

歩行経路記録映像における画像検索性能評価

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

E-mail: kameda@iit.tsukuba.ac.jp, ohta@iit.tsukuba.ac.jp

Abstract

歩行者ナビゲーションは社会的に高い需要があり、その中でも予め定められた経路をナビゲーションする形態を我々は想定している。この手法では経路に沿って一人称視点で事前撮影した歩行経路記録映像を利用する。現在位置を知りたい歩行者は、その場で問い合わせ用の画像を一枚撮影する。問い合わせ画像が歩行経路記録映像中のどのフレームに類似しているかを画像検索によって調べ、そのフレーム番号に対応する記録位置をもって推定位置とする。この手法では位置推定の精度は画像検索の精度によって直接影響を受ける。歩行経路記録映像における画像検索が失敗しうる要因としては様々なものが考えられるが、本稿では以下に挙げる3つの要因を取り上げる。

- ほぼ同じ地点からの撮影であっても一人称視点での撮影により多少見かけが異なること。
- 事前撮影の映像と問い合わせ画像とでは日照などの変動により見かけが異なること。
- 経路上で異なる地点でありながら一見すると類似して見える地点が存在しうること。

また、この画像検索に用いる局所特徴量に SIFT と SURF を利用した場合の画像検索性能の差についても検証する。

1 はじめに

我々の想定する歩行者ナビゲーションは、予め定められた特定の経路に沿って歩行者を誘導する形式のものである。

歩行者ナビゲーションを行う上で歩行者の位置を推定することは重要である。

歩行者の位置を推定する方法として GPS が広く普及しており、GPS を用いた歩行者ナビゲーションも既に広く普及している。

一方で GPS が利用できない状況を想定して、GPS を使わずに歩行者の位置を推定する研究も進められている。Kourogiri 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している。ロボットの分野でも Morales らの取り組み [3] のように位置を推定する 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する手法も提案されている。例えば、Hays ら [4] 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し、それらと入力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。本稿では画像を用いた位置推定の手法の一つを取り上げ [5]、その精度について検証する。

予め一人称視点で経路に沿って撮影した歩行経路記録映像があることを前提に、亀田ら [5] は画像特徴量を用いて事前撮影映像と歩行者がその場で撮影した問い合わせ画像とを比較・検索することで歩行経路上の位置推定が可能であることを示している。

この手法について図1を用いて模式的に説明する。図1の赤い線が歩行者を誘導する経路を表しているとする。この経路に沿って一人称視点で予め撮影された歩行経路記録映像をフレーム毎に分解することでデータベース画像集合を作成する。図中では赤いカメラが経路上での撮影地点を示している。図中では青い人物がナビゲーション対象者で、この人物は経路上で一人称視点に構えたカメラで問い合わせ画像を撮影する。ナビゲーション対象者が撮影した問い合わせ画像とデータベース画像集合間で画像検索を行う。検索結果画像の撮影地点がナビゲーション対象者の経路上での推定された位置となる。

この手法では画像検索の精度が歩行者の位置推定の精度に直接影響を与える。そのため、本稿ではこの手法を用いた際に起こると考えられる画像検索の性能を低下

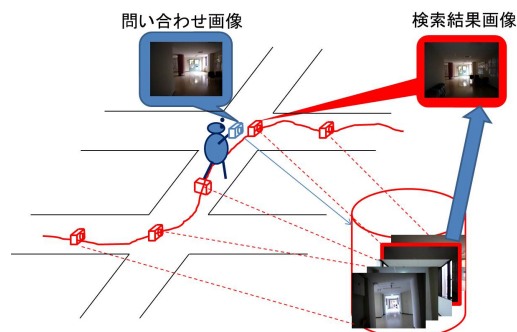


図1 画像検索による歩行者の位置推定の様子

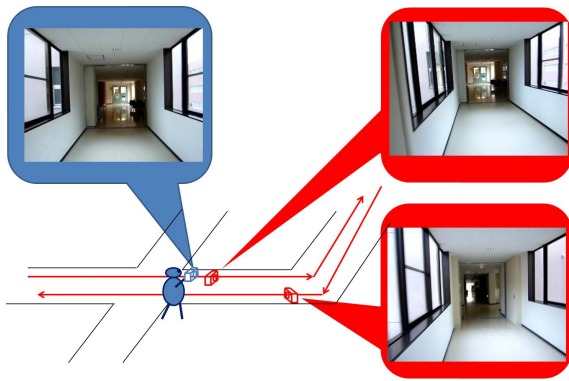


図2 経路上異なる位置で見た目の近い画像が撮影される様子

させる要因を3つ挙げ、それらの要因が実際に画像検索の性能にどれほどの影響を与えるか実験によって確認する。

また、この時に用いる局所特徴量として SIFT[6] と SURF[7] を取り上げこれらの違いが画像検索性能に与える影響についても述べる。

2 歩行経路記録映像での探索に含まれる精度低下の要因

本手法では位置推定にあたって特定の経路に沿って一人称視点で撮影された問い合わせ画像が用いられる。撮影された問い合わせ画像には下記に挙げるような画像検索を行う上で問題となる要因が含まれていると考えられる。

A. 撮影地点のずれ

ほぼ同じ地点からの撮影であっても一人称視点での撮影により多少見かけが異なること。

B. 撮影時間の違い

事前撮影の映像と問い合わせ画像とでは日照などの変動により見かけが異なること。

C. 異なる地点で類似画像が撮影されること

経路上で異なる地点でありながら一見すると類似して見える地点が存在しうること。(図2)

これらの要因に対処するために [5] では回転や全体的な輝度変化にロバストであり高い表現力をもつ局所特徴量を用いることを提案している。本稿では上に挙げた3つの要因を含む歩行経路記録映像に対して画像検索を行い位置推定に十分な精度を得られるか検証する。この時、SIFTとこれを高速化したSURFを用いて検証し、各特徴量ごとの違いについても検証する。

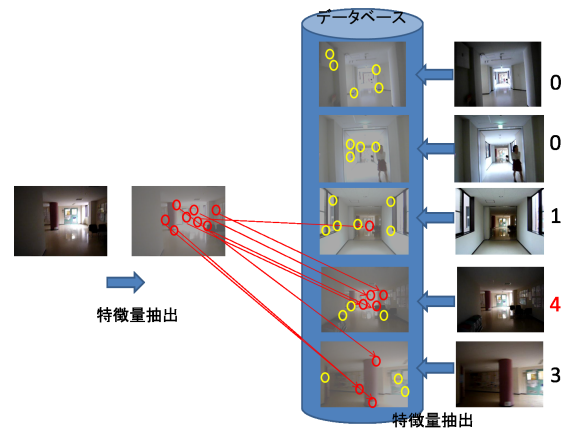


図3 局所特徴量を用いた画像検索の様子

3 局所特徴量を用いた画像検索

この章では [5] で用いている局所特徴量を用いた画像検索について説明する。図3のように前処理として事前撮影した歩行経路記録映像をフレーム毎に分解し、各々の画像において特徴量ベクトルを求めたデータベース画像集合を作成する。歩行者が撮影した問い合わせ画像からも特徴量ベクトルを求める。問い合わせ画像の各特徴量ベクトルに最も類似する特徴量ベクトルをデータベース画像集合の中から探索する。問い合わせ画像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画像検索の出力とする。なお、本稿では問い合わせ画像とデータベース画像集合間での類似特徴量ベクトルの数をマッチ数と呼称する。

図3の例では問い合わせ画像に対して最大のマッチ数である上から4番目の画像が検索結果となる。

4 画像検索の性能評価

歩行経路記録映像による画像検索精度を低下させる要因を含む映像を用意して実際に画像検索を行った。特徴量にはSIFTとSURFを用い、実装にはOpenCV2.4.2[8]を利用した。2節で挙げた3つの要因を含むようにしつつ、同一の経路に沿って一人称視点でほぼ同じ速度で移動しながら撮影した映像を2本用意する。このうち一方を問い合わせ画像、一方をデータベース画像集合として画像検索を行った。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問い合わせ画像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。

4.1 撮影地点のずれに対する性能評価

2節で一つ目に挙げた撮影地点のずれ(A)について検証する。同一の経路で1月16日17:16頃と1月16日17:18頃に撮影した2本の映像を用意して実験した。



図 4 4.1 節で画像検索に用いた映像

撮影は 30fps で行い、全フレームで分解した。撮影は茨城県つくば市のショッピングモールで行った。2本の映像は一人称視点で撮影されるため、ほぼ同一の地点であっても手振れ等の原因によって同一の映像を取得することは困難である。ショッピングモールのため人々を避けながらの撮影なので実際に通った位置も似てはいるが同じではない。また、2本の映像間で違う人物が写り込んでいたりする。これらによって検索精度が低下することが考えられる。

実際に撮影した画像の一例を図 4 に示す。図 5 に検索結果を示す。上が問い合わせ画像、下はそれに対して近いと思われるデータベース画像集合の 1 フレームである。横軸が問い合わせ画像のフレーム番号、縦軸が検索結果となったデータベース画像のフレーム番号を示している。ほぼ直線の関係性を維持しており高い検索精度を維持していることが分かる。また、図 6 は図 5 にマッチ数の軸を追加し、SIFT と SURF を同時にプロットしたものである。直線から外れている部分は正答の場合と比較してマッチ数が低く、閾値処理によってこれらを除去できると考えられる。

4.2 撮影時間の違いに対する性能評価

2 節で二つ目に挙げた撮影時間の違い (B) について検証する。同一の経路で 1 月 16 日 17:18 頃と 2 月 17 日 14:46 頃に撮影した 2 本の映像を用意して実験した。撮影は 4.1 節と同じ場所で行った。2 本の映像は 4.1 節で述べた要因に加え撮影時間が異なるため日照等の違いにより見た目が大きく異なり検索精度を低下させると考えられる。実際に撮影した映像の一例を図 7 に示す。図 8,9,10 に検索結果を示す。図 5 と図 8 のグラフを比

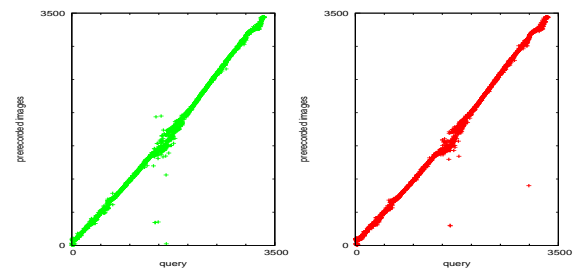


図 5 SIFT と SURF による撮影地点のずれに対する検証 (左:SIFT, 右: SURF)

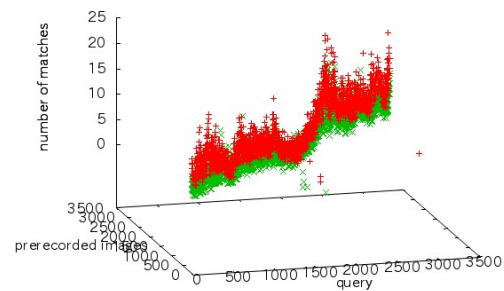


図 6 マッチ数を考慮した撮影地点のずれに対する検証 (緑: SIFT, 赤: SURF)

較すると撮影時間の違いにより大きく精度が低下していることが分かる。しかし、図 9,10 に示すようにマッチ数に着目すると正答とみなせる直線上以外の部分は概ねマッチ数が少ない傾向にあることが分かる。これにより閾値処理により検索精度の向上が見込めると考えられる。

4.3 異なる地点で類似画像が撮影されることに対する性能評価

2 節で三つ目に挙げた異なる地点で類似画像が撮影されること (C) について検証する。図 2 のように経路上の異なる地点で見た目の近い画像が撮影された 2 本の映像を用意して実験した。この映像は筑波大学の第三エリア M, L 棟内で撮影し、建物内の同じ廊下で往路と帰路で見た目の近い構造をしている場所を通過するような経路で撮影を行っている。異なる地点で見た目の近い場所があるため検索精度の低下が考えられる。実際に撮影した映像の一例を図 11 に示す。上が往路、下が帰路で撮影された画像である。図 12,13,14 に検索結果を示す。1400, 2100 フレーム付近でグラフが欠損しているのは経路上で極めて暗い地点があり、画像特徴量が取得できなかったことが原因である。用意した映像は 700 フレーム前後と 2600 フレーム前後で見た目の近い画像が撮影されている。そのため図 12 グラフでは 700, 2600 フレーム付近での誤検索が目立つ。しかし、



図 7 4.2 節で画像検索に用いた映像

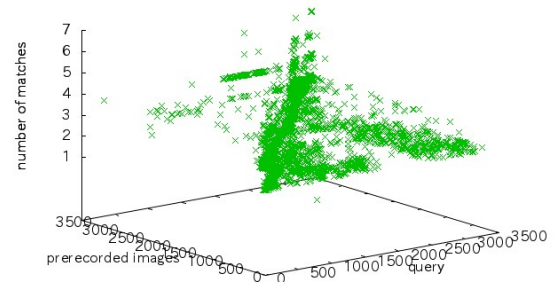


図 10 マッチ数による撮影時間の違いに対する検証 (SIFT)

図 13,14 のグラフが示すように直線上の以外の部分はマッチ数が少ない傾向であることが分かる．これから，要因 C についても閾値処理による検索精度の向上が見込めると考えられる．また，SIFT と SURF について比較すると SIFT のほうが誤検索された部分のマッチ数が安定して低い傾向がみられる．

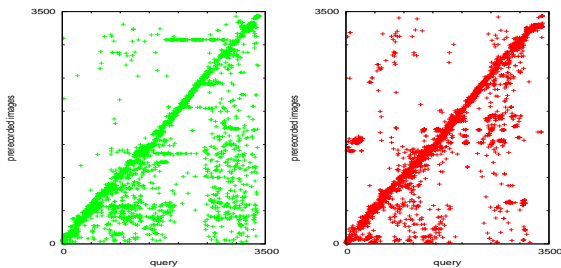


図 8 SIFT と SURF による撮影時間の違いに対する検証 (左:SIFT , 右:SURF)

5 おわりに

本稿では歩行経路記録映像を用いた歩行者の経路上での位置推定が可能であることを検証するために画像検索の性能評価を行った．また，SIFT と SURF によって画像検索を行い精度の違いについて検証した．2 章で挙げた撮影地点のずれに関しては位置推定を行うのに十分な精度を有していると考えられる．撮影時間の違いと異なる地点で類似画像が撮影されることについては検索結果に大きなばらつきが見られたもののマッチ数を用いた閾値処理により検索精度の改善が期待できることが分かった．

また，SIFT と SURF の比較では記述力の違いから SIFT が誤検索を低減させる上で有利であることが分かった．

今後は閾値処理による誤検索の低減を実際に実施し，検索精度向上の数量評価を行ってきたい．

参考文献

- [1] M. Kouroggi 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. Kouroggi, T. Ishikawa, T. Kurata, “A Method of Pedestrian Dead Reckoning Using Action Recognition,” In Proc. Position Location

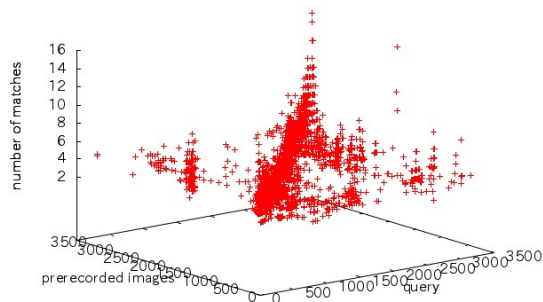


図 9 マッチ数による撮影時間の違いに対する検証 (SURF)



図 11 4.3 節で画像検索に用いた映像 (上: 往路 700 フレーム付近 下: 帰路 2600 フレーム付近)

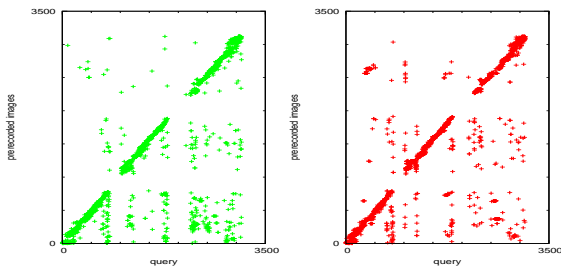


図 12 SIFT と SURF による異なる地点で類似画像が撮影されることに対する検証 (左:SIFT , 右:SURF)

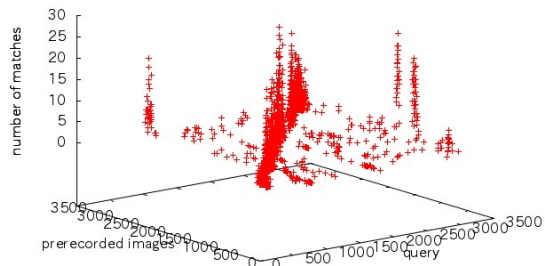


図 13 マッチ数による異なる地点で類似画像が撮影されることに対する検証 (SURF)

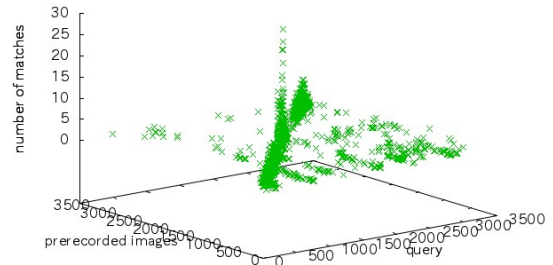


図 14 マッチ数による異なる地点で類似画像が撮影されることに対する検証 (SIFT)

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] 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.
- [6] D. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” Int. Journal of Computer Vision, vol.60, no.2, pp.91-110, 2004.
- [7] H. Bay, T. Tuytelaars, and L. V. Gool “SURF: Speeded up Robust Features” 9th European Conference on Computer Vision, pp.404-417, 2006.
- [8] <http://opencv.org>