撮影環境の異なる画像群間の類似画像検索による歩行者位置推定

釜坂 一歩‡　北原 格†　亀田 能成†

†‡筑波大学 〒305-8573 茨城県つくば市天王台 1-1-1
E-mail: ‡s1620768@tsukuba.ac.jp, †{kitahara, kameda}@iit.tsukuba.ac.jp

あらまし　我々が研究している視覚障害者を対象とした歩行ナビゲーションシステムでは，経路に沿って事前に撮影された映像をデータベースとする類似画像検索によって歩行者の位置を推定する必要がある．類似画像検索には局所画像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が，画像の撮影された時間や天候が異なる場合に検索の精度が低下するという問題がある．この問題に対し，本稿では撮影時間帯による日照条件の変化を対象として，複数の環境下で取得された映像を組み合わせたデータベースを作成し，環境の変化に頑健な歩行者位置推定手法を提案する．

キーワード　位置推定，画像検索，歩行者ナビゲーション，局所画像特徴量，単眼カメラ

1. はじめに

カメラのみを利用したナビゲーションシステムの実現には，事前に経路に沿った映像を撮影し，生成されたデータベースを利用した位置推定が必要不可欠な技術である．実現を困難としている問題として，データベースを生成する経路映像と，位置推定を行う問い合わせ画像の撮影環境が大幅に異なる場合に，推定精度が低下することが挙げられる．この問題を解決するため，撮影時間や季節などの環境変化にロバストな手法が提案されている[1, 2]．これらの手法では車両やロボットの位置推定を目的としているが，我々は視覚障害者のナビゲーションを目的として，カメラのみを用いて環境の変化にロバストな歩行者位置推定に取り組んでいる．

視覚障害者が外出する際には多くの場合介助者を伴うため，人的あるいは経済的な負担が大きいという問題がある．我々のシステムでは経路に沿った映像を事前に撮影するだけで良いため，少ない負担で視覚障害者の単独外出を支援することができる．カメラのみを用いた歩行者ナビゲーション手法として，亀田ら[8]は類似画像検索による位置推定手法を提案している．この手法では，経路に沿った事前撮影映像から生成されたデータベースに，歩行者の取得した画像を問い合わせ，検索結果画像の撮影位置を推定位置とする．類似画像検索には SIFT[9]局所特徴量のマッチングを用いた一般的な画像認識手法を用いている．しかし，データベース画像と問い合わせ画像が撮影された時間帯や天候が異なる場合，特徴量マッチングの精度低下に伴い，位置推定の精度も低下する問題がある．ナビゲーションを行う環境と同じ条件下で撮影された映像を用意することは現実的でないため，環境の変化に対するロバスト性が要求される．そこで，我々は複数の異なる環境下で撮影された映像から生成されるデータベースを統合し，問い合わせ画像の撮影環境にロバストな位置推定を実現する．

2 章では関連研究として，視覚障害者を対象としたナビゲーションシステムと，撮影環境の異なる画像シーケンスを用いた位置推定手法について述べる．3 章では類似画像検索を用いた歩行者位置推定手法の概要と，撮影環境が異なる場合の問題点について述べる．4 章では撮影環境の変化にロバストな位置推定として，撮影環境が異なる複数のデータベースを統合する手法について述べる．5 章では，異なる時間帯に撮影された屋内外を含む経路映像により統合データベースの有効性を検証し，最後に 6 章でまとめを述べる．

2. 関連研究

視覚障害者にとってナビゲーションシステムは，単独外出の有用な支援となる[3]．普及しているナビゲーションシステムの多くは GPS による測位に基づいており，視覚障害者を対象としたシステムもすでに商用化されている[4]．しかし，GPS は屋内や地下街といった電波の届かない場所では精度が低下するという問題がある．視覚障害者に対しては屋外のみならず，屋内でもナビゲーションが可能であることが望ましい．GPS 以外の歩行者位置を測位する手法として，経路に沿って配置された RFID（radio frequency identifier）タグを，単一[5, 6]あるいは GPS と併用[7]したナビゲーションシステムが提案されている．山下ら[7]は RFID タグと準天頂衛星を併用することで，屋外での測位精度を向上させ,屋内でのナビゲーションを可能とした．しかし，屋内で精度の高い位置推定を行うためには RFID タグの敷設が必要であるため,人口の少ない地域や使用頻度の低い施設への普及は困難である．我々の手法ではカメラのみを用いるため，屋内外を問わず利用可能である．

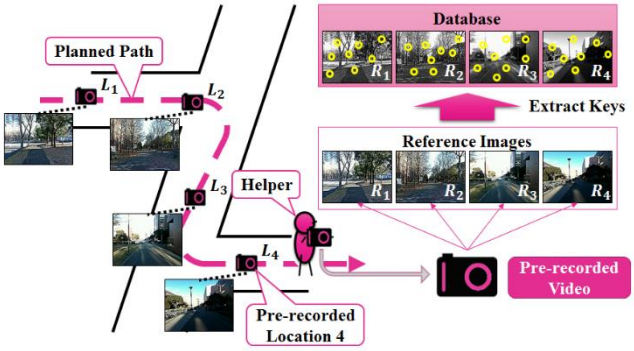

図 1　事前撮影映像からのデータベース生成[10]

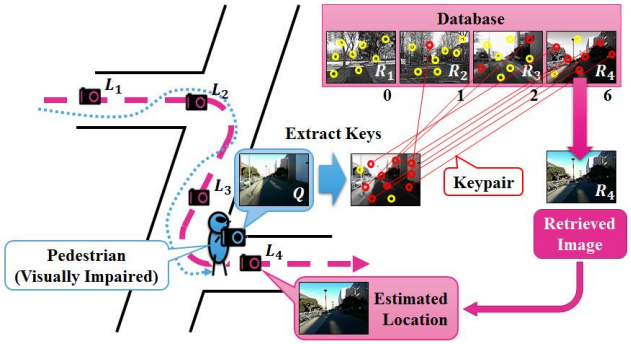

図 2　局所特徴量マッチングを利用した画像検索と歩行者位置推定[10]

撮影環境の異なる画像を用いた位置推定に関する研究は盛んに行われている．Milford らは連続画像により局所特徴量を用いない画像場所認識手法である SeqSLAM[1]を提案した．SeqSLAM では縮小した画像をパッチごとに正規化し，画像間の SAD(Sum of Absolute Difference)を計算することで，環境の変化に頑健な比較を行う．また，問い合わせ画像に最も類似する単一の画像をデータベースから検索するのではなく，過去数フレームの問い合わせ画像群との SAD の和が最小となる画像列を検索するため，一時的なカメラの遮蔽などの外乱に対するロバスト性が高い．しかし，データベースを構成する事前撮影映像と，問い合わせ画像を取得するカメラが，経路の同一部分では近い速度で移動することを前提としている．歩行ナビゲーションの際には歩行速度の個人差や，信号や横断歩道などの影響により，移動速度の前提を常に満たすことは困難である．

Naseer ら[2]は格子状のパッチに区切られた領域から抽出された HOG 特徴量を利用して画像間の相違度を計算し，その相違度をエッジのコストとしたグラフを構築した．このグラフの最小費用フロー問題を解くことによって画像シーケンスの最適な対応付けを行う．HOG 特徴量は画像の輝度変化に対し頑健である一方で，カメラの姿勢が異なる場合に画像の特徴量記述が大きく異なる．車線のように経路や進行方向が定まっている場合には問題無いが，歩行時には同じ経路でも位置や体の向きが異なる場合があり，歩行者ナビゲーションのための位置推定手法として適していない．

単一のデータベースを用いた歩行者の位置推定は[8, 10]で実現している．本稿では環境の変化にロバストにするため，複数の異なる環境下で撮影された経路映像から生成されたデータベースを統合する手法を提案する．近い地点で撮影された見た目の類似する画像が複数データベースに登録されると，類似画像検索の精度が低下する可能性がある．データベースを統合した際に類似画像をデータベースから除去することで検索の精度を向上させる．

3. 類似画像検索を用いた歩行者位置推定手法

亀田ら[7]の提案するカメラのみを用いた歩行者位置推定手法について説明し，データベース画像と問い合わせ画像の撮影環境が異なる場合の問題について述べる．

3.1. 歩行者位置推定手法の概要

視覚障害者の外出形態の特性からナビゲーションを行う経路が既知であることを前提とする．図 1 に示すように経路に沿って撮影された映像からフレームを抽出し，各フレームから検出された局所特徴量をデータベースに登録する．実際の環境で撮影される映像から，歩行者位置推定に適したフレームを選択する手法を[10]で提案した．本稿で扱うデータベースは[10]の手法を用いて生成する．

図 2 は生成されたデータベースを用いた歩行者位置推定の様子を示す．歩行者に備え付けられたカメラから一定の間隔で撮影された画像をデータベースに問い合わせ，最も類似する画像を検索する．検索結果画像の撮影された位置が経路上における推定位置となる．類似画像の検索には SIFT 特徴量のマッチングを用いる．問い合わせ画像から検出されたすべての SIFT キーについて，データベースから最も類似度の高い最尤 SIFT キーを探索し，その対をキーペアと呼ぶ．キーペアが最も多く検出されたデータベース画像を検索結果画像とする．図 2 では問い合わせ画像 Q に対する検索結果画像はR_4であり，推定位置はL_4となる．

3.2. 撮影環境の違いによる推定精度の検証

類似画像検索に用いる SIFT 特徴量は輝度変化に頑健な特徴量ではあるが，撮影時間帯や天候といった日照条件が大きく異なる場合にマッチングの精度が低下する．位置推定の精度は局所特徴量のマッチング精度

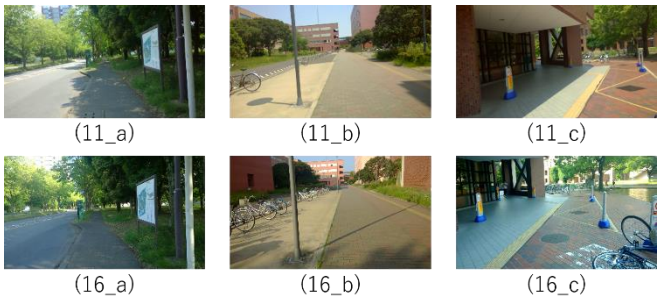

図 3　予備実験経路の一部
(上：11 時撮影映像，下：16 時撮影映像，a-c はそれぞれ類似地点で撮影された画像)

に依存しているため，撮影環境の変化によって位置推定の精度が低下すると考えられる．

撮影時間の違いによる位置推定精度への影響を調べるための予備実験を行った．2016 年 5 月 23 日の 9 時から 18 時までの 1 時間ごとに筑波大学構内の晴天下の屋外で，徒歩 4 分程度の経路映像を 2 本ずつ撮影した．経路の一部を図 3 に示す．T 時の 1 本目の映像からデータベースDB_Tを生成し，2 本目の映像を 3fps の時間間隔で抽出したフレームを問い合わせ画像群Q_Tとする．DB_{11}に対するQ_{11}及びQ_{16}の位置推定結果を図 4 及び図 5 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．キーペア数の少ない検索結果は信頼度が低いため閾値処理を行う．予備実験ではキーペア数が 5 未満の検索結果を棄却した．撮影時間帯が異なるQ_{16}を問い合わせたとき，得られるキーペアが減少するため，閾値処理により位置推定結果が得られるフレームが減少することがわかる．Q_qに含まれる全画像についてDB_dに対する類似画像検索を行ったとき，検索結果画像のキーペア数の平均を位置推定の精度を表す指標$S_{Q_qDB_d}$と定義する．従来手法で用いている単一データベースのロバスト性を検証するため，$S_{Q_qDB_d}(q = 9 \ldots 18, d = 9,12,15,18)$を求めた．図 6 は横軸が問い合わせ画像の撮影時間q，縦軸が平均キーペア数$S_{Q_qDB_d}$である．$q = d$となる撮影時間が同じ場合に$S_{Q_qDB_d}$は最大となり，撮影時間が離れるに従い小さくなる．実験結果から，晴れた日では撮影時間帯が約 2 時間以上異なると位置推定の精度が大幅に低下すると考えられる．複数の異なる環境下で撮影された映像から生成したデータベースを統合することで，環境の変化に頑健な歩行者位置推定を行う．

4. データベースの統合手法

撮影時間や天候などの環境変化に対するロバスト性を高めるため，複数の環境下で撮影された経路映像から統合データベースを生成する手法について述べる．

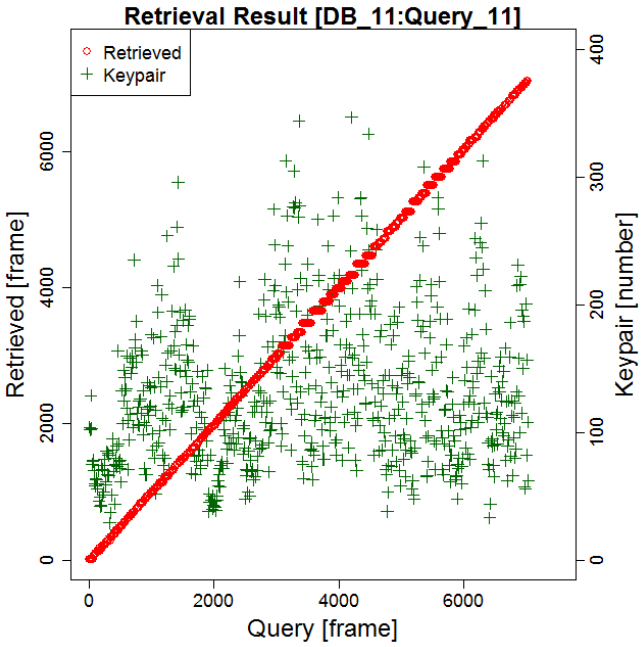

図 4　同一時間帯に撮影された映像間の位置推定 (DB_{11}，Q_{11})

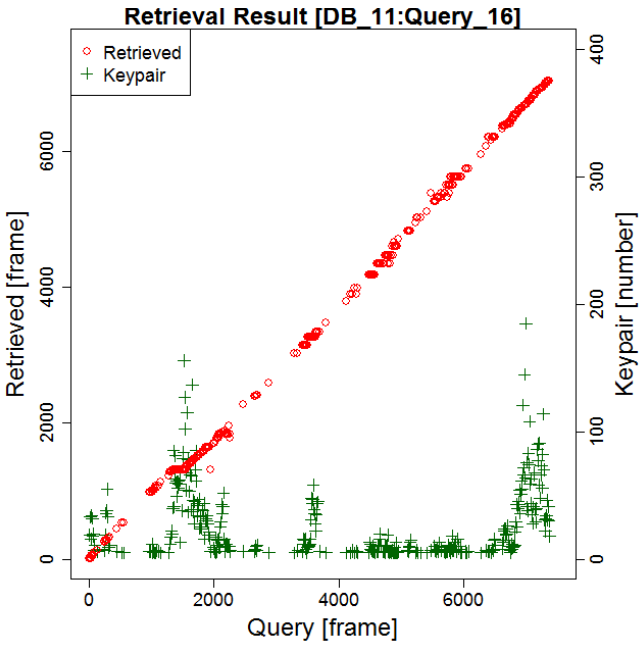

図 5　異なる時間帯に撮影された映像間の位置推定(DB_{11}，Q_{16})

4.1. 統合データベース

予備実験から，晴れた日の映像は撮影時間が 1 時間程度異なる場合でも位置推定が可能であることが分かった．しかし，予備実験のように時間ごとの経路映像を取得し，ナビゲーションを行う環境に合わせてデータベースを選択することは困難であると考えられる．そこで，ナビゲーション時の環境に最も近いと考えら

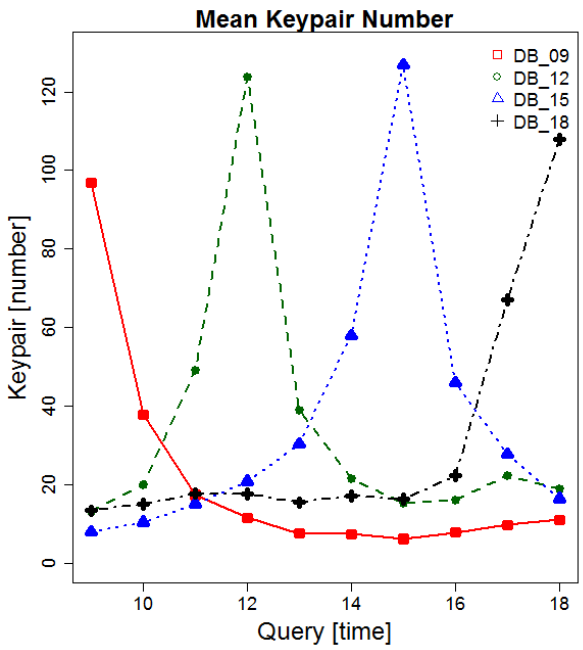

図 6　データベースと問い合わせ画像群の撮影時間帯別の平均キーペア数

れるメインデータベースと，撮影環境の異なるサブデータベースを組み合わせて利用することで，時間帯ごとの経路映像を用意する必要なしに，精度の高い位置推定を実現する．

メインデータベースとサブデータベースに含まれる画像を統合し，統合データベースを生成する．メインデータベースとサブデータベースを生成する経路映像が同じ移動速度で撮影された場合，検索結果画像のフレーム番号を経路上における推定位置として統合できる．しかし，実際の撮影状況で一定の速度で移動することは困難であり，また，信号などで停止することも考えられる．我々は，これらの状況を考慮したデータベース生成手法を[10]で提案している．しかしこの手法では，データベースに登録する画像が一定のフレーム間隔で選択されないため，検索結果画像のフレーム番号を単純に推定位置とすると，検索結果画像の含まれるデータベースによって推定位置が前後してしまう問題がある．そこで，サブデータベースに含まれる全画像とメインデータベースに含まれる画像の対応付けを行う．サブデータベースに含まれる画像が検索結果として選択された場合，その画像に対応するメインデータベースの画像を求め，メインデータベースの撮影経路上における推定位置を得る．予備実験において，撮影時間の違いによる見た目の変化が小さく，正確な画像の検索が可能である場合を確認した．図 7 に示す例のように，常に日陰である地点や，画像全体の輝度が変化する場合である．このような画像を共にデータベースに登録すると，キーペアが分散し検索の精度が

図 7　異なる撮影時間でも見た目の変化が小さい地点の例(左赤枠：16 時撮影画像，右黄枠：11 時撮影画像，緑線：キーペア)

低下する可能性がある．そこで，撮影時間が異なっていても正しい検索が可能な画像については，サブデータベースの画像を統合データベースに登録しないことで精度の向上を図る．

4.2. 撮影環境の異なる映像間の対応付け

メインデータベースとサブデータベースに含まれる画像間の対応付け手法について述べる．事前撮影はナビゲーションを行う経路に沿って撮影された映像であるため，撮影開始地点と終了地点は一致している．したがって，移動速度が完全に同じかつ，データベースに登録する画像を一定の時間間隔で選択していれば，単純にフレーム番号による対応付けが可能である．しかし，[10]による実際の撮影状況に即したフレーム選択では，ある 2 地点間のフレーム枚数がデータベースごとに異なる可能性がある．そこで，我々が位置推定に利用している類似画像検索を用いて画像の対応付けを行う．予備実験から撮影時間の異なる画像間の検索精度が低いことがわかっている．一方で環境の違いによる見た目の変化が小さい区間では精度の良い検索が可能である．図 6 に示すような撮影環境が異なっていても正確な検索が可能な問い合わせ画像と検索結果画像の組を画像ペアと呼ぶ．最初に画像ペアの算出方法について述べ，次に，画像ペア用いてすべてのデータベース画像の対応付けを行う方法を説明する．

図 8　実験経路の一部
(上：9 時撮影映像，中：13 時撮影映像，下：17 時撮影映像，a-c はそれぞれ類似地点で撮影された画像)

メインデータベースをDB_{Main}，構成するI枚の画像群を$\{Img_{Main}\}$，各画像を$Img_{Main.i}(i=1\ldots I)$と定義する．$SubDB$も同様に，$DB_{Sub}$，$\{Img_{Sub}\}$，$Img_{Sub.j}(j=1\ldots J)$とする．$DB_d$に対する$Img_q$の検索結果画像を$Img_{dq}$，検索の信頼度を示すキーペア数を$Key_{dq}$と表す．$Img_{Main.\hat{\imath}}$の$DB_{Sub}$に対する検索結果が$Img_{Sub.\hat{\jmath}Main.\hat{\imath}}, Key_{Sub.\hat{\jmath}Main.\hat{\imath}}$であるとき，$Img_{Main.\hat{\imath}}$と$Img_{Sub.\hat{\jmath}}$の組を画像ペアの候補とする．同様にして，$\{Img_{Sub}\}$を$DB_{Main}$に問い合わせる．このとき，$Img_{Sub.\hat{\jmath}}$の$DB_{Main}$に対する検索結果画像が$Img_{Main.\hat{\imath}Sub.\hat{\jmath}}$かつ$Key_{Sub.\hat{\jmath}Main.\hat{\imath}}$と$Key_{Main.\hat{\imath}Sub.\hat{\jmath}}$が閾値以上のとき，$Img_{Main.\hat{\imath}}$と$Img_{Sub.\hat{\jmath}}$の組を画像ペア$(\hat{\imath},\hat{\jmath})$とする．

次に，算出された画像ペアを用いてすべてのサブデータベース画像をメインデータベースの画像に対応付ける．ここで，各事前撮影映像の出発地点と到着地点が一致していると仮定し，$Img_{Main.1}$と$Img_{Sub.1}$及び$Img_{Main.I}$と$Img_{Sub.J}$の組を画像ペアとする．画像ペア$(\hat{\imath}1,\hat{\jmath}1)$と画像ペア$(\hat{\imath}2,\hat{\jmath}2)$が得られたとき，$Img_{Sub.j}(j=\hat{\jmath}1\ldots\hat{\jmath}2)$に対応する$Img_{Main.i}(i=\hat{\imath}1\ldots\hat{\imath}2)$を求める．この対応付けは線形補間で行う．これらの処理により，$\{Img_{Sub}\}$に含まれるすべての画像は$\{Img_{Main}\}$のある画像と対応付けられる．メインデータベースに複数のサブデータベースを統合する際には，すべてのサブデータベース画像について同様の処理を行う．

4.3. サブデータベースの統合と重複画像の除去

統合データベースは，$\{Img_{Main}\}$と$\{Img_{Sub}\}$及び，画像の対応情報から構成される．しかし，単に$\{Img_{Main}\}$と$\{Img_{Sub}\}$を結合するだけでは，図 7 に示すような見た目の変化が小さい画像が含まれる．類似画像がデータベースに含まれる際の問題点については[10]で述べた．この問題を解決するため，$\{Img_{Sub}\}$に含まれる画像のうち，画像ペアを構成する画像を統合データベースから除去する．すなわち，DB_{Main}とDB_{Sub}の間にP組の画像ペアが存在する場合，統合データベースを構成する画像数は$I+J-P$枚となる．

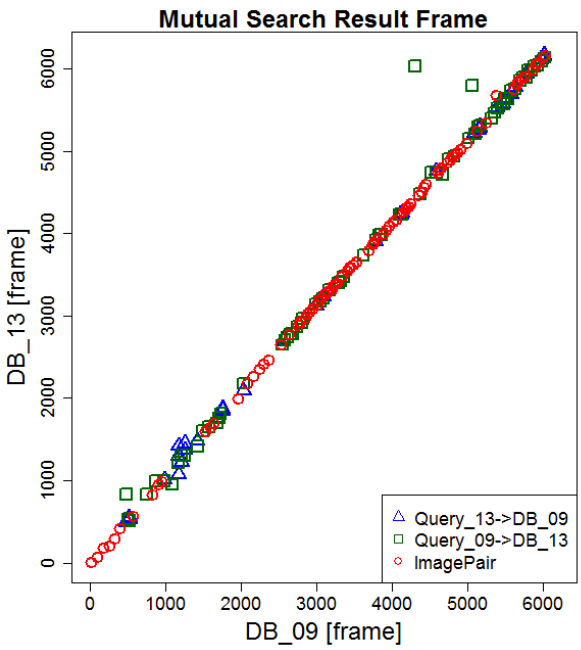

図 9　相互検索結果及び算出された画像ペア

5. 評価実験

想定する経路に沿って複数の時間帯で映像を取得し，提案手法により統合したデータベースの有効性を検証する．

5.1. 実験概要

筑波大学敷地内の屋内外を含む徒歩 3 分半程度の実験経路を設定した．2016 年 10 月 4 日の 8～17 時の 1 時間ごとに経路映像を FHD(1920x1080)で 2 本撮影した．このうち，8~14 時は晴れの映像，15~17 時は曇りの映像である．各映像を 320x180 に縮小し，3fps でサンプリングする．1 本目を事前撮影映像，2 本目をフレーム分割した画像群を問い合わせ画像群とする．9 時，13 時，17 時における経路の一部を図 8 に示す．

5.2. 実験結果

撮影時間の違いによる日照条件の影響を検証するため 9 時と 13 時の撮影映像を統合したデータベースDB_{9+13}を生成する．このときのデータベース間の画像の対応付け結果について述べる．また，統合データベースと単一データベースに対する位置推定精度の比較を示す．画像検索の際のキーペアの閾値は実験的に求め，5 に設定した．

5.2.1. データベース間の画像対応付け

前処理として 9 時と 13 時の事前撮影映像から[10]の手法により個別のデータベース，DB_9及びDB_{13}を生成した．DB_9は 191 枚，DB_{13}は 209 枚の画像より構成され

る．この二つのデータベースを統合するため，DB_9とDB_{13}の間で画像の対応付けを行う．DB_9とDB_{13}を構成する画像群をそれぞれQ_9，Q_{13}として相互検索を行った結果を図 9 に示す．縦軸横軸はそれぞれDB_9とDB_{13}を構成する画像群のフレーム番号である．緑の四角はDB_{13}に対するQ_9の検索結果，青い三角はDB_9に対するQ_{13}の検索結果を表し，互いに検索結果が一致する画像ペアを赤い丸で示す．両映像は可能な限り同速度で撮影したため，正しい検索結果は直線y = x上に並ぶ．図 9 から，一方向の検索では直線上から逸脱する誤検索が見られるが，双方向の検索結果を用いることで撮影時間の異なる画像間でも対応付けが可能である．DB_9とDB_{13}の間に算出された画像ペアは 83 組であった．

5.3. 統合データベースを用いた歩行者位置推定

まず，統合データベースによる有効性を検証するため，11 時に撮影された経路映像から 3fps の時間間隔で抽出した 613 フレームの画像を問い合わせ画像群Q_{11}とし，DB_{9+13}及びDB_9に対する位置推定を行う．統合データベースDB_{9+13}は，DB_9をメインデータベース，DB_{13}をサブデータベースとして生成した．統合データベースに対する位置推定結果を図 10，単一データベースに対する位置推定結果を図 11 に示す．どちらも一部の誤推定を除いて歩行者の位置が推定できている．また，平均のキーペア数は共に約 28 と大きな差は見られなかった．一方で 613 フレームの問い合わせ画像に対し，検索結果が棄却されず位置推定結果を得られたフレーム数は，DB_9の 489 フレームに対しDB_{9+13}では 558 フレームと，システムの返答率が向上した．

次に，サブデータベースを統合したことによる影響を検証するため，9 時に撮影された 620 フレームの画像を問い合わせ画像群とし，同様にしてDB_{9+13}及びDB_9に対する位置推定を行う．統合データベースに対する位置推定結果を図 12，単一データベースに対する位置推定結果を図 13 に示す．問い合わせ画像と同一時間帯に撮影された画像のみを含むデータベースに対し，サブデータベースを加えた統合データベースでは，検索結果画像に対するキーペア数がわずかに減少するものの，推定精度には大きく影響しないことが分かった．これらの実験結果から，メインデータベースに撮影環境の異なるサブデータベースを統合することで，同一時間帯における推定精度を低下させることなく，撮影時間の変化に対してロバストな推定が可能であることが示唆された．

6. おわりに

本稿では，局所特徴量のマッチングに依存した位置推定手法は，映像が撮影された環境が大きく異なる場合に推定精度が低下するという問題に対し，様々な環境下で撮影された映像を統合したデータベースを生成する手法を提案した．実験では特に晴天下の撮影時間の違いによる日照条件の変化を対象とし，統合したデータベースと単一データベースの位置推定精度について検証を行った．

撮影環境の異なる事前撮影映像から生成されたデータベースを統合することで，データベースを画像の撮影された環境と異なる条件下で撮影された問い合わ

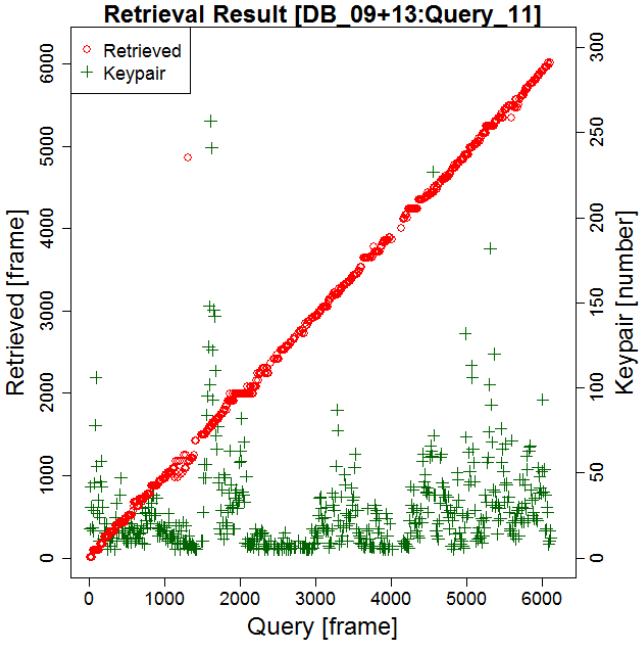

図 10　統合データベースに対する位置推定結果
(事前撮影映像と問い合わせ画像は異なる撮影時間)

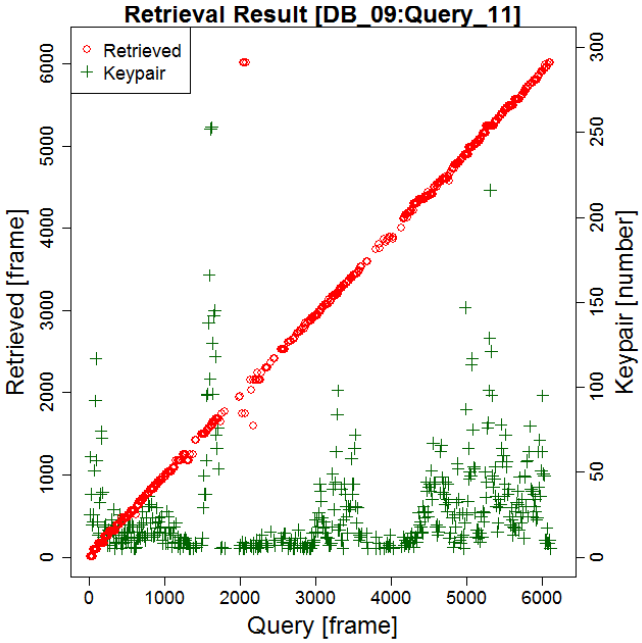

図 11　単一データベースに対する位置推定結果
(事前撮影映像と問い合わせ画像は異なる撮影時間)

せ画像に対して，位置推定の精度を低下させることなくシステムの返答率を向上させた．また，最も検索精度の良い同一時間帯の単一データベースと比較して，キーペア数はわずかに減少するものの，同等の位置推定精度を有する．実験の結果から，複数の異なる環境下で撮影された映像から個別のデータベースを生成し，それらを統合することで，環境の変化に対しロバストな位置推定が可能であることが示唆された．

今後は時間だけではなく，天候の変化についても本手法を適用することを考える．また，[1, 2]のような時系列情報を考慮した検索手法についても検討する．

本研究は JST-RISTEX の研究助成を受けた．

文　献

[1] M.J.Milford, G.F.Wyeth, "SeqSLAM: Visual Route-Based Navigation for Sunny Summer Days and Stormy Winter Nights," Proc. of the IEEE International Conference on Robotics and Automation, pp.1643-1649, 2012.

[2] T.Naseer, L.Spinello, W.Burgard, C.Stachniss, "Robust Visual Robot Localization Across Seasons using Network Flows," Proceedings of the AAAI Conference on Artificial Intelligence, pp.2564-2570, 2014.

[3] 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.

[4] 「トレッカーブリーズ」，エクストラ, 2012. http://www.extra.co.jp/breeze/

[5] H.Fernandes, V.Filipe, P.Costa, J.Barroso, "Location based services for the blind supported by RFID Technology," Procedia Computer Science, vol.27, pp.2-8, 2014.

[6] S.Chumkamon, P.Tuvaphanthaphiphat, P.Keeratiwintakorn, "A Blind Navigation System Using RFID for Indoor Environments," ECTI-CON, vol.2, pp.765-768, 2008.

[7] 山下晃弘，佐藤佳，佐藤俊太，松林勝志，「視覚障がい者ナビゲーションを目的とした RFID タグと準天頂衛星による位置測位システム」，情報処理学会研究報告, Vol.2016-AAC-1, no.2, pp.1-7, 2016.

[8] K.Yoshinari, O.Yuichi, "Image Retrieval of First Person Vision for Pedestrian Navigation in Urban Area," ICPR, pp.364-367, 2010.

[9] D.G.Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," International Journal of Computer Vision, vol.60, pp.91-110, 2004.

[10] 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.

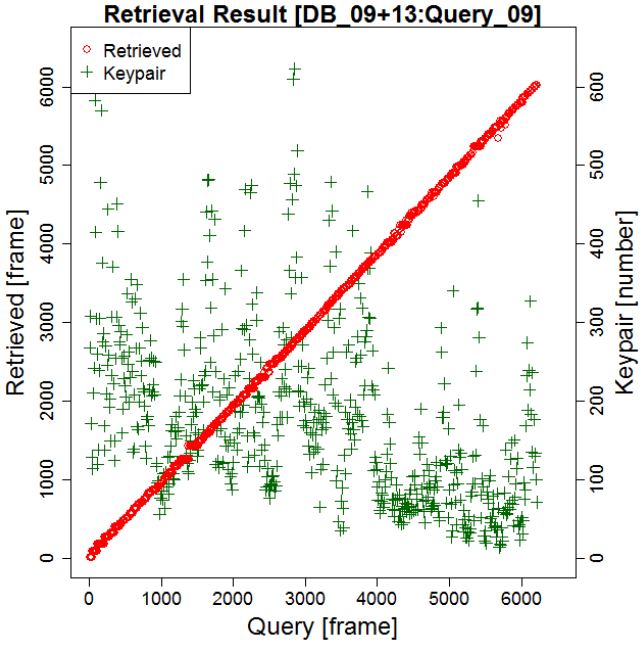

図 12　統合データベースに対する位置推定結果
(事前撮影映像と問い合わせ画像は同じ撮影時間)

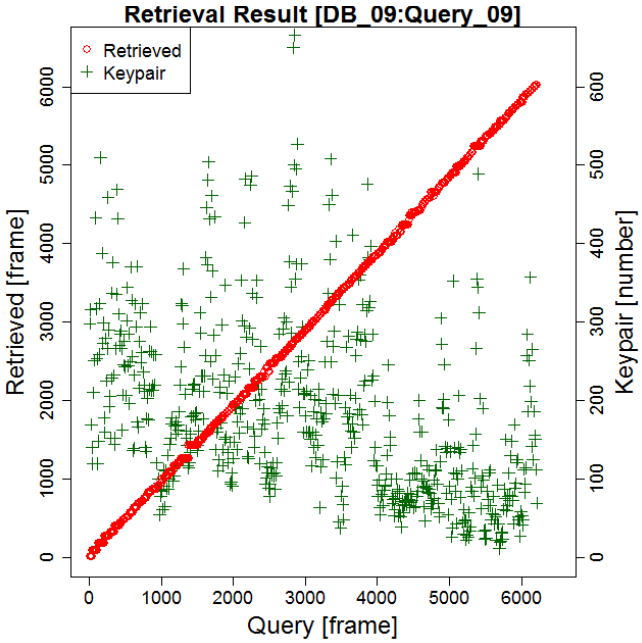

図 13　単一データベースに対する位置推定結果
(事前撮影映像と問い合わせ画像は同じ撮影時間)