

類似画像検索による歩行位置精度向上のための凸包の利用

小河原 洸貴[§] 宍戸 英彦[‡] 北原 格[†] 亀田 能成[†]

^{† ‡ §} 筑波大学 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: §ogahara.koki@image.iit.tsukuba.ac.jp, ‡shishido@ccs.tsukuba.ac.jp,

† {kitahara, kameda}@iit.tsukuba.ac.jp

あらまし 事前に撮影された映像に基づく歩行者ナビゲーションについての我々の先行研究では、事前撮影映像から抽出された参照画像とその局所特徴量をデータベースとし、問い合わせ画像との類似画像検索により参照画像を選出することで位置推定を行っている。本稿では、撮影位置の前後の距離差と画像中の局所特徴点の凸包面積の相関を弱透視投影モデルにおいて仮定し、参照画像に対する問い合わせ画像の凸包面積比を用いて位置推定の精度を向上させる手法を提案する。実際の経路における事前撮影映像を用いて、提案手法の有効性を確認した。

キーワード 歩行者位置推定, 類似画像検索, 局所特徴量, 凸包

1. はじめに

2016年時点で日本全国に31万人を超える視覚障がい者が存在する[1]。視覚障がい者が単独で歩行する際には、経路間違いを防ぐために同行援護やボランティア等晴眼者による歩行の補助や誘導を受けることが望ましい。しかし、補助者の絶対数が不足していることや経済的な負担が大きいことなどの理由から、全ての視覚障がい者が満足いく人的支援を受けることは難しい。そのため、視覚障がい者の単独歩行の際に目標となる経路に沿った誘導を行う歩行支援が必要とされている。

推定された位置と方向から視覚障がい者に適切な指示を与え、目標経路に沿った誘導を行うことのできるインタフェースの研究に我々は取り組んでいる。先行研究[2]では、視覚障がい者の単独外出の際に事前に計画される歩行計画が直進と方向転換の繰り返しというターンバイターン方式で構成されることに着目し、各状況に対応した異なる誘導指示を行う誘導インタフェースを提案した。本インタフェースでは、歩行者の位置推定手法として類似画像検索による位置推定手法[3]を用いている。

類似画像検索による位置推定手法では、目標経路に沿って歩行者が事前に撮影する事前撮影映像から参照画像を抽出し、あらかじめ検索用データベースを生成する。位置推定の際は問い合わせ画像に最も類似する参照画像を検索し、その撮影位置を歩行者の位置と推定する。参照画像はある程度の間隔を持って用意されていることから、求められる歩行位置の精度をその間隔より小さくすることはできない。

本稿では、撮影位置の前後の距離差によって画像中の特徴点群の分布が変化することに着目し、参照画像と問い合わせ画像中の局所特徴点群の凸包面積比を用いて位置推定位置の精度を参照画像の間隔から事前撮

影映像のフレーム間隔へ向上させる手法を提案する。理想的な状況として弱透視投影モデルを用いて、二つの画像の撮影位置間の距離と、各画像中に現れる局所特徴点群の凸包面積との相関を仮定する。従来の参照画像検索用データベースに加えて凸包面積比をあらかじめ求めておくことで、位置推定時に参照画像と問い合わせ画像との凸包面積比からより細かな精度での位置推定を可能にする。実際の目標経路として屋内外の経路を設定し、その仮定が成立することを確認した。

2. 関連研究

視覚障がい者向け誘導システムとしてGPSを用いた携帯情報端末[4]がすでに提案されている。しかし、電波が届きにくい高層ビル街や室内、地下などにおいてはGPSによる測位精度が低下し、十分に機能しないという問題がある。

この問題を解決するために、屋内向けのナビゲーションシステムとしてRFIDタグや、BLEビーコン等の設置型機器を用いた手法が提案されている。Draganらが開発しているBLEビーコンを用いた屋内誘導アプリケーションNavCog[5]では、既に国内の商業施設における実証実験を行っている。しかし、経路への機器の設置および維持コストが高く、設置可能な場所に限られるという問題がある。

これらの外部機器によって歩行者の位置を推定する手法に代わり、歩行者が身に付けるセンサによってその位置を推定する手法が研究されている。

自立移動ロボットの分野においては、SLAMと呼ばれる自己位置推定と環境地図の作成を同時に行う手法が提案されている。LIDERなどの三次元測距センサを用いるSLAMは様々な状況で利用できるが、センサ自体は高価である。SLAMの改良として、近年では比較的安価なRGB-Dカメラを使用する手法[6]や、単眼カ

メラのみで動作する Visual SLAM と呼ばれる手法[7]が研究されている．計算機の性能向上に伴い，スマートフォンやタブレット端末のような携行可能な機器の上でも SLAM の実時間実行が可能となったため，歩行者の誘導システムにおける位置推定手法として用いている研究も存在する[8]．しかしながら，依然として計算コストが大きいことや，特徴点が得られにくいテクスチャレスな環境下では安定した動作が望めないことなどの問題を抱えている．

3. 歩行位置精度向上のための凸包の利用

本節では，我々が利用する位置推定手法の概要とその歩行位置精度について説明し，精度向上のために利用する撮影位置と局所特徴点群の凸包面積との相関について弱透視投影モデルと実環境の両観点から述べる．

3.1. 類似画像検索による歩行位置推定精度

我々の研究では，基本とする歩行者の位置推定手法として釜坂らの提案する類似画像検索による位置推定手法[3]を用いる．

類似画像検索による位置推定手法では，支援者が胸部に取り付けた単眼カメラを用いて撮影する事前撮影映像から参照画像を抽出し，その SIFT[9]特徴量(SIFT キー)と合わせて類似画像検索用のデータベース(参照画像データベース)を生成する．この際，同一の SIFT キーが複数の参照画像内に現れないようにするため，参照画像はある程度のフレーム間隔を持って用意される．位置推定の際は，事前撮影時と同様のカメラを用いて問い合わせ画像を歩行者が撮影する．この問い合わせ画像中の SIFT キーを用いて，参照画像データベースから最も類似する参照画像を検索し，選ばれた参照画像の撮影位置を歩行者の位置推定結果とする．このとき，問い合わせ画像と参照画像との対応する SIFT キーの組み合わせ(キーペア)が複数得られる．

この手法は，必要とする機材が単眼カメラのみであり，撮影可能な経路であれば屋内外を問わない．三次元復元を行わないため SLAM 等と比較して計算コストが低いことから，モバイルコンピューティングに適している．しかし，先ほども述べたように参照画像はある程度の間隔を持って抽出されており，求められる歩行位置の精度に限度があるという問題がある．

3.2. 弱透視投影モデルにおける撮影位置と

局所特徴量の凸包との相関

本研究では，位置推定の精度向上のため，参照画像と問い合わせ画像の前後関係により各画像中のキーペアに含まれる SIFT キーの凸包面積の変化を利用する．

撮影位置と面積比の関係を求めるため，図 1 のような弱透視投影モデルの環境を考える．基準となる参照画像(Reference)の撮影位置からカメラの光軸方向に距

離 x だけ離れた点を問い合わせ画像(Query)の撮影位置とする．参照画像の撮影位置から距離 l の地点に投影面があるとし，問い合わせ画像の撮影範囲を w_1 ，参照画像の撮影範囲を w_2 とおく．このとき問い合わせ画像と参照画像中のキーペア分布の比率 R_x は

$$R_x = \frac{w_2}{w_1} = \frac{l}{l-x}$$

となる．垂直方向で R_y も同じく

$$R_y = \frac{l}{l-x}$$

とすれば，凸包面積比 R_S は

$$R_S = R_x \times R_y = \frac{l^2}{(l-x)^2}$$

となる．したがって， R_S の平方根が撮影位置と線形関係にある．

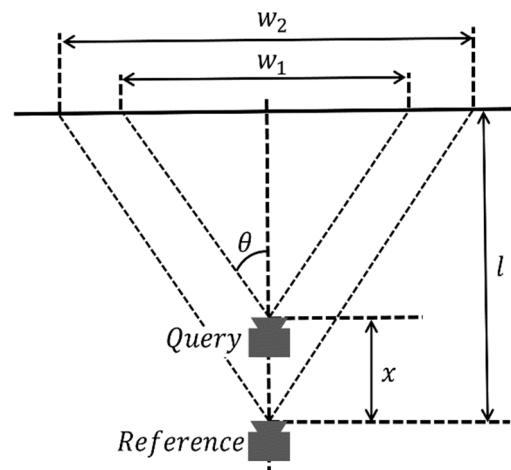


図 1 弱透視投影モデルにおける撮影位置と凸包面積の模式図

3.3. 凸包面積比を用いた位置推定

本節では，参照画像として抽出されなかった事前撮影映像のフレームについても凸包面積比の情報を求めることで位置推定の向上を行うことを提案する．事前記録時には事前撮影映像から参照画像データベースを生成したのちに事前撮影映像と参照画像データベースの類似画像検索を行い，各フレームの凸包面積比を記録した凸包面積比データベースを作成する．各参照画像における面積比の分布からそれぞれの近似直線を求め，これを凸包面積比データベースとして記録する．位置推定時には参照画像データベースによる推定後，凸包面積比データベースに記録されている近似直線を用いて事前撮影映像のフレーム番号を算出し，より高精度な位置推定を可能にする．

4. 凸包面積比を用いた歩行位置精度向上

本節では、3 節で求めた撮影位置と凸包面積比の相関を用いて、事前撮影映像から凸包面積比データベースを作成しそれを用いて位置推定を行う手法について述べる。

4.1. 凸包面積比データベースの作成

従来の類似画像検索による位置推定と同様に、凸包面積比データベースの作成には事前撮影映像を用いる。以下にその詳細な手順を述べる。

1. 従来手法により事前撮影映像から参照画像を抽出し、参照画像データベースを生成する。
2. 事前撮影映像中の各フレームを問い合わせ画像として手順 1 で生成した参照画像データベースに対する位置推定を行う。これにより、事前撮影映像中の各フレームに対してある一枚の参照画像が対応付けられ、画像間のキーペアが得られる。
3. 各問い合わせ画像と対応する参照画像のそれぞれにおいてキーペアに含まれる SIFT キーの凸包面積を計算する。問い合わせ画像の凸包面積を S_Q 、参照画像の凸包面積を S_R としたとき、凸包面積比 R_S を

$$R_S = \frac{S_Q}{S_R}$$

として求める。

4. 各参照画像に対応付けられた問い合わせ画像における凸包面積比の分布から、最小二乗法を用いて凸包面積比とフレーム番号の直線回帰式を算出する。3.1 節で述べたように凸包面積比とフレーム番号は非線形の関係であり、面積比の平方根が線形関係となる。しかし、投影面までの距離 l に対して撮影位置間の距離 x が十分に小さいことから凸包面積比も直線で近似可能と考えられる。そのパラメータである回帰係数と切片を凸包面積比データベースとする。

実際の事前撮影映像において、上記の手順により算出された凸包面積比とその回帰直線の例を図 2 に示す。これらの点と直線は対応する参照画像ごとに色分けされている。各参照画像の前後フレームで凸包面積比が右肩上がりの直線状に分布していることから、直線近似が行えると見込める。

4.2. 凸包面積比データベースを用いた位置推定

歩行時の位置推定において、参照画像データベースに加えて凸包面積比データベースを用いることによってその推定精度を向上させる。まずは従来手法にのっとり問い合わせ画像と参照画像データベースとの類似画像検索による位置推定を行う。その後、選ばれた参照画像に対する回帰直線のパラメータを凸包面積比データベースより読み込み、問い合わせ画像と参照画像の凸包面積比を回帰式にあてはめてフレーム番号を算出して位置推定結果とする。これにより、事前撮影映

像のフレーム単位での位置推定が可能となる。

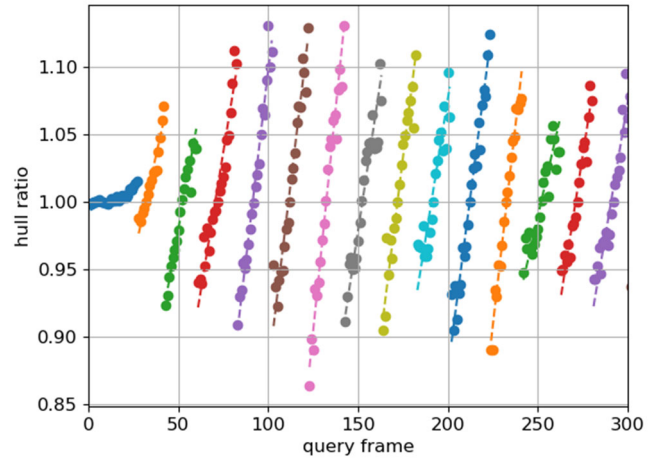


図 2 問い合わせ画像番号と凸包面積比

4.3. キーペアのフィルタリング

実際の位置推定においては、歩行者が目標経路から外れ、カメラの移動において回転や併進が生じる場合がある。このときは射影ひずみにより凸包面積比が正しく算出されない可能性がある。こうした射影ひずみによる影響を低減するため、凸包面積比データベースの作成に用いるキーペアのフィルタリングを行う。フィルタリングの様子を図 3 に示す。上段がフィルタリング無し、下段がフィルタリングありの場合で、各段右側が問い合わせ画像、左側が参照画像である。緑線で結ばれた二点がキーペアであり、アウトライアと判定されたキーペアを赤線で示す。各画像中の多角形が凸包である。射影ひずみの影響が大きい画像の外縁部分に存在するキーペアを除去するため、画像サイズを一定比率 R_{window} で縮小した矩形(図 3 青線)を設定し、この矩形内に含まれているキーペアのみを抽出する。フィルタリングによってキーペア数が 3 以下となり凸包の形成ができない場合、凸包面積比データベースによる位置推定は行わないものとする。

5. 位置推定精度の検証実験

5.1. 実験概要

筑波大学内の 3 つの経路(outdoor01, indoor01, indoor02)において凸包面積比データベースを用いて位置推定を行った。各経路の様子を図 4 に示す。各経路において、データベースの作成に用いる事前撮影映像(reference), reference と同条件の映像(query), 経路上でカメラを進行方向に対して左に向けた状態の映像(rotation), 経路から左に併進した状態の映像(translation)の 4 種類を撮影し、reference から作成したデータベースを用いて他の映像の位置推定を行った。実験には Surface Pro 4 (CPU: Core i7-6650U 2.20GHz,

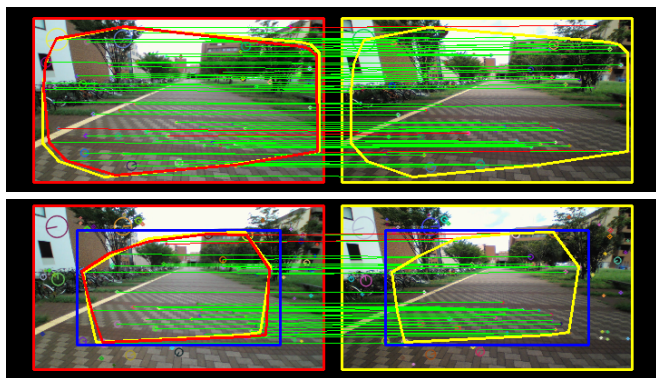


図3 矩形領域によるキーペアのフィルタリング
(上: フィルタリングなし,
下: フィルタリングあり, $R_{window} = 0.7$)



図4 各経路における reference の
0, 150, 300 フレーム付近の画像
(上: outdoor01, 中: indoor01, 下: indoor02)

RAM: 16GB)を使用し、映像は USB カメラ(BUFFALO BSW200MBK, 水平視野角 120 度)を Surface に接続して撮影した。

本実験では Structure from Motion の一手法である COLMAP[10, 11]の結果を 3 次元的な撮影位置を真値とした。撮影した 4 種類の映像から 5 フレーム間隔でフレームを抽出し、これらを COLMAP の入力画像群として 3 次元復元を行った。映像の撮影は一定の高さに固定されたカメラで行ったため、同一平面上で撮影されたという仮定をおき、得られた 3 次元カメラ位置をこの平面に投影することで 2 次元カメラ位置を求めた。reference の進行方向を x 軸方向、撮影開始地点を原点とおき、reference の経路長が実空間での距離となるように正規化した。他の種類の映像は x 座標の値を真値として扱った。入力画像群以外の撮影映像中のフレームについては、入力画像群の結果から線形補間を行い、真値とした。

5.2. 実験結果

各経路において提案手法および従来手法により位置推定結果を行い、真値との比較を行った。加えて、

R_{window} を変化させながらキーペアのフィルタリングを行った際の誤差の平均値と推定失敗率を求めた。

I. outdoor01

位置推定結果を図 5 に示す。query の結果では、提案手法による推定結果がほぼ真値に近く、目標経路に対してずれがない場合には位置精度の向上が確認できる。一方で、rotation および translation の結果では、どちらの手法においても推定値が前後する傾向が見られる。このように目標経路とのずれが生じ射影ひずみによる影響が存在する場合、凸包面積比が正しく算出されず正しく補正されにくい。

誤差の平均値と推定失敗率を表 1, 2 に示す。query, translation では提案手法による誤差の減少が認められた。フィルタリングについては、query, rotation では一部の値で誤差が減少した。一方、どの映像でも R_{window} を小さくすると推定失敗率が上昇する。

II. indoor01

位置推定結果を図 6 に示す。query の結果では、推定結果がほぼ真値に近く、提案手法による位置精度の向上が確認できる。rotation および translation の結果では、推定値が前後する傾向が見られるが、従来手法に比べ真値に近い値となっている。

誤差の平均値と推定失敗率を表 3, 4 に示す。すべての映像で提案手法による誤差の減少が認められる。フィルタリングについては、outdoor01 と同様の傾向が見られた。

III. indoor02

本経路ではテクスチャレスな環境であったために COLMAP による 3 次元復元に失敗し、真値を得ることができなかった。そのため、提案手法および従来手法によって得られた推定フレームの結果を図 7 に示す。query および rotation では、従来手法による推定結果より細かいフレーム間隔の位置推定結果が得られている。一方 translation では、提案手法の結果が大きく外れる場合があった。キーペア数の減少に伴い、凸包面積比が安定して算出されないためと考えられる。

6. おわりに

本稿では、参照画像と問い合わせ画像中のキーペアに含まれる SIFT キーの凸包面積比を用いて位置推定の精度を向上させる手法を提案した。二つの画像の撮影位置間の距離と、各画像中に現れる局所特徴点群の凸包面積との相関を弱透視投影モデルにおいて仮定した。従来の参照画像検索用データベースに加えて凸包面積比を記録した凸包面積比データベースを作成し、位置推定時には参照画像と問い合わせ画像との凸包面積比から事前撮影映像のフレーム間隔での位置推定を可能にした。

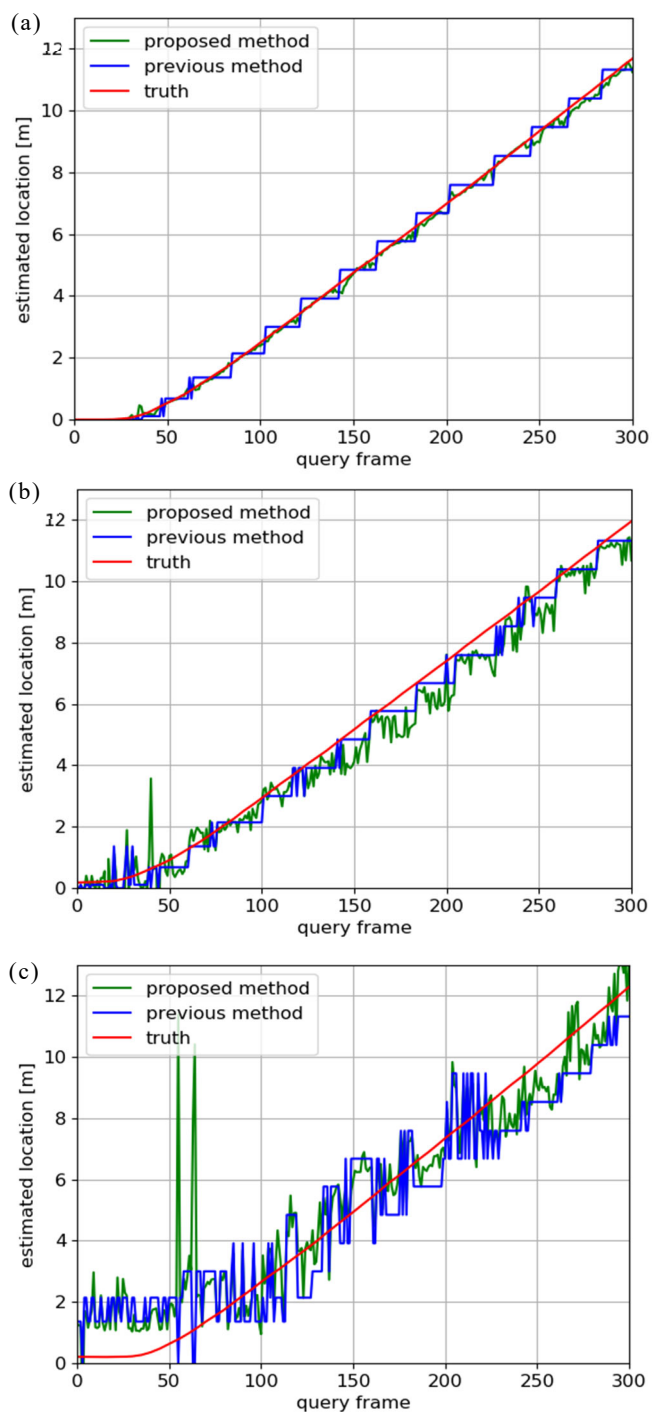


図 5 outdoor01 における位置推定結果
(a) query, (b) rotation, (c) translation

表 1 outdoor01 における誤差の平均値[m]

	query	rotation	translation
フィルタ無	0.09	0.65	0.99
$R_{window} = 0.8$	0.08	0.56	1.14
$R_{window} = 0.7$	0.09	0.63	1.13
$R_{window} = 0.6$	0.10	0.76	1.24
$R_{window} = 0.5$	0.13	0.64	1.17
従来手法	0.20	0.35	1.29

表 2 outdoor01 における推定失敗率

	query	rotation	translation
フィルタ無	0.000	0.000	0.000
$R_{window} = 0.8$	0.000	0.000	0.007
$R_{window} = 0.7$	0.000	0.167	0.083
$R_{window} = 0.6$	0.003	0.077	0.403
$R_{window} = 0.5$	0.003	0.353	0.777

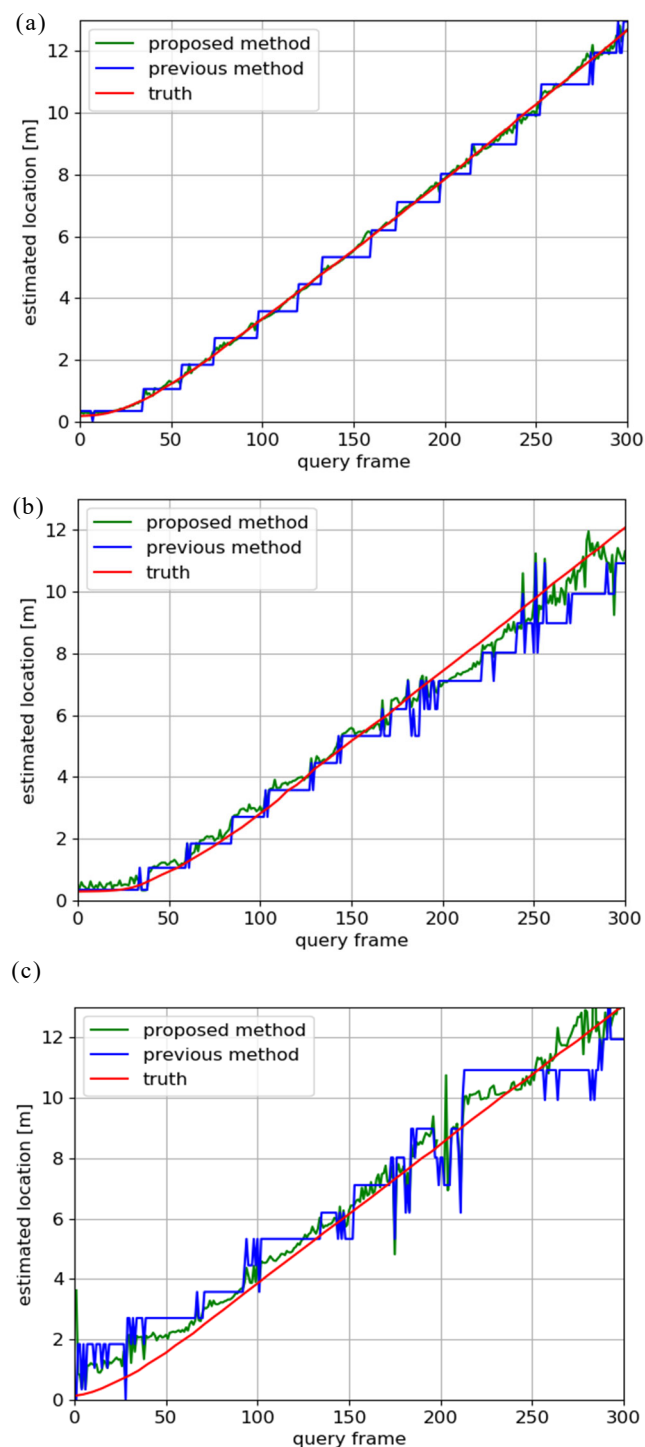


図 6 indoor01 における位置推定結果
(a) query, (b) rotation, (c) translation

表 3 indoor01 における誤差の平均値[m]

	query	rotation	translation
フィルタ無	0.07	0.34	0.47
$R_{window} = 0.8$	0.06	0.28	0.57
$R_{window} = 0.7$	0.05	0.29	0.54
$R_{window} = 0.6$	0.05	0.26	0.67
$R_{window} = 0.5$	0.07	0.26	0.68
従来手法	0.24	0.52	0.85

表 4 indoor01 における推定失敗率

	query	rotation	translation
フィルタ無	0.000	0.000	0.007
$R_{window} = 0.8$	0.000	0.000	0.070
$R_{window} = 0.7$	0.000	0.000	0.140
$R_{window} = 0.6$	0.000	0.010	0.330
$R_{window} = 0.5$	0.000	0.043	0.460

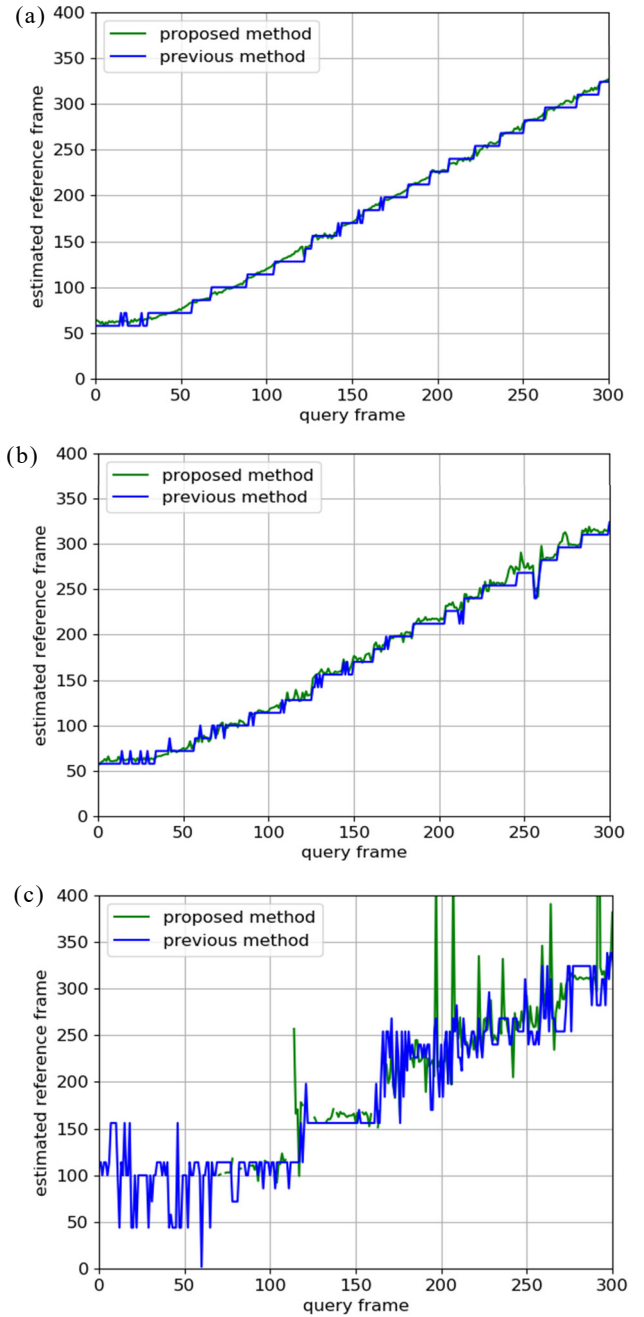
外縁部の SIFT キーを、矩形のフィルタを用いてフィルタリングを行った。

提案手法を用いて 3 種類の経路にて位置推定を行い、目標経路上において高精度な位置推定が可能であることを確かめた。回転や併進が生じている状況においても、従来手法と同程度の性能を維持することができた。

本研究の一部は科研費 17H01773 の助成を受けた。ここに謝意を表する。

文 献

- [1] 厚生労働省社会・援護局障害保健福祉部，“平成 28 年生活のしづらさなどに関する調査(全国在宅障害児・者等実態調査)結果”，2018.
- [2] 小河原 洸貴，穴戸 英彦，北原 格，亀田 能成，“ステレオ音と振動提示による視覚障害者誘導インタフェース”，信学技報，vol.117，no.48，pp.17-22，2018.
- [3] 釜坂 一步，北原 格，亀田 能成，大田 友一，“経路上の歩行者位置推定に適した事前撮影映像からのデータベース生成”，信学技報，vol.115，no.495，pp.19-24，2016.
- [4] 有限会社エクストラ，“トレッカーブリーズ”，2012，<http://www.extra.co.jp/breeze/>
- [5] D. Ahmetovic, et al. “Navcog: A Navigational Cognitive Assistant for the Blind.” Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services, pp.90-99, 2016.
- [6] M. Labbé, F. Michaud, “RTAB-Map as an Open-Source Lidar and Visual Simultaneous Localization and Mapping Library for Large-Scale and Long-Term Online Operation,” Journal of Field Robotics, pp.1-31, 2018.
- [7] R. Mur-Artal, J. M. M. Montiel, J. D. Tardos, “ORB-SLAM: a Versatile and Accurate Monocular SLAM System,” IEEE Transactions on Robotics, vol.31, no.5, pp.1147-1163, 2015.

図 7 indoor02 における位置推定結果
(a) query, (b) rotation, (c) translation

- [8] T. Gulde1, S. Kärcher, C. Curio1, “Vision-Based SLAM Navigation for Vibro-Tactile Human-Centered Indoor Guidance,” ECCV 2016 Workshops, Part II, LNCS 9914, pp. 343–359, 2016.
- [9] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” International Journal of Computer Vision, vol.60, issue.2, pp.91-110, 2004.
- [10] J. L. Schonberger, J. M. Frahm, “Structure-from-Motion Revisited,” The IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104-4113, 2016.
- [11] J. L. Schonberger, et al, “Pixelwise View Selection for Unstructured Multi-View Stereo,” European Conference on Computer Vision, pp.501-518, 2016.