

# 類似画像検索における歩行位置推定能力の实地検証

小河原 洸貴<sup>†</sup> 宍戸 英彦<sup>‡</sup> 北原 格<sup>‡</sup> 亀田 能成<sup>‡</sup>

**概要:** 我々は、事前に撮影された映像に基づく誘導インタフェースの研究を行っている。その位置推定には、事前撮影映像から抽出された参照画像と問い合わせ画像との類似画像検索による手法を利用している。推定精度向上のため、弱透視投影モデルにおける仮定により推定位置を補正する手法を提案してきた。本稿では、仮定が崩れた場合を考慮した改良案を考案する。これら3手法について、実際に想定される複数の経路における撮影映像を用いて位置推定能力を検証する。

## Field Test of Position Estimation Ability on Walking by Image Retrieval

KOKI OGAHARA<sup>†</sup> HIDEHIKO SHISHIDO<sup>‡</sup> ITARU KITAHARA<sup>‡</sup> YOSHINARI KAMEDA<sup>‡</sup>

### 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し[1]、その単独歩行の際に誘導を行う歩行支援が必要とされている。我々は、推定された位置と方向から適切な指示を与え、目標経路に沿った誘導を行うインタフェースの研究に取り組んでいる。本インタフェースで利用する位置推定手法[2]では、事前に経路上で撮影された参照画像との類似画像検索により問い合わせ画像の撮影位置を推定する。時間的かつ空間的にある程度の間隔を開けて撮影した画像を参照画像とするため、検索結果をそのまま利用する場合は推定位置の精度はその撮影間隔が上限となる。そこで我々は、弱透視投影モデルにおける仮定により推定位置の補正を行う手法を提案した[3]。本稿では、仮定が崩れた場合を考慮した補正法の改良案を考案する。

これら3手法を用いて、実際に想定される経路における撮影映像の位置推定を行う。その結果より、3手法の位置推定能力を検証し、それぞれの有効性を議論する。

### 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]やBLEビーコンを用いた屋内誘導アプリ[5]などが提案されている。このような外部機器を用いて位置推定を行う手法は、機器の設置や維持が必要で、使用場所も制限されるといった欠点がある。

近年ではRGB-Dカメラや、単眼カメラを用いたVisual SLAMと呼ばれる手法[6, 7]が研究されている。Visual SLAMには計算コストが大きいことや、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がある。

### 3. 類似画像検索による位置推定手法

#### 3.1 撮影地点法

類似画像検索による位置推定手法[2]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[8]特徴量(SIFTキー)を求めて参照画像データベースを生成する。位置推定の際は、歩行者が撮影する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SIFTキーを検索する。対応が得られたSIFTキーの組をキーペアとする。全てのSIFTキーを検索したのち、得られたキーペア数が最も多い参照画像の撮影地点を位置推定結果とする。

#### 3.2 弱透視投影に従う推定位置の補正

直進中はカメラの光軸方向と移動方向が一致するため、撮影位置の前後の変化がほぼ弱透視投影に従う。これに着目し、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推定位置の補正を行う手法を我々は提案した[3]。この弱透視投影補正法では、参照画像検索用データベースを作成したのちに事前撮影映像との類似画像検索を行い、各フレームにおける凸包面積比を算出する。ここで、凸包面積比とは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キーの凸包面積の比率である。各参照画像において凸包面積比の分布から回帰直線を求める。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を算出して回帰直線に当てはめ、事前撮影映像のフレーム番号を得る。

#### 3.3 弱透視投影補正法の改良

3.2節の弱透視投影補正法では、カメラの光軸方向と移動方向が一致することを前提としている。カメラの移動に回転や併進が生じた場合においては弱透視投影の仮定が崩れ、凸包面積比が正しく算出されない場合がある。そのため、経路によっては回転や併進時に補正を行うと精度が悪化する可能性がある。

弱透視投影の仮定が崩れた場合の精度悪化を防ぐために、弱透視投影補正法の改良を行う。各参照画像において回帰直線を求める際に凸包面積比の最大値と最小値を記録する。位置推定時には凸包面積比がその範囲内であるとき正しく算出されたとして補正を行い、それ以外は撮影地点法の結果を用いる。

<sup>†</sup> 筑波大学 大学院システム情報工学研究科  
Graduate School of Systems and Information Engineering,  
University of Tsukuba

<sup>‡</sup> 筑波大学 計算科学研究センター  
Center for Computational Sciences, University of Tsukuba

## 4. 検証実験

### 4.1 実験概要

つくば市内の商業施設において3節で述べた3手法を用いて位置推定を行った。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30fpsで撮影し, 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, 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[9, 10]を用いて求めたカメラ位置を真値とした。

### 4.2 実験結果

queryにおける位置推定の結果を図1に, 誤差の累積ヒストグラムを図2に示す。撮影地点法では, 推定位置が参照画像の撮影地点の間隔で得られるため, 階段状の結果となった。弱透視投影補正法では, おおむね真値に近い結果が得られた。補正改良法では, 84%の検索において補正が行われた。誤差については, 弱透視投影補正法は撮影地点法に比べて誤差の小さい結果の割合が多く, 推定精度が向上したといえる。補正改良法も同様に推定精度の向上が見られるが, 誤差が小さくとも補正を行わないという場合があったために弱透視投影補正法よりもやや誤差が大きい結果の割合が増えている。

rotationやtranslationにおける位置推定結果, および誤差の累積ヒストグラムは紙面の都合上割愛する。撮影地点法, 弱透視投影補正法ともに結果が真値から外れてしまう場合が多くみられた。特に弱透視投影補正法では弱透視投影の仮定が崩れたために大きな誤差が生じる場合があった。補正改良法では, このような大きな誤差が生じることを防ぎ, 撮影地点法と同程度の誤差にとどめることが可能であった。

## 5. おわりに

本稿では, 撮影地点法, 弱透視投影補正法, 補正改良法の3つの位置推定手法について実際に想定される経路における撮影映像を用いて位置推定を行い, これらの手法の位置推定能力を検証した。事前撮影映像と同様に直進している映像では, 弱透視投影補正法による精度向上が可能であった。回転や併進が発生した映像では, 弱透視投影補正法ではその仮定が崩れたために大きな誤差が生じる場合があった。こういった場合には, 補正改良法により悪化を防ぐことが可能であった。

### 謝辞

本研究の一部は科研費17H01773と18H03480の助成を受けた。ここに謝意を表す。

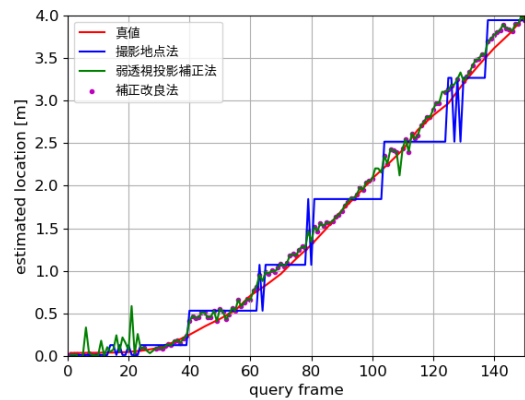


図1 queryにおける位置推定結果

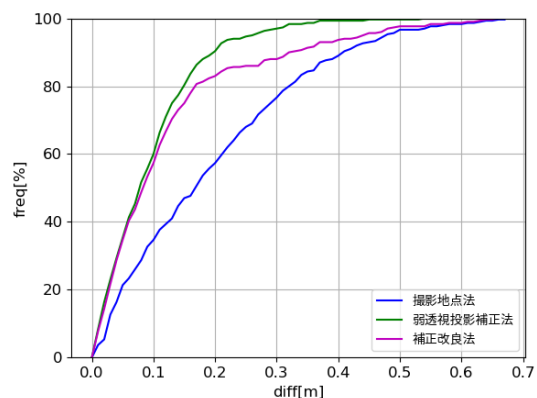


図2 queryにおける誤差の累積ヒストグラム

## 参考文献

- [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] 釜坂一歩, 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.
- [3] 小河原洗貴, 宍戸英彦, 北原格, 亀田能成. 類似画像検索による歩行者位置精度向上のための凸包の利用. 信学技報, vol. 118, no. 405, pp. 5-10, 2019.
- [4] 有限会社エクストラ. “トレッカーブリーズ”. <http://www.extra.co.jp/breeze/>, (参照 2019-04-01).
- [5] Ahmetovic, D. 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] Labbé, M. and Michaud, F. 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] Mur-Artal, R. J. et al. ORB-SLAM: a Versatile and Accurate Monocular SLAM System. IEEE Transactions on Robotics, vol. 31, no. 5, pp. 1147-1163, 2015.
- [8] Lowe, D. G. Distinctive Image Features from Scale-Invariant Keypoints. International Journal of Computer Vision, vol. 60, issue. 2, pp. 91-110, 2004.
- [9] Schonberger, J. L. and Frahm, J. M. Structure-from-Motion Revisited. The IEEE Conference on Computer Vision and Pattern Recognition, pp. 4104-4113, 2016.
- [10] Schonberger, J. L. et al. Pixelwise View Selection for Unstructured Multi-View Stereo. European Conference on Computer Vision, pp. 501-518, 2016.