

全方位カメラによる参照画像を用いた屋内における位置推定 — デジタルミュージアムでの鑑賞者の行動記録に向けて —

畑田 晃希† 河治 寿都† 山崎 俊彦† 相澤 清晴‡,†
† 東京大学大学院 情報理工学系研究科 ‡ 東京大学大学院 情報学環

1 まえがき

個人の行動履歴を取得、解析する技術に近年注目が集まっている [1]. 本稿では特に、博物館など展示会場における見学者の行動履歴に着目する. 博物館における行動履歴の解析は、観賞体験の共有や、展示や関連する情報の推薦など多くのアプリケーションが期待されている. そこで、行動履歴の取得に不可欠な、見学者の位置推定技術について述べる.

屋内における位置推定技術は、Wi-Fi, RFID, Bluetooth などを用いた手法が数多く研究されている. しかし、位置推定技術を博物館に適用する際、設備の導入コストや、博物館内の環境への影響、見学者の負担など様々な条件を満たす必要がある. この様な観点では、見学者が撮影した画像によって位置推定を行うことが、最も効果的であると言える.

展示物を画像から認識する試みとしては、Bay らの SURF 特徴量 [2] により画像マッチングを行う手法 [3] や、筆者らの Locality Sensitive Hashing (LSH) [4] と全方位パノラマ映像を用い、多数の展示物に対して高速かつ高精度な画像マッチングを実現した手法 [5] などがある.

本稿においては、筆者らの従来研究と同様に、全方位映像を用いた画像地図 (博物館内の位置参照情報が付加された画像データベース) の構築、局所特徴量と LSH を用いた高速かつ高精度な画像マッチングについて述べる. その上で、計算量と認識精度の観点から、展示物認識に有効な特徴量として、SIFT [6], PCA-SIFT [7], SURF の比較を行う.

2 画像地図の作成

画像マッチングにより位置推定を行う際、最も重要なことの一つは、位置を参照できる画像群のデータベースを構築することである. 本稿では、そのような画像データベースを画像地図と呼ぶこととする. この画像群と見学者が撮影した画像とで、特徴量のマッチングを行うことで、展示物の認識を行い見学者の位置推定をする. 画像地図は主に図 1 に示す全方位映像撮影システムを用いて、画像を撮影し、位置情報を付加して構築する. 全方位映像を利用することで、一回の撮影でより多くの展示物を網羅することができる. また、画

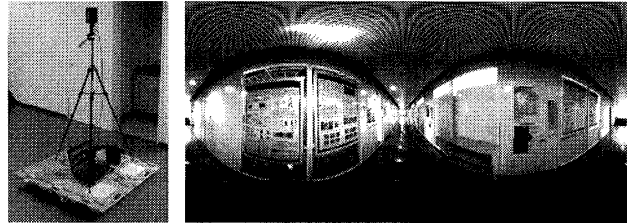


図 1: 全方位映像撮影システム (左), 取得画像 (右)

像地図は位置推定のみだけでなく、見学者が博物館の見学後、観賞体験を共有するなど、様々なアプリケーションへの応用が期待される. ただし、画像地図の構築の際、全方位映像撮影システムでは撮影できない展示物や撮影角度については、別途、一般的なカメラで撮影を行う.

3 展示物の認識

3.1 画像マッチング

画像マッチングは、局所特徴量の対応点探索により行う. 画像地図に含まれる参照画像は、あらかじめ特徴量を抽出しておき、見学者が撮影した画像と対応点探索を行う. そのうち、RANSAC を用いて対応点の射影変換を計算し、誤対応を除去する. 最終的にもっとも多くの対応点を得られた参照画像により、展示物を判別し位置推定とする.

3.2 LSH

参照画像が多数存在する場合、対応点探索を高速に行う技術が不可欠である. 最近傍探索を高速に行う手法には LSH, ANN [8] などがある. 特に、LSH は探索する次元が大きい (20 次元~) 場合においても、高速に探索が行えるため、画像検索などで用いられている [9]. LSH とは、近接するベクトルが同一のハッシュ値となるようなハッシュ関数を用いて、確率的に最近傍探索を行う手法である.

今回は、ユークリッド距離に基づいて近傍探索を行うため、Datar らの提案した E^2 LSH [4] を用いて、特徴量のハッシュ値を計算する. E^2 LSH を用いた場合、 $O(d\rho^c \log n)$ (ただし、 n はデータの数、 d はベクトルの次元、 $\rho(c) < 1/c$, $c \in (1, 10]$) の計算時間で、(R,c)-最近傍探索が可能である.

Fast and Robust Vision-Based Indoor Localization using Omnidirectional Panorama Images

†Koki Hatada †Hisato Kawaji †Toshihiko Yamasaki †Kiyoharu Aizawa †Department of Information and Communication Engineering, The University of Tokyo

‡Interfaculty Initiative in Information Studies, The University of Tokyo

3.3 局所特徴量

局所特徴量の対応点探索で画像マッチングすることにより、オクルージョンや、撮影角度の変化に対して、頑健性の高い展示物認識が可能であると期待できる。局所特徴量として、SIFT, PCA-SIFT, SURF などが提案されている。PCA-SIFT は SIFT を改良した特徴量として Ke らによって提案された、勾配情報を主成分分析によって次元を圧縮し、36 次元で記述する特徴量である [7]。SURF は、Haar-Wavelet 変換により 64 または 128 次元 (今回は 128 次元で用いる) で算出される特徴量で、SIFT, PCA-SIFT に比較して数倍高速に計算できる。また、E²LSH の計算量を踏まえると、PCA-SIFT が SIFT や SURF に比較して、高速な対応点探索が期待できる。

4 実験

4.1 条件

建物の 1 フロアの通路約 100m を対象とした実験を行った。全方位映像と全方位映像を補う画像をあらかじめ撮影し、位置参照情報を付加、特徴量の抽出を行った。展示物の数は 36、補助の画像は 800 × 600 pixel で 148 枚となった。展示は主にポスターである。テスト画像として、一般的なカメラで 85 枚の画像 (800 × 600 pixel) を撮影し、3 章に則り展示物の認識をした。テスト画像には、オクルージョンのあるものや、正面以外の角度から撮影したものを含めた。特徴量として、SIFT, PCA-SIFT, SURF を用いて精度・計算時間を比較した。実験は core2duo 2.40GHz, メモリ 4GB の PC 環境で行った。

4.2 結果

表 1 に画像 1 枚あたりの平均の計算時間と認識精度を示した。いずれの特徴量においても約 9 割の高い精度で、展示を認識できることが確認できた。特徴量による認識精度の差はほぼないと言える。特に、対応点探索によって得られた対応点の数に着目すると、平均して SIFT が 164.1 個、PCA-SIFT が 164.1 個、SURF が 33.0 個となった。これは、オクルージョンへの頑健性が、SIFT と PCA-SIFT は SURF より大きいと言える。

対応点探索の計算時間は、PCA-SIFT が、SIFT と SURF の約 3 倍速く、SIFT と SURF はほぼ同一である。これは E²LSH で、最近傍探索にかかる計算量が、次元に比例することに合致する。ただし、特徴量の抽出にかかる計算時間は、SURF が SIFT や PCA-SIFT より非常に小さく、総計算時間でみると、SURF が最も優位な特徴量であると言える。

特に SURF の場合、線形探索では 1 枚の画像あたり約 120 秒かかり、LSH により約 78 倍の高速化となった。E²LSH の計算量と線形探索の計算量 ($O(dn)$) を比較すればこの計算時間の差は、参照画像群が多くなるほど、大きくなると言える。

表 1: 精度と計算時間

特徴量	抽出	探索	合計	認識精度
SIFT	1.51 秒	1.60 秒	3.11 秒	91.9 %
PCA-SIFT	1.63 秒	0.56 秒	2.19 秒	90.8 %
SURF	0.22 秒	1.54 秒	1.77 秒	88.2 %

一方、PCA-SIFT は SIFT, SURF (128 次元) に比較して次元が小さく、メモリ空間の使用量の観点からは優位である。特徴量の抽出の高速化ができれば、非常に大量の参照画像を扱うような場面では有用である。

5 まとめ

本稿では、博物館において展示物を認識し見学者の位置を推定する手法について述べた。LSH を用いることで、大幅な高速化を実現した。画像マッチングに用いる特徴量として、SIFT, PCA-SIFT, SURF を比較し、計算時間の観点から SURF の優位性を確認した。今回の手法は、将来、博物館の利用者の行動履歴取得へ応用するという観点から、十分な精度と計算速度を実現した手法であると言える。今後、位置推定結果の可視化・応用の研究や、大量の位置参照画像を用いた大規模な実験を行っていきたい。

参考文献

- [1] 相澤清晴, “体験の記録とコンテキストに基づく効率的な検索,” システム制御学会誌 システム/制御/情報, Vol. 50, No 1, pp. 7-11 (2006)
- [2] H. Bay, et al, “SURF: Speed-Up Robust Features,” 9th European Conference on Computer Vision, pp.404-417 (2006)
- [3] H. Bay, et al, “Interactive Museum Guide: Fast and Robust Recognition of Museum Objects,” in Proceedings of the First International Workshop on Mobile Vision (2006)
- [4] M. Datar, et al, “Locality-sensitive hashing scheme based on p-stable distributions,” in Proceedings of the 20th Annual Symposium on Computational Geometry, pp. 253-262 (2004)
- [5] 畑田晃希, et al, “全方位カメラ画像を用いた博物館における展示物認識,” 電子情報通信学会 全国大会 (2010) (to appear)
- [6] D. G. Lowe, “Distinctive image features from scale-invariant keypoints,” International Journal of Computer Vision, 60, 2, pp. 91-110 (2004)
- [7] Y. Ke, et al, “PCA-SIFT: A More Distinctive Representation for Local Image Descriptors,” in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 511-517 (2004)
- [8] S. Arya, et al, “An optimal algorithm for approximate nearest neighbor searching fixed dimensions,” Journal of the ACM, Vol. 45, No. 6, pp. 891-923 (1998)
- [9] Y. Ke, et al, “Efficient Near-duplicate Detection and Sub-image Retrieval,” in ACM Conf. on MultiMedia, pp. 869-876 (2004)