

## Structure from Motion とマルチショット画像検索を用いた屋内位置推定

石渡翔太郎†

谷口行信†

東京理科大学†

## 1 はじめに

現在、位置推定の手段として主流になっている GPS には、電波が到達しない屋内施設、例えば大型ショッピングモールや博物館、地下鉄駅構内等において正確な位置推定が行えないという問題がある。GPS に代わる屋内位置測位技術として Wi-Fi, RFID(Radio Frequency Identification) タグ, 可視光通信, IMES(Indoor Messaging System) 等が挙げられるが、施設内に多くの無線機器を設置する必要があるため、構築する手間がかかるといった問題がある。そこで本稿では、画像認識技術により現在地を推定する手法に着目する。画像認識による位置推定技術は、予め施設内で撮影した位置情報付きの画像(参照画像)をデータベース(DB)化しておき、現在地を知りたいユーザーがその場で撮影した画像(クエリ画像)に類似する参照画像を検索することで、現在位置を推定するものである[1]。この技術には次の二つの問題がある：(i) 各参照画像に撮影位置、方向といった位置情報を人手で付与する必要があり、参照画像 DB 構築に手間がかかる。(ii) 見た目が似たシーンやテクスチャのない壁面が施設内の離れた場所にあると、位置推定に失敗することがある。以下、(ii) を類似誤検索問題と呼ぶ。

本稿では画像認識による位置推定技術の二つの課題である参照画像 DB 構築の手の削減、及び、類似誤検索問題の解決を目的に、Structure from Motion(SfM)の手法を用いて参照画像に撮影位置の情報を自動付与する；推定に失敗したときに、異なる方向からの撮影(マルチショット)をユーザに促し、複数枚のクエリ画像を利用した位置推定を行うことで、位置推定精度を高める手法を提案する。

## 2 従来研究

類似誤検索問題に対処するための従来研究として、SfM を用いた画像認識技術に Wi-Fi を組み合わせる手法がある[2]。この手法はクエリ画像を撮影した際に得られる Wi-Fi の受信強度情報を基におおよその撮影位置を推定し、推定した位置に近い位置情報を持つ参照画像を検索対象とすることで類似誤検索問題に対処している。しかし、Wi-Fi が設置されていない屋内施設では利用できない。Wi-Fi に代わる手間のかからない技術を用いて、類似誤検索問題を解決することが望ましい。

## 3 提案システム

本研究では、SfM を用いて位置情報付き参照画像 DB を少ない手間で構築し、複数のクエリ画像に類似する参照画像を絞り込みことで、類似誤検索による失敗を

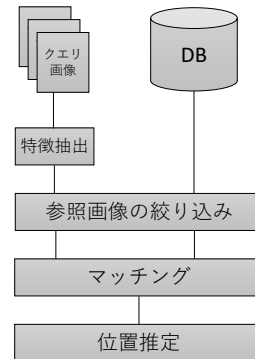


図 1: 提案システムにおける処理の流れ

低減する屋内位置推定法を提案する。本システムの構成を図1に示す。各項目の詳細を以下に示す。

**参照画像 DB 構築：** 参照画像 DB 構築の手順は以下のとおり：(i) 施設内で密に撮影した参照画像を入力する。(ii) SfM を用いて参照画像から抽出した AKAZE 特徴点を求め、三次元空間に投影する。(iii) 各参照画像に対して、以下の情報を DB に格納する：特徴記述子、特徴点に対応した3次元点、カメラ外部パラメータ  $[R|t]$ 、撮影位置  $c$ 。ただし、撮影位置  $c$  は  $c = -R^{-1}t$  である。(iv) 各参照画像から Bag of Visual Words(BoV) 特徴ベクトルを算出し DB に格納する。

**クエリ画像取得：** ユーザにマルチショットを促し複数枚のクエリ画像  $Q_i (i = 1, 2, \dots, N_q)$  を入力する。

**特徴抽出：** 各クエリ画像  $Q_i$  から AKAZE 特徴量を抽出し、特徴点を記述する。さらに BoV 特徴ベクトルを算出する。

**参照画像の絞り込み：** 各クエリ画像  $Q_i$  に類似する上位 10 枚の参照画像を以下の手順で選択する：(i) クエリ画像  $Q_i$  に対応する BoV 特徴ベクトルを  $\mathbf{v}_i^{(q)} (i = 1, 2, \dots, N_q)$ 、 $Q_i$  と類似する上位  $a$  番目の参照画像  $D_{ia}$  に対応する BoV 特徴ベクトルを  $\mathbf{v}_{ia}^{(r)}$  とする。このときの  $Q_i, D_{ia}$  間の距離  $s_{ia}$  を

$$s_{ia} = S(Q_i, D_{ia}) = \frac{\|\mathbf{v}_i^{(q)} - \mathbf{v}_{ia}^{(r)}\|^2}{\|\mathbf{v}_i^{(q)} - \mathbf{v}_{i1}^{(r)}\|^2} \geq 1,$$

と定義する。ただし、 $s_{ia}$  の値が小さいほど画像  $Q_i$  と  $D_{ia}$  が類似していることを示す。(ii) 選択された参照画像  $D_{ia}$  を位置情報に基づいて k-means 法でクラスタリングする。ただし、参照画像間の距離は、SfM で求めた撮影位置  $c$  から算出する。k-means 法によって生成されるクラスタを  $C_l (l = 1, 2, \dots, N_l)$  とする。(iii) クラスタ  $C_l$  に含まれる参照画像  $D_{ia}$  とクエリ画像  $Q_i$  の間の画像類似度  $S(Q_i, D_{ia})$  の平均が最も高いクラスタの番号 select

Indoor Localization Based on Multi-Shot Image Retrieval with Structure from Motion

†Shotaro Ishiwata †Yukinobu Taniguchi

†Tokyo University of Science

を、次式で選択する:

$$select = \arg \min_l \sum_{D_{ia} \in C_l} \frac{S(Q_i, D_{ia})}{|C_l|}. \quad (1)$$

**マッチング:** 選択されたクラス  $C_{select}$  に属するすべての参照画像とクエリ画像間で、特徴点マッチングを行う。

**位置推定:** マッチング点に対応する3次元点と  $i$  枚目のクエリ画像の特徴点から P3P-RANSAC を用いて、回転行列  $\hat{R}_i$  及び平行移動ベクトル  $\hat{t}_i$  を推定し、 $\hat{c}_i = -\hat{R}_i^{-1}\hat{t}_i$  によりクエリ画像の位置  $\hat{c}_i$  を算出する。位置推定に成功した推定位置  $\hat{c}_i$  の平均を現在位置として出力する。

## 4 実験条件

提案手法の有効性を示すために、研究室内で取得したデータを用いて実験を行った。参照画像 266 枚とマルチショットのクエリ画像を 4 枚 1 組、計 34 組用意した。撮影機器は Apple 社製の iPhone7 である。また、評価方法として次の二つを用いる: (i) 特徴点マッチングで抽出された特徴点のペアが 3 組未満となり、位置推定ができない割合をクエリ返却失敗率として定義する。(ii) クエリ画像の撮影位置 (正解位置) と推定した撮影位置の間の距離が  $d(m)$  以内のとき位置推定に成功したとみなし、推定に成功した割合を正答率として定義する。実験 1 でクエリ画像の枚数と位置推定精度の関係を評価する。実験 2 で提案手法を含めた三つの参照画像絞り込み手法を比較する。具体的には、比較手法 1 は、参照画像  $D_{ia}$  の絞り込みを行わない手法、すなわち BoV によって DB から抽出した全ての類似画像を位置推定に用いる手法である。比較手法 2 は、式 (1) のクラスタの選択の条件を、要素数が最も多いクラスタに変更した手法である。ただし、実験 2 ではマルチショットの枚数を 4 枚とする。

## 5 実験結果

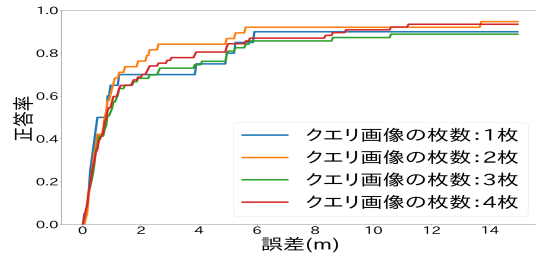
表 1: クエリ数とクエリ返却失敗率の関係

| クエリ画像の枚数 | 1    | 2    | 3    | 4    |
|----------|------|------|------|------|
| クエリ返却失敗率 | 0.41 | 0.26 | 0.12 | 0.12 |

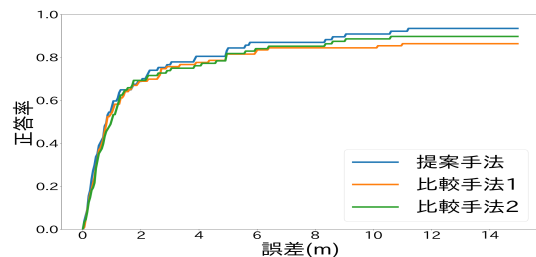
表 2: 実験 2 におけるクエリ返却失敗率

| 絞り込み手法   | 比較手法 1 | 比較手法 2 | 提案手法 |
|----------|--------|--------|------|
| クエリ返却失敗率 | 0.00   | 0.09   | 0.12 |

**実験 1:** クエリ画像の枚数を 1 から 4 枚に変化させながら、正答率を測定した。結果を表 1、図 2(a) に示す。表 1 によれば、クエリ枚数 3 枚と 4 枚のときクエリ返却失敗率が最も低かった。これは、位置推定に失敗するクエリ画像が含まれていても、残りのクエリ画像を用いて位置推定が可能になるため、シングルショットと比べて位置推定に成功する割合が増加する結果になったと考えられる。



(a) クエリ数と正答率の関係



(b) 実験 2 における正答率

図 2: 実験の結果

また、図 2(a) より、正答率はクエリ画像の枚数に大きく依存しないことが分かった。以上から、マルチショットの有効性を示した。

**実験 2:** 参照画像絞り込み手法を、提案手法、比較手法 1、2 としてクエリ返却失敗率と正答率を測定した。表 2、図 2(b) に実験結果を示す。比較手法 1 はクエリ返却失敗率が最も低かったが、正答率が提案手法より低くなった。これは、アウトライヤが混入した状態で位置推定を行ったため、正答率が低下したと考えられる。また、比較手法 2 における正答率は提案手法より低くなった。これは、選択されたクラスタ内に含まれている類似度が低い参照画像が位置推定に悪影響を与えたと考えられる。

## 6 まとめ

本稿では、施設内に類似したシーンが存在する状況に対して頑健な屋内位置推定手法の開発を目的とし、SfM とマルチショット類似画像検索を用いた屋内位置推定手法を提案した。今後の課題として、テクスチャの少ないクエリ画像が精度に悪影響を与えていることから、ユーザにテクスチャの多いシーンの撮影を促す手法を考案すること、また大型ショッピングモールや地下鉄等の屋内施設において実験・評価を行い、本研究の有用性をより明確化することが挙げられる。

## 参考文献

- [1] 沖谷卓哉, 島田裕, 谷口行信, “R-MAC 特徴ベクトルの加重平均による屋内現在地推定の精度向上,” 第 20 回画像の認識・理解シンポジウム (MIRU 2017), PS1-40, 2017.
- [2] Guoyu Lu and Jingkuan Song, “3D Image-based Indoor Localization Joint with WiFi Positioning,” ICMR, pp. 465–472, 2018.