

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

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 〒305-8573 茨城県つくば市天王台 1-1-1

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

あらまし 歩行者に搭載されたカメラの映像を利用して、その歩行者位置を推定する手法を提案する。

歩行者ナビゲーションには歩行者の位置推定が不可欠であるが、都市部や地下街など様々な場所を歩く歩行者の位置推定には、GPS だけでは不十分である。本手法は、事前に撮影済みの同一経路上の映像を手がかりに、現在の位置を画像認識のみによって推定する。歩行者カメラから得られる 1 人称映像では、揺動が多く、障害物が頻出するなど、他にはみられない条件が存在する。提案手法では、SURF ベースの一般画像認識をベースにしながらも、1 人称映像ならではの拘束を利用することで、信頼度の高い位置推定を実現する。歩行者ナビゲーションが要求される状況の中でも比較的困難な状況と考えられる、東京駅周辺の地下から地上に至る 1 km ほどの歩行経路で実証実験を行い、その有効性を確認した。

キーワード 1 人称映像, SLAM, ウェアラブルカメラ, SURF, KD ツリー

Pedestrian Location Estimation by First-Person View Camera

Yoshinari KAMEDA[†] Yuichi OHTA[‡]

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

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

Abstract We propose a new pedestrian location method that exploits only a human-borne camera and runs on line. Pedestrian localization is an essential technology for pedestrian walking navigation. Unfortunately, commonly used GPS is not available, or unreliable in underground shopping mall, indoor paths, and urban streets with tall buildings. Our proposed method can estimate current position only by one image of the first person vision camera, assuming that a video that is taken by walking through the same path is given to our system in advance. As a first person camera is mounted on a body of a pedestrian, images taken by this camera always have weaving motion and obstacles such as other pedestrians. The proposed method adopts SURF based general image retrieval, but it exploits some conditions that are specific to first person vision camera so that it can produce reliable localization results. We take up one of the typical and difficult paths for experiment and evaluation; a path that starts from underground level of Tokyo station area to the ground level where tall buildings cover the sky. The result shows the proposed method can work well even in urban areas like there.

Keyword First Person Vision, SLAM, Wearable camera, SURF, KD tree

1. はじめに

歩行者ナビゲーションのためには、歩行者の位置推定が不可欠であるが、現在一般的に用いられている GPS や携帯電話の基地局に対する電波状況の情報だけでは、位置推定が不可能な状況があったり、その推定精度が歩行者ナビゲーションの役に立たないレベルであったりする。この傾向は、都市部で顕著であり、GPS 人工衛星からの電波にマルチパスを引き起こす高い建物が多い街路や、そもそも電波が入らない屋内や地下街などでは現状では歩行者ナビゲーションは難しいとされている。

一方、特に視覚障害者においては、都市部における歩行者ナビゲーションは歩行の自由度を確保するため強く望まれている。GPS ベースの歩行者ナビゲーション支援システムはすでに商用化されているもの[1]もあるが、現状では都心部ではその精度に限界がある。

こうした問題を解決するために、これまでも様々な歩行者位置推定方法が提案されている。例えば、無線 LAN のアクセスポイント情報を利用する方法は日本国内では PlaceEngine という名前で商業化されている。GPS 自体も様々な改良が提案されており、特に人工衛星からの電波のみで精度向上が可能な Differential GPS

の一種である Satellite Based Augmentation System (SBAS)は、日本国内では Multi-Functional Satellite Augmentation System (MSAS)というサービスで現在すでに利用可能である。しかしながら、これらの位置推定誤差は都心部では電波状況の悪さもあって数mから数十mにおよび、歩行者ナビゲーションを行えるだけの精度には不足である。

一方で、多数のビーコンを対象地区に埋め込んで歩行者ナビゲーションを行う手法が提案され実証もされている[2][3]。例えば、通常の GPS と組み合わせられる屋内 GPS(Indoor messaging system)や、電波ビーコン等を利用した定位手法などが挙げられる。しかし、いずれもビーコンを埋め込むコストとその維持管理が必要であるため、費用対効果が期待できる場所や、大規模博物館内の道案内など利用者が数多く見込める一部の状況を除き、今後も一般的な普及は期待しにくい。

本稿では、歩行者視点カメラ1台だけを用いて、現在の歩行位置を推定する手法を提案する。歩行者視点カメラでは自動車に搭載される車載カメラと異なり、歩行に伴う揺動が比較的大きいこと、前方の歩行者などで視界が大きく遮られることなど、固有の現象がみられる。提案手法は基本的には一般画像認識の手法をベースにしつつ、歩行者視点カメラ固有の問題に対処できるよう改良がなされている。

また、具体的なアプリケーションとして、視覚障害者に対する歩行支援を想定する。これまで我々は、厚生労働省による「画像・GPS等のセンサ統合による日常利用可能な屋内外視覚障害者歩行支援システムの開発」(研究代表者：静岡県立大学 石川准)に参画してきた。その経験をもとに、歩行者ナビゲーションに必要な条件についても本稿で議論する。

2. 関連研究

都市部のような擾乱の多い環境下での歩行者に対する測位システムとして、Kouroggi らは歩行に伴う加速度変化や磁気変動から現在位置を慣性航法によって推定する手法を提案している[4]。本手法は短距離では目覚ましい効果を発揮するが、その性質上長時間運用を継続するとドリフト誤差が蓄積する。その解消のために、他の大域的定位手段の併用が必須である。

歩行者視点カメラに近い設定状況としては、ITS 研究分野における車両搭載カメラがある。この分野では路面状況の観察と慣性センサのみから走行距離を割り出す研究[5]があるが、これは外部の電波状況に影響は受けないものの、測位には不向きである。あらかじめ用意した道路平面のパターン情報と照合することで位置推定を行う研究も提案されている[6][7]が、歩道にはコントラストの高い十分な量のテクスチャが存在す

る保証がないことから、歩行者定位への適用は難しいと考えられる。

一方、画像を用いた測位システムとしては、過去に撮影された画像に撮影地点の情報を合わせて記録し、入力画像に対して一般画像認識を行う取り組みが進められてきている [8][9]。歩行者ナビゲーションの場合、歩行者視点カメラの映像には似たような画像が多く含まれることになるが、これらの手法ではそうした歩行者視点カメラ特有の事情に対応できるかどうかは示されていない。また、別の取り組みとして、カメラを用いた SLAM では一般に事前の観測データが広角であるほど測位がしやすくなる傾向があり、このことを利用して博物館等で歩行者の定位を可能にする研究[10]が提案されている。ただし、提案手法では事前に全方位カメラを使ってシーンの画像集合を獲得する必要がある。事前に収集する画像集合を通常のカメラで行う取り組みとしては、[11]が挙げられる。この研究では車道上の経路に沿って事前に撮影した映像に対して現在位置を求めることに成功しているが、リアルタイム処理には不適である。また、歩行者ナビゲーションの場合、歩行経路に自由度が高いため、事前撮影の映像に対して自然特徴の対応点数がエピソード拘束推定できるほど多くない状況が多いこともこの手法の適用を難しくしている。

3. 歩行者ナビゲーション問題の定義

3.1. 外出の形態

本研究では、歩行者の経路は[11]と同じく、予めほぼ既知であるものとする。視覚障害者の外出の一形態として、目的地や歩行する範囲をある程度計画してから外出に臨むことがある。目的地がわかっていたとしても、その途中の経路に関する知識が不確かであったりする場合、歩行支援システムによってナビゲーション情報を正確に提示することは、視覚障害者への有用な支援となる。このことは、カメラに依る位置推定方法の開発にあたっては、全く未知の経路で位置推定を行うのではなく、ある程度歩行範囲や経路を予め絞り込んでおくことを意味する。

その一方で、一般に目的地や経路選択は多岐に渡ると考えられるので、カメラ位置推定において準備に手間がかかることは避けるべきである。これは、特に多くの人を通るわけではない主要ルートなどが歩行範囲として想定される場合は重要な要因である。

3.2. 歩行者視点カメラ

本研究では、歩行者の胸部前面に搭載したカメラ(歩行者視点カメラ)で撮影した風景画像を手がかりに、予め規定された歩行範囲内において現在位置を推

定する。実際に視覚障害者の歩行支援システムに組み込むことを考えると、カメラを頭部に装着すると歩行に支障があるとの意見が視覚障害者らから出たため、実際には視点位置より低い高さへの取り付けが前提となった。参考までに、図1に歩行者視点カメラ装着の例を示す。図中の左側に見える立方体から円筒が歩行方向に突き出しているのがカメラであり、比較的低い位置であることがわかる。このため、得られる映像は必然的に前を歩く歩行者などによって視野を覆われることが多くなる。



図1：歩行者視点カメラ装着の例

歩行者位置推定手法の研究にあたっては、歩行中の視覚障害者に支援を行えるようにするため、携帯可能なノートパソコン程度の性能でもリアルタイムに実行できることが求められる。

4. 歩行者位置推定

歩行者位置推定の手順は、まず事前に撮影した画像集合から現在観測されているシーンに最も近い画像を求め、それが実際に画像として似ているかどうかを歩行者視点カメラならではの7つの基準で検定する。検定を通った場合、その参照画像の撮影位置と撮影方向を歩行者位置推定結果として利用者に示す。

4.1. 参照画像集合

外出計画に含まれる歩行経路には様々な経路が考えられるため、事前の準備にかかる作業量は小さいほどよい。そこで、本研究では、カメラで撮影する風景画像全体を手がかりに使うものとする。これによって、予め蓄積しておくべき画像については、経路に沿って単に撮影して回るだけで収集することができるようになるので、画像収集に対する作業量を小さくできる。こうして集められた画像に撮影位置と撮影方向の情報を予め付与した上で、画像データベースに登録していく。外出前には、これから歩行する可能性のある経路

上の画像を集めて画像集合を生成する。これを参照画像集合と呼ぶ。

4.2. 検索による第一候補画像の決定

いま、参照画像集合には、経路に沿った様々な地点で撮影された N 枚の画像が含まれているとする。各画像は番号 R ($1 \leq R \leq N$) で参照する。検索のベースとなる画像 R の SURF キー k_{iR} は予め全て求めておく[12]。入力画像が与えられたとき、第一画像候補 R^* を以下の式によって求める。ここで、 $\text{CountSURFPairs}(Q, R)$ は画像 Q と R の間で類似する SURF キーのペア数を求める関数であり、 K は第一画像候補に対するそのペア数である。実現にあたってこの検索は、KD ツリー探索によって高速化する。

$$R^* = \max_R (\text{CountSURFPairs}(Q, R))$$

$$K = \text{CountSURFPairs}(Q, R^*)$$

4.3. 第一画像候補の検定

SURF ペア数 K は、画像間の類似度を示す指標であり、撮影地点間の違いや撮影方向の違い、撮影時点での障害物の有無、日照条件等の環境の違いに影響を受ける。特に、歩行者視点カメラで都心部での撮影による実験では、SURF ペア数は同じ日に同じ経路で撮影してもその数は多くなく、一般的なエピソードを利用したような手法に必要な数十ペアにも届かないことが大半である(表1参照)。そのため、第一画像候補であっても2枚の画像が実際に近い地点で撮影されているという保証はない。そのため、この2枚の画像の類似性を検定するための7つの基準を導入する。この中でも、(1)(5)(6)(7)はペア数が1桁程度しか見つからないような状況で効果的である。

(1) 最低ペア数

もし2ペア以下しか見つからない場合は、第一画像候補を棄却する。

(2) 大きさ整合性

それぞれの SURF キーはもともとシーン中の同じ部分に相当しているはずであるので、もし2枚の撮影カメラが同じで撮影位置も同じなら、対となる SURF キーはすべて同じ大きさのはずである。カメラが同じでも撮影位置が異なると、物体までの距離に反比例してその見かけの大きさは変化することになる。実際には、歩行者視点カメラは前方に向けられ、通常は歩行方向の正面に物体が来ることは少ないので、結果として対象物は比較的遠方に位置することが多くなる。このことから、ここでは近似として、SURF キーの大きさの正規化をその画像中に存在する SURF キーの平均の大きさで行う。この正規化のあとでペアを成す SURF キーどうしのサイズが一致しているかどうかを

確認する。画像Pについて、正規化された SURF キーの大きさ s'_{ip} はもとのサイズ s_{ip} とその画像P中の SURF キーの大きさの平均 \bar{s}_p を用いて、

$$s'_{ip} = s_{ip} / \bar{s}_p$$

となる。2枚の画QとR*の間の大きさ相違度 E_{size} を次式によって定義する。この値が0.0のとき最も整合が取れていることになる。

$$E_{size} = \frac{1}{K} \sum_{1 \leq i \leq K} |s'_{iq} - s'_{ir*}|$$

(3) 方向整合性

SURF キーペアのそれぞれの SURF キーの方向は、カメラのシーンに対する光軸周りの角度が同一であれば、すべて同じはずである。実際には歩行者視点カメラは上体に対してしっかり装着できないこともあり、カメラの水平は必ずしも保たれていない。そのため、SURF キーの方向についても、画像中の SURF キーの平均をとり、そこからの相対角度を求めることで方向整合性の検定を行う。方向相違度 E_{dir} は次式によって与えられる。ただし、 \bar{d}_p は画像Pに含まれる SURF キーの方向の平均である。

$$E_{dir} = \frac{1}{K} \sum_{1 \leq i \leq K} |d'_{iq} - d'_{ir*}|$$

$$d'_{ip} = d_{ip} - \bar{d}_p$$

(4) 2次元アフィン拘束

通常、対応点はシーン中の剛体上で発見されると考えられるので、SURF キーペアの画像中の点群はそれぞれエピポーラ拘束に本来従う。しかしながら、 $\langle a \rangle$ ペア数Kは小さな値になりうること、 $\langle b \rangle$ 対応として正しくないペアが見つかること、 $\langle c \rangle$ 参照画像を撮影したカメラと歩行者視点カメラが同じでないなど内部パラメータが未知の状況も想定されること、などからエピポーラ拘束をここで直接用いるのは困難である。そこで、本手法では、それを緩和した2次元アフィン拘束を考える。2次元アフィン行列Hは3ペアあれば求めることができるので、ペア数が小さい時にも有効である。

ここでは、2次元アフィン相違度 E_{affine} を SURF キーどうしの再投影誤差によって以下のように与える。ここで、 x_{ip} を画像P中の SURF キーiの位置とする。

$$E_{affine} = \frac{1}{K} \sum_{1 \leq i \leq K} |H x_{iq} - x_{ir*}|$$

(5) 被覆領域

SURF ペアを校正する SURF キーが一部に集中している時、それらの点群はシーン中の非常に小さい領域にしか対応していないことになる。このような場合、例えばあちこちに出てくる看板や類似したテキストに反応している可能性が高く、第一画像候補は検索結果として不適切である可能性が高くなる。このことか

ら、被覆領域の大きさを表す指標 E_{area} を用意し、これが小さい場合は第一画像候補を棄却する。ここで、関数 $area(\{x_{ip}\})$ は SURF キー集合 $\{x_{ip}\}$ が被覆する凸多角形の面積を求めるものとする。

$$E_{area} = \min(area(\{x_{iq}\}), area(\{x_{ir*}\}))$$

(6) 2次元アフィン行列の整合性

ペア数が少ない時、2次元アフィン拘束(4)で求めた2次元アフィン行列Hに関して、さらに分析する。いま、ある画像の直交軸の単位ベクトル u と v がHによって u' と v' に写像されるとする。歩行者視点カメラの装着に関して、しっかりと上体に装着されていないとはいえ、2枚の画像の撮影において、カメラの光軸が 90° 以上回転するというのは考えにくいし、画像どうしが鏡像反転することも生じえない。このことは、 $Hu \cdot u \geq 0.0$ (v についても同様)を検証することで調べることができる。結果が負であった場合、第一画像候補を棄却する。

(7) 三角形領域の方向整合性

表1からもわかるように、入力画像に対して、しばしば3ペアしか見つからない場合がある。このため、3ペアだけしかない状態でも検索結果の正当性を検証できることが望ましい。ここでは、各画像上でその3点から三角形を構成し、対応する辺どうしが 90° 以上回転していないことを検証する。1辺でも違反した場合は、第一画像候補を棄却する。この条件は、(4)(6)が効かない場合にも発火することに注意する。

5. 実験

検証実験として、東京駅八重洲地下中央口付近から地下街を約300m歩き、そこで地上に出てほどなく高島屋に到達、そのあと東京駅八重洲北口近辺まで戻る地上経路約600mの合計約900mを外出計画における経路として設定した。経路図を図1に示す。図中の左端が東京駅であり、図中では概ね反時計回りに歩行することになる。最初の赤色の矢印は地下経路に相当し、引き続く次の緑色の矢印が地上経路に相当する。地下経路はショッピング街であり、多数の店が並んでいる。地上へはエスカレータで登る。北東に向かう経路は片側2車線の大通りであり、図中右上が高島屋の正面入り口に相当する。それに対して北西に向かう通りは狭く、かつ人の往来が比較的多い。

撮影には、図1で例示したカメラではなく、パナソニック DMC-FX37 ないしアップル社 iPod nano を用いた。

参照画像集合は、晴れた冬の平日の12:15から、DMC-FX37で12分かけて経路を歩行することで撮影した。焦点距離は35mm換算で28mmである。撮影時はVGAの動画であり、これを3.0fpsでフレームに分

解すると共に QVGA のサイズにして参照画像とした。参照画像数 N は 2,100 である。これは、経路に沿って最大でも 0.5m の間隔で撮影したことに相当する。参照画像では平均で 333.1 個の SURF キーが抽出された。SURF キーの記述には 64 次元を用い、KD ツリー検索を行うため、本手法を試作したプログラムは実行時に 300 MB 以下程度のメモリを使用する。

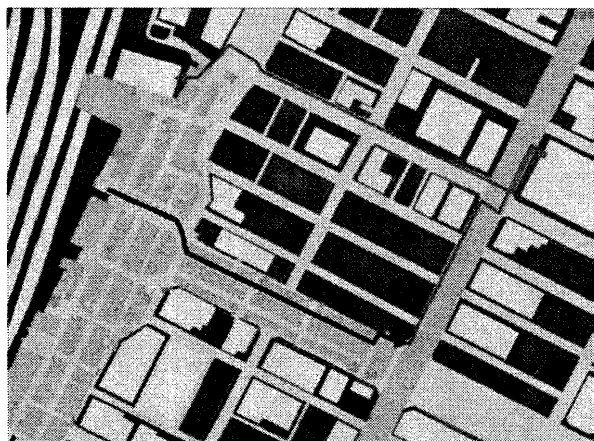


図 1: 実証歩行経路

4 回の歩行者位置推定実験結果を表 1 に示す。ビデオ 1 は参照画像獲得歩行の直後に同じカメラで撮影した。ビデオ 2 は参照画像獲得日に対して 4 7 日前の年末の平日午後のものであり、ビデオ 3 は参照画像獲得日 6 日後の午後に撮影した。ビデオ 4 は参照画像獲得日と同日であるが、iPoD nano で撮影したものである。ここでは検証のため、11 分～16 分程度のこれらの歩行に対して、30fps で撮影した映像の全フレームを QVGA にして入力画像とした。歩行者位置推定のパラメータとして、 E_{size} に 0.2、 E_{dir} に 45.0° 、 E_{affine} には 10.0 画素、 E_{area} には 50.0 画素を閾値として設けた。この結果、第一画像候補が受理されたのは 13.8% から 27.5% となった。これは、平均すると 4～8 フレームごとに有効回答が得られることに相当する。なお、実行時間は Intel Core2Duo U9400 (1.4GHz) を搭載した実際のノート PC で計測した。

図 2、図 3 は、参照画像と入力画像の撮影位置を歩行開始位置からの線距離で表した結果である。図中で受理を濃点（赤色）で、不受理を薄点（水色）で示す。不受理の場合はほとんどが不正解といえる状況に対して、受理時には誤位置推定が少ないことがわかる。図 4 では、ビデオ 1 において、比較として単純に第一画像候補を受理する方法との相違を示す。ただし、受理率が同等になるように、ペア数に閾値を設けている。図中で pairs と表記された点が比較手法による推定結果であり、対角線から外れた誤推定と見なされる領域に散在していることがわかる。

表 1: 実験結果

ビデオ番号	1	2	3	4
入力画像数	21030	28260	24060	25110
受理画像数	5782	3911	5854	4151
受理率 [%]	27.5	13.8	24.3	16.5
# 平均キー数	344.6	297.8	366.2	229.6
SURF 抽出 [msec]	126.3	119.2	137.0	132.4
検索 [msec]	230.8	211.8	256.9	220.9
0-1 ペア [%]	0.2	0.1	0.1	1.3
2 ペア [%]	4.5	12.2	12.2	23.2
3 ペア [%]	20.2	36.1	36.1	26.6
4 pairs [%]	16.7	18.4	18.4	10.7
5 pairs	10.5	10.1	10.1	6.3
6-10 pairs	25.7	17.0	17.0	16.1
11- pairs	22.3	6.3	6.3	15.9

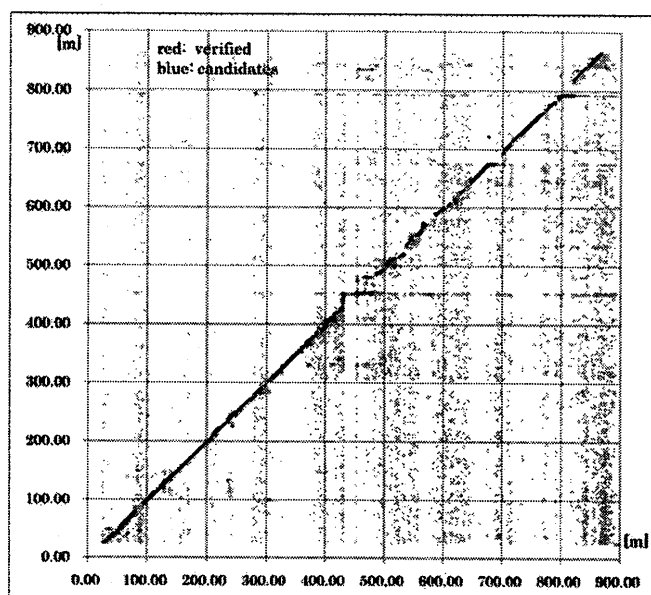


図 2: ビデオ 1 の実験結果

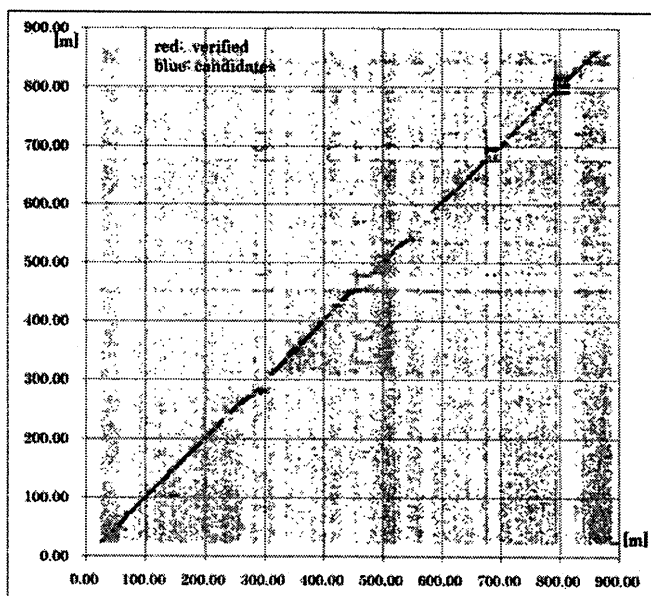


図 3: ビデオ 2 の実験結果

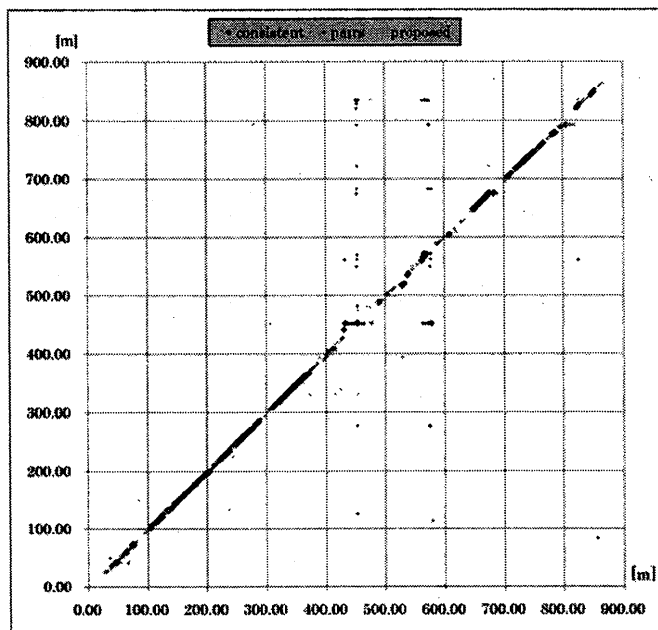


図 4: ビデオ 1 における比較手法との比較



図 5: 画像照合の様子

図 5 に画像照合の様子を示す。第一列と第三列が入力画像、第二列と第四列が第一画像候補である。第一画像候補上に太線が引かれている場合、それらは棄却されたことを示す。

6. おわりに

本稿では、比較的低い位置に取り付けられた歩行者視点カメラから歩行者位置を推定する方法を提案した。

提案手法は、歩行者視点カメラからの映像では SURF にもとづく対応ペアが多くは見つからないということから、少ないペア数でも第一画像候補を検定で

きるようにすることで、位置推定精度を保ちつつ誤推定を低減させることに成功した。

謝辞

本研究の一部は、厚生労働省の障害者保健福祉推進事業（障害者自立支援機器等研究開発プロジェクト）で採択された「画像・GPS 等のセンサ統合による日常利用可能な屋内外視覚障害者歩行支援システムの開発」（研究代表者：静岡県立大学 石川准）の成果である。

文 献

- [1] “GPS ナビ,” 株式会社エクストラ, 2009.
<http://www.extra.co.jp/gpsnavi.html>
- [2] “国土交通省自立移動支援プロジェクト,” 委員会等資料, 第 7 回推進委員会, 資料-2, 2010.
http://www.mlit.go.jp/seisakutokatsu/jiritsu/seisakutokatsu_jiritsu_tk_000006.html
- [3] “国土交通省自立移動支援プロジェクト,” 委員会等資料, 第 6 回推進委員会, 資料-2, 2009.
http://www.mlit.go.jp/seisakutokatsu/jiritsu/seisakutokatsu_jiritsu_tk_000006.html
- [4] 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.
- [5] N. Nourani-Vatani, J. Roberts, and M. V. Srinivasan, “IMU Aided 3D Visual Odometry for Car-Like Vehicles,” Australasian Conference on Robotics and Automation, 2008.
- [6] O. Pink, F. Moosmann and A. Bachmann, “Visual Features for Vehicle Localization and EgoMotion Estimation,” Intelligent Vehicles Symposium, pp.254-260, 2009.
- [7] 野田雅文, 高橋友和, 出口大輔, 井手一郎, 村瀬洋, 小島祥子, 内藤貴志, “空撮画像と時系列車載カメラ画像の照合による自転車位置推定,” 信学技報, vol.109, no.306, PRMU2009-118, pp.177-182, 2009.
- [8] J. Hays and A. A. Efros, “IM2GPS: Estimating Geographic Information from a Single Image,” pp.1-8, CVPR, 2008.
- [9] 小野善太郎, 堀田政二, 馮堯楷, 内田誠一, “局所特徴量を用いた事例参照に基づく位置同定,” 信学技報, vol.109, no.418, PRMU2009-212, pp.31-36, 2010.
- [10] 畑田晃希, 河治寿都, 山崎俊彦, 相澤清晴, “全方位カメラによる参照画像群を用いた屋内位置推定 - デジタルミュージアムでの鑑賞者の行動記録に向けて -,” 信学技報 vol.109, no.466, MVE2009-141, pp.57-62, 2010.
- [11] 久徳遙矢, 出口大輔, 高橋友和, 日加田慶人, 井手一郎, 村瀬洋, “過去の車載カメラ映像との道路面差分による不特定障害物の検出,” 信学技報, vol.109, no.470, PRMU2009-273, pp.235-240, 2010.
- [12] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, “SURF: Speeded Up Robust Features,” Computer Vision and Image Understanding, vol.110, no.3, pp. 346-359, 2008.