

物体の対称性に着目した詳細類似画像検索 Fine-Grained Similar Image Retrieval Focused on Symmetry of Objects

藤森 真綱[†] 望月 貴裕[†] 佐野 雅規[†]
Naotsuna Fujimori Takahiro Mochizuki Masanori Sano

1. はじめに

放送局では、大量に蓄積された素材映像や資料画像から特定の被写体を検索する技術が求められている。機械学習による画像認識で付与されたメタデータに基づいて検索する手法が主流であるが、付与できるキーワードの種類と数に限りがあり、任意の被写体の検索は難しい。そこで、機械学習を用いないアプローチとして、画像同士を直接比較し、見た目の似た画像を探し出す類似画像検索技術が研究されている[1]。従来の類似画像検索手法の多くは、大まかに似た画像を探すことを目的としており、同じ被写体を含む画像を精度良く検索することが困難であった。そこで我々は、被写体特有の特徴として「対称性」に着目し、この課題の解決を図った。対称性は自然物および人工物ともに多く観察されるもので、人間が物体を認識する重要な手がかりの一つである。本稿では、対称性を類似画像検索に利用することとし、画像から得られる特徴点の対称性に基づいた新しい画像特徴の適用により、検索精度を向上させる手法を提案する。さらに、対称性の顕著な蝶のデータセットおよび一般的な建物のデータセットを用いて提案手法の精度を評価する。

2. 提案手法

提案する類似画像検索手法の概要は以下のとおりである。まず、検索対象画像及びクエリ画像から特徴点を抽出する。続いて、特徴量の類似性と位置関係に基づき、対称性のある特徴点で構成される「対称性グループ」を生成する。そして、画像全体から算出した特徴の類似度と、対称性グループ毎に算出した特徴の類似度との統合により、2つの画像の類似度を求める。以降の各節で処理の詳細を述べる。

2.1 特徴点の抽出

特徴点は画像の濃淡が2方向に変化している点とし、回転不変なものを用いる。各特徴点の周辺領域における色、テクスチャ、勾配などをベクトル化したものを局所特徴として求める。本稿では、特徴点抽出にはSIFT[2]と同様にガウシアンピラミッドを用いた。また、特徴点の記述には回転およびスケール不変性に加えて鏡面変換不変性を有するMIFT[3]を利用した。

2.2 対称性の検出

2.1節で抽出した特徴点について、特徴空間上で距離の近い組を列挙し、同じ対称性を持つ組ごとにグループ分けすることにより対称性を検出する。ここで得られた特徴点のグループを対称性グループと呼ぶ。本稿では鏡面对称のみに着目し、1枚の画像から複数の鏡面对称を検出することができる[4]の方法を用いた。抽出された対称性グループの例を図1に示す。○は対称性グループに所属する特徴点

[†] 日本放送協会

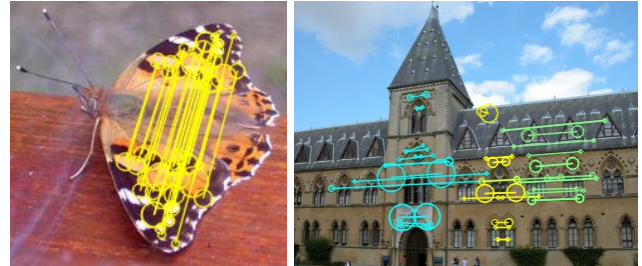


図1 抽出された鏡面对称の例

を表す。鏡面変換により対応する特徴点が線で結ばれている。左の画像では対称性グループが1つ、右の画像では複数の対称性グループが抽出されている（見やすさのため、3つのグループのみ表示）。

2.3 特徴量の計算

全ての検索対象画像から抽出された特徴点を特徴空間上でクラスタリングし、その代表ベクトルをコードブックとする。本稿ではクラスタリングにk-meansを用い、コードブックサイズは $k = 10000$ とした。得られたコードブックを利用して以下で述べるように画像全体の特徴ベクトル（以下「全体特徴ベクトル」）および対称性を考慮した詳細特徴ベクトルを求める。

2.3.1 全体特徴ベクトル

全体特徴ベクトルは、2.1節で得られた局所特徴を2.3節で求めたコードブックによりベクトル量子化し、その出現頻度のヒストグラムを正規化することで得られる。画像 I に対する全体特徴ベクトルを $\mathbf{V}_G(I)$ とする。これはBag of Visual Words (BoVW)と呼ばれ、画像認識や画像検索で広く用いられている特徴ベクトルである。本稿の実験ではtf-idfを用いたL2正規化を行った。

2.3.2 詳細特徴ベクトル

詳細特徴ベクトルは、2.2節で得られた対称性グループに対して、各グループに所属する特徴点の局所特徴を2.3.1節と同様にベクトル量子化、正規化することで得られる。したがって、1枚の画像に対して検出された対称性の数だけ詳細特徴ベクトルが得られる。画像 I に対する詳細特徴ベクトルを $\mathbf{V}_S(I, 1), \dots, \mathbf{V}_S(I, M)$ とする。ただし、 M は検出された対称性グループの個数を表す。本稿では所属する特徴点の数が10以上となる対称性グループのみを用いた。

2.4 類似度の計算

クエリ画像 I_Q と検索対象画像 I_D の類似度は以下のように計算する。まず、2.3.1節で求めた全体特徴ベクトルについて、その類似度 $S_G(I_Q, I_D)$ を

$$S_G(I_Q, I_D) = \mathbf{V}_G(I_Q) \cdot \mathbf{V}_G(I_D) \quad (1)$$

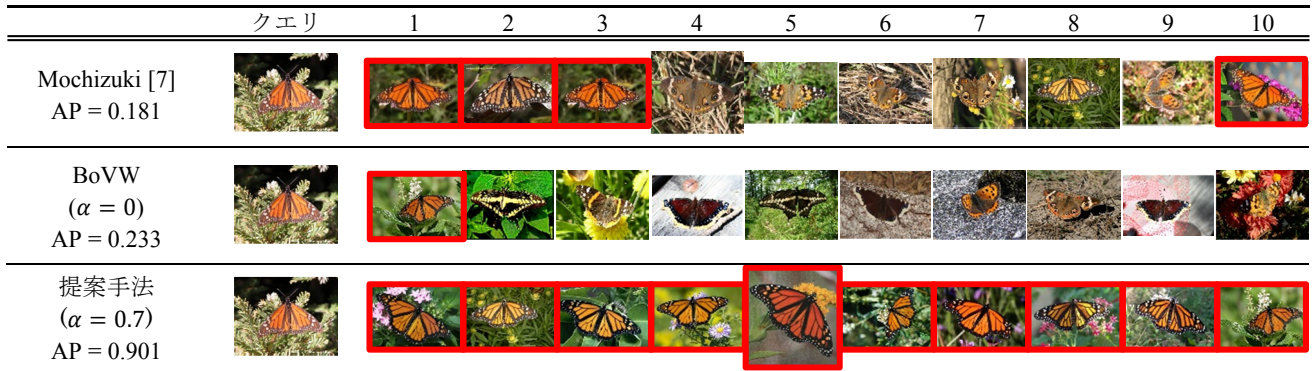


図 2 検索結果例

として求める。同様に、2.3.2 節で求めた詳細特徴ベクトルについて、その類似度 $S_S(I_Q, I_D)$ を

$$S_S(I_Q, I_D) = \frac{1}{K} \sum_{k=1}^K \mathbf{V}_S(I_Q, i_k) \cdot \mathbf{V}_S(I_D, j_k) \quad (2)$$

として求める。ここで、 K は類似度を考慮する対称性グループのペアの総数であり、 i_k, j_k は $\mathbf{V}_S(I_Q, i_k) \cdot \mathbf{V}_S(I_D, j_k)$ が最大となる対称性グループの組み合わせを順次選ぶことで決定する。ただし、 $\mathbf{V}_S(I_Q, i_k) \cdot \mathbf{V}_S(I_D, j_k) = 0$ となる組み合わせは除外する。最後に、画像の類似度 $S(I_Q, I_D)$ を

$$S(I_Q, I_D) = (1 - \alpha) S_G(I_Q, I_D) + \alpha S_S(I_Q, I_D) \quad (3)$$

として求める。ここで、 α は対称性の寄与を示すパラメータである。

3. 実験

対称性が顕著な蝶のデータセット (Leeds Butterfly Dataset [5]) および、より一般的な建物のデータセット (The Oxford Building Dataset [6]) を用いて提案手法の精度を評価した。蝶のデータセットは 10 種類の蝶に分類された 832 枚の画像であり、1 種類あたりの枚数は 55~100 である。また、建物のデータセットは Flickr から集められた 5062 枚の画像からなり、11 種類の建物に関してアノテーションがつけられている。各クエリ画像に対する平均適合率 (AP: Average Precision) の平均値 (MAP: Mean Average Precision) により評価することとし、Mochizuki らのテキスト特徴を用いた類似画像検索手法 [7] および、式 (3) において $\alpha = 0$ とおいた、全体特徴ベクトルのみを用いた手法 (BoVW を用いた方法に対応する) を比較手法とした。

表 1 に実験結果を示す。中列は蝶のデータセットでの実験結果であり、全ての画像をクエリとした検索結果の MAP である。 α は予備実験により 0.7 とした。また、1 枚の画像につき被写体の蝶は 1 頭のみ写っているため、所属する特徴点数が最大の対称性グループのみを用いた。蝶での検索結果例を図 2 に示す (赤枠が正解画像)。比較手法ではクエリ画像と背景のテキストが似た画像が検索結果の上位に来ているが、提案手法では背景の影響を受けずに同じ被写体の画像を検索できている。表 1 の右列は建物のデータセットでの実験結果であり、11 種類の建物についてそれぞれ指定された 5 枚のクエリ画像による計 55 回の検索結果の MAP である。 α は予備実験により 0.4 とした。

対称性が顕著な蝶のデータセットにおいては、提案手法の MAP は比較手法に対して大きく改善している。これは、

手法	蝶	建物
Mochizuki [7]	0.244	0.137
BoVW ($\alpha = 0$)	0.397	0.338
提案手法	0.517 ($\alpha = 0.7$)	0.349 ($\alpha = 0.4$)

表 2 実験結果

対称性グループに着目することで背景による影響を効果的に除去できたためと考えられる。一方、必ずしも対称性が含まれていない建物のデータセットにおいては、MAP は改善しているものの、大きな精度向上は得られなかった。これは、正しく対称性が検出できない、もしくは誤った組み合わせの対称性グループ同士で類似度を計算しているなどの原因が考えられる。

4. まとめ

本稿では、被写体の対称性に着目した新しい画像特徴を用いた類似画像検索を提案した。対称性が顕著な蝶のデータセットおよび一般的な建物のデータセットを用いた実験により提案法の有効性が確かめられた。今後の課題として、対称性の抽出精度の向上、複数の対称性グループのより効果的な扱い、鏡面对称以外の対称性の利用が挙げられる。

参考文献

- [1] R. Datta, D. Joshi, J. Li, J. Z. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age," *ACM Computing Surveys*, Vol. 40, No. 2, pp. 1-60, 2008.
- [2] D. G. Lowe, "Object recognition from local scale-invariant features," *Proc. of IEEE International Conference on Computer Vision (ICCV)*, pp. 1150-1157, 1999.
- [3] X. Guo, X. Cao, "MIFT: A framework for feature descriptors to be mirror reflection invariant," *Image and Vision Computing*, vol. 30, no. 8, pp. 546-556, 2012.
- [4] Z. Tang, P. Monasse, J. Morel, "Reflexive Symmetry Detection in Single Image," *Proc. of the 8th International Conference Curves and Surfaces*, pp. 452-460, 2014.
- [5] J. Wang, K. Markert, M. Everingham, "Learning Models for Object Recognition from Natural Language Descriptions," *Proc. of the 20th British Machine Vision Conference (BMVC)*, 2009.
- [6] J. Philbin, O. Chum, M. Isard, J. Sivic, A. Zisserman, "Object retrieval with large vocabularies and fast spatial matching," *Proc of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007.
- [7] T. Mochizuki, Y. Kawai, M. Sano, H. Sumiyoshi and M. Fujii, "Visual-Based and Object-Conscious Image Retrieval by Block Reallocation into Object Region," *IEEJ Transactions on Electrical and Electronic Engineering*, vol. 11, no. S2, pp. S44-S52, 2016.