最近傍Q学習に基づくランドマーク観測行動計画

○若山和樹（福井大学）　吉田光希（福井大学）　森下裕大（福井大学）
山本稜悟（福井大学）　田中完爾（福井大学）

1. 緒言

視覚移動ロボットの分野において，クロスドメイン自己位置推定の研究が行われている．一般に自己位置推定とは，これまでに収集された記録画像に対して，視野画像を入力とし，ロボットの状態を推定する問題である．この自己位置推定問題は，ドメイン（時刻，天候，季節）が，記憶画像とライブ画像との間で異なる場合には挑戦的なものになる．この問題に対し，深層学習を用いて，画像から不変特徴量を検出する研究が行われている．具体的には，これらのドメインの動きに関して，頑強な道標（ランドマーク）検出器の設計や，その訓練方法の研究である．([1]，[2]，[3]) しかし，これらは訓練データにない，未知のドメインへ対応できないという課題があった．そこで，代替の方法として，不変物体（例：壁 [4]，道路 [5]，電柱 [6]）を ランドマークとして利用する方法に注目が集まっている．特に，本研究では，屋内外を問わずに存在する柱状ランドマークに注目する．(図 1)

既存研究の多くは，巡回警備やレスキューなど，外部タスクのために決められた行動を行う，受動的な観測者（ロボット）を想定しており，視点や観測者の制御のような問題は考慮されてこなかった．しかし，通常のランドマークと比べると，不変ランドマークは空間的に疎なため，受動的な観測者にとっては，自己位置推定は不良設定問題になってしまう．そこで，本研究では，能動的な観測者によるアクティブ自己位置推定を考える．これは，屋外の開けた場所や不変物体が見えないような，ランドマークの発見ができない不都合なシーンを避けたり，情報量の多いシーンに効率的に移動することで，センシングや計算の効率を向上させることが目的である．

本研究では，最近傍 Q 学習（Nearest Neighbor Q-Learning：NNQL）を用いて，季節の変化に対応することのできるランドマーク観測計画法を提案する．NNQL は，Q 関数を，最近傍検索エンジンにより近似する方法である．これにより，Q テーブルにおける次元の呪いを回避することができる．また，最先端の検索エンジンをもちいることで，計算効率の向上が期待できる．また，未経験の状態を，最も類似する既知状態により近似して，頑強に対応することができる．

実験では，学習を１種類のドメインを用いて行い，それとは異なるドメインでテストを行った．その性能をランドマークの観測数と，行動回数で評価を行った．その際，ランダムな行動を行う場合，最小の行動を行う場合，最大な行動を行う場合と比較を行った．

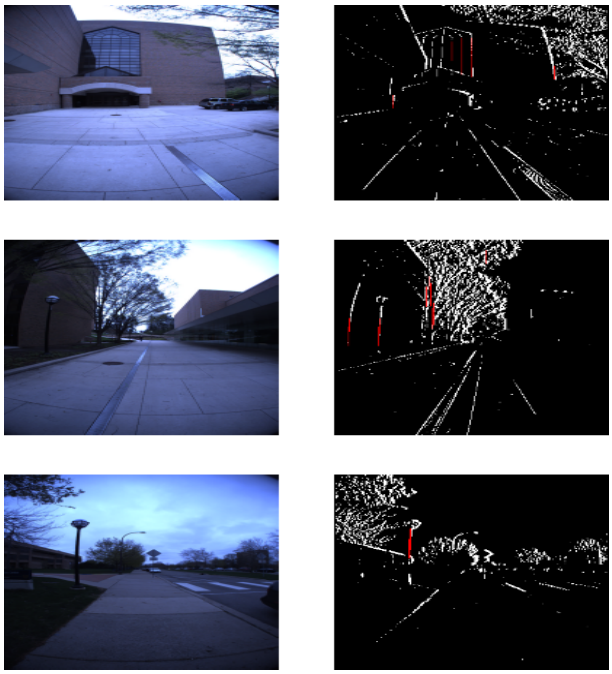

図 1　ポール観測例

2. 従来研究

ロボットの自己位置推定は，画像検索 [7] 画像分類 [8] などの様々な定式化を伴う長年の問題である．その中で，クロスドメイン自己位置推定は，ロボットの自己位置推定の中でも重要で困難な問題である．多くの研究では，ドメイン不変の視覚的特徴の検出器の設計や学習に焦点が当てられている [7] [9]．しかし，現在観測している画像が以前に見たことがないドメインの場合，ドメイン不変の特徴検出器を設計することは本質的に不向きである．そこで，事前に得られたドメインの知識を元に，ドメイン不変のランドマークオブジェクトを道標とする方法がある．我々の研究では，このドメイン不変のランドマークを利用する自己位置推定について考えた．また，情報量の多い場所に効率的に移動することができれば，自己位置推定の精度が向上すると考えた．つまり，ドメイン不変のランドマークを多く観測する方法を提案する．最も関連する研究として，マシンビ ジョンの分野における NBV(Next-Best-View) 問題を上げることができる．[10] しかし，ドメインシフト下での NBV 問題は十分に研究がなされているとはいえない．そこで，今回提案するドメインシフトに対応した手法を用いてランドマークを観測することで，有効的な NBV への入力になると考えられる．そのために，ドメインシフトに左右されないランドマークの

効率的な観測を行うことができるのかが今回の課題となっている．

3. 手法

提案手法は，移動ロボットが行動して観測した画像から柱状ランドマーク検出を行い，ランドマークの数が多くなることを目標として，NNQL による行動計画を行う．本研究で提案した，NNQL に基づくランドマーク観測行動計画システムの構成を図 2 に示す．システムは，知覚モジュールと計画モジュールと行動モジュールから構成される．知覚モジュールでは，入力画像から柱状ランドマークを検出し，特徴量を抽出する．計画モジュールでは，特徴量から NNQL を用いて行動計画を行う．NNQL とは，強化学習の手法である．行動モジュールでは，計画モジュールで計画された行動を実行する．

3.1 強化学習

強化学習 [11] とは，与えられた環境の中で，将来の報酬を最大化するように，エージェントを学習させる枠組みである．強化学習のサイクルはマルコフ決定過程として，(S,A,R,S',p_t) で定義される．S は，各時刻における状態の集合 $\{s^n_{t=0}\}$，A は行動の集合 $\{a^n_{t=0}\}$，R は状態と行動に依存して決まる報酬の集合 $\{r^n_{t=0}\}$，S'は次の状態の集合 $\{{s'_{t=0}}^n\}$ であり，p_tは状態の遷移確率（$S \times A \times S'$）である．しかし，強化学習ではタイムステップ T において得られる報酬 r_Tを最大化する行動を選択した場合に，将来得られる報酬和の最大化に繋がるわけではない．割引率 $\gamma \in (0,1]$ を用いて，現在割引和を求めている．割引率を含め，式 (1) で与えられる J を最大化するような行動を選択することを考えている．

$$J = \sum_{t=T}^{n} \gamma^{t-T} r_t \tag{1}$$

つまり，強化学習では現在だけではなく将来の報酬も加味した上で行動を選択している．Q 学習と呼ばれる強化学習の手法の一つでは，状態と行動に付随する価値をこれまでの行動と獲得報酬の組である経験の蓄積を基にして学習を行う．その価値を最大化することで，近似的に式 (1) で表される報酬和の最大化を行っている．

3.2 知覚モジュール

本研究での知覚モジュールでは，エッジ検出を用いた柱状ランドマーク検出を行う．はじめに，Sobel フィルタを用いて入力画像の y 軸方向のエッジを多数検出する．次に，多数検出したエッジのうち，y 軸方向の長さがある一定の値以上のものを柱状の物体としてランドマーク検出を行った．柱状ランドマーク検出を行った画像の x 座標において，柱状ランドマークがある座標を 1，無い座標を 0 にして，各 x 座標を二値化し，二値ベクトルへ変換を行った．この二値ベクトルを特徴量とした．

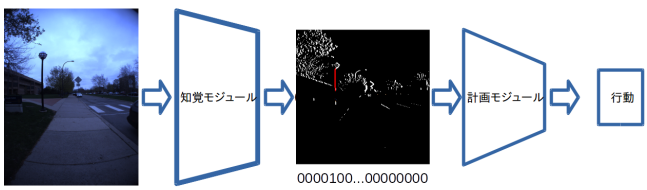

図 2　提案手法概要

3.3 計画モジュール

計画モジュールでは，知覚モジュールで観測した特徴量を入力とし，NNQL を用いて行動計画を行う．NNQL は強化学習の手法の一つである．強化学習とは，ロボットが現在の環境の状態を観測して，一つの行動を実行する．その状態と行動により，環境は新しい状態に遷移する．その新たな状態に応じて報酬をロボットに与え，それに基づいてロボットはあるタスクのための目的行動を学習する．本研究では，状態を知覚モジュールで観測した特徴ベクトルとし，行動はロボットが前進 1m,2m, 10m の合計 10 個とした．学習の目的は，ロボットが柱状ランドマークをより多く観測出来るように行動計画をすることであり，また，その時の行動を伴うコストは最小になることを目指す．よって，報酬はロボットが観測した柱状ランドマークの本数とした．

3.3.1 NNQL

NNQL は Q 学習の代替手法であり，Q テーブルを K 近傍探索（K Nearest Neighbor : KNN）によって代替している．Q 学習では，状態を全て Q テーブルに記憶する必要があるが，NNQL ではロボットが実際に行動して観測した状態のみデータベースに記憶する．ロボットが行動して観測した現在の状態と同じ状態をデータベースから探し，その状態の Q 関数 Q(s,a) を使用して行動計画を行う．しかし，現在の状態がデータベースにない場合がある．その場合は，Q 関数を KNN を用いて近似する．ここで，s は状態，a は行動を表す．近似の方法は，まず，現在の状態に近い状態を KNN を用いてデータベースから K 個探す．そして，近い状態の Q 関数の平均を現在の状態の Q 関数とする．KNN によって探した状態における Q 関数の集合を N(s,a)，その要素を Q(s',a') として，近似の式を以下に示す．

$$Q(s,a) = \frac{1}{|N(s,a)|} \sum_{s',a' \in N(s,a)} Q(s',a') \tag{2}$$

Q(s,a) を用いて，行動がとられる毎に，Q 関数の更新を行う．Q 関数の更新式は以下のようになる．

$$\mathrm{Q}(s_t,a_t) \leftarrow Q(s_t,a_t) + \alpha[r_{t+1} + \gamma max Q(s_{t+1},a) - Q(s_t,a_t)] \tag{3}$$

ここで，r は状態 s で行動 a をとったときの報酬，α は学習率，γ は減衰係数を表す．NNQL の学習ステップは，まず，現在の状態がデータベースにあるか調べる．ある場合は，現在の状態 s_tを使い $Q(s_t,a)$ を求め，ない場合は，(2) 式から $Q(s_t,a)$ を求める．次に，$\max_a Q(s_t,a)$ となる行動 a を選択し実行する．最後に，(3) 式に従い Q 関数の更新を行う．これを，繰り返すことにより目的行動を学習する．

3.4 行動モジュール

行動モジュールでは，計画モジュールで決定した行動計画に従い，行動する．この時，IMU データを用いる．IMU データとは，ロボット内部のセンサーから，行動後，ロボット自身がどれだけ進んだのかを，記録したデータであり，行動後の位置を IMU データを記録したメモリから引き出し，次の状態の観測につなげる．行動後，知覚モジュールにて画像の観測を行い，計画モジュールを用いて，観測した画像から類似した画像を NNQL を用いて次の行動を計画し，行動モジュールにおいて，その行動を実行する．このサイクルを繰り返していく．

4. 実験

本研究では，提案手法の有効性を，異なるドメインごとの柱状ランドマーク検出数をランダム手法，最大行動手法，最小行動手法とで比較して評価を行った．この時，ランダム手法とは 1，2，3 … 10m の行動を無作為に選択し行動する手法である．最大行動手法は 10m，最小行動手法は 1m の行動を常に実行する手法である．我々は，公開されている NCLT データセット [12] を使用した．NCLT データセットは，ミシガン大学の北キャンパスでセグウェイビークルのロボットプラットフォームによって収集されたロボット研究のための大規模かつ長期的な自律性データセットである．研究に使用したデータは，Ladybug3 の正面に取り付けたカメラによって取得されたロボットの起動に沿った視野画像と，GPS 情報を含んでいる．本実験では，NCLT データセットの 2012/1/22 2012/3/31 2012/8/4 の 3 種類を用いた．3 つのデータセットは，それぞれ 26208 枚，26365 枚，24138 枚の画像で構成されている．NNQL の訓練には，1 月，3 月，8 月の 3 つの訓練データセットで行った．訓練データは，柱状ランドマークの正解データを作成して使用した．正解データの作成にあたって，観測モジュールの方法では，一定の長さ未満の柱状ランドマークを検出しないため，ロボットと柱状ランドマークの距離が離れていると検出しないという問題がある．そのため，遠くにある柱状ランドマークを認識をするような正解データを作成をする必要がある．通常，ロボットが前進すると，遠くにあった柱状ランドマークは近づいて，大きく見える．これを逆再生すると，近くにあった柱状ランドマークは遠ざかり，小さく見える．このことを利用して，トラッキングをすることで遠くにある柱状ランドマークを検出した．テストは，訓練データセットのドメインとは異なる残りの 2 種類のドメインでそれぞれ行った．図 3 は行動前に観測した画像と，計画，行動を行った田とに観測された画像の例である．このようによりポールの観測数が多いシーンに移動する計画を行っている．

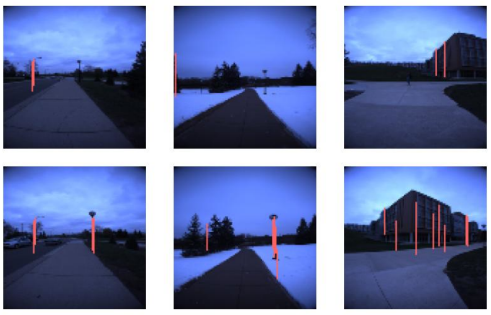

図 3 行動前 (上) と行動後 (下) に観測した画像

表 1 1 月のドメインで訓練した時の柱状ランドマークの平均観測数と平均移動距離

	3 月		8 月	
	観測数	移動距離	観測数	移動距離
NNQL	3.57	33.85	3.21	33.92
最小	1.76	10	1.59	10
最大	1.7	100	1.37	100
ランダム	1.79	55.67	1.49	54.91

表 2 3 月のドメインで訓練した時の柱状ランドマークの平均観測数と平均移動距離

	1 月		8 月	
	観測数	移動距離	観測数	移動距離
NNQL	2.89	92.27	3.89	91.33
最小	1.7	10	1.17	10
最大	1.14	100	1.53	100
ランダム	1.4	55.42	1.44	54.11

表 3 8 月のドメインで訓練した時の柱状ランドマークの平均観測数と平均移動距離

	1 月		3 月	
	観測数	移動距離	観測数	移動距離
NNQL	4.22	49.98	2.3	49.8
最小	1.3	10	0.66	10
最大	1.67	100	1.01	100
ランダム	1.9	55.03	0.91	54.35

5. 結果と考察

性能評価方法について説明する．まず，ランダムなスタート地点から，その手法によって決められた行動や学習された行動を行う．この行動を１０回繰り返し，その間に観測された柱状ランドマークの観測数と移動距離を求める．これを１サイクルとし，１００サイクル行い，その観測数と移動距離の平均の結果を表 1，表 2，表 3 に示す．表 1 は，1 月のデータセットで訓練を行い，3 月，8 月のデータセットでテストを行った結果，表 2 は 3 月のデータセットで訓練を行い，1 月，8 月のデータセットでテストを行った結果，表 3 は 8 月のデータセットで訓練を行い，1 月，3 月のデータセットでテ

ストを行った結果である．この時，NNQL，最小，最大，ランダムとは，提案手法，最小行動手法，最大行動手法，ランダム手法のことである． これらの結果より，NNQL は最大と比べ行動回数が少なく, 他の行動全てと比べ柱状ランドマークの観測数が多い．このことから，NNQL を用いた提案手法では移動にかかるコストを最大にすることなく，効率的に柱状ランドマークを観測していることがわかる．つまり，NNQL を用いた学習はドメインシフトの影響下でのランドマークの観測において有効であると言える．

6. 結論

本論文では，ドメインシフトに対応したランドマーク観測における学習方法について提案した．提案手法では，Q テーブルを用いることなく，NNQL によって学習することで，柱状のランドマークという本質的に安定した自己位置推定のランドマークを検出することができるようになるであろうと仮説を立てた．その仮説に基づき，NNQL を用いて，従来の Q 学習とは異なり大規模なデータである Q テーブルを用いることなく学習を行い，安定したランドマーク検出ができるようにした．その結果，NNQL を用いることによって，異なったドメインで訓練していても効率的にランドマークの観測を行うことができることがわかった．

参 考 文 献

[1] H. Hu, H. Wang, Z. Liu, C. Yang, W. Chen, and L. Xie：“Retrieval-based localization based on domain-invariant feature learning under changing environments, ”IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS)，2019,pp. 3684-3689.

[2] R. Arandjelovic, P. Gronat, A. Torii, T. Pajdla, and J. Sivic: “NetVLAD: CNN architecture for weakly supervised place recognition, ”IEEE Conference on Computer Vision and Pattern Recognition, 2016.

[3] N. Merrill and G. Huang: “CALC2.0: Combining appearance, semantic and geometric information for robust and efficient visual loop closure, ”IEEE/RSJ Int. Conf. Intelligent Robots and Systems (IROS), Macau, China, Nov. 2019.

[4] M. Drumheller：“Mobile robot localization using sonar, ”IEEE trancactions on pattern analysis and machine intelligence, no.2,pp.325-332, 1987.

[5] M. A. Brubaker, A. Geiger, and R. Urtasun：“Lost! leveraging the crowd for probabilistic visual self-localization, ”Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2013, pp.3057-3064.

[6] R. Spangenberg, D. Goehring, and R. Rojas：“Pole-based localization for autonomous vehicles in urban scenarios, ”2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2016, pp.2161-2166.

[7] M. Cummins and P. Newman: “Fab-map: Probabilistic localization and mapping in the space of appearance, ”Int. J. Robotics Research, vol. 27, no. 6, pp. 647-665, 2008.

[8] G. Kim, B. Park, and A. Kim: “1-day learning, 1-year localization:Long-term lidar localization using scan context image, ”IEEE Robotics and Automation Letters, vol. 4, no. 2, pp. 1948-1955,2019.

[9] M. J. Milford and G. F. Wyeth: “Seqslam: Visual route-based navigation for sunny summer days and stormy winter nights, ”2012 IEEE INt. Conf. Robotics and Automation. IEEE, 2012, pp. 1643-1649.

[10] R. Pito：“A solution to the next best view problem for automated sur face acquisition, ”IEEE Transactions on pattern analysis and machine intelligence, vol. 21, no. 10, pp. 1016-1030, 1999.

[11] R. S. Sutton and A. G. Barto: “Reinforcement Learning, ”An Introduction, Bradford Books, 1988.

[12] N. Carlevaris-Bianco, A. K. Ushani, and R. M. Eustice：“University of michigan north campus long-term vision and lidar dataset, ”The International Journal of Robotics Research, vol. 35, no. 9, pp. 1023-1035, 2016.