

早期棄却と早期認識による文書画像検索の高速化

宮沢 嘉希[†] 岩田 基[†] 岩村 雅一[†] 黄瀬 浩一[†]

[†] 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

E-mail: †miyazawa@m.cs.osakafu-u.ac.jp, ††{iwata,masa,kise}@cs.osakafu-u.ac.jp

あらまし 高速な文書画像検索手法の一つに、LLAH がある。LLAH は高速・高精度な文書画像検索を実現できる一方、図表やグラフを多く含む文書を対象としたとき、精度が著しく低下するという欠点がある。一方、幅広い対象に適用可能な画像検索手法として、SIFT などの局所特徴量を用いた手法があるが、低速であるという欠点がある。本稿では、局所特徴量を用いた画像検索手法に、早期認識と早期棄却の二つの概念を導入することで、幅広い対象に適用可能で、なおかつ高速・高精度な文書画像検索手法を提案する。早期認識は、計算コストの少ない処理を繰り返し、検索に十分な証拠が得られた時点で処理を終了することで達成する。早期棄却は、少ない計算コストの処理から、検索に有効ではない特徴抽出の範囲を決定し、それを破棄することで実現する。実験の結果、1 万ページのデータベースに対して、精度 94%、処理時間 45.6[ms] で検索可能であることが分かった。

キーワード 文書画像検索, 画像検索, LLAH, 特徴量

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できれば、関連情報を得ることができ有益であると考えられる。

このような文書画像検索を実現する手法の一つに LLAH(Locally Likely Arrangement Hashing) [1] がある。LLAH は、高速かつ高精度な検索が可能な手法であるものの、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る。文書画像検索の実用化を考えた場合、このような文書も検索できることが望ましい。

この問題を解決する一つの方策は、画像検索や物体認識に使用される SIFT [2] などの汎用の特徴量を使うことである。実際に用いてみると、文書の内容によらず高い検索精度を得ることができる。しかし、LLAH の特徴抽出と比較して、SIFT は特徴抽出に約 20 倍もの時間がかかるという問題がある。これは、文書画像が文字などの微細な構成要素を多数含み、これらから大量の特徴量が抽出されるためである。

SIFT を用いた文書画像検索の様子をみると、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わかる。むしろ、ほんの一部を使うだけでも、十分な精度で検索できる。そこで本稿では、検索に有効な領域から必要な分だけの特徴量を取り出して照合することにより、文書画像検索を大幅に高速化する手法を提案する。

高速化の処理は、早期棄却と早期認識の 2 つの概念に基づく。ここでいう早期棄却は、文書画像から特徴量を取り出せない領

域を計算コストの低い処理により発見し、特徴抽出の対象から除外する処理である。これにより計算コストのかかる特徴抽出を、特徴量を取り出せる領域に限定できる。さらに、早期認識の概念に基づき、抽出対象となる領域に順序を付け、順番に特徴抽出・検索・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処理を打ち切ることによってさらなる高速化を実現する。実験の結果、精度 94%、検索質問 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精度、LLAH と同程度の処理時間であることから、両者の利点を併せ持つことが分かった。

以下、2 節では関連研究、3 節では提案手法、4 節では実験、5 章では全体のまとめを述べる。

2. 関連研究

2.1 LLAH

文書画像検索の手法の一つに、LLAH がある。LLAH は、単語の連結成分の重心を特徴点とし、リアルタイム検索が可能な手法である。また、射影歪みや隠れ、紙面の湾曲にロバストであるという特徴もある。しかし、LLAH は単語の連結成分の重心を特徴点としているため、写真や絵、グラフといった文章以外の要素が多くを占める文書を検索できないという短所がある。

2.2 局所特徴量

画像検索に用いられる特徴量の一つに、SIFT や、SURF [3] などの局所特徴量がある。SIFT は、ガウシアン平滑化画像の差分である DoG 画像を複数枚用いて、特徴点を検出する。スケール変化や、回転にロバストな特徴量であるが、DoG 画像の生成に多くの計算コストがかかり、特徴抽出に多くの時間がかかるという欠点がある。SURF は近似計算を導入して SIFT

を高速化した特徴量であるが、リアルタイム検索に十分なほどではなく、スケール変化や照明変化に対する頑健性が低下している。

バイナリコードによって表現される特徴量に FREAK [4] や、BRISK [5] がある。これらの特徴量は、省メモリで特徴量を表現でき、また特徴量間の距離はハミング距離で定義されるため、類似度を高速に計算できる。一方で、スケール変化に弱く、特徴抽出時間もリアルタイム検索をする上では、十分に高速とはいえない。

2.3 早期認識, 早期棄却

早期認識とは、計算コストの少ない処理を優先し、認識に十分な証拠が得られた段階で、処理を打ち切るという考え方である。このような早期認識を用いた手法に、野口らの手法 [6] がある。野口らの手法は、ハッシュの多段階化によって近似最近傍探索の探索範囲を絞り込み、少しずつ探索範囲を広げることで、高速・高精度、かつ大規模な検索を可能とした。

早期棄却とは、計算コストの少ない処理を優先し、残りの処理が必要ないと判断された場合は、残りの処理を棄却するという考え方である。このような早期棄却を用いた手法に、顔検出などに用いられる Viola-Jones の手法 [7] がある。Viola-Jones の手法は、Haar-Like 特徴量による弱識別器を連結させた、カスケード構造の識別器を用いる。早い段階の弱識別器によって検出対象がないことを判断できるため、後の識別器の処理を、棄却でき、高速な検出が可能となる。

3. 提案手法

3.1 設計指針

検索に有効な領域から、必要な分だけの特徴量を取り出して照合することで、大幅な高速化を図る。本稿では、検索に有効な領域を判断する処理として早期棄却を、必要な分だけの特徴量を用いて認識する処理として早期認識を導入する。早期棄却では、画像中の検索に有効な領域を判断することで、検索に役立たない領域の特徴抽出を棄却し、処理時間を削減する。早期認識では、早期棄却において検索に有効と判断された領域から、検索に必要な分だけ特徴抽出することで、特徴抽出の計算コストを削減する。

3.2 構成

全体の構成を以下に示す。まず早期棄却では、検索質問画像から検索に有効な領域を得る。続いて早期認識では、早期棄却で得られた検索に有効な領域から少しずつ特徴抽出し、その都度照合する。照合は、各特徴量について近似最近傍探索を用いてデータベースから最近傍を求める。そして、最近傍となった特徴量が属する画像に投票する。投票結果から、認識結果の信頼性を測り、十分な証拠が得られた時点で、認識処理を終了し、検索結果である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。本稿では、特徴量として SIFT、近似最近傍探索として BDH [8] を用いる。

3.3 早期棄却

3.3.1 早期棄却の方針

早期棄却では、全体の処理時間を削減するために、十分に高

速な処理によって、検索に有効な領域を判断する必要がある。特徴抽出は多くの計算コストを必要とするが、その処理時間は画像の解像度に大きく依存する。よって、少ない計算コストで得られる低解像度の特徴量を用いて、検索に有効な領域を判断する。

3.3.2 早期棄却の流れ

早期棄却の流れを以下に示す。まず、検索質問画像を縮小した低解像度の画像を生成する。縮小の倍率は m とする。低解像度の画像で特徴量が得られた座標を、元の検索質問画像に対応させ、その周辺を検索に有効な領域とする。具体的には、低解像度で得られた特徴量の座標を (x, y) 、検索に有効な領域の中心の座標を (X, Y) とする。このとき、検索に有効とされる領域の中心の座標は以下のように定める。

$$(X, Y) = m \times (x, y) \quad (1)$$

低解像度の画像で特徴が全く得られない領域は、文章の余白など、高解像度でも特徴量が得られないことが多いため、このような方法で検索に有効な領域を定めている。低解像度で得られた特徴量の大きさは、そのままその特徴量が表現する範囲の大きさを示す。そこで、検索に有効な領域の大きさは、低解像度の画像から得られた特徴量の大きさに基いて定める。また、検索に有効な領域は、実装の都合上正方形とした。具体的な領域の大きさは、低解像度の画像で得られた特徴量の大きさを s 、検索に有効な領域の大きさを S とし、以下の用に定める。

$$S = s \times K \times m \quad (2)$$

ここで、 K は定数である。低解像度で得られた特徴量を、データベースと照合し、その投票結果のみで十分な証拠が得られた場合は、その時点で認識を終了する。十分な証拠が得られたかの判断の方法は、3.4.2 に後述する。

3.4 早期認識

3.4.1 早期認識の方針

早期認識のポイントは、検索の精度と特徴抽出の計算コストのバランスを保つことである。従来の特徴抽出は、一度に全ての特徴量を得る。そのため、特徴抽出では必ず多くの計算コストが必要となり、処理時間が増加してしまっていた。しかし、本稿では、少量の特徴量を得る処理を繰り返し、その都度照合する。そして、投票結果から十分な証拠が得られた時点で認識処理を終了するため、計算コストを低く抑えられる。

少量の特徴量を得る処理を繰り返すとき、少ない計算コストでより多くの特徴量が得られる特徴抽出が優先されるのが望ましい。そのために、以下の二つの指標を用いる。一つ目は、早期棄却で得られる検索に有効な領域の大きさ、 S である。特徴抽出の計算コストは、対象の画像の大きさに大きく依存する。よって、小さな領域から優先的に特徴抽出することで、処理時間の増加を防ぐ。この一つ目の処理を、領域の絞り込みと呼ぶ。二つ目は、特徴点検出に用いる DoG 画像のスケールである。解像度の低い DoG 画像を優先することで、少ない計算コストで特徴点を得られる処理が優先される。この二つ目の処理を、スケールの絞り込みと呼ぶ。

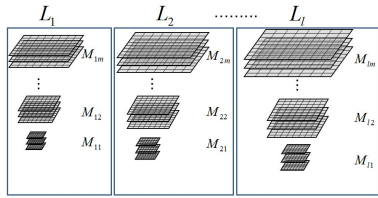


図1 L, M の関係

3.4.2 早期認識の流れ

早期認識の具体的な流れを以下に示す。まず、早期棄却で得られた検索に有効な領域を、大きさの小さい順にソートして得たものを、 $L_1, L_2, \dots, L_i, \dots, L_l$ とする。続いて、 L_i から得られる DoG 画像を解像度の低いものから順に、 $M_{i1}, M_{i2}, \dots, M_{ij}, \dots, M_{im}$ とする。 L, M の関係を図1に示す。はじめに、 L_1 から、 M_{11} を生成し、特徴抽出と照合をする。十分な証拠が得られなかった場合、 M_{12} の生成と特徴抽出をし、以後このような処理を繰り返す。 M_{1m} まで処理が進んでも十分な証拠が得られない場合は、 L_2 へと移行する。 L_2 でも、 L_1 と同様の処理を繰り返し、十分な証拠が得られない場合、 L_l まで続ける。 L_l の照合まで終了した場合、得票数1位の画像を検索結果とする。また、終了判定で十分な証拠が得られたと判断された場合、その時点での得票数1位の画像を検索結果とする。

十分な証拠が得られたかを調べる終了判定は、検索の精度と、全体の処理時間を左右する重要な処理である。終了判定を高速に行うために、次のような方法を用いる。得票数1位と2位の画像の得票数を、それぞれ u_1, u_2 とする。これらが $u_1 > T$ と $Ru_1 > u_2$ をともに満たすとき、その時点で認識処理を終了する。ここで、 T と R はあらかじめ定めておく定数である。この式は、1位の得票数が十分大きく、かつ2位との差が十分あれば、十分な証拠が得られていると判定するものである。

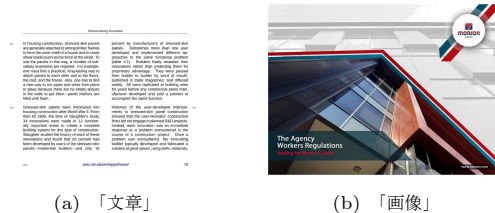
4. 実験

4.1 実験条件

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、文章を多く含む論文のような形式の文書である。もう半分は、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、LogicoolR 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、「文章」、「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。それぞれのクラスに属するデータベースの画像の例を図2に示す。

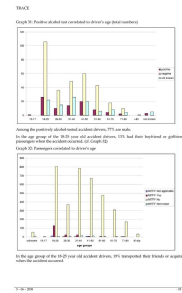
実験に用いた計算機の CPU は、Opteron6238 2.6GHz、メモリは512GBである。予備実験より提案手法では、 $m = 4$ 、 $K = 15$ とした。

実験1では、SIFTと最近傍探索を組み合わせた手法に対して、近似最近傍探索、領域の絞り込み、スケールの絞り込み、を順番に加え、どの処理がどの程度、検索精度と処理時間に寄与しているのかを調べた。



(a) 「文章」

(b) 「画像」



(c) 「グラフ」

(d) 「表」



(e) 「絵」

図2 データベース画像の例

表1 実験1の精度、時間の比較、時間の単位は [ms]

手法	特徴抽出時間	検索時間	総処理時間	精度 [%]
SIFT+NN	503.2	4.82×10^6	4.82×10^6	99
SIFT+ANN	502.8	208.2	711.0	98
Area	43.8	14.1	57.9	95
Proposed	36.8	8.8	45.6	94

実験2では、文書画像検索の一手法である LLAH、省メモリかつ高速なマッチングが可能とされるバイナリコードで表される特徴量を用いた検索手法と提案手法を比較した。また、本実験では、10[ms]以内の計測時間を誤差として、計測した。

4.2 実験1

実験1の結果を、表1に示す。処理時間の単位は [ms]、精度は [%] である。ここで、精度とは検索質問画像100枚のうち、正しい検索結果が得られた割合である。表1では、SIFTと最近傍探索を組み合わせたものを SIFT+NN、SIFTと近似最近傍探索を組み合わせたものを SIFT+ANN、と表す。また、近似最近傍探索と領域の絞り込みのみをしたものを Area、近似最近傍探索と領域、スケールの絞り込みを両方用いたものが提案手法に相当し、Proposed、と表す。

精度について述べる。提案手法は、精度94%となり、SIFT+ANNと比較しても、大きな精度の減少はなかった。よって、提案手法では、検索に必要なところから、必要な分だけの特徴量を十分に得られていることが分かる。

特徴抽出時間について述べる。提案手法は、SIFT+ANNと比較して、大幅に特徴抽出時間を削減できた。これは、提案手法において、早期認識と早期棄却が有効に働いているからで

表 2 実験 2 の精度に関する比較 [%]

手法	文章	写真	グラフ	表	絵	総合
Proposed	100	100	95	75	100	94
LLAH	80	15	60	50	35	48
SIFT+ANN	100	100	100	90	100	98
ORB	50	100	90	85	100	85

表 3 実験 2 の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
Proposed	36.8	8.8	45.6
LLAH	32.8	3.9	36.7
SIFT+ANN	502.8	208.2	711.0
ORB	371.8	1.16×10^5	1.16×10^5

ある。また、領域の絞り込みとスケールの絞り込みの両方の処理が特徴抽出時間の削減に寄与していることが、表 1 から分かる。

検索時間について述べる。提案手法は、SIFT + ANN と比較して、検索時間が大幅に削減できた。提案手法は、SIFT + ANN と同様の近似最近傍探索を用いる。この検索時間の減少の原因は、検索に使用する特徴量の数を削減できたからである。早期認識では十分な証拠が得られた時点で検索を終了する。その結果、検索に用いる特徴点数が大幅に減少する。検索時間は、特徴量の数の減少によって線形的に短くなる。また領域の絞り込みとスケールの絞り込みの両方が、検索時間の削減に有効であることが分かる。

総処理時間について述べる。提案手法は、SIFT + ANN の約 94% の総処理時間を削減した。そして、提案手法は、リアルタイム処理が十分に可能であると言える。

以上の実験結果、考察から、提案手法は、SIFT+ANN と比べて、精度を 94% に維持しながら、全体の処理時間を約 94% 削減できることが分かった。

4.3 実験 2

実験 2 では、既存の文書画像検索手法や画像検索の特徴量を用いた検索手法と、提案手法を比較した。比較手法として、文書画像検索手法の一つである LLAH、画像検索の手法である SIFT に近似最近傍探索を組み合わせたもの、バイナリコードで表される特徴量である ORB [9] を用いた検索手法を用いた。ORB は、パッチ内の 2 点の輝度差を用いてバイナリコードを生成するが、2 点のペアが学習によって選択されるという特徴がある。ORB における、特徴量のマッチングには、最近傍探索を用いた。検索精度と、処理時間を、表 2、表 3 に示す。表 2、表 3 においては、ORB と最近傍探索を組み合わせたものを ORB と表す。

表 2 に示すように、検索精度は、提案手法が SIFT + ANN に次いで高かった。LLAH は特に、写真や絵を多く含む文書に対して、正しく検索するのが困難であった。これは、LLAH が単語の連結成分の重心を特徴点としているため、写真や絵からは安定した特徴点を得られなかったからだと考えられる。ORB を用いた場合は、論文のような文章を多く含む検索質問画像の精度が 50% と低くなった。図 2(a) は、ORB での検索に失敗

したもの一つである。この画像と正解のデータベース画像を特徴抽出し、最も距離の近い特徴量同士を照合させたところ、画像中の異なる場所同士が数多くマッチングした。風景や物体同士の場合と異なり、文書画像に対しては、ORB は正しく対応付けるのが困難であることが分かった。一方で、提案手法の「表」に対する精度は 75% で、SIFT + ANN 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データベース画像が共通する部分を持つことが多い。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の失敗が生じやすいと考えられる。検索質問画像全体では、精度 94% で、他の手法と比較しても高い精度を実現できた。

表 3 に示す特徴抽出時間より、ORB 以外の手法は、非常に高速であった。提案手法は、特徴抽出の範囲を減らすことで、LLAH のような特徴抽出が高速な手法に匹敵するほど処理時間を削減できた。

検索時間は、表 3 に示すように、ORB が他の手法と比較して処理時間が長かった。このことから、ORB は、ハミング距離を用いることでユークリッド距離と比較して、マッチングの高速化を図っているが、それ以上に他の手法の近似最近傍探索が高速であることが分かる。また、提案手法は、検索に使用する特徴量の数が少ないため、検索時間を削減することができた。

表 3 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している。

以上の実験結果と考察から、幅広い対象に適用可能で、かつ高速・高精度という点において、提案手法が最も優れている。

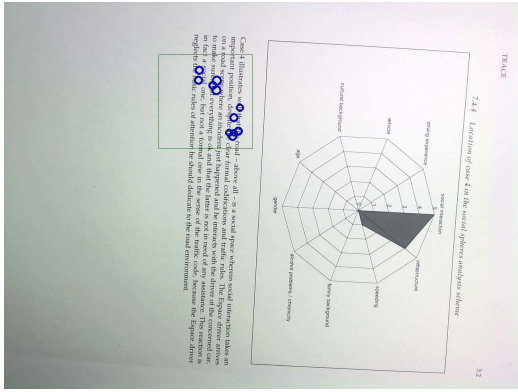
4.4 考察

提案手法の早期棄却、早期認識、終了判定といった各処理が、精度や処理時間にどのように寄与したのかを以下に示す。また、提案手法の処理時間が何に依存して変化するかを調べた。

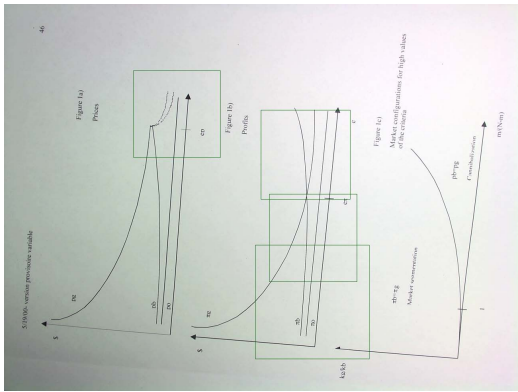
4.4.1 早期棄却

検索質問画像中の大部分を文字や写真が占め、特徴量が多く発生する文書であれば、その一部の特徴量を用いることで、大幅な処理時間の短縮ができる。実際に早期棄却が成功した検索質問画像を、図 3(a) に示す。画像中の矩形は、早期棄却で検索に有効とされ特徴抽出された領域を示す。矩形は画像全体と比較して、ごく一部の領域である。しかし、この領域から得られる特徴量だけでも検索には十分であり、大幅な特徴抽出時間の削減ができた。

一方で、早期棄却が有効に働かなかった検索質問画像の例を、図 3(b) に示す。図 3(b) は、早期棄却で検索に有効と判断された全ての領域から特徴抽出をしているが、正しく検索できなかった。この画像が検索に失敗した原因は、早期棄却によって検索に必要な特徴量を含む領域を棄却したことである。提案手法は、このような低解像度で特徴量が出ない場所に、検索に必要な特徴量が含まれる画像に対しての対応が必要である。特に現在の早期棄却の、低解像度で特徴量が出た領域を有効とするという指標以外の、検索に有効な領域をより正確に判断する方法が必要である。



(a) 有効な画像



(b) 有効ではない画像

図 3 早期棄却について

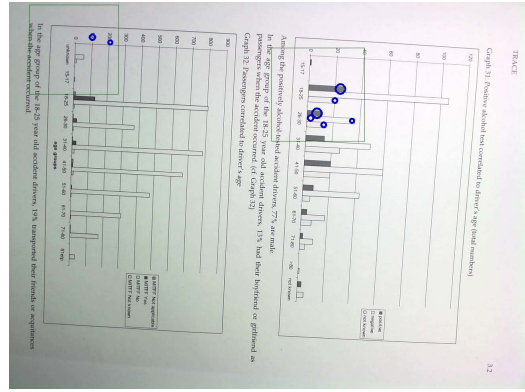
4.4.2 早期認識

実際に早期認識が有効に働いた検索質問画像を図 4(a) に示す。この画像では、左上の矩形からは検索に有効な特徴量あまり得られなかった。そこで処理を継続し、中央上部の矩形で多くの正解の特徴量が得られた時点で、早期に検索を終了し、処理時間を削減できた。早期棄却で得られた領域のうちから、検索に必要な分だけの特徴抽出をしたことが分かる。

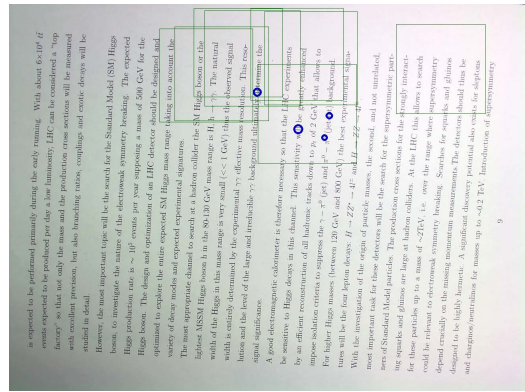
しかし、提案手法は特徴抽出する領域の重複について考慮していない。つまり、検索に有効だと判断された領域が密接して存在し、そこから検索に有効な特徴量が得られない場合、何度も検索に役立たない特徴抽出を繰り返す。その結果、検索終了までに多くの処理時間を要してしまう。図 4(b) は、そういった例である。この画像では、中央上部で多くの矩形が重なっている。つまり、重複した領域から何度も同じ特徴量を抽出し、誤ったデータベース内の画像に投票を繰り返したということである。そのため、この画像の検索には、約 200[ms] 秒もの処理時間が必要となった。SIFT+ANN では、この画像の検索には約 700[ms] の時間がかかる。よって高速化はできているが、重複した部分からの特徴抽出を防ぐことで、さらなる処理時間の短縮が可能である。また、このような重複部分が原因の検索誤りもあったため、精度向上にも有効である。

4.4.3 終了判定

4.3 節の検索精度に関する実験結果の項で触れたように、「表」のような特徴を持つ画像ではデータベース内に検索質問画像と



(a) 有効に働いた画像



(b) 有効に働かなかった画像

図 4 早期認識について

類似した見た目の部分を含む不正解の画像があることが多く、正解の画像とその画像に票が分散し、検索を失敗しやすい。

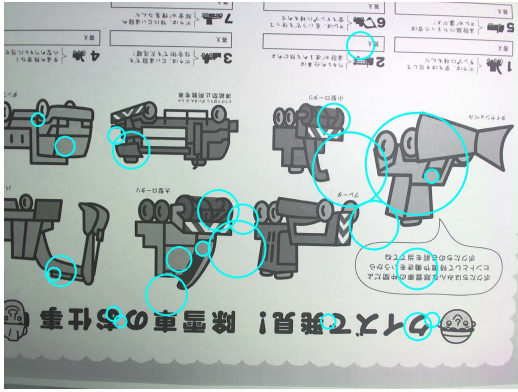
「絵」や「写真」では見た目が類似したデータベースの画像がなく、終了判定の失敗はなかった。例えば、図 5(a) は低解像度の特徴量のみで、終了判定の条件を満たした検索質問画像の例である。処理時間は 10[ms] 以下であり、終了判定が有効に機能している。

しかし、「表」では似通った見た目の部分が複数のデータベースの画像に含まれ、そこから類似した特徴量が発生し、検索の失敗につながった。図 5(b) はその例である。データベースに、検索質問画像と同様のフォント、表の形態を持つ画像が含まれ、そこに誤投票が集中した。局所的には見た目が類似した部分を持つ画像同士でも、レイアウトや画像内での特徴量の位置関係は異なる。検索の失敗への対応として、単純な票数の多さだけでなく、そういった点を終了条件に加える事で、検索の失敗を防ぐことができる。

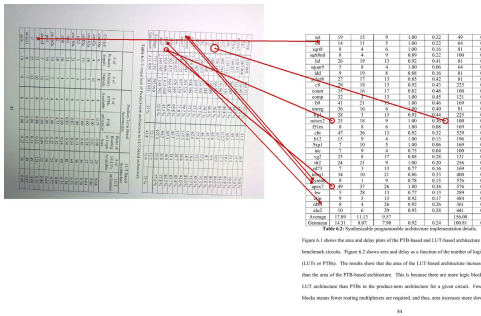
4.4.4 処理時間

提案手法の処理時間が何に依存しているのかを示すために、図 6 と図 7 を用いる。図 6 の横軸は検索質問画像から得られた特徴点数、縦軸は処理時間を示し、全ての検索質問画像における値をプロットしたものである。図 6 より、特徴量の数が多いと、線形的に処理時間が増加する傾向があることが分かる。

また図 7 の横軸は、早期認識において特徴抽出をした領域の面積（以下、処理面積）の総和、縦軸は処理時間を示し、全て



(a) 有効な画像



(b) 有効ではない画像

図5 終了判定について

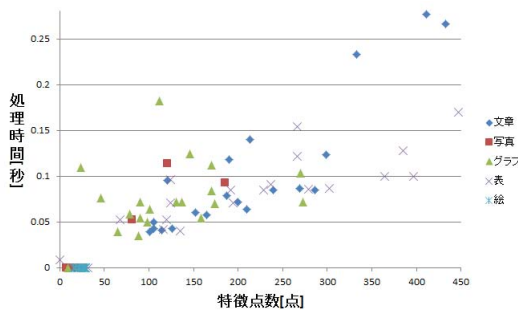


図6 特徴点数と処理時間の関係

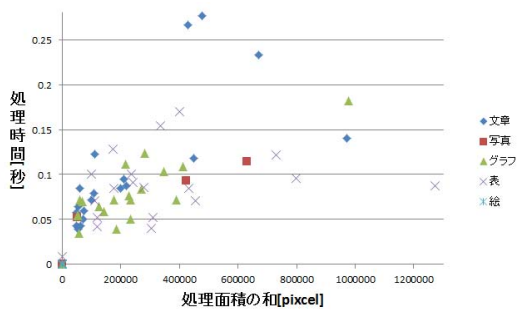


図7 処理面積と処理時間の関係

の検索質問画像における値をプロットしたものである。早期棄却における低解像度の特徴抽出の面積は全ての検索質問画像について共通であるため、図7の処理面積には含めない。処理面積が大きくなるほどに、処理時間が増加する傾向にあることが分かる。

図6, 図7より、検索の高速化には、精度を保てる範囲内で、より少ない特徴量と少ない処理面積を用いて検索する必要があることが分かる。

5. まとめ

本稿では、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、画像検索が一部の特徴量のみで十分に可能であることに着目し、画像中の必要箇所から、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。

SIFTと最近傍探索を組み合わせた最もシンプルな検索手法に、一つずつ処理を加え、それぞれの高速化の工夫がどのように検索精度と処理時間に寄与しているのかを調べた。その結果、提案手法は精度94%を維持しながら、1枚あたり、処理時間45.6[ms]を実現できることが分かった。また、文書画像検索手法、画像検索手法、バイナリコードで表される特徴量を用いた場合の検索手法と提案手法を比較実験した。その結果、幅広い文書を対象とした高速・高精度な検索手法という要件に対して、提案手法が最も優れているということが分かった。今後の課題には、パラメータの値を適用的に変更すること、早期棄却のための解像度の多重化などがある。

謝辞 本研究の一部は、JST CREST および日本学術振興会科学研究費補助金基盤研究(A)(25240028)の補助による。

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] H. Bay, T. Tuytelaars, and L. Van Gool, “Surf: Speeded up robust features,” Computer Vision–ECCV 2006, pp.404–417, 2006.
- [4] A. Alahi, R. Ortiz, and P. Vandergheynst, “Freak: Fast retina keypoint,” Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on IEEE, pp.510–517 2012.
- [5] S. Leutenegger, M. Chli, and R.Y. Siegwart, “Brisk: Binary robust invariant scalable keypoints,” Computer Vision (ICCV), 2011 IEEE International Conference on IEEE, pp.2548–2555 2011.
- [6] 野口和人, 黄瀬浩一, 岩村雅一, “近似最近傍探索の多段階化による高速特定物認識,” 電子情報通信学会論文誌 D, vol.J92-D, no.12, pp.2238–2248, 2009.
- [7] P. VIOLA and M.J. JONES, “Robust real-time face detection,” International Journal of Computer Vision, vol.57, no.2, pp.137–154, 2004.
- [8] 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.
- [9] 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.