

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

釜坂 一步[‡] 北原 格[†] 亀田 能成[†] 大田 友一[†]

[†] [‡] 筑波大学 〒305-8573 茨城県つくば市天王台 1-1-1

E-mail: [‡] s1211104@tsukuba.ac.jp, [†] {kitahara, kameda, ohta}@iit.tsukuba.ac.jp

あらまし 我々は、事前に撮影された映像に基づく歩行者ナビゲーションシステムの研究を行っている。その位置推定手法として、事前撮影映像から生成したデータベースに対する類似画像検索が用いられる。先行研究では、事前撮影映像の一定間隔でのフレーム抽出により参照画像を選択し、参照画像から検出した局所特徴量をデータベースに登録している。本稿では、実際の状況で撮影された動画から選択され得る、参照画像として適切でないフレームについて述べ、それらのフレームが参照画像として選択されないよう、フレーム抽出の間隔を適応的に決定する手法を提案する。実際の経路において撮影した映像を用いて提案手法の有効性を確認した。

キーワード 位置推定, 画像検索, 局所特徴量, データベース

Database Building from Pre-Recorded Video Suitable for Estimating the Location of Pedestrian

Kazuho KAMASAKA[‡] Itaru KITAHARA[†] Yoshinari KAMEDA[†] and Yuichi OHTA[†]

[†] [‡] University of Tsukuba 1-1-1 Tennoudai, Tsukuba, Ibaraki, 305-8573 Japan

E-mail: [‡] s1211104@tsukuba.ac.jp, [†] {kitahara, kameda, ohta}@iit.tsukuba.ac.jp

Abstract We are studying a walker navigation system that guides based on a pre-recorded video. Image retrieval of a snapshot on a path to the database that is built from the frames of the pre-recorded video plays an important role in this system. The conventional approach selects frames at a constant time interval from the video. However, this may include unexpected frames that may degrades the performance of the image retrieval. We define the frames that should be rejected on building the database and propose a new method to build the database by selecting desirable frames. The experimental results show that the improvement of the image retrieval performance.

Keywords Location Estimation, Image Retrieval, Local Feature, Database

1. はじめに

近年普及してきている歩行者ナビゲーションシステムの多くは健常者を対象としており、視覚障害者が利用できる例はまだ少ない。視覚障害者にとって使い方が簡単で、どこでも使用可能であり、1人で行けるという条件を満たすナビゲーションシステムが望まれている[10].

視覚障害者を対象とした、GPSの測位に基づくナビゲーションシステム[11]はすでに商用化されている。しかしGPSを利用する性質上、高層ビルの多い都心部や、電波の届かない屋内及び地下街では十分に機能しない問題を抱えている。この問題の解決策として、経路に沿って配置されたRFID(radio frequency identifier)タグ[4, 7]やBLE(Bluetooth Low Energy)ビーコン[12,

13]などを利用して歩行者位置を測位するシステムが提案されている。しかしこれらのシステムは、タグやビーコンの配置と維持に費用がかかることや、使用可能な場所が限られることが問題となる。

本稿では、亀田ら[6]の提案するカメラのみを利用した歩行者位置推定手法を前提に、その推定をより安定させる手法について提案する。この手法では、先行者が経路に沿って事前に撮影した映像からデータベースを作成し、歩行者が撮影した問い合わせ画像と類似する画像を検索する。検索結果の画像が撮影された位置を歩行者の経路上における位置と推定する。類似画像検索にはSIFT[1]局所特徴量のマッチングを利用した一般的な画像認識手法を用いている。以後、画像から検出されたSIFT局所特徴量をSIFTキーと呼称する。

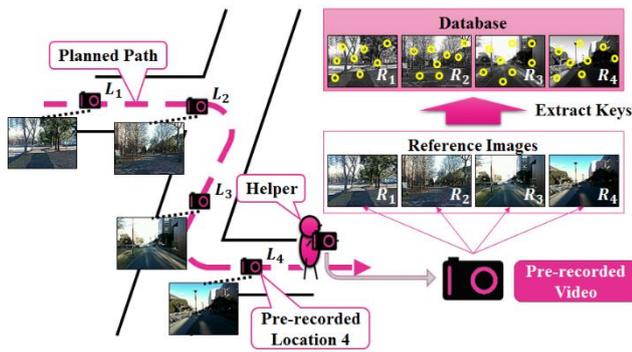


図 1 事前撮影映像からのデータベース生成

亀田らの手法では、30fps で撮影された事前撮影映像を 3fps の時間間隔でフレーム抽出し、画像検索の参照画像集合とする。データベースには参照画像集合から検出された SIFT キーを登録する。検索結果の画像の撮影位置を歩行者の推定位置とするため、参照画像は経路上の異なる地点で進行方向を向いて撮影されたフレームであることが好ましい。しかし、事前撮影映像の一定の時間間隔でのフレーム抽出では、事前撮影映像を撮影する先行者の挙動によって、歩行者位置推定に適さない参照画像が選択され得る。本稿では、参照画像として適切なフレームを事前撮影映像のフレーム間の見た目の違いに応じて選択する手法を提案する。

2. 関連研究

視覚障害者を対象とする、GPS 以外の歩行者位置推定手法を用いたナビゲーションシステムが多数研究されている。例として、経路に沿って配置された RFID タグを、白杖の先端などに装着した RFID リーダーで読み取り現在位置を取得するシステム[4, 7]がある。しかし、視覚障害者一人一人の使用する経路は多岐に渡り、そのすべてに RFID タグを配置することは現実的でない。また、駅などの施設に複数設置された BLE ビーコンにより、歩行者位置を測位する取り組みが各地で行われている[12, 13]。この手法はビーコンの設置された特定の施設においてのみ利用可能であり、人口の少ない地域や使用頻度の低い施設への普及は見込めない。

事前に撮影された画像群から作成したデータベースを用いて、問い合わせ画像の撮影位置を推定する研究が進められている。Schindler ら[3]は、街中で撮影された 1700 万枚に及ぶ参照画像集合からデータベースを作成し、問い合わせ画像の撮影位置を推定する取り組みを行った。しかし、データベースの作成には全方位 LIDAR センサやパノラマ動画撮影カメラといった機材を要するため、一人一人の要求する経路のデータベースを網羅的に作成することは困難である。

本研究では局所特徴量として SIFT を利用する。SIFT

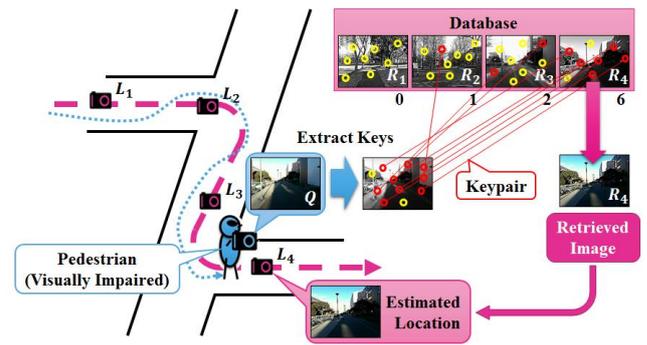


図 2 局所特徴量マッチングを利用した画像検索と歩行者位置推定

の拡張として、計算コストが高い代わりにアフィン変換に頑健な ASIFT[5]が提案されている。今後、画像検索の精度を向上するため、ASIFT を用いることも検討している。

動画から特徴的なフレームを選択する取り組みとして、佐藤ら [8]や Klein ら[2]の研究が挙げられる。例として、Klein らの PTAM は、環境から検出された特徴点のマッピングと、カメラ姿勢のトラッキングをリアルタイムに行う手法であり、マーカレス AR などに利用される。フレーム間のカメラ姿勢の変化は小さいため、キーフレームと呼ばれる特徴的なフレームのみでマッピング処理を行い、同一の三次元位置から検出される特徴点がマッピングされるのを避けることで処理を高速化している。

3. カメラのみを用いた歩行者位置推定手法

本章では、我々が利用する歩行者位置推定手法の概要と問題点について述べ、それを改善する手法を提案する。

3.1. 歩行者位置推定手法の概要

歩行者位置推定手法として亀田ら[6]の手法を用いる。この手法では図 1 に示すように、前処理として予め経路に沿って撮影された一人称視点映像からデータベースを作成する。一人称視点映像とは、歩行者の前面に取り付けられたカメラから撮影される映像である。この事前撮影映像からフレームを抽出することで参照画像集合を作成する。参照画像集合から検出された SIFT キーをデータベースに登録する。

歩行者位置推定の過程を図 2 に示す。位置推定の際には問い合わせ画像として、歩行者の一人称視点映像を 1 枚撮影する。撮影された問い合わせ画像から SIFT キーを検出し、各キーについてデータベース上から最も類似する最尤 SIFT キーを探索する。探索された特徴点の対をキーペアと呼び、キーペア数が最大となる参照画像を検索結果画像とする。検索結果画像の撮影

された位置が歩行者の経路上における推定位置であり、キーペア数はその信頼度を表す。図2の例では、画像 R_4 が問い合わせ画像 Q に対する検索結果画像であり、歩行者の推定位置は L_4 となる。

3.2. 一定の時間間隔のフレーム抽出の問題点

亀田らの歩行者位置推定手法では、一定の時間間隔でのフレーム抽出により参照画像を選択している。しかし実際の経路では様々な要因により、常に一定の速度と向きで歩行することは困難であるため、参照画像として適切でないフレームが選択され得る。歩行者位置推定に適切でない参照画像として、距離的に近い地点で撮影された見た目の類似するフレームと見直し動作中のフレームの問題点を述べる。

実際の経路では、信号待ちや他の歩行者などの要因により、先行する撮影者が停止あるいは歩行速度を落とすことは十分に考えられる。このとき一定の時間間隔でのフレーム抽出では、事前撮影映像から見た目の類似するフレームが参照画像として複数選択される。類似画像から検出される相違度の小さい SIFT キーがデータベースに登録されると、その付近で撮影される問い合わせ画像のキーペアが複数の参照画像に分散する。キーペアの分散により検索結果画像のキーペア数が減少すると、同一地点で撮影された問い合わせ画像に対しても、検索結果が変動しやすくなる。また、本手法では誤検索をキーペア数の閾値処理により除去するため、問い合わせ画像に対して正しい位置推定結果が得られる割合が減少するという問題がある。

安全や経路の確認のために先行する撮影者が見直し動作を行った場合、経路の進行方向を向いていないフレームが参照画像として選択される可能性がある。SIFT キーのマッチングにより位置推定を行うため、データベースに登録される SIFT キー数の増加に伴い、問い合わせ画像の撮影から位置推定結果を得るまでの時間も増加する。

これらの問題から、事前撮影映像から一定の時間間隔で抽出された参照画像は、類似画像検索による位置推定に適していない。本稿では、フレーム抽出の間隔を事前撮影映像の見た目の違いに応じて決定する手法を提案する。

3.3. 適応的な参照画像の選択

提案手法では、事前撮影映像の見た目の違いに応じて参照画像を選択する。予め各フレーム間の類似度を算出し、その変化を見た目の違いと定義する。類似度の算出方法については4.1節に述べる。

見た目の違いと経路上の進行の関係を明らかにするため撮影実験を行った。経路映像は本学キャンパス内の屋外を、3fps で約 1000 フレームかけて徒歩で撮影し、500 フレーム目から 200 フレーム程度の停止と、

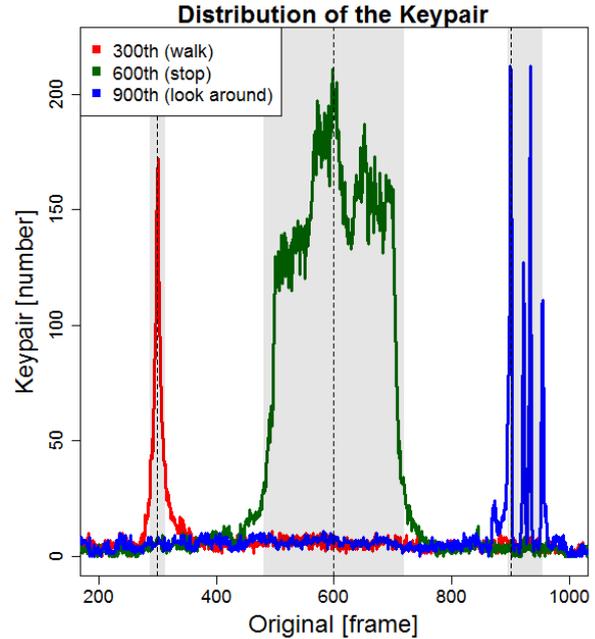


図3 300, 600, 900 フレームを基準とする他フレームとの類似度の変化と類似範囲

900 フレーム目から 50 フレーム程度の見直し動作を含む。経路映像の 300, 600, 900 フレーム目を基準とした、他フレームとの類似度の変化を図3に示す。横軸は経路映像のフレーム番号、縦軸はフレーム間の類似度である。基準となるフレームが点線で示されており、300, 600, 900 フレーム目はそれぞれ、移動中、停止中、見直し動作中のフレームである。歩行中はカメラの移動に従ってフレームの見た目が変わるため、ある一瞬の 300 フレーム目を基準とした他フレームとの類似度は、撮影位置が基準となるフレームに近いほど高くなる。停止していた区間に含まれる 600 フレーム目を基準とすると、停止中のフレーム間の類似度は大きく変化しない。見直し動作中の 900 フレーム目に対する類似度は、他フレームを撮影したカメラの向きと、900 フレーム目を撮影したカメラの向きとの差が大きくなると減少する。

図3の灰色の領域は、基準となるフレームに対し類似地点で撮影されたと考えられるフレームの範囲を示しており、本稿では類似範囲と呼称する。類似範囲が重複しないよう参照画像を選択することで、経路上の異なる地点で進行方向を向いて撮影されたフレームのみを抽出する。

4. 事前撮影映像からのデータベース生成

本章では、参照画像を事前撮影映像の見た目の違いに応じて選択する手法について述べる。

4.1. 2画像間の類似度

2画像間の類似度を評価する方法を述べる。式(1)に

より，画像 I_1 から検出された SIFT キー $K_{I_1,a}$ に対する最尤 SIFT キーを，別の画像 I_2 から検出された SIFT キー集合 $\{K_{I_2}\}$ から求める．ここで $d(K_{I_1,a}, K_{I_2,b})$ は SIFT キー間の相違度を求める関数であり， $K_{I_1,a} \cdot d_{min}$ は最尤 SIFT キーとの相違度である．

$$K_{I_1,a} \cdot d_{min} = \underset{\{K_{I_2}\}}{\operatorname{argmin}} (d(K_{I_1,a}, K_{I_2,b})) \quad (1)$$

このとき， $K_{I_1,a}$ と $K_{I_2,b}$ の組がキーペアであり，キーペア数を2画像間の類似度とする．SIFT キーの単純なマッチングでは図4の1, 2列目に示すように，誤った対応を含み得る．誤対応のキーペアは，画像間の類似度を測る指標として適さないため，次に示す3つの基準によるフィルタリングで除去する．

I. 相違度

相違度の大きいキーペアを除去する．

$$K_{I_1,a} \cdot d_{min} > D_{th} \quad (2)$$

II. 類似度係数

SIFT キー $K_{I_1,a}$ に対し，1番目と2番目に小さい相違度が $K_{I_1,a} \cdot d_{min1}, K_{I_1,a} \cdot d_{min2}$ であるとき，その比 r_d を式(3)により求める．

$$r_d = K_{I_1,a} \cdot d_{min1} / K_{I_1,a} \cdot d_{min2} \quad (3)$$

SIFT キー $K_{I_1,a}$ が画像中に広く分布する路面の模様などから検出されている場合，誤った対応点のキーペアが生じやすい．このとき r_d の値は大きくなるため，閾値 R_{th} を設ける．

$$r_d > R_{th} \quad (4)$$

III. クロスチェック

SIFT キー $K_{I_2,b}$ が，複数の SIFT キー $K_{I_1,a}$ の最尤 SIFT キーとして探索されるとき，そのうち正しい対応は最大で1つしか存在しない．したがって，複数のキーペアのうち $K_{I_1,a} \cdot d_{min}$ が最小のペアのみを残し，他のキーペアを除去する．

これらの処理により，誤対応である可能性の高いキーペアを除去した様子を図4の3, 4列目に示す．フィルタリング後のキーペア数を $KeyPairNumber_{I_1, I_2}$ と表記する．キーペア数が多い2画像は類似度が高く，類似地点で撮影された画像であると考えられる．データバ



図4 フィルタリング前後のマッチング結果

(1,2列)フィルタリング前，(3,4列)フィルタリング後
ース生成の前処理として，事前撮影映像の全フレーム間の類似度を算出する．その分布を利用した参照画像の選択手法について次の4.2節で述べる．

4.2. 参照画像の選択手法

最初に，事前撮影映像をフレーム分割し，事前撮影画像集合を作成する．事前撮影映像が N フレームある時，各フレームを $P_n (0 \leq n < N)$ ，その撮影位置を L_n と表記する．また，事前撮影画像集合を $\{P_n\}$ と表記する．すべての事前撮影画像 P_n に対し， $P_l (0 \leq l < N, l \neq n)$ との間の類似度 $KeyPairNumber_{P_n, P_l}$ を4.1節に述べた手法で算出する．

3.3節に示したように，移動中のフレームでは $|n-l|$ の減少にしたがって，キーペア数は増加する．そこで， $\{P_n\}$ について n の両側から $KeyPairNumber_{P_n, P_l}$ を走査し，キーペア数が閾値以上となる l の範囲を求める．閾値はフレームによって検出される SIFT キー数が異なることを考慮し， P_n から検出される SIFT キー数の P_{th} 倍とした．

すべての事前撮影画像について類似範囲を求めた後，類似範囲が重複しないよう参照画像を選択する．このとき経路上で撮影された問い合わせ画像は，最大でも2枚の参照画像の類似範囲にしか含まれないため，従来手法よりキーペア数が増加する傾向になる．

5. 評価実験

本実験で想定する経路に沿って撮影した事前撮影映像からデータベースを作成し，提案手法の有効性を検証する．

5.1. 実験概要

検証実験として，筑波大学敷地内の屋内外を含む経路を通学路として設定した．経路の一部を図5に示す．経路に沿った2本の映像を，同じ天候，時間帯の異なる日に3fpsのQVGA(320×240)で撮影した．前者を事前撮影映像とし，後者をフレーム分割した画像群を問い合わせ画像群とする．

実験には SurfacePro3 (CPU: Corei5-4300U 1.9GHz,

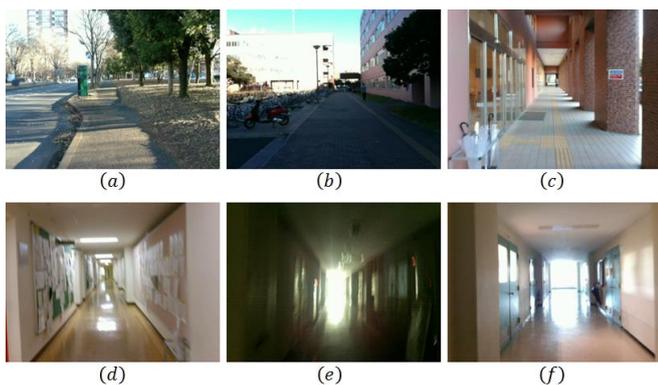


図5 実験経路の一部 (a)10 フレーム, (b)500 フレーム, (c)990 フレーム, (d)1120 フレーム, (e)1500 フレーム, (f)2035 フレーム

RAM: 8GB) を用いた. 画像をリアカメラから取得し, すべての処理は同端末上で行った. 事前撮影映像から亀田らの手法と提案手法によるデータベースを作成し歩行者位置推定を行う.

5.2. 実験結果

最初に, 事前撮影映像から選択される参照画像集合について述べる. 続いて, それぞれのデータベースを用いて類似画像検索による位置推定を行った結果について述べる.

5.2.1. 参照画像集合

事前撮影映像は 3fps で 2050 フレームであり, およそ 1000 フレーム目で経路が屋外から屋内へと変わる. また, 500 フレーム目と 1500 フレーム目から 200 フレーム程度の停止, 900 フレーム目と 1950 フレーム目から 50 フレーム程度の見直し動作を含む.

データベースの作成に使用したパラメータは実験的に求め, $D_{th} = 200, R_{th} = 0.9, P_{th} = 0.2$ である.

事前撮影映像から選択された参照画像数と, データベースに登録された SIFT キー数についての比較を表 1 に示す. また, 図 6 に参照画像として選択されたフレーム番号の比較を示す. 横軸は選択された参照画像の番号, 縦軸は事前撮影映像におけるフレーム番号である. 赤点は提案手法により選択された参照画像である.

提案手法では, 図 6 の灰色の領域内に赤点が見られないことから, 停止あるいは見直し動作を行っている間のフレームが参照画像として選択されていないことがわかる. また, 事前撮影映像のうち歩行中のフレームは約 1050 フレームである. 提案手法により歩行中の 348 フレームのみが選択されたため, 歩行中のフレームで参照画像として選択されるフレーム数は約 23% まで削減された. このことは, 将来さらに必要なフレーム数の削減が可能であることを示唆している.

5.2.2. 歩行者位置推定

それぞれのデータベースに対し, 問い合わせ画像群

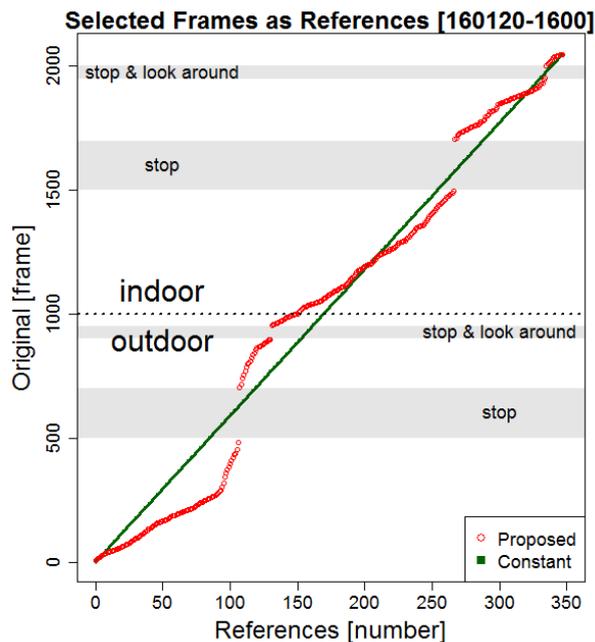


図6 提案手法により選択された参照画像

表1 事前撮影映像から選択された参照画像

	提案手法	一定時間間隔
参照画像数	348	2050
SIFTキー数	124111	622946
サイズ [MB]	66.5	434

に含まれる 1480 フレームの撮影位置を, 亀田らの手法 [6] によって推定する. ただし, 検索の際に用いるキーペアに 4.1 節に述べたフィルタリングを適用する. 提案手法により作成したデータベースに対する検索結果を図 7 左, 一定の時間間隔でのフレーム抽出により作成したデータベースに対する検索結果を図 7 右に示す. 横軸は問い合わせ画像のフレーム番号, 縦軸左は検索結果画像のフレーム番号, 縦軸右はキーペア数である. 問い合わせ画像集合及び事前撮影映像の停止と見直し動作を除く区間は, 概ね一定の速度で歩行しながら撮影を行ったため, 検索結果が正しければ赤い点は直線状となる. 図 7 は共に直線から逸脱する点が見られないことから, 誤検索が存在しないことがわかる. また緑の点の分布から, 提案手法のデータベースを利用することで問い合わせ画像に対するキーペア数が全体的に増加していることがわかる.

データベース別の検索結果数とキーペア数の頻度を表 2 に示す. 提案手法では問い合わせ画像に対して得られる平均のキーペア数が上昇している. キーペア数の増加により, 問い合わせ画像に対して位置推定結果が得られる割合が約 8.6% 向上した.

各データベースに対する処理速度の比較を表 3 に示す. ここで処理速度とは, 1440 枚の問い合わせ画像すべての推定位置を算出するのに要した時間である. 提

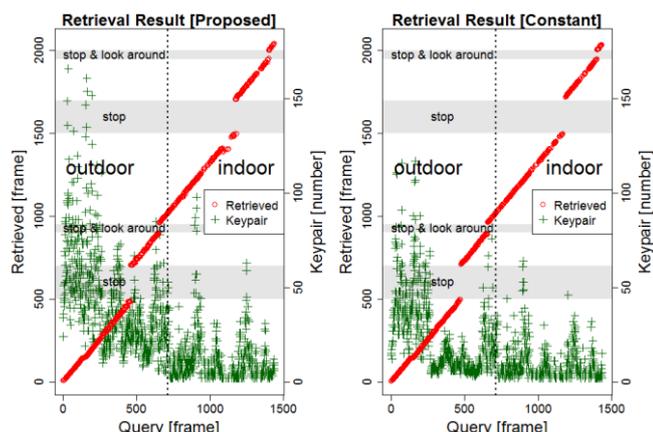


図 7 位置推定結果 (左)提案手法 (右)一定時間間隔でのフレーム抽出

案手法ではデータベースに登録される SIFT キー数の減少により処理速度が向上している。

6. おわりに

本稿では、事前撮影映像に基づく歩行者位置推定手法に用いるデータベース生成手法について、参照画像集合を事前撮影映像から適応的に選択する手法を提案し、実際の経路で比較実験を行った。距離的に近い地点で撮影された類似するフレームと見回し動作中のフレームを参照画像に適さないフレームと考え、提案手法によってこれらのフレームから検出される SIFT キーを登録しないデータベースを生成した。

提案手法によるデータベースを用いることで問い合わせ画像に対するキーペア数が増加し、検索結果を得られる割合が増加した。また、データベースに登録される SIFT キー数の減少により処理時間が短縮されたため、より長い経路を対象とすることや、マッチングの精度が高い代わりに計算コストの高い ASIFT の導入が可能になると考えられる。

今後は歩行者へ提示する移動方向の算出を行い、音声でのナビゲーションを考える。また、実際の視覚障害者を対象とする実証実験により、システムの評価と検証及び改良を行う。

本研究は JST-RISTEX の研究助成を受けた。

文 献

- [1] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision*, vol. 60, issue. 2, pp. 91-110, 2004.
- [2] G. Klein, D. Murray, "Parallel Tracking and Mapping for Small AR Workspaces", *ISMAR*, pp. 225-234, 2007.
- [3] G. Schindler, M. Brown, R. Szeliski, "City-Scale Location Recognition", *CVPR*, 2007.
- [4] H. Fernandes, V. Filipe, P. Costa, J. Barroso, "Location based services for the blind supported by RFID Technology", *Procedia Computer Science*, vol.

表2 検索結果の割合とキーペア数の比較

	提案手法	一定時間間隔
問い合わせ画像数	1440	1440
検索結果数	1236	1112
割合 [%]	85.8	77.2
0-1 pair [%]	7.8	10.4
2 pairs [%]	8.6	6.9
3 pairs [%]	4.2	5.7
4 pairs [%]	3.3	5.3
5 pairs [%]	2.4	6.3
6-10 pairs [%]	10.9	23.6
11-30 pairs [%]	31.2	25.9
31- pairs [%]	31.7	15.9
Average [pairs]	25.3	15.3

表3 各データベースに対する処理速度の比較

	提案手法	一定時間間隔
処理速度 [sec]	128.8	147.6

27, pp. 2-8, 2014.

- [5] J. M. Morel, G. Yu, "ASIFT: A new framework for fully affine invariant image comparison", *SIAM Journal on Imaging Sciences*, vol. 2, issue. 2, pp. 438-469, 2009.
- [6] K. Yoshinari, O. Yuichi, "Image Retrieval of First Person Vision for Pedestrian Navigation in Urban Area", *ICPR*, pp. 364-367, 2010.
- [7] S. Chumkamon, P. Tuvaphanthaphiphat, P. Keeratiwintakorn, "A Blind Navigation System Using RFID for Indoor Environments", *ECTI-CON*, vol. 2, pp. 765-768, 2008.
- [8] T. Sato, M. Kanbara, H. Takemura, N. Yokoya, "3-D Reconstruction from a Monocular Image Sequence by Tracking Markers and Natural Features", *In Proc. 14th Int. Conf. on Vision Interface*, pp. 157-164, 2001.
- [9] Y. Zhi, C. Xin, L. Zhu, "Video based mobile location search with large set of SIFT points in cloud", *MCMC*, pp. 25-30, 2010.
- [10] 北川博巳, 横山哲, 船場ひさお, 「視覚障害者を対象とした歩行誘導システムのニーズに関する研究」, 土木計画学研究発表会・講演集, no. 30, 2004.
- [11] 「トレッカーブリーズ」, エクストラ, 2012.
<http://www.extra.co.jp/breeze/>
- [12] 「東京駅構内ナビ」, JR 東日本, 2016.
<http://www.jreast-app.jp/s-navi/>
- [13] "Wayfindr", RLSB, 2016.
<http://www.rlsb.org.uk/tech-hub/wayfindr>