

地理情報を埋込んだ画像特徴量に基づく写真画像の位置情報推定

Estimating the location of photo images by embedding geographic information in image features

パオ ワンナー† 立間 淳司‡ 青野 雅樹‡
Pav Vanna Atsushi Tatsuma Masaki Aono

1. はじめに

近年、ソーシャルメディアの普及により、利用者の趣味嗜好に基づく情報を自由に発信することができる環境が整いつつある。その一つに写真画像があり、インターネット上に膨大に存在している。代表的な写真共有サイト Panoramio では、投稿された写真に位置情報が付与されており、関心のある地域の写真を、地図上に配置して閲覧できる。一方で、ウェブ上の画像には、位置情報が付与されていないものも大量にある。そこで、本研究では特定の建造物や世界遺産などの写真画像から、位置情報を推定することを目的とする。

写真画像の位置情報を利用した研究は、いくつか存在する。和多田ら[5]は、位置情報とタグを有した画像を利用し、タグが付与されていない画像に対して、自動でタグを推定するシステムを提案した。和多田らのシステムでは、与えられた画像を検索質問として、類似画像検索を行い、得られた類似画像群に付与されたタグを、推定結果としてユーザに提示する。このとき、テスト画像の位置情報から、検索対象画像の絞り込みを行う。また、Hays ら[3]の研究では、与えられた画像を検索質問として、類似画像を Flickr から検索し、検索結果の位置情報付き写真を利用して、ユーザから与えられた画像の撮影位置を推測する。

2. 提案手法

本研究では、利用者から与えられた、メタデータが付与されていない建造物や世界遺産など（以降ランドマークと呼ぶ）の写真画像に対して、位置情報を推定する手法を提案する。

提案手法の概要を図 1 に示す。本研究では、位置情報として、ランドマークの名称を推定する。提案手法は、大きく、訓練段階と推定段階とに分けられる。

訓練データセットに含まれる写真画像には、各ランドマークの航空写真と説明文書（以降地理情報と呼ぶ）が紐付けられている。訓練段階では、まず、訓練データセットの写真画像から画像特徴量を抽出する。次に、紐付けられた地理情報からも特徴量を抽出し、写真画像の画像特徴量とペアにして、正準相関分析により共通の部分空間を推定する。写真画像の画像特徴量を、推定された部分空間に射影することで、地理情報を埋め込む。そして、ランドマークの名称をラベルとし、部分空間に射影した画像特徴量を用いて、分類器の訓練を行う。

推定段階では、まず、利用者から与えられた写真画像から、訓練段階と同様にして画像特徴量を抽出する。次に、正準相関分析により得られた射影ベクトルを用いて、部分空間に射影する。最後に、訓練した分類器を用いて、与えられた写真画像のランドマークの名称を推定する。

以降、各処理の詳細について述べる。

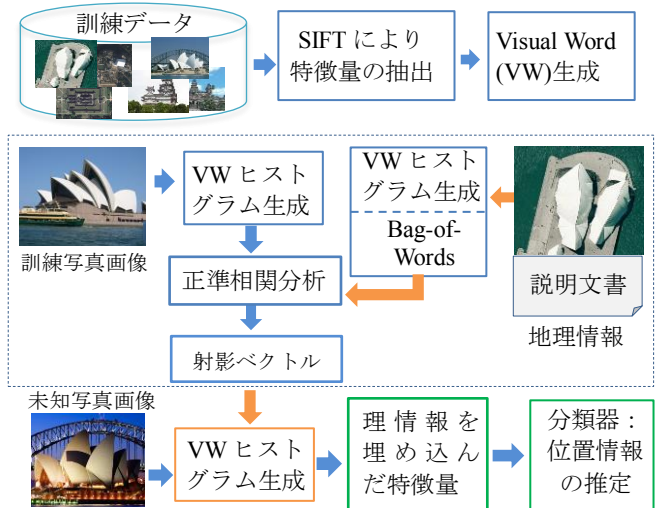


図 1. 提案手法の概要

2.1. Bag-of-Visual-Words による画像特徴量

本研究では、画像特徴量に、画像認識の分野で標準的な特徴量となっている Bag-of-Visual-Words (BoVW) ヒストグラムを用いる。BoVW は、文書分類で使用されている Bag-of-Words のアナログであり、画像を Visual Word の出現頻度によるヒストグラムで表現する。本研究では、BoVW ヒストグラムを、以下の様にして作成した。

1) Visual Word の生成

- (1) 画像の局所特徴量を抽出する。
- (2) 特徴量群に対して k-means 法を用いて量子化し、その代表ベクトルを Visual Word とする。

2) ヒストグラムの作成

- (1) 画像から抽出した局所特徴量を、最近傍にある Visual Word に割り当てる。
- (2) 割り当てられた局所特徴量を数え上げ、ヒストグラムとして表現する。

ここで、局所特徴量には、SIFT (Scale-Invariant Feature Transform) 記述子[2]を用いた。

2.2. 正準相関分析による地理情報の埋込み

正準相関分析(Canonical Correlation Analysis, CCA)[4]は、二種類の変数間に共通する部分空間を推定する手法である。

本研究では、訓練データセットに含まれる n 個の写真画像の画像特徴量 $X = \{x_i\}_{i=1}^n \in \mathbb{R}^{p \times n}$ と、航空写真の画像特徴量もしくは説明文書の Bag-of-Words $Y = \{y_i\}_{i=1}^n \in \mathbb{R}^{q \times n}$ とをペアとして、CCA により共通の部分空間を推定する(図 2)。二種類の変数をそれぞれ射影ベクトル w_x と w_y により、共通の部分空間に射影することを考えるとき、CCA による部分空間の推定は、以下の最大化問題で表される。

† 豊橋技術科学大学大学院 工学研究科 情報・知能工学専攻

‡ 豊橋技術科学大学大学院 工学研究科 情報・知能工学系

$$J = \underset{w_x, w_y}{\operatorname{argmax}} \frac{w_x^T C_{xy} w_y}{\sqrt{w_x^T C_{xx} w_x w_y^T C_{yy} w_y}}$$

ここで、 C_{yy} , C_{xy} , C_{xx} は X と Y から得られる分散共分散行列である。この最大化問題を解くことで、射影ベクトル w_x と w_y を得る。テストデータの写真画像が与えられた場合、その画像特徴量を、射影ベクトル w_x により部分空間に射影することで、メタデータを付与されていないテストデータにも、航空写真や説明文書で表される地理情報を埋め込むことができる。



図 2. CCA における写真画像との地理情報のペア

3. 評価実験

提案手法の有効性を確認するため、評価実験を行った。本研究では、世界文化遺産[1]の写真画像と航空写真および説明文書を Web から収集してデータセットを作成した(図 3)。

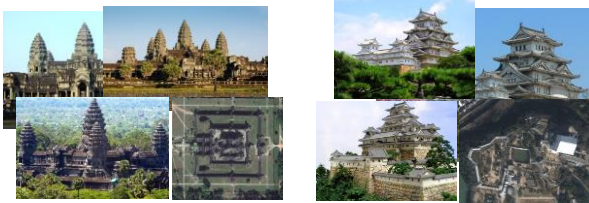


図 3. データセットの例 (写真画像、航空写真)

3.1. 実験方法

評価実験では、50 種類の世界文化遺産 (WCH) をクラスとし、位置情報を推定する実験を行った。各クラスには 55 枚の写真画像、1 枚の航空写真、1 つの説明文書が含まれている。1 つクラスの写真画像を訓練用として 35 枚、テスト用として 20 枚に分けた。分類器には、RBF カーネルによる Support Vector Machine を用いた。

位置情報の推定性能の評価尺度には、総合正解率 (Overall Accuracy) と κ 係数 (Kappa Coefficient) を用いた。観測した画像数を b 、クラス数を C とする。また、推定ラベル w_i について、 w_i を正解とする画像数を N_{ii} とする。このとき、 N_1, N_2 をそれぞれ正解画像数と分類器が w_i 付与した画像数とすると、総合正解率と κ 係数は、それぞれ以下の式で定義される。

$$\text{Accuracy}(P_0) = \frac{\sum_{i=1}^C N_{ii}}{b}, \quad \text{Kappa} = \frac{P_0 - P_e}{1 - P_e}$$

ただし、 P_e は、次の式で求める。

$$P_e = \sum_{i=1}^C N_1 N_2 / b^2$$

3.2. 実験結果

写真画像の位置情報を推定する実験を、(1) 地理情報を用いずに写真画像のみを用いる手法 (ベースライン)、(2) 写真画像と航空写真をペアとして CCA を行う手法 (CCA-SAT)、(3) 写真画像と説明文書をペアとして CCA を行う手法 (CCA-DOC)、(4) 航空写真と説明文書をペアとして CCA を行ったものと、写真画像とをペアとして再度 CCA を行う手法 (CCA-SAT-DOC) を用いて行った。

実験結果を図 4 に示す。CCA を用いて地理情報を埋め込む提案手法が、ベースライン手法と比較して、優れた推定性能を得られていることがわかる。航空写真を用いた CCA-SAT が最も優れた推定性能となった。これは、航空写真に、写真画像だけでは捉えられない、ランドマークを判別する特徴が含まれおり、その特徴を CCA により写真画像の画像特徴量に埋め込むことができたためと考える。

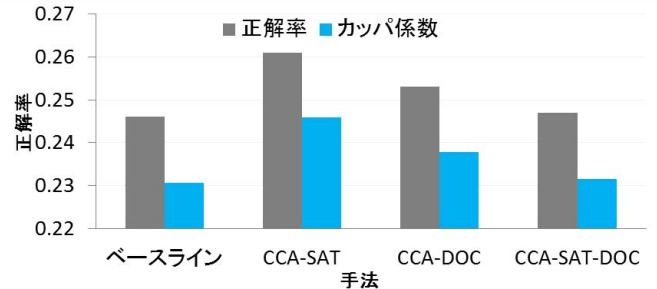


図 4. ラベル推定の結果

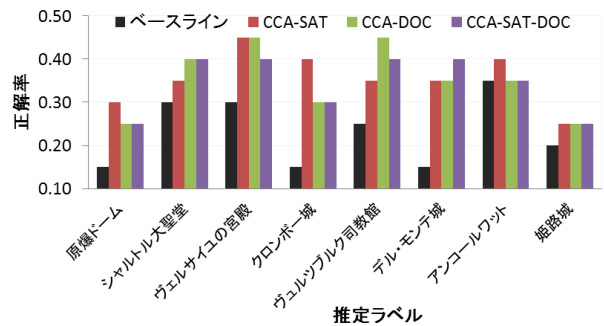


図 5. クラス別の推定性能

図 5 は、8 種類の世界文化遺産での正解率である。提案手法が、ベースラインと比較して、優れた推定性能となっていることがわかる。

4. おわりに

本研究では、メタデータが付与されていない建造物や世界遺産などの写真画像から位置情報を推定するシステムを提案した。優れた推定性能を得るために、航空写真や説明文書で表される地理情報を、正準相関分析を用いて、写真画像の画像特徴量に埋め込む手法を提案した。提案手法の有効性を検証するため、世界文化遺産によるデータセットを作成して比較実験を行った。地理情報を埋め込む提案手法は、画像特徴量のみを用いるベースラインと比較して、優れた推定性能を得られた。

今後では、写真画像のラベル推定に加えて、経緯度も推定できる手法を提案する。

参考文献

- [1] 世界文化遺 2014 年 : <http://whc.unesco.org/en/list/>
- [2] D. Lowe, "Distinctive Image Feature from Scale-Invariant Keypoints", *International Journal Computer Vision*, Vol.60, No.2, pp91-110, 2004.
- [3] J. Hays, Alexei A. Efros, "IM2GPS: estimating geographic information from a single image", *Proceedings of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2008.
- [4] H. Hotelling, "Relations Between Two Sets of Variates", *Biometrika*, Vol.28, No.3-4, pp321-377, 1936.
- [5] 和多田 吉樹, 鎌原 淳三, 長松 隆, 田中 直樹, "位置情報付き類似画像を用いた未知画像のタグ推定における精度評価実験", *DEIM Forum B9-1*, 2012.