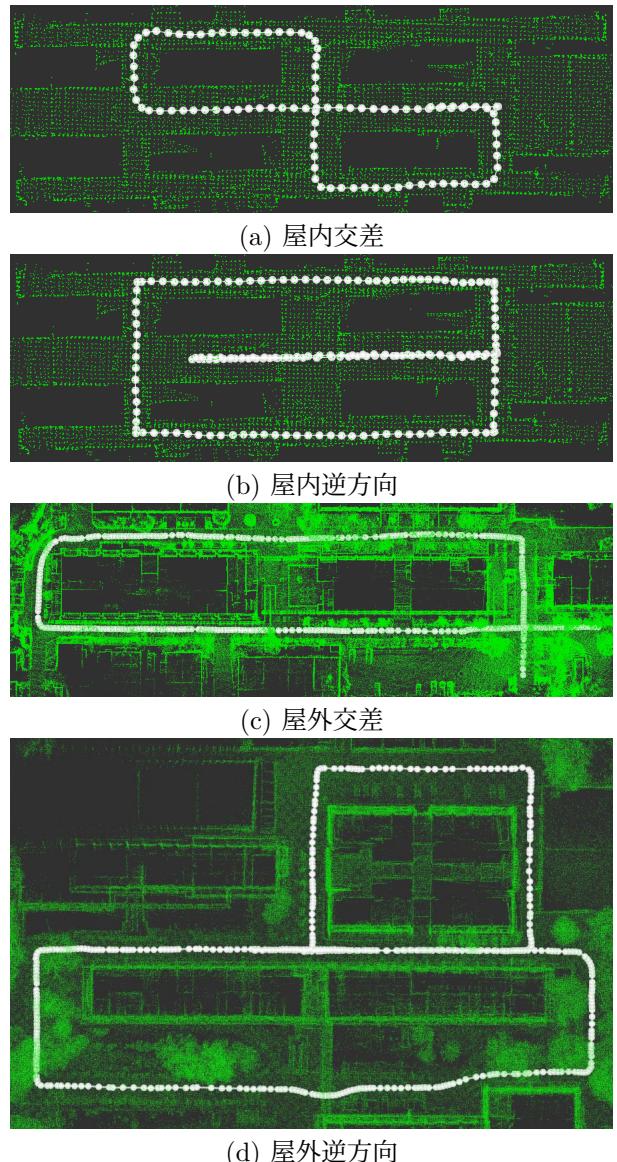


Fig. 6: メトリックマップ上のノードが作成された位置

TABLE II: 定量評価：屋内交差

	Simple Th.	P. Shift & Simple Th.	Ours
TP	14	18	9
FP	16	34	0
TN	72	58	81
FN	2	0	9
Precision ↑	47%	35%	100%
Accuracy ↑	83%	69%	91%
F1 Score ↑	61%	51%	67%

を向上させ、交差や逆方向での再訪においてもループ検出が可能となった。さらに、空間的一貫性チェックにより誤検出を抑制し、実環境との位相構造の整合性を持つトポロジカルマップを構築できることを示した。

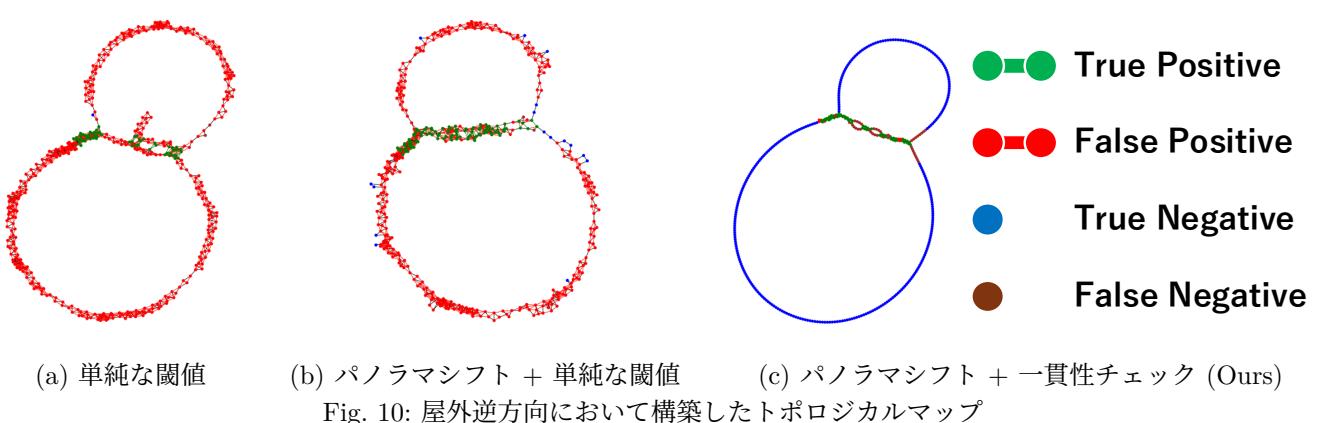
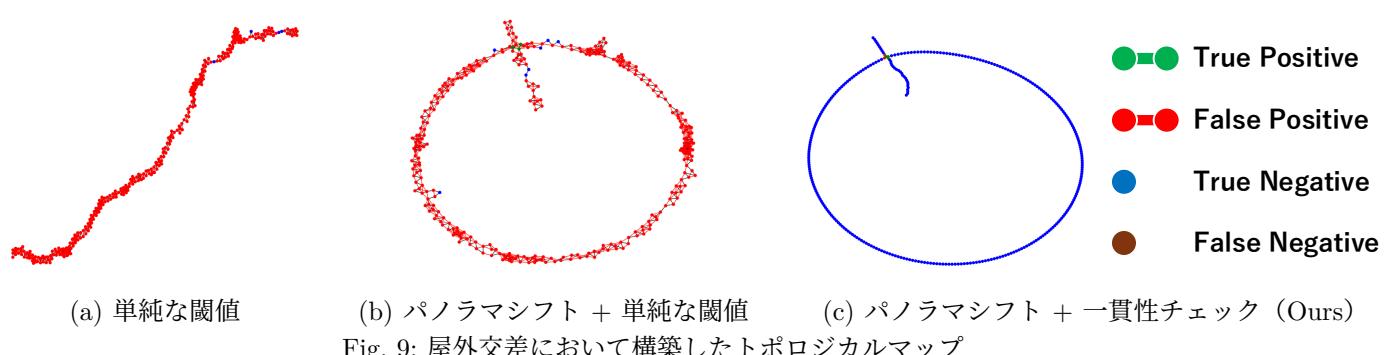
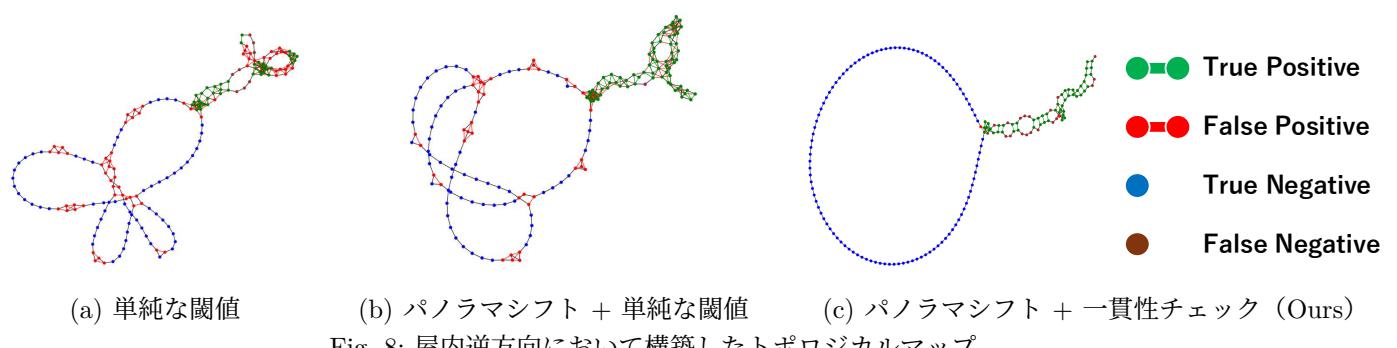
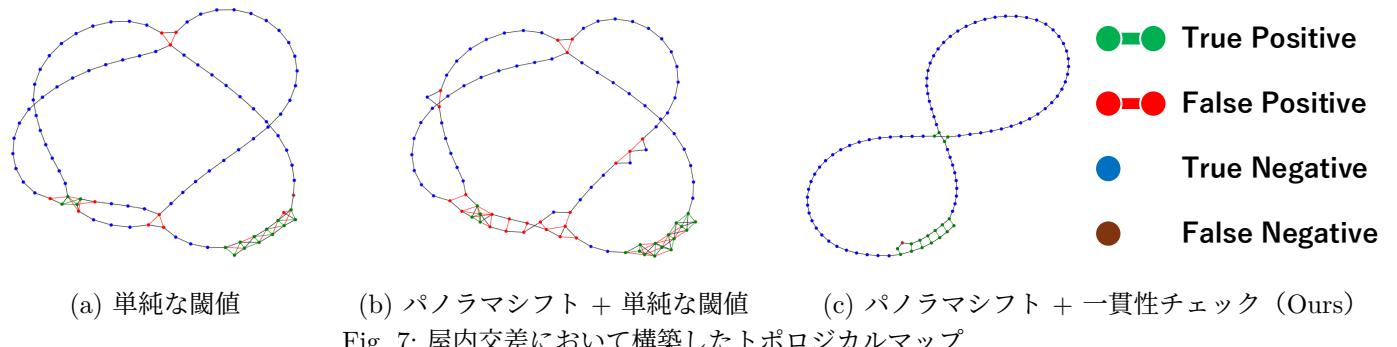


TABLE III: 定量評価：屋内逆方向

	<b>Simple Th.</b>	<b>P. Shift &amp; Simple Th.</b>	<b>Ours</b>
TP	49	87	34
FP	146	125	6
TN	64	66	105
FN	49	9	62
Precision ↑	25%	41%	<b>85%</b>
Accuracy ↑	37%	53%	<b>67%</b>
F1 Score ↑	33%	<b>56%</b>	50%

TABLE IV: 定量評価：屋外交差

	<b>Simple Th.</b>	<b>P. Shift &amp; Simple Th.</b>	<b>Ours</b>
TP	0	2	2
FP	821	716	1
TN	4	7	314
FN	4	2	2
Precision ↑	0%	0.28%	<b>67%</b>
Accuracy ↑	0.48%	1.2%	<b>99%</b>
F1 Score ↑	0%	0.55%	<b>57%</b>

TABLE V: 定量評価：屋外逆方向

	<b>Simple Th.</b>	<b>P. Shift &amp; Simple Th.</b>	<b>Ours</b>
TP	38	69	42
FP	1383	1124	54
TN	1	16	317
FN	50	15	42
Precision ↑	2.7%	5.8%	<b>44%</b>
Accuracy ↑	2.6%	6.9%	<b>79%</b>
F1 Score ↑	5%	11%	<b>47%</b>

## 謝辞

本研究は、明治大学自律型ロボット研究クラスターの下で実施された。ここに御礼申し上げる

## REFERENCES

- [1] Daichi Ogiwara, Naoki Takahashi, Kosuke Kimizuka, Takaya Nakao, Yoshitaka Hara, and Yoji Kuroda: “Topological Mapping using Visual Place Recognition”, *Proc. of Robotics Symposia (ROBOSYM)*, 2025. (in Japanese)
- [2] Takaya Nakao, Yoshitaka Hara, and Yoji Kuroda, “Topological Mapping with Constrained Optimization based on Visual Place Recognition and Orientation Constraints”, *Proc. of IEEE/SICE Int. Sympo. on System Integration (SII)*, 2026.
- [3] Nikhil Keetha, Avneesh Mishra, Jay Karhade, Krishna Murthy Jatavallabhula, Sebastian Scherer, Madhava Krishna, and Sourav Garg: “AnyLoc: Towards Universal Visual Place Recognition”, *IEEE Robotics and Automation Letters*, vol. 9, no. 2, pp. 1286–1293, 2023.
- [4] Benjamin Kuipers: “The Spatial Semantic Hierarchy”, *Artificial Intelligence*, vol. 119, no. 1–2, pp. 191–233, 2000.
- [5] Sebastian Thrun: “Learning Metric-Topological Maps for Indoor Mobile Robot Navigation”, *Artificial Intelligence*, vol. 99, no. 1, pp. 21–71, 1998.
- [6] Kurt Konolige, Eitan Marder-Eppstein, and Bhaskara Marthi: “Navigation in Hybrid Metric-Topological Maps”, *Proc. of IEEE Int. Conf. on Robotics and Automation (ICRA)*, 2011.
- [7] James Philbin, Ondrej Chum, Michael Isard, Josef Sivic, and Andrew Zisserman: “Object Retrieval with Large Vocabularies and Fast Spatial Matching”, *Proc. of IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [8] Hervé Jégou, Matthijs Douze, Cordelia Schmid, and Patrick Pérez: “Aggregating Local Descriptors into a Compact Image Representation”, *Proc. of IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2010.
- [9] Relja Arandjelović, Petr Gronat, Akihiko Torii, Tomas Pajdla, and Josef Sivic: “NetVLAD: CNN Architecture for Weakly Supervised Place Recognition”, *Proc. of IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2016.
- [10] Filip Radenović, Giorgos Tolias, and Ondrej Chum: “Fine-Tuning CNN Image Retrieval with No Human Annotation”, *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 41, no. 7, pp. 1655–1668, 2019.
- [11] Gabriele Bertoni, Carlo Masone, and Barbara Caputo: “Rethinking Visual Geo-localization for Large-Scale Applications”, *Proc. of IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR)*, 2022.