計測自動制御学会東北支部 第 317 回研究集会 (2018.7.19)
資料番号 317-2

ビジュアルランドマークに基づく自律移動のためのベンチマークデータセットの構築と評価

Construction and evaluation of benchmark data set for autonomous movement based on visual landmark

○福士遼真 *，間所洋和 *，佐藤和人 *

○ Ryoma Fukushi*, Hirokazu Madokoro*, Kazuhito Sato*

*秋田県立大学大学院システム科学技術研究科

*Graduate School of Systems Science and Technology, Akita Prefectural University

キーワード： 移動ロボット (mobile robot)，単眼カメラ (monocular camera)，ビジュアルランドマーク (visual landmark)，ベンチマークデータセット (benchmark datasets), 顕著正マップ (saliency maps)

連絡先： 〒 015-0055　秋田県由利本荘市土谷海老ノ口８４－４秋田県立大学 システム科学研究化 機械知能システム学専攻 脳情報工学研究室

福士遼真，Tel.: 0184-27-2000，Fax.: 0184-27-2180，E-mail: m20a022@akita-pu.ac.jp

1. 緒言

ロボットが自律移動を行う際に，一般的には自己位置推定，周辺環境認識，経路計画，経路追跡の 4 項目をロボット自身が正確に行う能力が求められる．また，人間社会と同じ空間で活動させるには，人間の識別，影響への対策が必要である．そのため，周辺環境認識は不可欠な要素であり，認識機能として地図構築と自己位置推定がある [1]．地図構築は障害物やランドマークなどを記した地図を自動生成する技術であり，自己位置推定は地図上のロボットの位置を推定する技術である．

我々は，ロボットにビジュアルランドマークを自動的に抽出させることにより，自律移動の実現を目指した研究に取り組んでいる．先行研究 [2] では，人物に影響されないビジュアルランドマークの検出と自己位置の意味的認識を実現した．取得したオリジナルデータセットによる性能評価を実施した本研究では，先行研究 [2] で提案された意味的シーン認識法を踏襲し，大規模データセットの構築と評価を試みる．その際に，区間ごとのデータ量の差異が，精度に与える影響の評価に加えて，カテゴリマップへの写像結果から，GT（Ground Truth）の定義へのフィードバックについて，精度の比較を通じた区間変更の有用性に関する検証を行う．

2. 関連研究

ビジュアルランドマークに基づくロボットの自己位置認識に関する研究は，多様な手法によって進められている．ランドマークを事前に設置するアプローチ [3][4] や，画像内における新規のランドマーク抽出を目的とした研究 [5] が進められているが，閉鎖的空間の実験環境や評価実験

Fig. 1　単眼カメラ PIXPRO を搭載した移動ロボット Double

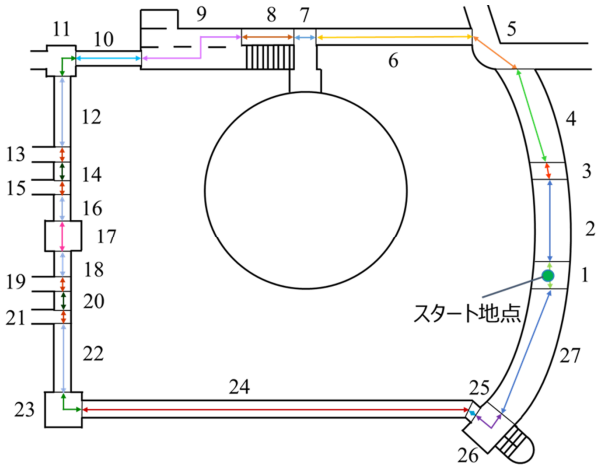

Fig. 2　GT に基づいた実験環境と区域の地図

の不足など課題が残されている．さらに，人物干渉を考慮し，SLAM による移動ロボットのナビゲーションの研究も進められている．[6)]

3.　データセット構築

3.1　使用機器

研究で使用した移動ロボット (Double; iPresence) と単眼カメラ (PIXPRO sp360; OLYMPUS) の外観を図 1 に示す．先行研究 [2)] では，頭部に装着する iPad の内蔵カメラを用いて画像を取得した．しかし，撮影した画像は iPad 内には保存されず，無線 LAN を通じてリアルタイムに送信された．ただし，無線 LAN の電波状況によっては，フレーム落ちが発生した．そこで本実験では，高解像度に加えて，広域視野に対応するために，単眼カメラを当該ロボットに搭載し，時系列の正面シーン画像を取得した．

3.2　実験環境

図 2 に，実験環境および GT として分割した区間を示す．時系列画像データは本施設を 1 周しながら取得した．移動距離は約 392m である．取得した映像は，CW，CCW それぞれ 1 周を 1 データセットとし，各周回方向を 3 回施行した．左右の周回方向について，同一の区間においても，視覚的にシーンの見え方に変化が発生する．本研究では，学習を行う際に，左右方向を考慮し，各 3 データセットを使用する．ロボットの移動は，自律走行でなく，操作用として用意したノートパソコンのキーボードを使って手動により走行させた．取得画像を地図情報付きベンチマークデータセットとして，視界の変化に基づいて 27 ゾーンを分類した．図 3 は，各区間における代表的な画像を示す．分割された領域には実験者の主観性が含まれているが，評価実験より，カテゴリマップを使用して統合と分割を行った．なお，ロボットを定速撮影を行ったが，各ゾーンの長さ依存して画像数の差が生じた．

3.3　区間統合

我々のベンチマークデータセットは，区間の長さによって，各区間における画像枚数が異なっている．この差は対向伝播ネットワーク（Counter Propagation Networks:　CPN）のマッピング結果として認識精度に影響する．一方，カテゴリを生成する際に淘汰されるラベルが存在する．GT による時系列シーン変化の適用によれば，図 2 の左側領域の区間 13 と区間 19 が，カテゴリマップへの写像の際に淘汰された．淘汰された

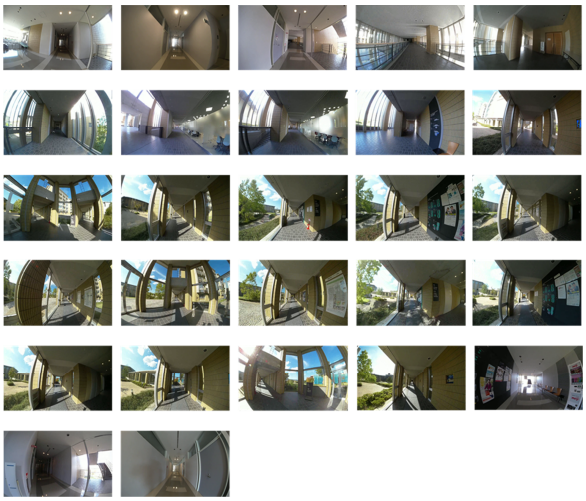

Fig. 3　27 区間の画像サンプル

Fig. 4　CCW における区間 13(左)，19(右) のサンプル画像

両区間におけるデータについて，主観的に比較したところ，類似したシーンであった．図 4 に両区間のサンプル画像を示す．両区間は，人間が視覚的に判断する際にも類似度が高く，ビジュアルランドマークによる自己位置の判断が難しい．そこで，両区間を同じカテゴリとして統合する．これ以外にも，シーンが類似している区間が存在し，以下に示す．

(i) 区間 13，15，19，21
(ii) 区間 14，20
(iii) 区間 12，16，18，22，24
(iv) 区間 2，27

構築したデータセットの画像の総数は 2,300 枚である．区間によってデータ量に極端な差が存在する．特に，区間 13,15,19,21 の画像数が少なく，学習不足の可能性がある．そのため，CPN に学習に使用する画像の量を増やすためにゾーンを統合し，学習データを量的に的に確保した．

3.4　GT 定義

本実験は，映像の取得からビジュアルランドマークを自動的に抽出し，意味的シーン分類を行う．提案手法における CPN による写像学習では，ネットワークが保持している結合荷重とラベルに加えて，新たにカテゴリ候補のラベルを用いる．27 回のシーン変化に対する区間統合による精度向上を目的として、学習によるラベルの淘汰や類似区間の統合を踏まえて，複数の GT を定義した．表 1 に定義した GT を示す．作成した GT に基づき，交差検定 [8] によって評価し，精度を比較する．

4.　評価実験

4.1　特徴抽出

図 5 に，先行研究 [2] で提案された方法と同様の特徴抽出方法の全体構造を示す． 最初に，注目度の高い視覚的ランドマークの候補に用いられる関心領域を抽出するための原画像に顕著性マップを適用する．次に AKAZE の特徴を抽出する． 人物を含む画素の集合は顕著性が高いため，人物領域は HOG マスクを用いて抽出される．最後に，人間の領域をマスク処理した後，関心領域からの AKAZE 記述子を使用して特徴を抽出する．

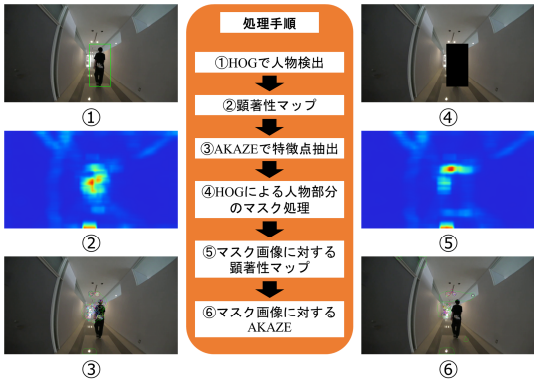

Fig. 5　特徴抽出手順

Table 1　区間統合による GT 定義

GT	統合区間	統合区間	統合区間
1	-		
2	13,19	-	
3	13,15,19,21	-	
4	13,15,19,21	14,20	-
5	13,15	19,21	-
6	13,15	19,21	14,20

4.2　抽出結果

評価方法は，テスト用データを 1 セット残した状態で，他のデータセットを学習用データとして用いる 1 リーブアウトの交差検定とした．表 1 に示す GT1〜6 に対して，交差検定による認識精度を表 2(CW)，表 3(CCW) に示す．学習結果として生成されたカテゴリマップ表現を図 6，図 7 に示す．各マップの右側に示すカラーバは，区間 1 から区間 26 までの位置カテゴリと対応している．カテゴリ数は統合した区間の発生に応じて変動する．色温度が低い (青色) カテゴリが区間 1 となっており，色温度に応じて順にカテゴリが対応付いており，最も高温 (赤色) のカテゴリが区間 27 となる．通常のカテゴリマップでは，近傍と競合学習により，各カテゴリは類似性に基づいて集合を形成することが多いが，本データセットのように，特徴量が複雑になる場合には，同じカテゴリでも複数の部分集合に分散して分布するという特性を示す．

Table 2　CW における認識精度

GT	認識率 [%]	精度差 [%]	カテゴリ数
1	70.76	0	26
2	71.92	1.17	25
3	70.42	-0.33	24
4	69.55	-1.21	23
5	69.55	-1.21	25
6	70.8	0.05	24

Table 3　CCW における認識精度

GT	認識率 [%]	精度差 [%]	カテゴリ数
1	72.24	0	25
2	73.22	0.98	26
3	72.11	-0.13	24
4	73.85	1.62	23
5	74.54	2.3	25
6	72.15	-0.09	24

4.3　考察

定義した 6 種の GT の精度を比較した．GT2 の統合の場合において，両周回方向ともに精度が向上した．よって，学習の際に淘汰される区間の統合は，精度向上に結び付くことを示唆している．各区間の認識率の精度は，混同対照表を用いて分析する．混同対照表のヒートマップ表現を図 8 と図 9 に示す．類似区間の統合である，GT3〜GT6 認識率を算出すると，精度が低下する GT が発生した．画像データの比較より，画像の 3 分の 1 を占める外観が大きく変化していた．人間が屋内環境を認識する基準は，屋内を注視し認識する．しかしながら，区間 13，15，19，21 では，屋内の構造物から抽出された特徴は少なく，大部分の特徴点がガラス越しの屋外環境から抽出されていた．教師データの特徴から，屋外環境の特徴点群から学習を行ったと考えられる． GT6 は GT5 の CCW において精度向上の結果を得たため，区間 14，20 を追加した．結果は CCW における統合後の認識率が低下した．混同対照表による認識の詳細を確認したと

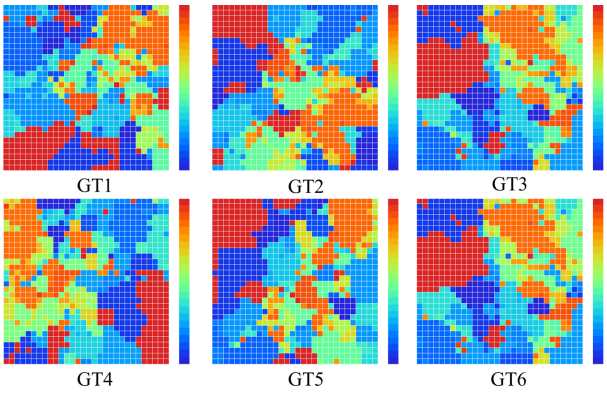

Fig. 6　生成したカテゴリマップ (CW)

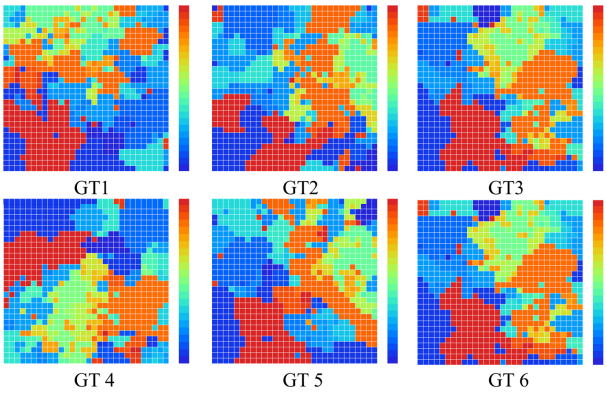

Fig. 7　生成したカテゴリマップ (CCW)

ころ，他区間のデータが統合された区間を認識している場合が増加していた．すなわち，統合区間の数を増やしたことで，従来のラベルの認識が新しく統合された区間へ誤認識として発生した．

生成されたカテゴリマップの分布特性は，GT3 と GT6 が，CW と CCW とも極めて類似している．結合荷重がランダムに初期化されるカテゴリマップは，このような類似性を示すことは稀である．Grossberg 層ユニットと Kohonen 層ユニットとの結合荷重は，ランダムに初期化される．ここで，GT の割り当てを変更しても，コードブックによるヒストグラム特徴量の入力パターンには変わりがないため，ランダム性を支配する乱数の初期値が同じなら，入力層と Kohonen 層の結合荷重は同じ値に導かれる．一方，GT3 と GT6 で異なるラベルが割り振られているということは，Grossberg 層と Kohonen 層は，乱数の初期値が同じでも，教師信号のパターンが異なるため，GT のパターン毎に違う値へと学習が進む．混同対照表から認識結果の詳細を確認したところ，統合区間の対象となる区間 13，15，19，21 と区間 14，20 以外の認識精度が全く同じ結果となった．よって，GT の変更が全体の GT 数と比較して微小の場合，入力特徴量とネットワークの動作を支配するパラメータが同一の場合は，カテゴリマップでの分布の変化として出現しないことを示唆する結果であった．認識率の向上の底上げとなるのは，画像枚数の多い区間であり，画像枚数が少ない区間では誤認識の割合が高い．移動距離の短い区間においても，正しい認識を行うためには，画像枚数を確保してデータの量的側面を確保しなければならないと考えている．

5.　結言

本研究では，ロボットと人間が共生する実環境における，ビジュアルランドマークに基づく自律移動ロボットの自己位置の意味的認識を目的として，大規模データに対する区間ごとのデータ量の差による影響を評価した．加えて，カテゴリマップへの写像結果から，GT の定義へのフィードバックについて，精度の比較を通じた区間変更の有用性を検証した．CW，CCW の各 3 データセットを用いて，精度評価を行った結果，区間内のデータ量が少ない場合，学習によるラベル付けの際に，淘汰されるラベルの存在が明らかになった．GT の定義を変更し，淘汰された区間を統合することで学習によるラベルの分類について改善と精度の向上が期待された．実験より，カテゴリマップの写像結果に加えて，パーツベースの特徴点を抽出したシーン画像を分析することで，人間と，コンピュータビジョンとしての機械学習に基づくロボットの環境認識の差異について検討した．その中で，取得画

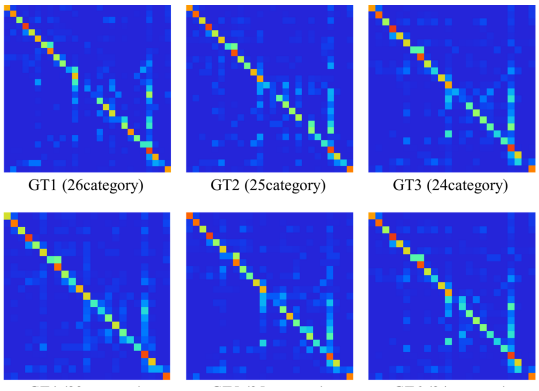

Fig. 8　CW における混同対照表

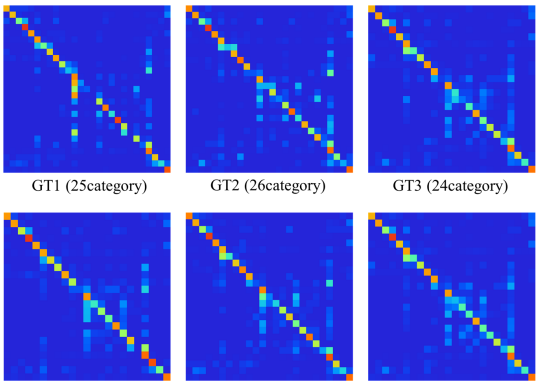

Fig. 9　CCW における混同対照表

像にガラスの壁面を通して屋外環境が見えていた場合に，提案手法では屋内外を区別するメカニズムを有していないため，屋外の構造物に対して特徴点を算出し，ビジュアルランドマークとする傾向が判明した．

今後の実験では，屋内環境を対象とし，日中の時間帯にデータを取得した．実験環境は壁面はガラス構造の区間があり，取得画像における屋外環境の情報が大きい場合，屋外の構造物に対して，特徴点を算出する傾向がある．この傾向の検証を目的として，夜間の時間を対象とした同じ実験環境におけるデータセットを作成する．したがって，取得データセットより，顕著性及び，特徴量を屋内環境に集中させた，区間における意味認識の検証実験を行う予定である．

参考文献

1) 友納 正裕 “ 移動ロボットの環境認識 : 地図構築と自己位置推定 (「不確実性に挑むロボティクス」特集号)” システム制御情報学会誌 システム／制御／情報 vol.60(12), pp 509-514, 2016.

2) Y. Ishikoori, H. Madokoro, and Kazuhito Sato, “Semantic Position Recognition and Visual Landmark Detection with Invariant for Human Effect,” Proc. IEEE/SICE International Symposium on System Integration (SII), 2017.

3) D. Chen, Z. Peng, and X. Ling, “A low-cost localization system based on Artificial Landmarks,” IEEE International Conference on Robotics and Biomimetics, 2014.

4) J. Buhl, T.A. Jensen, and H.A. Rasmussen, “Autonomous Robot Navigation Based On Natural Landmark Triangulation,”2005.

5) P. Sala, R. Sim, A. Shokoufandeh, and S. Dickinson, “ Landmark Selection for Vision-Based Navigation,” IEEE Trans. Robotics, vol.22, no. 2, pp.334-349, Apr. 2006.

6) 森岡博史, 李想揆, T. Noppharit, 長谷川修, “ 人の多い混雑な環境下での SLAM による移動ロボットのナビゲーション,” 第 28 回日本ロボット学会学術講演会, 2010.

7) 間所洋和, 佐藤和人, 石井雅樹, “視野画像列を用いた世界像の獲得と自己位置の推定,” 電子情報通信学会論文誌 D–II, vol.J83–D–II, no.12, pp.2587–2596, Dec. 2000.

8) R. Kohavi, “ A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection, ” Proc. International Joint Conference on Artificial Intelligence, vol.2, pp.11371143, 1995.