

早期棄却と早期認識を用いた文書画像検索の提案

宮沢 嘉希^{1,a)} 岩田 基^{1,b)} 岩村 雅一^{1,c)} 黄瀬 浩一^{1,d)}

1. はじめに

私たちの生活には、雑誌や図鑑、小説などといった様々な文書があふれている。また、スマートフォン等の普及により、誰もが手軽に望む場面で、写真を撮影することができる。そこで、スマートフォンで撮影した文書を検索にかけて撮影対象を特定する文書画像検索が実現すれば、関連情報を得ることができ、有益である。

このような文書画像検索を実現するアプローチは二つある。一つ目は、LLAH(Locally Likely Arrangement Hashing) [1] などの既存の文書画像検索を用いること。二つ目は画像検索や物体認識などに用いられる SIFT [2] などの汎用の特徴量を利用した検索を用いることである。ただし、前者には絵や写真を多く含む文書を正しく検索できないという問題があり、後者には文書画像検索の約 20 倍もの処理時間を要するという問題点がある。

前者は、文書が多くのテキストを含むと仮定し、その性質を利用することで、高速・高精度を実現している。そのため、同様の性質を持たない絵や写真に適応させるのは困難である。一方、後者が特徴抽出に長い時間を要する原因は、文書画像から得られる特徴量の多さ（数千から数万）である。さらに、SIFT を用いた文書画像検索を調べると、正しい検索のためには、抽出される大量の特徴量は必ずしもすべて必要ではないことが分かった。そこで本稿では、検索に必要なところから必要な分だけの SIFT 特徴量（以下、特徴量と呼ぶ）を取り出して、データベースに登録された特徴量と照合し、文書画像検索を大幅に高速化する手法を提案する。

一万枚の文書画像をデータベースとして用いた実験の結果、検索精度 94%、検索質問画像 1 枚あたりの処理時間 45.6[ms] を得た。この結果は、SIFT と最近傍探索を組み合わせた手法と同程度の検索精度の高さ、LLAH と同程度の処理時間の短さであることから、両者の利点を併せ持つことが分かった。

2. 提案手法

提案手法では、早期棄却と早期認識の 2 つの概念に基づいて高速化を実現する。ここでいう早期棄却とは、文書画像から検索に用いる領域を計算コストの低い処理により限定し、用いない領域を特徴抽出の対象から除外する処理である。また早期認識とは、抽出対象となる領域に順序を付け、順番に特徴抽出・照合・信頼度判定を繰り返し、信頼に足る検索結果が得られた段階で検索を終了し、その結果を出力する処理である。

提案手法の流れを以下に示す。まず早期棄却によって、検索質問画像から検索に用いる領域を決定する。早期棄却を低い計算コストで実現するために、提案手法では、検索質問画像を低解像度化した画像から特徴量を抽出し、特徴量が得られた周辺の領域以外を除外する。これは、低解像度の画像の特徴量ならば非常に少ない計算コストで得ることができ、なおかつ特徴量が得られた周辺の領域は、高解像度でより多くの特徴量が得られる傾向があるためである。

続いて早期認識では、早期棄却で得られた領域から、少ない計算コストで特徴量が得られる領域を優先して、逐次的に特徴抽出と照合を繰り返す。領域の順序付けには二つの基準を用いる。一つ目は、領域の小ささである。これは、小さな領域ほど特徴抽出の計算コストが小さいからである。二つ目は、特徴点検出に用いる多重解像度画像の解像度の低さである。これは、解像度の低い画像からは少ない計算コストで特徴量が得られるからである。上記で得られた各特徴量とデータベース内の特徴量とを近似最近傍探索を用いて照合し、最近傍となった特徴量が属する各画像に投票する。得票数第 1 位の得票数が十分大きく、かつ第 2 位との差が十分についた時点で検索を終了し、検索質問画像に対応するデータベースの画像のファイル名を出力する。

3. 実験

提案手法と、既存の文書画像検索と画像検索とを、検索精度と処理時間の観点から比較する。既存の文書画像検索法として LLAH を用い、既存の画像検索法として SIFT と近似最近傍探索を組み合わせた手法 [3] (以後 SIFT+BDH) ならびにバイナリコードで表される特徴量 ORB [4] によ

¹ 大阪府立大学 大学院工学研究科 〒 599-8531 大阪府堺市中区学園町 1-1

a) miyazawa@m.cs.osakafu-u.ac.jp

b) iwata@cs.osakafu-u.ac.jp

c) masa@cs.osakafu-u.ac.jp

d) kise@cs.osakafu-u.ac.jp

表 1 検索精度に関する比較 [%]

手法	テキスト	写真	グラフ	表	絵	総合
提案手法	100	100	95	75	100	94
LLAH	80	15	60	50	35	48
SIFT+BDH	100	100	100	90	100	98
ORB	50	100	90	85	100	85

表 2 処理時間に関する比較 [ms]

手法	特徴抽出時間	検索時間	総処理時間
提案手法	36.8	8.8	45.6
LLAH	32.8	3.9	36.7
SIFT+ BDH	502.8	208.2	711.0
ORB	371.8	1.16×10^5	1.16×10^5

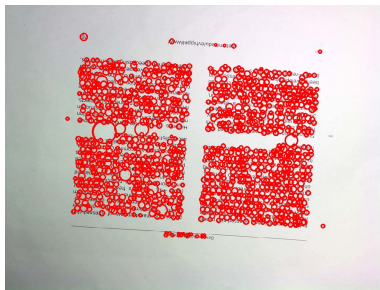


図 1 通常の SIFT によって得られる特徴量

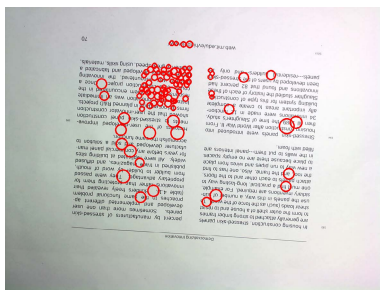


図 2 提案手法で検索に必要な特徴量

る手法を用いた。ORB の照合には、ハミング距離による全探索を用いた。データベースは、1 万枚の文書画像を用いて構成した。その半分はテキストを多く含む論文のような形式の文書であり、残りの半分は写真や絵を多く含む文書である。検索質問画像には、データベースの画像のうち、100 枚を印刷し、それを Web カメラ、Logicool HD Pro Webcam C920t によって正面から撮影したものを用いた。100 枚の検索質問画像は、「テキスト」、「画像」、「グラフ」、「表」、「絵」の計 5 クラスからなり、1 クラスあたりの枚数は 20 枚である。実験に用いた計算機の CPU は、Opteron6238 2.6GHz、メモリは 512GB である。ここで、本実験環境では 10[ms] 以内の計測誤差が生じる。

表 1 に示すように、検索精度は、提案手法が 94% で SIFT + BDH に次いで高かった。しかし、提案手法の「表」に

対する検索精度は 75% で、SIFT + BDH や ORB よりも低かった。表を多く含む文書の場合、罫線やその中に収まる要素名など、異なる画像であるにも関わらず部分的には一致することがある。提案手法は、検索質問画像の一部が十分にデータベース画像と類似する場合、その時点で検索を打ち切るため、そういった部分一致による誤りが生じやすい。表 2 に示す特徴抽出時間ならびに検索時間より、LLAH と提案手法はそれぞれ 40[ms] 以下、10[ms] 以下を記録しており、他の手法と比較して非常に高速であった。この結果より、提案手法が導入した特徴抽出の範囲を減らすことによる特徴抽出時間の削減や、検索に使用する特徴量数を減らすことによる検索時間の削減が有効に働いていることを確認できる。図 1 は、検索質問画像から一般的な SIFT によって得られる特徴量を全て図示したものである。そして、図 2 は、提案手法で検索に成功した際に抽出された特徴量を図示したものである。提案手法は、通常の SIFT と比べて、ごく一部の特征量だけで検索に成功していることがわかる。表 2 に示す総処理時間より、提案手法と LLAH はリアルタイム性を十分に有する。

4. まとめ

本稿では、早期棄却と早期認識を用いて、高速・高精度かつ、幅広い対象に適用可能な文書画像検索手法を提案した。提案手法の特長は、文書画像検索が一部の特征量のみで十分に可能であることに着目し、画像中の必要などから、必要な分だけの特徴抽出をすることで、高精度かつ高速な文書画像検索を実現した点である。

また、他の検索手法と提案手法を比較実験した。提案手法は処理時間 45.6[ms]、検索精度 94% となった。その結果幅広い文書を対象とした高速・高精度な検索という要件に対して、提案手法は、最も高速な文書画像検索と同程度の時間で、最も高精度な SIFT+BDH に匹敵する検索精度を得た。

参考文献

- [1] 中居友弘, 黄瀬浩一, 岩村雅一, “特徴点の局所的配置に基づくデジタルカメラを用いた高速文書画像検索,” 電子情報通信学会論文誌 D, vol.J89-D, no.9, pp.2045-2054, 2006.
- [2] P.C. Ng and S. Henikoff, “Sift: Predicting amino acid changes that affect protein function,” Nucleic acids research, vol.31, no.13, pp.3812-3814, 2003.
- [3] M. Iwamura, T. Sato, and K. Kise, “What is the most efficient way to select nearest neighbor candidates for fast approximate nearest neighbor search?,” Proc. 14th International Conference on Computer Vision (ICCV 2013), pp.3535-3542, dec 2013.
- [4] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, “Orb: An efficient alternative to sift or surf,” Computer Vision (ICCV), 2011 IEEE International Conference on IEEE, pp.2564-2571 2011.