

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認識

Semantic Positional Recognition and Visual Landmark Detection without Effect of Pedestrians

○石郡祐司, 間所洋和, 佐藤和人

○Yuji Ishikori, Hirokazu Madokoro, Kazuhito Sato

秋田県立大学大学院 機械知能システム学専攻

Graduate school of Systems Science and Technology,
Akita Prefectural University

キーワード: ビジュアルランドマーク (Visual landmark), 認識率 (Recognition), HOG 特徴量 (Histograms of Oriented Gradients), YOLO (You Only Look Once), 人物検出 (Human detection)

連絡先: 〒015-0055 秋田県由利本荘市土谷字海老ノ口 84-4
秋田県立大学大学院 システム科学技術研究科 機械知能システム学専攻 脳情報工学研究室
石郡祐司, E-mail: M19A003@akita-pu.ac.jp

1. 緒言

ロボット産業の進展に伴い, 工場等の限定的な作業環境だけでなく, 人間の生活環境で利用される自律移動ロボットの開発が活発になっている。このようなロボットが, 人間と同じ空間で作業や役割を行う場合には, ロボット自身が合目的かつ自律的に判断し行動する機能が求められる。その中で, ロボットの自律移動には様々な研究が行われているが, 地図の事前準備や構築を必要としないビジュアルランドマークに基づく方式は, 人間の知覚に即したアプローチであることから, 意味認識との親和性が高く, 実用面での応用範囲が広い。一方, 環境の変動に対して, 頑強性と安定性を併せ持つ抽出方法の構築は, 困難な研究課題となっている。本研究では, ロボットと人間が共生する実環境におい

て, 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認識を目的として, 人物の影響に対して頑強なビジュアルランドマークの抽出ならびに記述方法を提案する。

2. 関連研究

ビジュアルランドマークに基づく自己位置の認識に関して, 多種多様な手法が提案されている。Chen ら¹⁾は, 簡易で安価な位置同定システムを構築している。Livatino ら²⁾は, ドアの取っ手, ポスタ, 電気のスイッチなどをビジュアルランドマークと定義し, 4種類の特徴量(エッジ, コーナ, 強度, 局所対称性)を用いて特徴的な領域を抽出しているが, 評価実験は限定的な環境に留まっており, 汎化性が十分に示されていない。このように従来手法はビジュアルラ

ンドマークが安定して検出されるという前提のもとで、位置推定やナビゲーションについて精度を評価しており、部分的なオクルージョンやコラプションの発生は考慮されていない。

人間と共生する実環境では、人物干渉が避けられないため、提案手法と評価実験ではこれらを考慮しなければならない。地図を持たずSLAM (Simultaneous Localization and Mapping) によって環境地図を構築する手法では、森岡ら³⁾は、人が多く混雑する環境下でのSLAMによる移動ロボットのナビゲーションを行った。複数枚の連続画像からSIFTを抽出し、連続的にマッチングを取り、全ての画像でマッチングが取れた特徴点のみを抽出するPIRF (Position Invariant Robust Feature) 特徴量と、オドメトリから3D-PIRFを抽出することで環境地図を構築し、ナビゲーションを行った。このように人の影響に対する研究は進められているが、ビジュアルランドマークに基づくロボットの自律移動については取り組み事例が少ないのが現状である

3. 提案手法

提案手法の全体構成を図1に示す。始めに、顕著性マップ⁴⁾を用いてビジュアルランドマークの候補となる顕著度の高い領域を抽出する。この領域に対して、AKAZE(Accelerated KAZE)⁵⁾の特徴量を抽出する。画像中に人物が含まれている場合は、顕著度が高くなる可能性があるため、HOG(Histogram of Oriented Gradients)⁶⁾により人物領域を抽出し、マスク画像としてその部分を排除する。すなわち、顕著性マップとHOGのマスク画像の積集合の領域が、ビジュアルランドマークの候補となり、AKAZEによる特徴記述が施される。

4. 予備実験

自己位置の意味的認識を目的とする評価実験に先立ち、HOGの人物検出性能を評価するた

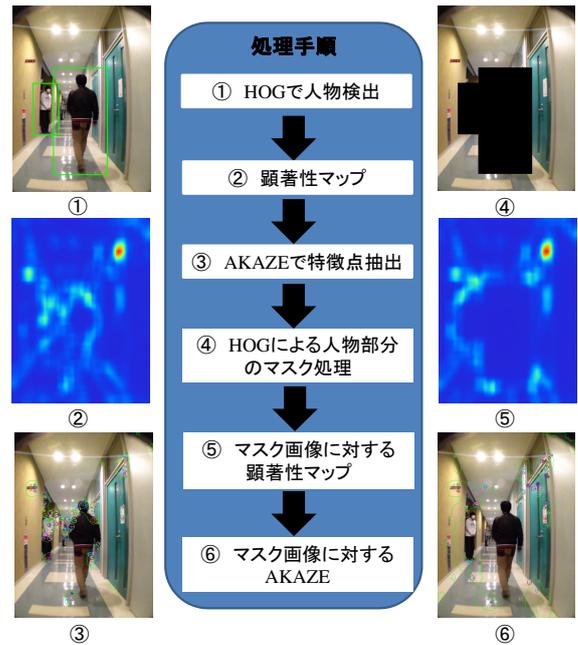


Fig. 1 実験の処理手順.

めの予備実験を実施した。

4.1 ベンチマークデータセット

本実験では、EPFL (スイス連邦工科大学ローザンヌ校) のコンピュータビジョン研究室が公開しているベンチマークデータセットを用いて、HOGの人物検出の精度を評価した。

人物検出の評価では、認識率と認識漏れ率の2指標を用いた。HOGの検出領域に対して、「人物あり」の領域を「人物」と検出(真陽性)した場合の個数をA、「人物あり」の領域が未検出(偽陰性)の場合の個数をB、「人物なし」の領域を「人物」と検出(偽陽性)した場合の個数をCとする。

認識率は、AとCを用いて次式として定義される。

$$\frac{A}{A+C} \times 100 \quad (1)$$

認識漏れ率は、AとBを用いて次式として定義される。

$$\frac{B}{A+B} \times 100 \quad (2)$$

なお、人物1名に対して、検出枠が n 個ある場合はAに1を加算し、Cに $n-1$ を加算した。同一枠内に n 人が含まれている場合には、Aに1を加算し、Bに $n-1$ を加算した。

4.2 実験結果

評価対象とした全画像に対する平均認識率は85.4%，平均認識漏れ率は41.6%であった。HOGは事前に設定された領域内で特徴量を算出するため、スケール変化に対して不変ではない。このため、本実験で用いた画像に対しては、人物の全体像が相対的に大きい(同一人物の場合はカメラに近い)と比較的認識されやすい傾向にあった。他方で、人物の全体像が小さい(同一人物の場合はカメラから遠い)場合は認識漏れとなった。また、オクルージョンやコラプションが発生している場合に関しても、認識漏れとなった。

5. 評価実験

5.1 実験条件

本実験では、屋内環境におけるロボットのナビゲーション及び位置推定用として、移動ロボットのDoubleを用いて、ベンチマーク用のオリジナル動画データセットを構築した。Doubleの外観を図2(a)、走行経路を図2(b)に示すロボットの全高は119 cmであり、首の部分が31 cmの範囲で上下する。本実験では、最も低い位置に固定して、実験環境内を移動させた。

Doubleには専用のカメラが搭載されておらず、頭部に装着するiPadの内蔵カメラを用いて画像を取得する。画像の解像度は 472×628 pixelで、フレームレートは30 fpsである。

本大学のキャンパスにおいて、校舎内の廊下の一部を実験環境とした。この校舎には、中央に吹き抜けがあり、廊下を挟んで、外側に教員室と研究室が並んでいる。本走行実験では、ロ

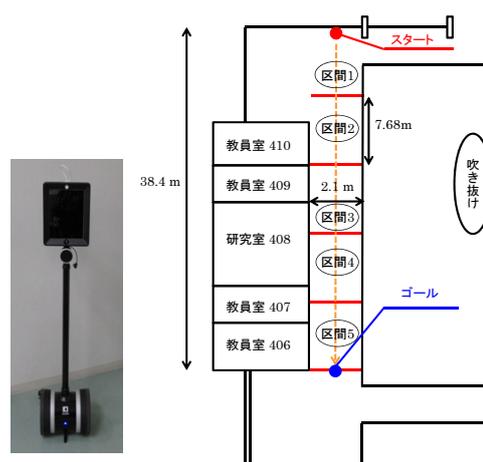


Fig. 2 (a) Double, (b) 走行環境.

ボットの進行方向に対して、左側に吹き抜けがあり、右側に教員室と研究室が並んでいる、38.4 mの直線区間を走行させた。この区間を5等分して、位置認識の対象区間とした。各区間のサンプル画像を図3に示す。右側にポスターやドア等のビジュアルランドマークの候補となるオブジェクトの存在が確認できる。

5.2 データセット

Doubleで取得した動画データを、処理時間とメモリ容量を考慮して、10 fps毎に静止画像として切り出した。実験環境に人物が存在する場合と、存在しない場合について、2種類のデータセットを構築した。本データセットでは、環境変動による頑強性を考慮するために、撮影日時を変更して動画データを取得した。人物が存在するデータセットは、延べ4日間で12パターンのデータセットを取得した。

5.3 特徴抽出結果

提案手法による特徴抽出結果を図4に示す。区間毎の結果を、順に(a)原画像、(b)顕著性マップ、(c)顕著度の高い領域内のAKAZE、(d)HOG、(e)HOGマスク領域外の顕著性マップ、(f)HOG



Fig. 3 取得データの代表画像.

マスク領域外と顕著度の高い領域の AKAZE となる。すなわち、(f) の特徴量が提案手法の結果になり、人物に影響されないビジュアルランドマークの特徴記述となる。

図4に示す実験結果から AKAZE は、人物だけでなくポスターなどの掲示物に対しても特徴点が密になる傾向がある。人物をマスク処理した場合、図4に示すように、掲示物に顕著性の高い領域が移り、ビジュアルランドマークとして特徴づけられるようになった。

5.4 パラメータ最適化

提案手法は機械学習に基づく手法であるため、様々なパラメータに支配されている。パラメータの設定値によっては、認識精度に直接的に影響を与えることから、評価実験に先立って、パラメータの最適化の予備実験を実施した。

最適化対象としては、コードブックモジュールのコードブック数となる SOM(Self-Organizing Maps) のマッピングユニット数、マッピングモジュールのカテゴリマップの分解能と表現力を

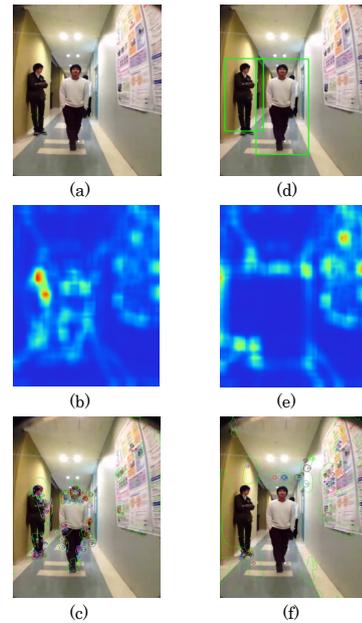


Fig. 4 区間4の抽出結果. (a)原画像, (b)顕著性マップ, (c) 顕著度の高い領域内の AKAZE, (d)HOG, (e)HOG マスク領域外の顕著性マップ, (f)HOG マスク領域外と顕著度の高い領域の AKAZE

支配するマッピング層のユニット数、学習の収束に影響を与える CPN(Counter Propagation Networks) の学習回数を、主要な3つのパラメータとした。本実験では人物の出現頻度やパターンによる影響を考慮し、走行環境中に人物が含まれていない2データセットを、最適化対象の動画データとして用いた。評価方法は、図3に示す5区間として、2データセットの交差検定により、位置認識率を算出した。

最適化結果から、コードブック数の最適化は64ユニット、マッピング層ユニット数の最適化は30×30ユニット、CPNの最適学習回数は10,000回に決定した。

5.5 実験結果

12データセットを用いて、移動ロボットにおける自己位置の意味的認識実験を実施した。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

Table 1 交差検定結果の精度 [%].

データセット	提案手法	HOGなし
a	51.6	54.2
b	38.2	40.9
c	56.9	47.1
d	49.3	40.0
e	43.1	52.9
f	40.0	35.1
g	60.9	54.2
h	51.6	46.2
i	54.7	47.1
j	58.2	48.4
k	43.6	47.6
l	50.7	46.7
平均	49.9	46.7

用いる1リーブアウトの交差検定とした。すなわち、12パターンのデータセットでは、12通りの組み合わせによる評価となる。

交差検定による認識精度を表1に示す。提案手法の結果に加えて、HOGマスク処理が組み込まれていない場合、すなわち人物領域を含めてビジュアルランドマークを抽出する場合についても、併せて精度を示している。提案手法による平均認識精度は49.9%であった。テストデータセットgの場合に、認識精度が最高の60.9%となった。一方、テストデータセットbの場合に、認識精度が最低の38.2%となった。両者間には、22.7%の差異があり、同じ環境においても、日時や人の行き交いによってバラツキが発生することが示唆される。比較対象としたHOGマスク処理なしによる平均認識精度は46.7%であった。同一のデータセットで、同一の識別器を使つての判定であるが、人物の影響を受けているため、提案手法より3.2%下回る結果となった。

5.6 考察

誤認識に関して、混同対照表を用いて分析する。認識結果の混同対照表を表2に示す。全12

データセットの交差検定結果を総和した結果となる。本実験では、ロボットの移動を等速直線運動に限定しているため、誤認識は隣接する区間で発生する傾向であることを示唆する結果となった。本ロボットに搭載されている視覚センサとしてのカメラは、移動方向に向けて取り付けられている。このため、区間1では、後方に続く区間がなく、前方に続く区間が徐々に視界の中に含まれる比率が増すという画像特性があり、本来は5区間の中で認識が最も難しい。しかしながら、本データセットの特性として、研究室前となる区間3及び区間4において、シーン内に現れる人物が増えることに加えて、ロボットの視界範囲内に写る人物のサイズが相対的に大きくなっている。逆に、その影響が相対的に小さくなる区間1において、掲示物等のビジュアルランドマークと人物が重複する割合が減ったため、認識率が相対的に他の区間よりも上回ったと考えられる。

5区間の中で、最も識別精度が低かったのは、区間4であった。全540フレームのうち、区間4として正しく認識されたのは216フレームであった。図3に示す代表画像からも確認できる通り、ロボットが教員室と研究室を通過する区間2から区間4は、比較的単調な情景が続く。また、図4に示す通り、ビジュアルランドマークとしてポスタやドアなどが抽出されているが、ポスタは描かれている図面の内容が違っても、ドアは各部屋とも同じ構造とパターンのもが使われている。このため、吹き抜けに架かる通路と自動販売機や椅子などが置かれた休憩スペースが視界に大きく入ってくる区間5と比較して、相対的に認識の難易度が高かったと考えられる。

6. まとめ

本研究では、ロボットと人間が共生する実環境において、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

Table 2 提案手法の混同対照表 (総計)

区間	1	2	3	4	5
1	367	61	42	39	31
2	66	219	94	79	82
3	54	92	259	83	52
4	36	93	103	216	92
5	39	83	73	59	267

的として、人物の影響に対して頑強なビジュアルランドマークの抽出ならびに記述方法を提案した。「人物あり」の12データセットを用いて、交差検定による性能評価を実施した結果、提案手法の認識率は49.9%であった。同じデータセットを用いて、HOGマスク処理をしない場合の認識率を算出したところ46.7%となり、提案手法は3.2%の比較優位性が得られた。混同対照表による各区間での誤認識の分析では、区間1が最も認識精度が相対的に高いことを示すとともに、誤認識は隣接する区間が多く、区間が離れるごとに少なくなるという傾向を明らかにした。今後の課題として、ベンチマークデータセットの取得範囲の拡大、また、YOLOを用いて交差検定を行い、HOGとの精度の比較が挙げられる。

参考文献

- 1) D. Chen, Z. Peng, and X. Ling, “A low-cost localization system based on Artificial Landmarks,” IEEE International Conference on Robotics and Biomimetics, 2014.
- 2) S. Livatino and C.B. Madsen, “Automatic Selection of Visual Landmark for Mobile Robot Navigation,” Proc. The 7th Danish Conference in Pattern Recognition and Image Analysis, 1998.
- 3) 森岡博史, 李想揆, T. Nopparit, 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.

- 4) L. Itti, C. Koch, and E. Niebur, “A Model of Saliency-Based Visual Attention for Rapid Scene Analysis,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.
- 5) Pablo F. Alcantarilla, Jesus Nuevo, Adrien Bartoli, “Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Space,” School of Interactive Computing Georgia Institute of Technology Atlanta.
- 6) Dalal N., Triggs B.: “Histograms of Oriented Gradients for Human Detection.” IEEE Conference on Computer Vision and Pattern Recognition, vol.1, pp.886-893, 2005.
- 7) Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016.