

経路上での事前の移動撮影映像を用いたスナップショット撮影位置の推定

樽見 佑亮[†] 亀田 能成[†] 大田 友一[†]

[†] 筑波大学 大学院システム情報工学研究科, 茨城県つくば市天王台 1 丁目 1-1

E-mail: [†]s1320844@u.tsukuba.ac.jp, ^{††}{kameda,ohta}@iit.tsukuba.ac.jp

あらまし 我々は、経路上を移動しながら撮影した映像を用いて、経路上で撮影されたスナップショットの撮影位置を推定する研究を行っている。本手法では、事前に撮影した映像から局所特徴量を抽出したデータベースを用意しておき、スナップショットから抽出した局所特徴量を利用して類似画像検索を行うことでスナップショットの撮影位置の推定を行う。検索対象として経路上で撮影した映像を用いることから、似たようなテキストチャが多く含まれる大量の画像データからスナップショットと最も類似した画像を検索する必要がある。そこで本稿では、検索に用いた SIFT 特徴量の類似度と、画像検索の精度との間に存在する関係について調査を行った。

キーワード SIFT, 位置推定, 画像検索, OpenCV

Location Estimation of a Snapshot by Referring Pre-Recorded Route Video

Yusuke TARUMI[†], Yoshinari KAMEDA[†], and Yuichi OHTA[†]

[†] Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tenoudai
, Tsukuba, Ibaraki, Japan

E-mail: [†]s1320844@u.tsukuba.ac.jp, ^{††}{kameda,ohta}@iit.tsukuba.ac.jp

Abstract We propose a location estimation method of a snapshot by referring pre-recorded route video. We make the database that is a set of detected key features in pre-recorded route video. Location of a snapshot is estimated by counting the number of key-pairs at every frame of the pre-recorded video, and the frame that gives the largest number of key-pairs indicates the location of the snapshot. As the similarity of key-pairs make much influence on the accuracy of frame selection, we investigate the relation between precision of retrieval and similarity degree of the image key features.

Key words SIFT, location estimation, image retrieval, OpenCV

1. はじめに

我々は、特定の経路を移動しながら事前撮影した映像を用いて、経路上で撮影したスナップショット位置を推定する研究を行っている。我々の想定している経路は、人物や移動ロボットが短時間で移動可能な長さで、かつ屋内屋外であることを制限しないものである。具体例を挙げると駅から店舗までの移動や大学の教室から教室の移動における経路である。これにより人物に対する経路に沿ったナビゲーションが可能である。また、近年警備や搬送などの今まで人間が行っていた作業をロボットに代行させるために自律移動ロボットの研究が行われている。移動ロボットの研究において位置推定は重要であり、屋内屋外を問わずに位置推定が可能な我々の手法は移動ロボットの研究にも利用可能であろう。

現在、広く普及している位置推定手法の一つとして GPS が挙げられる。GPS を利用した歩行者ナビゲーションやカーナビゲーションなどは既に多くの人々が利用している。しかし、GPS はその特性上屋内での利用ができないという問題がある。

一方で、GPS を利用できない状況を想定した位置推定に関する研究も進められている。Kouroggi ら [1] [2] は加速度変化や地磁気変動を利用して歩行者の位置推定を行っている。ロボット分野では Morales らの取り組み [3] のように自己位置推定と環境地図作成を行う SLAM の研究が進められている。

一方で、カメラのみを利用して位置推定を行う研究 [6] [7] も行われている。この手法は画像シーン全体の局所特徴量を利用した類似画像検索を用いることで位置推定を行っている。本稿では、画像シーン全体を利用した類似画像検索を行った際の局所特徴量の類似度と検索精度について検証する。また、局

所特徴量は SIFT [8] を用いて実装には Ubuntu Linux [9] 上の OpenCV2.4.2 [10] を利用した。

2. 関連研究

画像を利用した位置推定の手法として Hays ら [4] の提案する手法が挙げられる。この手法では予め画像に位置情報を付加して記憶した大量の画像を地球規模で用意する。予め用意した画像と位置推定を行いたい位置で撮影した画像とのマッチングを取ることで大まかな位置推定を実現している。

また、池田ら [5] の手法では事前に GPS の位置情報と全方位カメラ映像を元に structure-from-motion によって作成したランドマークデータベースを予め用意し、このランドマークデータベースと現在撮影した映像との対応をとることで現在位置の推定を行っている。

しかし、これら二つの手法は屋外での使用を想定しており、我々の想定する経路上での位置推定には適さないと考えられる。

亀田ら [6] [7] は事前に経路に沿って一人称で撮影した事前撮影映像があることを前提にして、歩行者がその場で撮影した画像と事前撮影映像とを画像特徴量を用いて比較・検索することで歩行経路上での位置を推定できることを示している。

3. 事前撮影映像を用いた位置推定

本節では [6] [7] で提案されている位置推定手法について説明する (図 1)。まず予め進行方向前方を撮影するようにカメラを持ち経路に沿って撮影した映像を用意する。撮影された映像をフレーム毎に分割することで事前撮影画像群を作成する。次に図中の人物のように経路上でスナップショット撮影する。このスナップショットと事前撮影画像群との類似画像検索を行うことでスナップショットの撮影位置を推定する。

続いて類似画像検索について説明する (図 2)。まず事前撮影画像群から特徴量ベクトルを抽出してデータベースを作成する。スナップショットからも特徴量ベクトルを抽出する。スナップショットから抽出した各特徴量ベクトルと最も類似した特徴量ベクトルをデータベース中から探索する。スナップショットの特徴量ベクトルと類似する特徴量ベクトルを最も多くもつ画像を検索結果とする。なお、本稿では類似特徴量ベクトルの数をマッチ数と呼称する。図 2 の例ではスナップショットに対して最大のマッチ数となっている上から 4 番目の画像が検索結果となる。

この様にこの手法では大量の類似した画像からシーン全体の特徴量を抽出し、これを利用することで類似画像検索を行っている。近年、シーン中から「人」や「道路標識」といった個別物体を識別する一般物体認識の研究は多くなされているがこの様な状況に対する研究はまだ少ない。[8] では 1 つの SIFT 特徴量に対する見かけの変化に対する考察はされているが、シーン中に表れる SIFT 特徴量全体の振る舞いについては考察がなされていない。そこで本稿では検索精度と特徴量の類似度の関係性について実験により検証する。

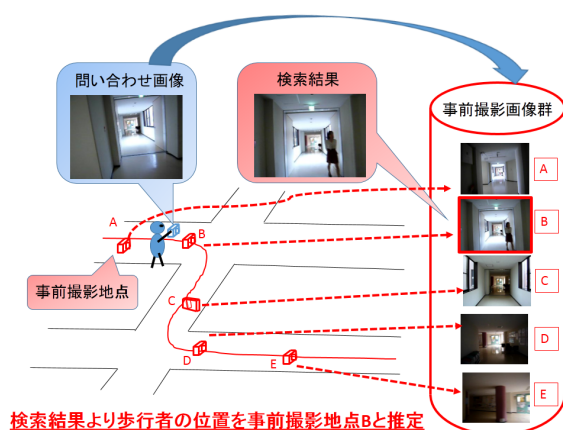


図 1 事前撮影映像を用いた位置推定

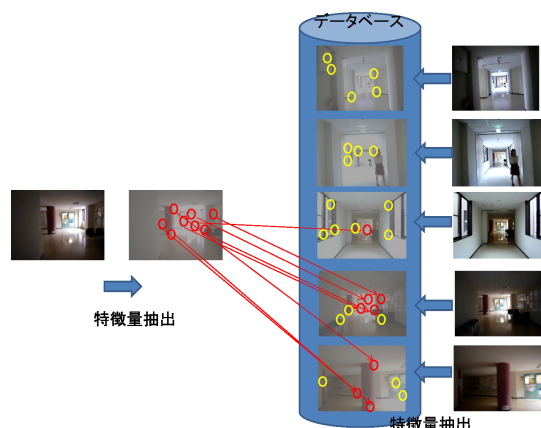


図 2 局所特徴量を用いた類似画像検索の概念

4. 検索精度と類似度の関係性の検証

4.1 検証方法

検証のためにまず同一の経路に沿って一人称視点でほぼ同じ速度で徒歩で移動しながら撮影した映像を 2 本用意する。また、2 本の映像はほぼ同じ日時に撮影を行うことで日照等による影響を極力抑えた 30fps で 2 分程度の映像である。この映像をフレーム毎に分割してこのうち一方をスナップショット、一方をデータベース画像集合として画像検索を行う。この映像はショッピングモール近辺での道案内などを想定して、つくば市内の商業地区周辺で撮影を行った。この際、利用する局所特徴量は SIFT を用いて比較器には BluteForce 法を用いた。2 本の映像は同一経路上に沿って一人称視点かつ同程度の速度で撮影された映像であるため、位置推定に十分な精度を有していると仮定するとスナップショットのフレーム番号と検索されたデータベース画像のフレーム番号はほぼ一致すると考えられる。そこでスナップショットのフレーム番号の前後 30 フレームまでを正しい検索結果 (inlier) として扱い、それ以外を誤った検索結果 (outlier) とする。

図 3 に横軸にスナップショットのフレーム番号、縦軸をデータベース画像のフレーム番号とした画像検索の結果の一つを示す。また、3 はスナップショットのフレーム番号の前後 30 フ

レームの inlier とする範囲を示している。

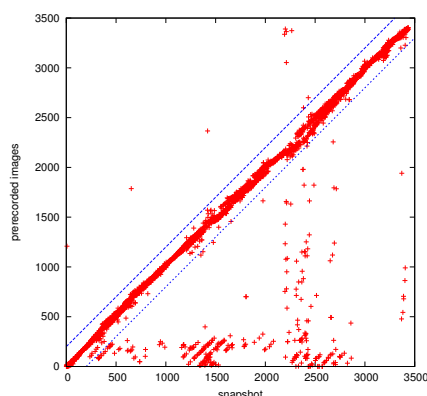


図 3 検索結果と inlier の様子

4.2 類似度と inlier 及び outlier の関係性

横軸を類似度、縦軸をその類似度をもつマッチ数としたものを図 4,5 に示す。類似度は 0 に近づくほど 2 つの特徴量ベクトルが類似していることを示す正の値である。類似度については OpenCV の `cv::DescriptorMatcher::match` メソッドによって求められた値を利用する。またこの際の画像検索の結果を図 6 に示し、図 6 にマッチ数の軸を追加したものを図 7 に示す。図 6,7 から、期待通りスナップショットのフレーム番号とデータベース画像のフレーム番号が一致したものが多くこれらはマッチ数も高い結果であることが分かる。図 4,5 から inlier は類似度 60 前後に一つのピークをもつ分布となっており、outlier については類似度 60 前後と 270 前後とに二つのピークを持つ分布となっていることが分かる。outlier の類似度 60 前後の分布は本来 inlier と判定されるべき画像が outlier 側にも多く含まれてしまっていることが原因と考えられる。一方で図 5 の類似度 270 前後の分布は inlier 側には見られない分布形状である。これによりこの周辺の類似度をもつ特徴点ペアは検索精度に良い影響を与えないという仮説が考えられる。

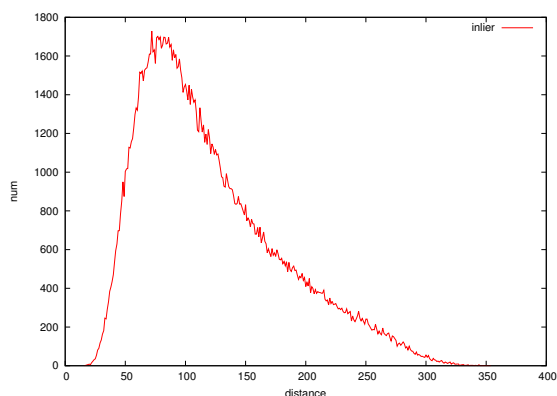


図 4 類似度と inlier の関係

4.3 低類似度と検索精度の関係性

4.2 節での仮説を検証するために実験を行った。まず全く異なる経路に沿って一人称視点でほぼ同じ速度で徒歩で移動しながら撮影した映像を 2 本用意する。これらは 30fps で撮影した 2 分程度の映像である。この 2 本の映像は経路は全く異なるが、

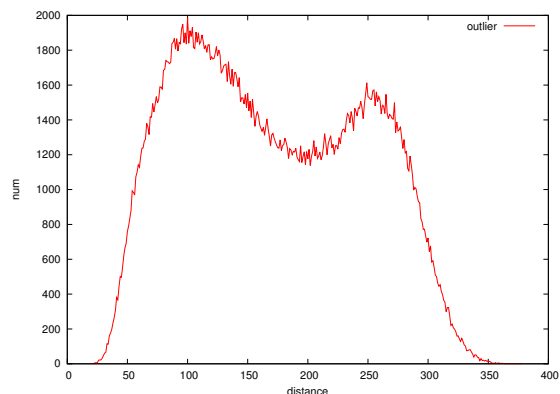


図 5 類似度と outlier の関係

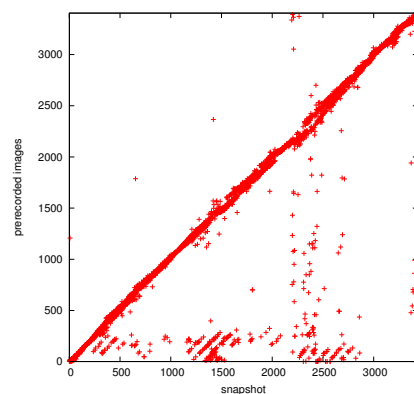


図 6 画像検索結果

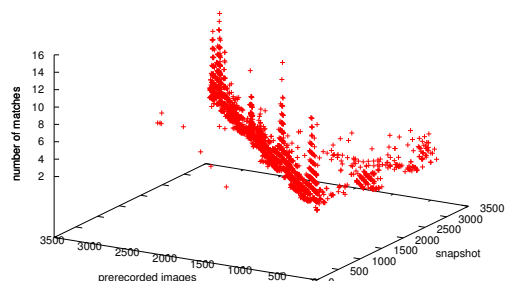


図 7 マッチ数を加えた図 6 の画像検索結果

似たような印象の経路を選んでいる。また、撮影の日時及び移動方向と太陽との向きがほぼ同じ状態になるように撮影することで日照等による影響を抑えている。スナップショット及びデータベース画像集合の内の一つを図 8 に示す。図 8 のように実際は異なる経路上で撮影した映像であっても見た目が似ている映像となっている。この 2 本の映像を用いて同様に画像検索を行う。inlier と outlier の境界についても同様に前後 30 フレームとしている。横軸を類似度、縦軸をその類似度をもつマッチ数としたものを図 9,10 に示す。またこの際の画像検索の結果を図 11 に示し、図 11 にマッチ数の軸を追加したものを図 12 に示す。図 11,12 は図 6,7 と比較すると検索結果が大きく散らばっており、マッチ数も全体的に低い値となっていることが分かる。また、図 9,10 ともに類似度 100 以上からマッチ数を増やし類似度 270 前後でピークに達する類似した分布をしていることが

見られる．画像検索に用いた 2 本の映像は全く異なる経路上で撮影した映像であるためこれらの検索結果は全てが誤った検索結果として扱われるべきものであることが分かる．このことと図 9,10 のような分布になることから類似度が 100 以上の特徴点ペアは画像検索の精度を維持する上で利用すべきではないと考えられる．



スナップショット データベース

図 8 異なる経路上での撮影画像の一例

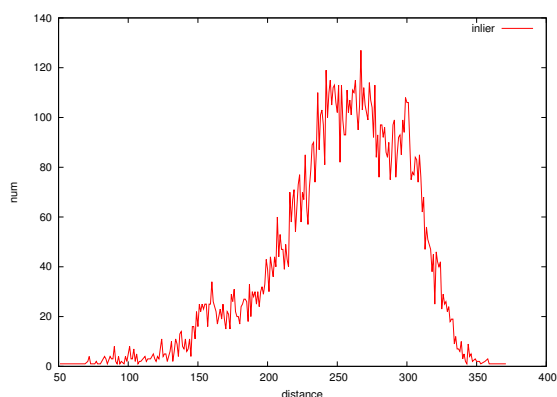


図 9 低類似度と inlier の関係

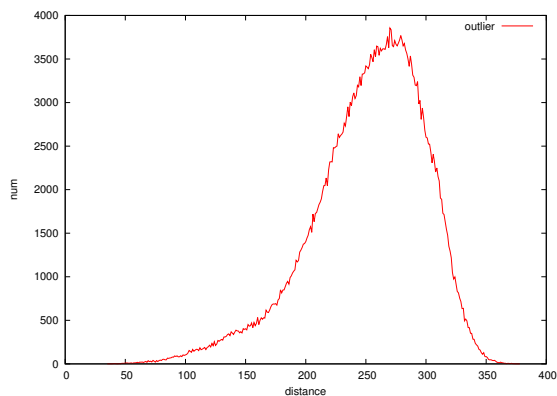


図 10 低類似度と outlier の関係

5. おわりに

本稿では経路上で事前撮影した画像と経路上撮影したスナップショットとの局所特徴量を用いた類似画像検索による位置推定手法について取り上げ、検索精度と特徴量ペアの類似度との関係について検証した．実験により低い類似度の特徴量ペアは検索精度に良い影響を与えないであろうということが考えられる．本稿では高い類似度の特徴量ペアと検索精度についての検

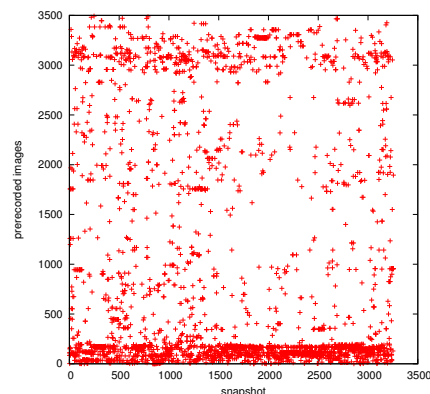


図 11 異なる経路での画像検索結果

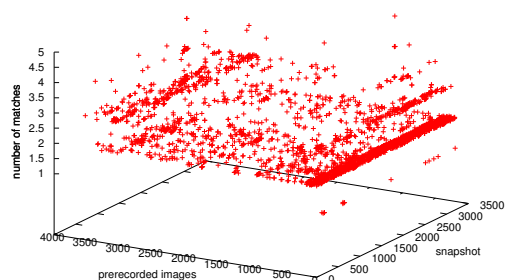


図 12 マッチ数を加えた図 11 の画像検索結果

証が不十分であるため、今後はより検索精度と類似度の関係について調査していきたいと考えている．また、今回の検証では類似度と検索精度の関係についてのみ行ったが今後は画像内の SIFT 特徴量の位置、スケール、回転などからなる相関関係を利用することで検索精度の向上を図ることができないか検証しようと考えている．

謝 辞

本研究の一部は JSPS 科研費 23300064 の助成を受けた．ここに謝意を表する．

文 献

- [1] M. Kourogi and T. Kurata, "Personal Positioning based on Walking Locomotion Analysis with Self-Contained Sensors and a Wearable Camera," Int. Symposium on Mixed and Augmented Reality, pp.103-112, 2003.
- [2] M. Kourogi, T. Ishikawa, T. Kurata, "A Method of Pedestrian Dead Reckoning Using Action Recognition," In Proc. Position Location and Navigation Symposium (PLANS), pp.85-89, 2010.
- [3] Y. Morales, E. Takeuchi and T. Tsubouchi, "Vehicle Localization in Outdoor Woodland Environments with sensor fault detection," Proc. of IEEE Int. Conf. on Robotics and Automation, pp.449-454, 2008.
- [4] J. Hays and A. A. Efros, "IM2GPS: Estimating Geographic Information from a Single Image," pp.1-8, CVPR, 2008.
- [5] 池田 聖, 佐藤 智和, 山口 晃一郎, 横矢 直和, "全方移動画像と GPS 位置情報からのランドマークデータベースの構築," MIRU, pp.1129-1134, 2007.
- [6] 亀田能成, 大田友一, "歩行者視点カメラによる歩行者位置オンライン推定の取り組み," 信学会 PRMU, vol.110, no.27, pp.67-72, 2010.

- [7] 亀田能成, 大田友一, “街中での歩行者カメラによるオンライン位置推定のための検討,” MIRU, pp.364-369, 2010.
- [8] D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” Int. Journal of Computer Vision, vol.60, no.2, pp.91-110, 2004.
- [9] Yoshinari Kameda, Itaru Kitahara, and Yuichi Ohta, “Uniform Software Environment for AR Performance Evaluation based on USB Boot Linux,” International Workshop on Benchmark Test Schemes for AR/MR Geometric Registration and Tracking Method (TrakMark)(in conjunction with ICPR2012), 2012.
- [10] <http://opencv.org>