

画像検索を用いた図書館内での自己位置推定

風間 光^{1,a)} 川本 一彦^{2,b)} 岡本 一志^{3,c)}

概要: 図書館における利用者の行動調査の効率化のために、ウェアラブルカメラを用いた自己位置推定法を提案する。提案手法では、撮影位置が既知な画像群の中から、ウェアラブルカメラによって得られたクエリ画像と最も類似した画像を検索し、検索結果として出力された画像の撮影位置をクエリ画像の撮影位置とする。画像間の類似度は、局所特徴点の対応点数によって計算する。このような自己位置推定は、特徴点の3次元的な配置を考えないため、最適化計算を伴う他の手法と比べて、特徴点の誤対応に対して安定した位置推定が行える。実際に大学図書館で撮影した画像を用いた実験において、最適化計算を伴う手法ではクエリ画像19枚中2枚しか撮影位置を求められなかったのに対し、提案手法を用いた場合は19枚全ての撮影位置を求めることができた。さらに、クエリ画像が時系列に沿って入力されることを仮定し、パーティクルフィルタによる局所探索を導入することで、自己位置の不自然なジャンプを抑制できることを示す。

キーワード: 自己位置推定, Structure from Motion, 画像検索, 情報探索行動

Visual localization in libraries using image retrieval

KAZAMA HIKARU^{1,a)} KAWAMOTO KAZUHIKO^{2,b)} OKAMOTO KAZUSHI^{3,c)}

Abstract: We propose a method for visual localization using an wearable camera in order to streamline the survey of Information-Use. Our method retrieves the most similar image to a query image taken by wearable camera from image database. We define the camera position of a query image as the camera position of the image outputted as search result. The similarity is computed by the number of matched local features. Unlike other methods that involve optimization, our method doesn't consider the geometry of the features and it enables stable localization. In experiments with real images taken at a library, we show that our method can localize all 19 query images while the method that involve optimization can localize only two query images. Furthermore, we assume that query images are time series data and we can suppress unusual jump of location using particle filter.

Keywords: Visual Localization, Structure from Motion, Image Retrieval, Information-Seeking Behavior

1. はじめに

図書館情報学の分野において、図書館の利用者がどのように図書館を利用しているのかを調査する「情報利用・探索行動調査」と呼ばれる調査がある [12]。これは、例えば「環境問題について調べてきなさい」と指示を受けた学生がどの書棚へ向かうのか、あるいはコンピュータを使いに行くのか、といったことを調査するものである。この調査において被験者の位置情報は非常に重要な意味を持つ。し

¹ 千葉大学大学院融合科学研究科
Graduate School of Advanced Integration Science, Chiba University

² 千葉大学総合メディア基盤センター
Institute of Media and Information Technology, Chiba University

³ 千葉大学アカデミック・リンク・センター
Academic Link Center, Chiba University

a) kazama@chiba-u.jp

b) kawa@faculty.chiba-u.jp

c) okamoto.kazushi@chiba-u.jp

しかし、現状では実験者が目視でそれを調査している。調査の効率化のためには、この利用者の位置情報の取得を何らかの方法で自動化することが好ましい。

図書館のような屋内環境では、GPS による位置情報取得は利用できない。また、図書館の各区画にあらかじめ設置した小型の電子タグを利用する自己位置推定法は、環境準備に伴う人的・経済的コストが大きくなり、導入が容易でない。

一方で、画像という情報を利用して自己位置推定を行う手法が提案されている。画像ベースの自己位置推定手法のひとつに、対象環境に設置した複数の定点カメラを用いた方法がある [8]。また、画像から環境の 3 次元復元とカメラの位置姿勢を同時に求める手法として、Visual Simultaneous Localization And Mapping (SLAM) [2] や Structure from Motion (SfM) [5] がある。しかし、図書館には背の高い棚が数多く並んでいるため、死角の無いように定点カメラを設置するには、非常に多くのカメラが必要になってしまう。Visual SLAM では被験者ごとに異なる環境地図が作成されてしまうため、複数人の被験者による調査結果の比較が難しい。SfM のような最適化計算を含む手法は、ピクセル精度での撮影位置推定を行うことが可能だが、自然特徴点の誤対応などのノイズに影響を受けやすく、推定結果が不安定になりやすいということが報告されている [13]。特に図書館では利用者による本の移動が頻繁に起こるため、自然特徴点の配置の幾何学的な整合性が取れず、復元に失敗する可能性がある。情報利用・探索行動調査のための位置推定であればピクセル精度での位置推定を行う必要は無く、1メートル程度の誤差までは許容範囲内としてよい。

本研究では、画像ベースの自己位置推定技術を用いて図書館情報学における調査の支援を行う。背の高い棚の間でも位置推定ができるように、被験者にウェアラブルカメラを装着させ、そのカメラから得られた画像の撮影位置を推定する。利用者による本の移動を考え、自然特徴点の配置の変化に頑健な画像検索技術を用いた手法を提案する。本手法は大きく分けて、環境地図の作成と自己位置推定の 2 段階に分けられる。

環境地図は、対象環境を撮影した画像群と、対象空間を復元した 3 次元点群で構成される。各画像はその撮影位置を保持しており、各 3 次元点はそれがどの画像から復元されたかという情報と、その画像上での局所特徴量ベクトルを保持している。このような環境地図を SfM を用いて事前に作成する。このとき、図書館内で撮影を行う事ができる時間は限られているため、効率よく撮影するために、全方位カメラを使用する。環境地図の概念図を図 1 に示す。

被験者に装着したウェアラブルカメラが撮影する画像をクエリ画像と呼ぶ。作成した環境地図からクエリ画像と最も類似した画像を検索し、ヒットした画像の撮影位置をその時点での被験者の位置とすることで位置推定を行う。類似度の計算には、局所特徴点の対応点数を使用する。SfM

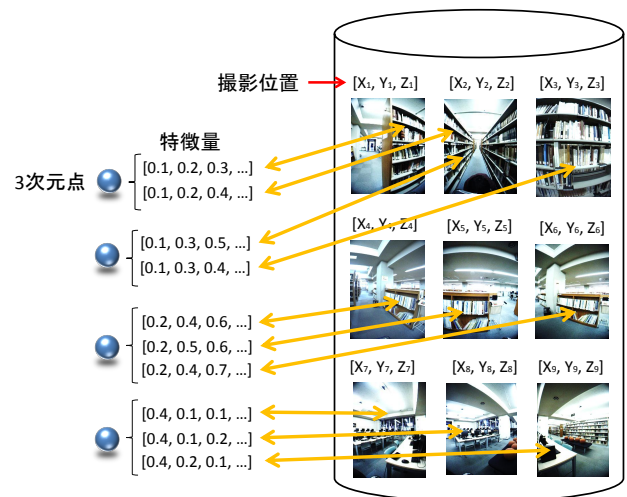


図 1 環境地図の概念図。環境地図は、対象環境を撮影した画像群と、対象空間を復元した 3 次元点群で構成される。各画像はその撮影位置情報を保持している。各 3 次元点は、どの画像から復元されたかという情報と、その画像上での局所特徴ベクトルを保持している。

等とは異なり、本手法で用いる画像検索技術は特徴点の配置を考慮せず最適化計算も含まないため、本の移動や画質の違いなどによる特徴点誤対応に対して頑健である。さらに、クエリ画像間の時間的連続性を利用し、カメラの撮影位置に対してパーティクルフィルタによるフィルタリングを施すことによって、被験者の位置が突然遠くにジャンプしたような結果が出る事を防ぐ。

本論文の構成は以下の通りである。2. では、自己位置推定のための環境地図の作成について述べる。3. では、作成した環境地図からクエリ画像と類似した画像を検索する方法を説明する。また、パーティクルフィルタを用いた誤推定の抑制について述べ、そのための状態空間や状態遷移モデル、尤度関数の定義を行う。4. では、提案手法によって、SfM による手法よりも安定した自己位置推定を行う事ができることを、実験により示す。

2. SfM に基づく環境地図の作成

ここでは、自己位置推定に用いる環境地図を作成する方法について説明する。本研究では、Structure from Motion を Snavely ら [5] の手法で解くことで、環境地図を作成する。

2.1 特徴点の検出と対応付け

まず、空間を撮影した画像から局所特徴点を検出する。局所特徴点の検出と記述には、Scale Invariant Feature Transform (SIFT) 法 [3][9] を用いる。SIFT は画像の回転・スケール変化・照明変化等に頑健な特徴点の検出・記述方法であり、その頑健性のために、大量画像からの 3 次元復元を行うシステムの多くは SIFT を採用している [10]。ある特徴点は、その特徴ベクトルと最もユークリッド距離の小

さい特徴ベクトルを持つ他の特徴点と対応付けることができる。各画像から SIFT 特徴点を検出した後、各画像ペアに対して特徴点の対応付けを行う。このとき、画像上の全特徴点を線形探索すると非効率的であるため、特徴ベクトルをノードに持つ k-d 木を各画像ごとに構築し、それに対して近似最近傍探索を行うことで、効率的に特徴点を対応づける [1]。

ここで対応づけられた特徴点は、一般に誤対応を含んでいる可能性がある。そこで、[5] と同様に、以下の方法でこれらの誤対応を取り除く。

まず、k-d 木の探索の際に最近傍点だけでなく 2 番目に近い点まで求めておき、それぞれに対する距離を d_1, d_2 としたとき、

$$\frac{d_1}{d_2} \leq T \quad (1)$$

となる場合のみ対応付ける。ここで、 T はしきい値である。この処理によって、最近傍点までの距離が 2 位以下と比べて十分小さい場合のみ対応付けることになり、対応付けの曖昧性を除去することができる。今回は $T = 0.6$ とした。

次に、Random Sample Consensus (RANSAC) のループの中で 8 点アルゴリズムを用いて各画像ペア間の基礎行列を計算し、外れ値となった対応点を除外する。この処理によって、対応付けの正当性を幾何学的な側面から検証することができる。外れ値を除外した結果対応点の数が 20 点未満になった場合、その画像間の対応付けを全て除去する。

また、ある特徴点が他の画像 1 枚の中の 2 点以上と対応付けられている場合、つまり特徴点の対応が枝分かれている場合、そこから辿れる一連の特徴点すべてを除外する。この処理によって対応付けの曖昧性が除去され、一連の特徴点が空間中の 3 次元点 1 点と対応付けられる。

2.2 空間中の 3 次元点とカメラ位置姿勢の復元

前節のマッチング処理によって、3 次元空間中のある 1 点と、それを投影した 2 次元画像中の特徴点が対応付けられた。その対応を用いて、空間の 3 次元復元とカメラパラメータの推定を行う。おおまかな流れは以下の通りである。

- (1) 適切な初期復元ペアを選択
- (2) そのペア間の相対的なカメラ位置を求める
- (3) 3 次元点の復元を行いバンドル調整 [7]
- (4) 共有視野を持つ画像を加えて復元・バンドル調整
- (5) 復元できる点が無くなるまで 4 を繰り返す

バンドル調整における非線形最適化計算の結果は初期値に大きく依存し、初期値によっては誤った局所解に陥ってしまうおそれがある。大量の画像から復元を行う際は全ての画像を同時に扱うのではなく、理想的な初期復元ペア 2 枚の復元から始め、そこに少しずつ画像を加えていく。理想的な初期復元ペアとは、特徴点の対応数が多く、なおかつカメラ間の距離が離れたペアである。



図 2 対象環境（図書館）を撮影した画像の一例。

初期復元ペアとして選ばれた 2 台のカメラパラメータを推定する。カメラパラメータは、画像間の対応点が 5 点以上あれば 5 点アルゴリズム [4] によって求められる。このときの対応点 5 点の選択を RANSAC のループの中で行う。カメラパラメータが求まれば、三角測量 (triangulation) によって両方の画像に写った 3 次元点を復元することができる。その後、今計算されたカメラパラメータと 3 次元点の座標値を初期値としてバンドル調整を行う。

次に、既に復元された 3 次元点を多く写している画像を復元画像群に追加する。復元済みの 3 次元点と追加された画像上の特徴点との対応が 6 組以上とることができれば、DLT 法を用いることで、追加した画像のカメラパラメータが線形計算で求められる。この 6 点の選択の際にも RANSAC が用いられる。またこの段階で、新しく加えられたカメラのパラメータのみを調整するバンドル調整が行われる。最後に、新しく加えられた画像上に写っている特徴点のうち、まだ復元されておらず、かつ他の復元済みカメラに写っている点を三角測量によって復元する。そして、全体に対するバンドル調整が行われ、また新しく画像を追加する。以上の処理を、復元済みの 3 次元点を 20 点以上写している画像が無くなるまで繰り返す。

3 次元復元完了後、推定された各画像の撮影位置を表す 3 次元ベクトルを k-d 木に格納することで、環境地図が構築される。実際に図書館を撮影した画像の一例を図 2 に、それらを用いて SfM を解いた結果の画像を図 3 に示す。撮影した図書館内の一区画が 3 次元復元され、撮影したカメラ位置が赤い点で示されている。

3. 画像検索による自己位置推定と時系列フィルタリング

ここでは、作成した環境地図を用いて実際に自己位置推定を行う手法について説明する。まず画像検索方法とそれに基づく自己位置推定法について述べ、次にパーティクルフィルタを用いた誤推定の抑止について説明する。

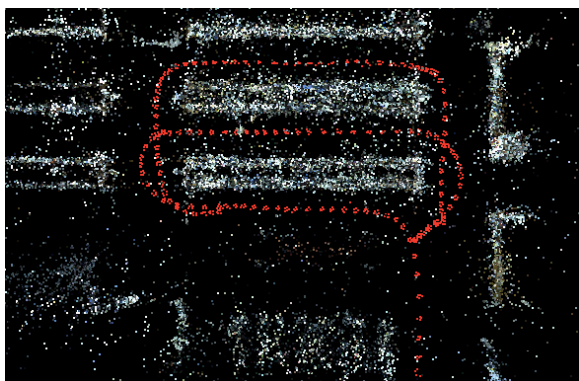
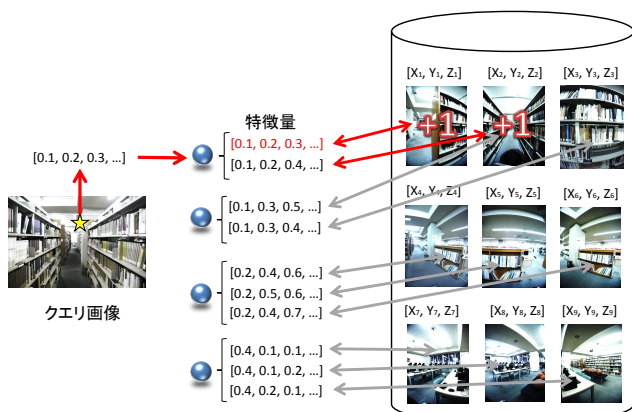
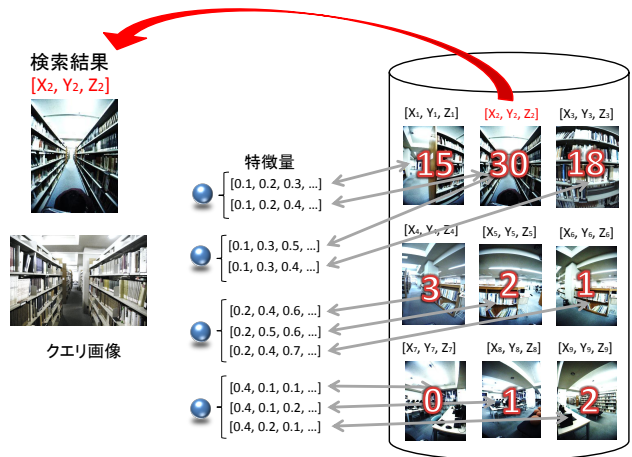


図 3 環境地図作成のための SfM の実行結果．図書館の棚を真上から見た図．赤い点はカメラの撮影位置を表す．



(a) まず，クエリ画像上の特徴点と環境地図中の 3 次元点を対応付け．次に，対応付けられた 3 次元点の復元に使用された画像全てに 1 票投票する．



(b) クエリ画像上の全ての特徴点に関して投票を行い，最終的に最も多くの票を得た画像を画像検索結果として出力する．得られた検索結果画像の撮影位置をクエリ画像の撮影位置とすることで自己位置推定を行う．

図 4 画像検索による自己位置推定の概念図．

3.1 画像検索による自己位置推定

画像検索による自己位置推定法の概要を図 4 に示す．まず，クエリ画像上の SIFT 特徴点を検出・記述し，環境地

図中の 3 次元点を持つ全ての特徴ベクトルに対して対応付けを行う．以下ではこのような「環境地図中の 3 次元点を持つ特徴ベクトルに対する対応付け」を「環境地図中の 3 次元点に対する対応付け」と略記することがある．このとき，式 (1) と同様にして誤対応を除去する．しかし，1 つの 3 次元点は，その点の復元に使用された画像枚数分の特徴ベクトルを持ち，それらの間のユークリッド距離は非常に小さいはずである．したがって，探索された最近傍特徴ベクトルと 2 番目に近いベクトルとが，同じ 3 次元点に属してしまう場合がある．この場合，対応付け自体は正しいにも関わらず，式 (1) の左辺が大きくなり，対応付けが除去されてしまう．そこで，2 番目に近い特徴ベクトルを，最近傍点の属する 3 次元点とは異なる 3 次元点から探すようにする．誤対応を除去した後，対応付けられた 3 次元点の復元に用いられた画像全てに対して 1 票を投じる．

この投票を，クエリ画像上の全特徴点に対して行い，最終的に最も多くの票を得た画像を，クエリ画像に対する類似画像検索結果として出力する．環境地図中の画像はその撮影位置を保持している．出力された画像の撮影位置をクエリ画像の撮影位置とする．

提案手法では環境地図中の画像の撮影位置が選択肢となり，その中から現在位置を選択することしかできない．したがって，環境地図作成用の画像は十分に高密度・広範囲で撮影する必要がある．一方，SfM では環境地図中の撮影位置以外の場所も求めることができるが，SfM は特徴点の誤対応のようなノイズに弱い [13]．本研究が目的としている情報利用・探索行動調査のためであればそれほど高精度の位置推定を行う必要は無いため，提案手法のような大まかな位置推定手法を用いても構わない．また，SfM による手法とは異なり，提案手法では投票の際に対応付けた特徴点が「どこにあるか」という幾何学的配置を考えず，ただ「写っているかどうか」のみを考慮する．そのため，図書館内の本の位置が移動した場合でも，安定した位置推定を行うことができる．

3.2 パーティクルフィルタによる誤推定の抑止

画像検索による位置推定を行った場合，特徴点の誤対応によって，被験者の位置が突然離れた場所にジャンプしたような誤推定が起きる可能性がある．特に図書館には，類似した見た目を持つ棚や机が，複数の離れた場所に存在することがある．そのため，クエリ画像が棚や机を多く含んでいる場合に，自己位置の誤ったジャンプが起こりうる．そのような誤推定は，クエリ画像の時系列性を利用することで抑止できるはずである．そこで，パーティクルフィルタ [11] と呼ばれる時系列フィルタを導入する．

環境地図中の画像の撮影位置の集合を $L \subset \mathbb{R}^3$ とする．また，クエリ画像全体の集合を Q とする．本研究では，位置推定対象者の位置を，環境地図中の画像の撮影位置で近

似的に表現する．したがって，対象は時刻 t において状態 $\mathbf{x}_t \in \mathbf{L}$ を持ち，観測 $\mathbf{y}_t \in \mathbf{Q}$ を得る，と定義できる．また， $\mathbf{Y}_t = \{\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_t\}$ とする．

時系列フィルタは， \mathbf{Y}_t が観測されたときの事後確率 $p(\mathbf{x}_t | \mathbf{Y}_t)$ を求めることである．この事後確率 $p(\mathbf{x}_t | \mathbf{Y}_t)$ は，ベイズの法則より

$$p(\mathbf{x}_t | \mathbf{Y}_t) = \frac{p(\mathbf{y}_t | \mathbf{x}_t)p(\mathbf{x}_t | \mathbf{Y}_{t-1})}{p(\mathbf{y}_t | \mathbf{Y}_{t-1})} \quad (2)$$

のように求められる．ここで $p(\mathbf{y}_t | \mathbf{x}_t)$ は，ある状態 \mathbf{x}_t のときに観測 \mathbf{y}_t を得る確率，すなわち尤度であり，この後で定義する観測モデルによって与えられる．分母の $p(\mathbf{y}_t | \mathbf{Y}_{t-1})$ は状態 \mathbf{x}_t とは無関係な項であり， $\int p(\mathbf{x}_t | \mathbf{Y}_{t-1}) d\mathbf{x}_t = 1$ となるような正規化定数である．また， $p(\mathbf{x}_t | \mathbf{Y}_{t-1})$ は，時刻 t における \mathbf{x}_t の事前確率であり， \mathbf{x}_t のマルコフ性より

$$p(\mathbf{x}_t | \mathbf{Y}_{t-1}) = \int p(\mathbf{x}_t | \mathbf{x}_{t-1})p(\mathbf{x}_{t-1} | \mathbf{Y}_{t-1})d\mathbf{x}_{t-1} \quad (3)$$

となる．ここで $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ は時刻 $t-1$ と t の間の状態遷移確率であり，後で定義する状態遷移モデルにより与えられる．

\mathbf{x}_{t-1} から \mathbf{x}_t への状態遷移モデルを考える．図書館内における歩行者の移動方向は予測が難しいため，本研究では局所的な一様分布モデルを採用する． \mathbf{x}_{t-1} の示す位置から半径 r の中に含まれる環境地図画像撮影位置の集合を $S(\mathbf{x}_{t-1}, r)$ とする．次の時刻 t における位置 \mathbf{x}_t を

$$\mathbf{x}_t \sim U(S(\mathbf{x}_{t-1}, r)) \quad (4)$$

として決定する．式 (4) は，集合 $S(\mathbf{x}_{t-1}, r)$ の中から一様分布に従ってランダムに 1 点サンプリングして \mathbf{x}_t とすることを意味する．したがって，状態遷移確率 $p(\mathbf{x}_t | \mathbf{x}_{t-1})$ は

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \frac{1}{|S(\mathbf{x}_{t-1}, r)|} \quad (5)$$

のように定式化できる．ここで $|S(\mathbf{x}_{t-1}, r)|$ は集合 $S(\mathbf{x}_{t-1}, r)$ の要素数を表す．探索を効率よく行うために，ノードが撮影位置で，撮影位置の 3 次元座標をキーとして探索できる k-d 木を構築しておく．

次に，観測モデルを考える．観測モデルは，ある状態 \mathbf{x}_t が観測 \mathbf{y}_t にどれだけマッチしているか，すなわち状態 \mathbf{x}_t の観測 \mathbf{y}_t に対する尤度を計算すればよい．状態 \mathbf{x}_t の示す環境地図画像とクエリ画像 \mathbf{y}_t との SIFT 特徴点の対応点数を $M(\mathbf{x}_t, \mathbf{y}_t)$ と定義する．本研究では尤度 $p(\mathbf{y}_t | \mathbf{x}_t)$ を次式

$$p(\mathbf{y}_t | \mathbf{x}_t) \propto \exp(M(\mathbf{x}_t, \mathbf{y}_t)) \quad (6)$$

で定義する．

パーティクルフィルタでは，式 (2) で示される事後確率分布をパーティクルと呼ばれる多数の重み付きサンプルで離散的に近似する．本手法では，パーティクルは連続的な

3 次元空間上に存在するのではなく，環境地図画像の撮影位置にのみ配置される．またパーティクルの重みは式 (6) で計算する．パーティクルフィルタの全体の流れは以下の通りである．

(1) 初期化

N 個のパーティクル $\{\tilde{\mathbf{x}}_0^{(i)}\}_{i=1}^N$ を撮影位置全体に一様分布させ， $t \leftarrow 1$ とする．

(2) 予測

N 個のパーティクル $\{\tilde{\mathbf{x}}_t^{(i)}\}_{i=1}^N$ を式 (4) の状態遷移モデルに従って移動させる．

(3) 重み付け

各パーティクル $\tilde{\mathbf{x}}_t^{(i)}$ ， $i = 1, \dots, N$ に関して式 (6) に従って重み $\tilde{w}_t^{(i)}$ を計算する．

$$\tilde{w}_t^{(i)} = \frac{\exp(M(\tilde{\mathbf{x}}_t^{(i)}, \mathbf{y}_t))}{\sum_{i=1}^N \exp(M(\tilde{\mathbf{x}}_t^{(i)}, \mathbf{y}_t))} \quad (7)$$

(4) リサンプリング

N 個のパーティクルを，その重みに比例する割合でリサンプリングする．各パーティクルの重みを 0 とする． $t \leftarrow t+1$ とし，2 に戻る．

投票処理において，その画像の撮影位置にパーティクルが存在する場合のみ投票を行う．投票できる画像が 1 つも無かった場合，最低でも 1 枚の画像に投票が行えるようになるまで，3 次元点への対応を付け直す．

ある時刻での自己位置として，パーティクル全体の位置の加重平均をとった位置 (MMSE 推定値) を出力とする方法と，最も大きい重みを持つパーティクルの位置 (MAP 推定値) を出力とする方法の 2 つが考えられる．前者はある時刻の自己位置がただ 1 カ所に決まるが，例えばクエリ画像が似た見た目を持つ棚を写していた場合，パーティクルが離れた複数箇所に分布してしまう可能性がある．この場合，加重平均をとると，現在位置が棚の中となるような起こり得ない結果が出力されてしまうおそれがある．一方後者は，複数のパーティクルが同票 1 位を獲得した場合，ある時刻の自己位置が複数箇所となってしまうが，自己位置が通路上になることが保証される．本手法では，自己位置を 1 つに決めることを優先し，重み付き平均をとる手法を採用する．

4. 大学図書館内画像データを用いた実証実験

千葉大学の図書館で撮影を行い，提案手法が SfM よりも安定した位置推定が行える事を実験で確かめた．ここではその結果を示す．

4.1 実験手順

撮影を効率よく行うため，図 5 に示した Point Grey 社の全方位カメラ「Ladybug3」と台車を用いて図書館内を撮影し，環境地図を作成した．このとき，SfM の計算はすべ



図 5 全方位カメラ
 Ladybug3.



図 6 ウェアラブルカメラ
 CONTOUR.

て bundler[6] によって行った．環境地図は 755 枚の画像から作成され，Intel Core i7-970 プロセッサ 3.20GHz，主記憶 12GB の計算機でおよそ 65 時間を要した．

その約 7 週間後，つまり本が自然に移動したあとに，環境地図を作成した範囲を歩きながらクエリ画像を撮影した．撮影には図 6 に示したウェアラブルカメラ「CONTOUR」と，通常のデジタルカメラ「PENTAX *ist DS」を用いた．撮影したクエリ画像の撮影位置を，

- 環境地図作成時の 3 次元復元データにクエリ画像を追加し，SfM を行う
- 提案手法において，
 - － パーティクルフィルタを用いる
 - － パーティクルフィルタを用いず，全範囲を探索する

といった 3 通りの方法で計算し，結果を比較した．式 (1) におけるしきい値 T は，環境地図作成時と同様の 0.6 を使用した．式 (4)，(5) における半径 r の値は，パーティクルが柵をまたいで移動しない値として実験的に求めた 14pixel を使用した．

4.2 結果と考察

まず，通常のデジタルカメラ「PENTAX *ist DS」を用いて撮影したクエリ画像で行った実験結果を示す．図 7 は，クエリ画像の撮影位置を各手法で求めた結果である．図書館内部を真上から見下ろしたもので，求めた撮影位置が赤い点で示されている．図 7(a) は実際にクエリ画像を撮影したルートである．図 7(b) は SfM によって求められたクエリ画像の撮影位置である．図 7(c) は提案手法においてパーティクルフィルタを導入せずに求めた位置推定結果で，同票 1 位になった点は全て表示している．図 7(d) はパーティクルフィルタを導入して求めた結果である．SfM による方法ではクエリ画像 19 枚中 2 枚の撮影位置しか求めることができなかったが，提案手法では 19 枚すべての撮影位置を求めることができていた．また，パーティクルフィルタを導入する前は撮影位置にジャンプが見られるが，パーティクルフィルタ導入後はジャンプが解消されている．

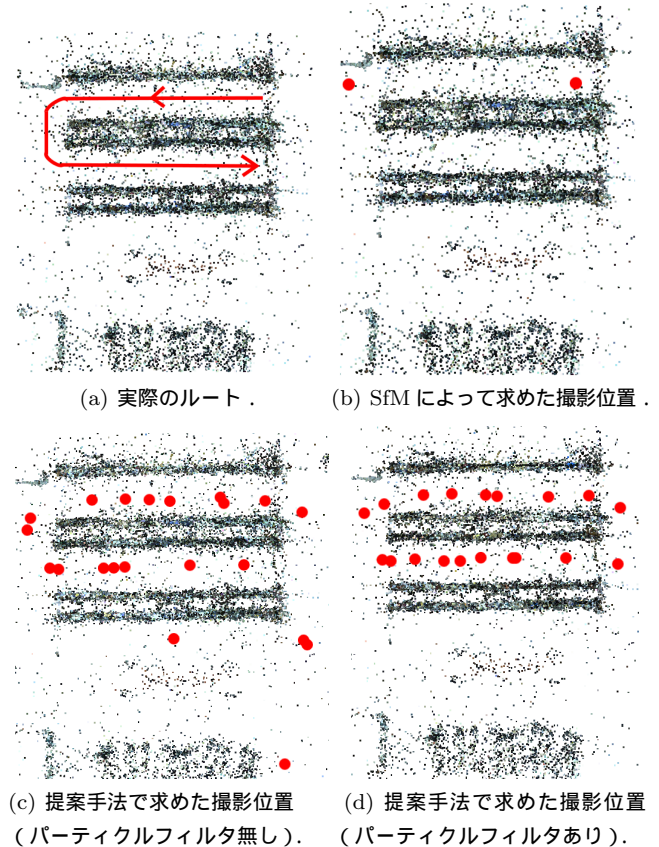


図 7 実際のルートと各位置推定結果の比較．赤い点が推定された撮影位置を表す．

図 8 は，提案手法によって画像検索を行った例である．検索の際はパーティクルフィルタを用いず，全探索した．図 8 で示したクエリ画像 2 枚は，SfM による方法でも成功した 2 枚のクエリ画像である．同じ場所を写した正しい画像が出力されているのがわかる．得票数はそれぞれ 208，46 であった．また表 1 は，各クエリ画像から検出された特徴点の数と，各クエリ画像に対して出力された検索結果画像が獲得した票数である．

図 9 は実際のパーティクルフィルタの動作を表す．赤く示された点はその時点での検索範囲で，黄色く示した点はその時点で最も高い票を得た撮影地点である．過去の情報に基づいて探索範囲を適切に限定できているのがわかる．

画像検索が正しく行えているため，特徴点の対応付け自体は成功しているといえる．しかし，提案手法における各画像の得票数は 10 票程度に留まるものが多かった．これは，全方位カメラと通常のカメラの間の画質の違いによる影響であると思われる．そこにエピソード拘束等を用いた SfM における幾何学的な検証を行えば，特徴点の対応点数はさらに減少する．そのため，SfM による手法では位置推定に失敗したものと考えられる．

次に，ウェアラブルカメラ「CONTOUR」で撮影した動画をフレーム毎に切り出し，25 フレーム毎に間引いたものをクエリ画像を用いて，同様の実験を行った．各パラ



図 8 提案手法による画像検索の例．ここで示した 2 枚のクエリ画像は，SfM による方法でも位置推定できた 2 枚である．正しい検索結果が出力されている．得票数はそれぞれ 208，46．

表 1 通常のデジタルカメラで撮影した各クエリ画像から検出された特徴点の数と，各クエリ画像に対して出力された検索結果画像が獲得した票数．

クエリ画像	特徴点数	得票数
1	3123	208
2	3273	11
3	5693	68
4	4834	46
5	4079	10
6	3530	10
7	3290	33
8	1963	7
9	2612	46
10	3135	48
11	3761	18
12	4126	10
13	3962	11
14	3649	16
15	3062	9
16	3209	14
17	2852	19
18	3146	14
19	2416	24

メータはすべて同様の値を用いた．図 10 は，実際に撮影を行ったルートと，提案手法によって求められた撮影位置を示したものである．ウェアラブルカメラから得られたクエリ画像を用いた場合，SfM による手法では撮影位置が 1 つも求まらなかった．また，提案手法においてもパーティクルフィルタが正しく動作していない．表 2 は，各クエリ画像から検出された特徴点の数と，各クエリ画像に対し

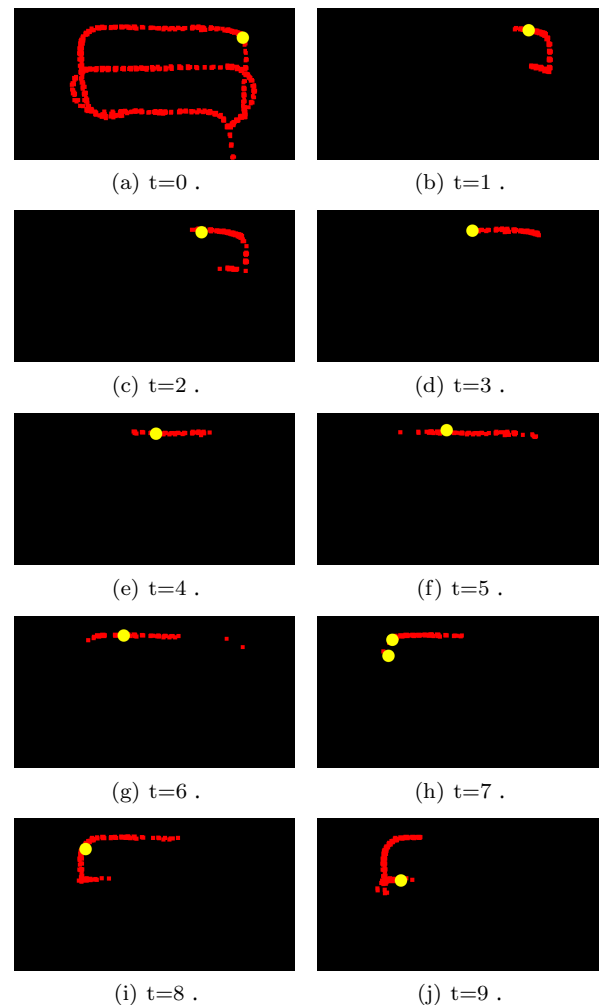


図 9 パーティクルフィルタの動作の様子．赤い点がその時刻での検索範囲，黄色い点がその時点で最も多い票を得た撮影位置である．

て出力された検索結果画像が獲得した票数である．ウェアラブルカメラから得られた画像は，元が動画データであることやブレの影響等によって画質が悪く，特徴点の対応付けに失敗する事が多い．そのため，通常のカメラでの実験時以上に対応点が少なくなり，SfM による位置推定が行えなかったものと考えられる．票が広く浅く散らばってしまった結果，少ない票数で同着 1 位となる画像が多くなり，パーティクルフィルタのリサンプリングが正しく動作しなかったものと思われる．

5. おわりに

本研究では，画像検索技術を用いた自己位置推定の方法を提案した．局所特徴点の対応付けによって類似画像検索を行い，位置推定が行える事を実験によって確かめた．SfM による方法ではクエリ画像 19 枚中 2 枚しか位置推定できなかったのに対して，提案手法では 19 枚すべての位置を推定することができた．また，状態空間モデルを定義し，パーティクルフィルタを用いることで，自己位置のジャンプを防ぐことができた．

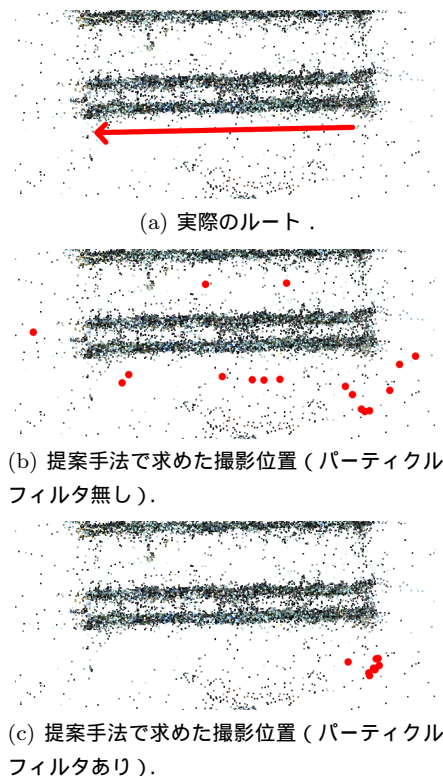


図 10 実際のルートと各位置推定結果の比較 . 赤い点が推定された撮影位置を表す .

表 2 ウェアラブルカメラで撮影した各クエリ画像から検出された特徴点の数と、各クエリ画像に対して出力された検索結果画像が獲得した票数 .

クエリ画像	特徴点数	得票数
1	1187	14
2	1285	10
3	1013	4
4	1453	3
5	1529	6
6	1656	5
7	1645	6
8	1477	4
9	1206	4
10	1244	6

しかし、環境地図作成用の全方位カメラとウェアラブルカメラとの間の画質の違いは、特徴点の対応付けの際に問題となった。本研究では SIFT 特徴点の対応を用いて画像検索を行ったが、色の分布のような他の情報を用いた画像検索方法を考えることは、今後の課題のひとつである。また、探索範囲の拡大も今後の課題である。今回の実験で用いた画像枚数は 755 枚で、図書館内の棚 3 つ分に相当するが、それでも環境地図の作成にはおよそ 65 時間かかった。将来的に図書館全体を対象とするには、環境地図の作成に時間がかかりすぎる。対象範囲の拡大による問題は、時間的な側面だけではない。対象範囲が広がれば、それだけ特徴点の誤対応も多くなり、復元が難しくなることが予想さ

れる。したがって、図書館内をいくつかのグループに分割し、それらを復元した後で復元結果を合成するような方法をとるか、あるいは SfM 以外の方法によって環境地図を構築する必要がある。

謝辞 本研究は科研費 (23700189) の助成を受けたものである。データ収集に関してご協力いただいた千葉大学附属図書館に感謝します。

参考文献

- [1] S. Arya, D.M. Mount, N.S. Netanyahu, R. Silverman, and A.Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM (JACM)*, Vol. 45, No. 6, pp. 891-923, 1998.
- [2] A.J. Davison. Real-time simultaneous localisation and mapping with a single camera. In *Proceedings of the Ninth IEEE International Conference on Computer Vision*, Vol. 2, pp. 1403-1410, 2003.
- [3] D.G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, pp. 1150-1157, 1999.
- [4] David Nistér. An efficient solution to the five-point relative pose problem. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 26, pp. 756-777, June 2004.
- [5] N. Snavely, S. M. Seitz, and R. Szeliski. Modeling the world from Internet photo collections. *International Journal of Computer Vision*, Vol. 80, No. 2, pp. 189-210, November 2008.
- [6] Noah Snavely. Bundler: Structure from motion (sfm) for unordered image collections., 2008. <http://phototour.cs.washington.edu/bundler/>.
- [7] 岡谷貴之. コンピュータビジョン最先端ガイド 3. CVIM チュートリアルシリーズ, 第 1 章. アドコム・メディア, 2010.
- [8] 小林貴訓, 杉村大輔, 平澤宏祐, 鈴木直彦, 鹿毛裕史, 佐藤洋一, 杉本晃宏. パーティクルフィルタとカスケード型識別器の統合による人物三次元追跡. 電子情報通信学会論文誌. D, 情報・システム, Vol. 90, No. 8, pp. 2049-2059, 2007-08-01.
- [9] 藤吉弘巨. Gradient ベースの特徴抽出: SIFT と HOG. 電子情報通信学会技術研究報告. PRMU, パターン認識・メディア理解, Vol. 107, No. 206, pp. 211-224, 2007-08-27.
- [10] 鳥居秋彦, 岡谷貴之, 延原章平. 多視点 3 次元復元の研究動向. 情報処理学会研究報告. CVIM, Vol. 2011, No. 1, pp. 1-22, 2011-03-10.
- [11] 加藤丈和. パーティクルフィルタとその実装法 (チュートリアル). 情報処理学会研究報告. CVIM, Vol. 2007, No. 1, pp. 161-168, 2007-01-11.
- [12] 田村俊作. 情報探索と情報利用. 図書館・情報学シリーズ. 勁草書房, 2001.
- [13] 松久亮太, 川崎洋, 小野晋太郎, 阪野貴彦, 池内克史. 因子分解法とバンドル調整による全方位画像列からの形状および位置姿勢の同時推定手法. 第 11 回 画像の認識・理解シンポジウム論文集 (MIRU2008 論文集), pp. 1610-1617, 2008.