社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

画像検索とのハイブリッド方式による文書画像検索の性能向上

杉本　恭隆† 　岩田　　基†† 　黄瀬　浩一††

† 大阪府立大学 工学部
〒 599–8531 大阪府堺市中区学園町 1–1
†† 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
E-mail: †sugimoto_k@m.cs.osakafu-u.ac.jp, ††{iwata, kise}@cs.osakafu-u.ac.jp

あらまし　文書画像検索の手法の 1 つとして，LLAH というものがある．LLAH を用いると，文書を撮影した画像から写っている文書を高速で検索できる．しかし，LLAH では撮影した画像に図や表などが多く含まれる場合，検索が失敗するといった問題がある．主な原因は，LLAH は質問画像が文章であるといった前提があるからである．検索時に利用する特徴点は単語の重心を用いているため，文書でない図や表の部分からは特徴点が出ない．この問題を解決する方法として，画像検索をすることが挙げられる．画像検索を用いることによって，文書だけでなく図や表を入力として検索できる．しかし，LLAH と比べ処理時間がかかるといった問題もある．本稿では，検索対象に対して効率の良い検索方法を適用することで，高精度かつ高速な検索を可能にする手法を提案する．そして LLAH と画像検索のそれぞれを比較し，性能の向上を示す．

キーワード　文書画像検索，LLAH，画像検索，局所特徴量

1. はじめに

近年，デジタルカメラやカメラ付き携帯電話の普及により，大量の画像を手軽に撮影・活用できるようになり，持ち運びも簡単にできるようになった．それに伴い，カメラを用いて撮影した画像を情報デバイスとして利用することができれば有用である．情報デバイスとしての利用法として，Layered Reading [1] という情報サービスがある．これは新たな読書体験とビジネスを創出する電子書籍アイデアである．電子書籍の紙面上に新しいレイヤを設け，付加情報を重ねて表示するというものである．また，カメラペンを用いて，紙文章に書いた文字をデジタルデータで扱うサービスがある [2]．このように，撮影画像を情報デバイスとして利用することによって，日常生活をより豊かなものにできると考えられる．こういった場合では，まず撮影画像から対応する文書を検索する必要がある．このような検索については様々な研究がなされている．その中でも，撮影した画像に文章が含まれており，どの文書をどのように撮影したのかを検索する文書画像検索という分野が注目されている．

文書画像検索の 1 手法として，高速かつ正確に検索できる方法として Locally Likely Arrangement Hashing (LLAH) [3] という手法がある．LLAH では文章の単語の重心を特徴点とし，他の近傍点との関係より特徴量を算出して検索する．よって，計算量が少なく処理時間が高速である．しかし，LLAH では正しく検索できない文書画像も存在する．たとえば，文書を撮影した画像には，図や表などの部分が含まれることがある．文書の図や表の部分を撮影した画像から検索する場合，正しい結果が得られないことがある．LLAH は文書を撮影した画像のような質問画像が文章であるという前提があり，文章の単語の重心を特徴点とし，他の近傍点との関係より特徴量を算出し検索するため，文章以外の領域からは検索に有効な特徴点を得ることができない．

この問題を解決する方法として，文書の図や表の領域には画像検索をすることが挙げられる．画像検索とは，一般的な画像を検索できる手法であり，最も単純な手法は投票に基づくものである [4] [5] [6]．そのような画像検索を用いることによって，文書の図や表の部分にも対応することができる．しかし，LLAH と比べ処理時間がかかるという問題がある．

本稿では，LLAH と一般的な画像検索とを組み合わせ，お互いの利点を活かす新たな手法を提案する．この手法では，まず画像が与えられたとき，単語の並びを求めて並んでいる単語に直線を引く．その直線の数によって LLAH と画像検索のどちらを利用するかの切り替えをすることにより，検索の効率化を図る．また，精度を向上させるため，画像を分割してそれぞれに切り替えをする．実験によって，精度と速度を従来法と比較し，性能向上を確認した．

以下，2 章で関連手法について述べ，3 章で LLAH と画像検索の処理の流れ，4 章で LLAH を用いた特徴量抽出，5 章で画像検索を用いた特徴量抽出について述べる．そして 6 章で提案する組み合わせ方式について述べ，7 章で実験と結果，8 章でまとめとする．

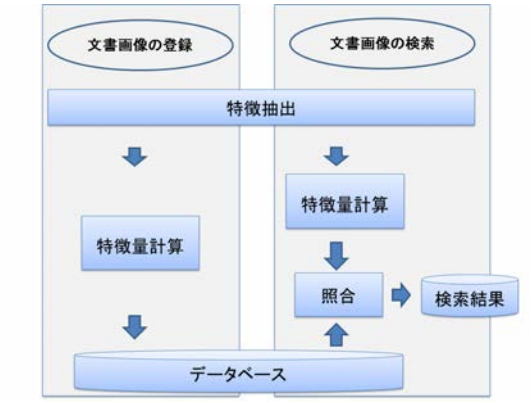


図 1　LLAH の処理の概要

2. 関 連 手 法

文書画像検索の手法では，LLAH の他に様々な研究がなされている．しかし，LLAH と他の文書画像検索 [7] [8] [9] [10] とで大きく異なる点は，画像が受ける種々の歪みに頑健であるということである．カメラにより撮影した文書画像は射影変換による歪みを被る．また，デジタルカメラ，特に携帯電話に付属のものの特性を考えると，部分的に取得された文書画像からも検索可能 (部分検索が可能) でなければならない．こういった点で，本手法では LLAH を使用するのが最適である．

また画像検索の手法でも，局所特徴量を使用するほかにも，研究がなされている．画像の照合に用いられる特徴量は大きく大域特徴量と局所特徴量に分類できる．大域特徴量とは，画像全体から抽出される特徴量であり，局所特徴量とは，画像の一部分から抽出される特徴量である．有名な例として SIFT(Scale-Invariant Feature Transform) [11] が挙げられる．局所特徴量を抽出するには，特徴量を取り出す局所領域を決定し，その領域から特徴量を抽出するという 2 段階の処理が必要になる．大域，局所の両者とも特徴量はベクトル（特徴ベクトル）として表現される．これを照合することによって，画像検索が可能となる．局所特徴量の局所領域の決定に対応する処理が大域特徴量の抽出では不要であるため，大域特徴量の抽出は計算量的に有利である．また，特徴量が画像 1 枚あたり 1 つであるため，画像の索引付けに用いるデータ量としても大幅に少なくて済むという利点もある．一方で，画像の一部分が隠れによって得られない場合，大域特徴量ではもはや同じ値を得ることは不可能という問題もある．局所特徴量を用いると，隠れによって得られない特徴量があっても，隠れていない部分から依然として同じ特徴量を得ることができるという利点がある．これは，文書を撮影した際に，文書の一部分が欠けていても，欠けていない画像と同じ場所から同じ特徴量が得られるということである．よって，文書を撮影した画像を質問画像とする画像検索に有効である．以上より，本研究では局所特徴量を用いている野口らの手法 [6] を用いる．

3. LLAH と画像検索の処理の流れ

まずは，LLAH の処理の流れを説明する．概要を図 1 に示す．登録処理は，データベース作成時の処理であり，文書をデータベースに登録する処理である．検索処理は，質問画像とデータベースを照合し，検索結果を求める処理である．文書画像を特徴点の集合に変換し，特徴量を計算する．計算された特徴量は，データベース作成時は登録処理，質問画像から検索時は検索処理に入力される．登録処理ではハッシュ表を用い，文書の ID，特徴点の ID，対応する特徴量を 1 つの組としてデータベースに登録する．検索処理ではデータベースにアクセスし，データベースに登録されている特徴量と，質問画像から得られる特徴量を比較し，投票処理によって対応する文書画像を検索結果として出力する．

次に，画像検索の処理の流れを説明する．概要の流れは，特徴点や特徴量の求め方を除いて，図 1 の LLAH と同じである．まず，画像から局所特徴量を求める．そして，登録時は，検索対象の画像から得た特徴量をデータベースに記録する．検索時は，得られる特徴量とデータベースに登録されている特徴量を照合する．そして，照合結果に基づく投票により検索する．

4. LLAH を用いた特徴量抽出

4.1 特徴点の抽出

特徴点抽出の処理を以下に示す．まず，入力画像（図 2(a)）は適応 2 値化され，2 値画像（図 2(b)）に変換される．次に，ガウシアンフィルタで画像をぼかし，再度適応 2 値化を施すことで，単語ごとに結合された画像（図 2(c)）が得られる．最後に，単語領域の重心（図 2(d)）が特徴点として抽出される．

4.2 特徴量計算

LLAH の処理に用いる特徴量は，以下の 2 つの条件を満たす必要がある．1 つは, 同一の特徴点からは, 同一の特徴量が得られなければならないというものである．同一の特徴点であるにもかかわらず, 文書を登録している処理と質問画像から検索する処理で異なる特徴量が得られた場合, それらは違うと特徴点と判断され，対応付けられない．そのため, 正しい検索結果を得ることができなくなる．もう 1 つの条件は, 異なる特徴点からは異なる特徴量が得られなければならないというものである．もし異なる特徴点から同一の特徴量が得られた場合, その異なる特徴点が対応付けられるので，誤ったものも検索結果として得られることになるからである．この 2 つの条件を満たす方法として，アフィン不変量を用いる方法がある．

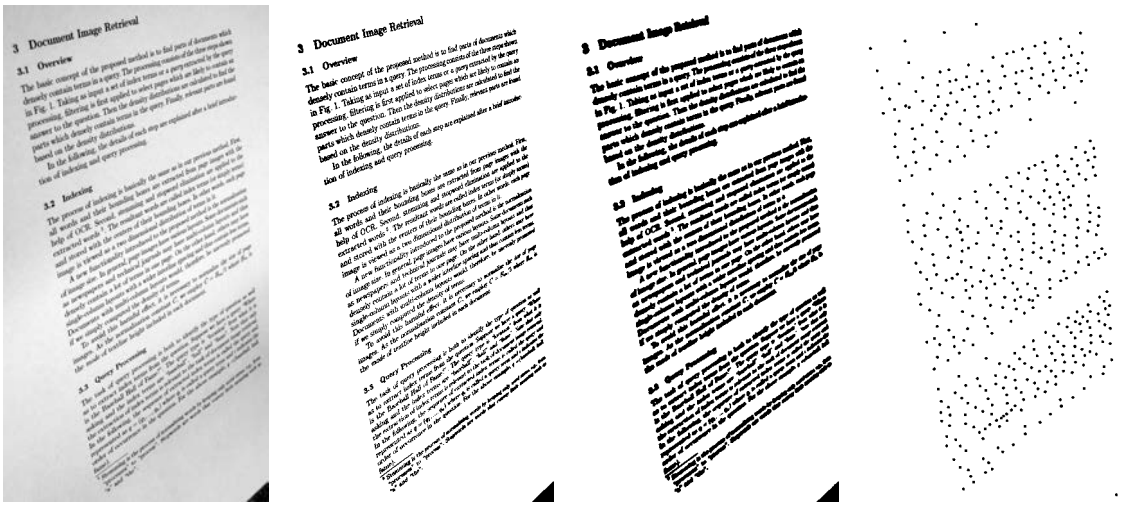
アフィン不変量とは，ある特徴点の近傍 4 点から求められる値である．近傍 4 点を $ABCD$ とするとき，アフィン不変量は以下の式で導出される．

$$\frac{P(A,C,D)}{P(A,B,C)} \tag{1}$$

ここで，$P(A,B,C)$ とは，ABC を頂点とする三角形の面積である．よって，アフィン不変量は三角形の比によって導出される．

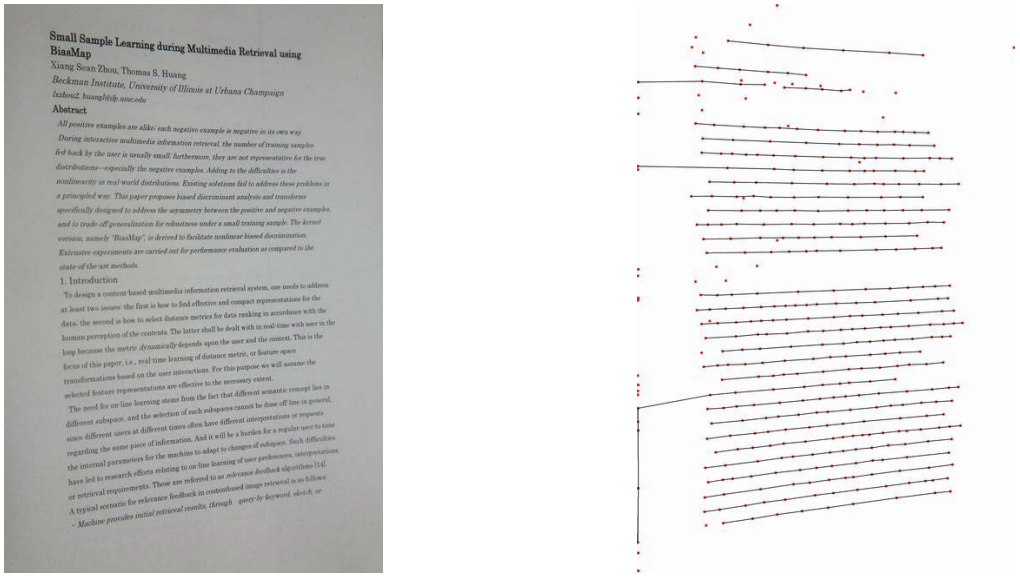
4.3 登録処理と検索処理

登録処理では，まず，得られた特徴量をハッシュ関数によってハッシュ表のインデックスに変換する．そして，データベースに文書 ID，特徴点 ID と特徴量の組をインデックスとして



(a) 入力画像　(b) 2 値画像　(c) 連結成分　(d) 特徴点

図 2　特徴点抽出



(a) 文書画像　(b) 単語の並び

図 3　画像検索に用いる行

ハッシュ表に登録する．この処理によって，得られた特徴量に文書と特徴点の位置の情報を付加している．

次に検索処理について説明する．検索処理では，まず，ハッシュ関数によって特徴量からハッシュ表のインデックスを求め，ハッシュ表にアクセスする．そして，登録されている文書 ID に投票する．この処理を全ての点について繰り返して，最終的に最も得票数を得た文書を検索結果とする．

5.　画像検索に用いる特徴量抽出

本章では，野口らの手法で用いる特徴量抽出について説明する．まず検索に使用する局所特徴量について述べ，登録と検索の順で説明する．

5.1　局所特徴量計算

ここでは，局所特徴量と計算法について説明する．局所特徴量とは，画像の局所的な領域から求められる特徴量である．局所特徴量を求める手法として，PCA-SIFT を用いる．PCA-SIFT とは，特徴量を 36 次元のベクトルとして抽出する手法である．基本的に 1 画像からは数百から数千個の特徴ベクトルが得られる

まず画像から特徴点を抽出する．そして，PCA-SIFT を用いて特徴点の周りの局所的な領域から特徴量を求める．

5.2　登録処理と検索処理

登録処理では，メモリ削減のために，PCA-SIFT によって得られた特徴量の各次元を 2 値化し，ビットベクトルを作成する．得られたビットベクトルをハッシュ関数によってハッシュ表のインデックスに変換する．そして，特徴量と画像 ID の組をハッシュ表のそのインデックスに登録する．

検索処理では，以下の手順に従う．質問画像から得た各特徴ベクトルに対して，登録時と同様にインデックスを計算して，ハッシュ表にアクセスする．そして，アクセスしたハッシュ表に登録されている特徴量を参照し，近似する特徴量を求め，対応する画像 ID に投票する．投票数が一番多かった画像を検索

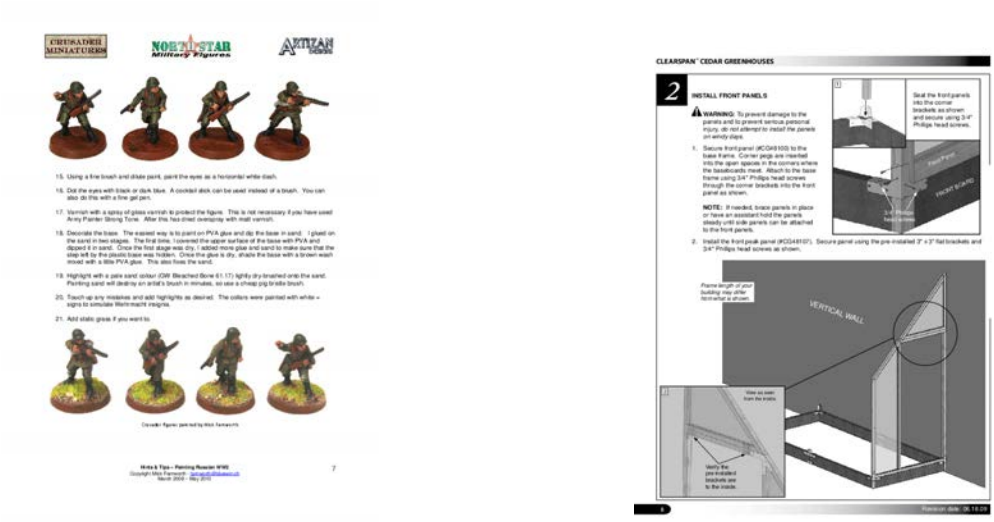


図 4　データベース画像例

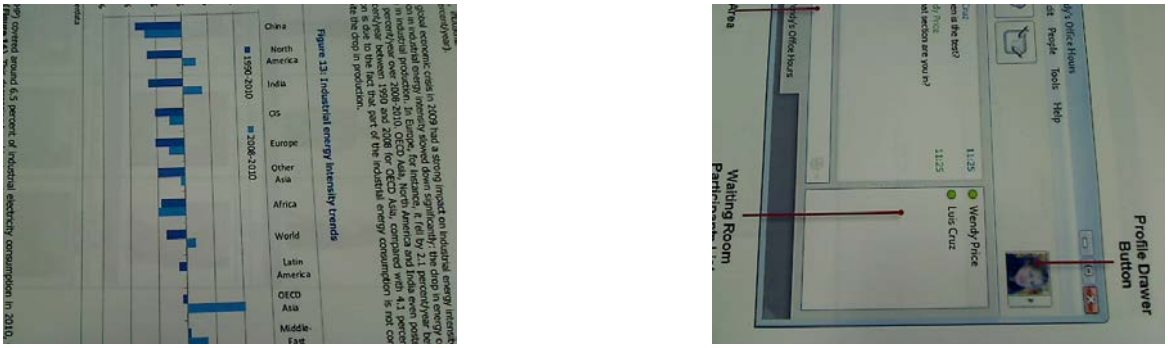


図 5　質問画像の例

結果とする．

6. 提案手法

本章では，文書画像検索と画像検索のハイブリッド方式について説明する．図 3(a) のような文書画像を入力として，初めに特徴点を抽出する．この特徴点とは文章の単語ごとの重心から算出されるもので，LLAH でも用いられている．この特徴点を用いて，最小全域木を計算することによって文字の並びを切り出す．最小全域木とは，いくつかの頂点とその辺の長さが定義されているときに，全ての頂点を含む木で辺の長さの総和が最小となるものである．ただし，木は環状にはならない．最小全域木は，辺の長さを昇順にソートし，小さい順に採用していくという処理である．このとき，採用することで木が環状になる場合は除去する．そして，頂点数-1 本の辺を採用するまで処理を繰り返す．一般的な文章では同じ行の単語の重心は直線状に配置されると考えられる．そういった仮定より，得られた最小全域木から極端に木が曲がる部分を除去する．そこから残った線をすべて単語の並びとする．図 3(b) に単語の並びを求めた図を示す．次に画像を分割し，それぞれの部分の単語の並びの行数を調べる．得られた行数が閾値以上ならばその特徴点を用いて LLAH による文書画像検索を適用し，閾値以下ならばその範囲には画像検索を適用する．そうすることにより，分割したそれぞれの結果が導出される．その結果の多数決により，最終的な結果を導出する．結果がすべて別々のものである場合や，結果が等しく分散しどの文書が正解であるか判断できない場合は検索失敗とする．

表 1　実験結果

	(1)LLAH	(2) 画像検索	(3) 提案手法
精度 (%)	81.0	94.0	88.0
処理時間 (ms)	12.4	86.4	60.1

7. 実　　験

7.1 実験条件

データベースの画像は文書画像 1000 枚，質問画像は文書画像を撮影した 100 枚で実験をした．本実験では簡単のため，質問画像はカメラと文書との角度はつけないように撮影したものを使用した．図 4 にデータベースの画像の例を示し，図 5 に質問画像の例を示す．ハイブリッド方式で画像検索に切り替えるかどうかを判別する行数の閾値は 3 と設定した．そして，画像の分割は縦横均等に二つずつ分け，四分割となるようにした．以上の条件の下，(1) 質問画像を LLAH によって検索した場合，(2) 画像検索によって検索した場合，(3) 提案手法によって検索した場合，のそれぞれ 3 種類の検索法での精度，処理時間を調

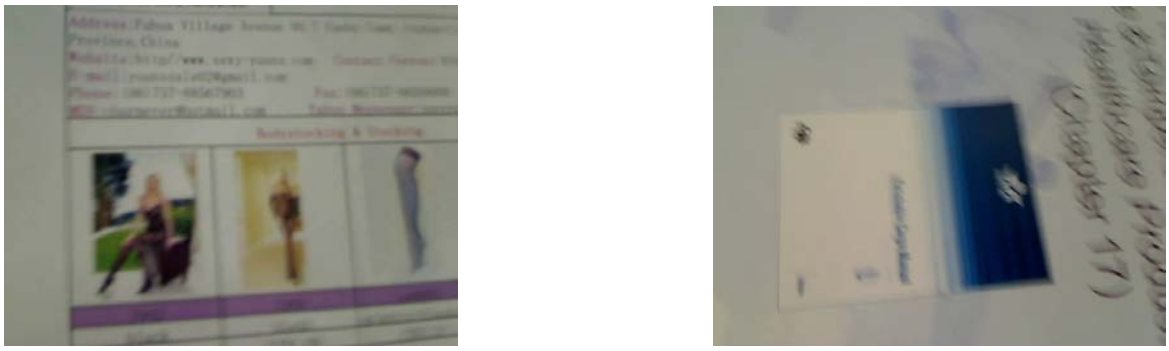


図 6　全手法で検索失敗した例

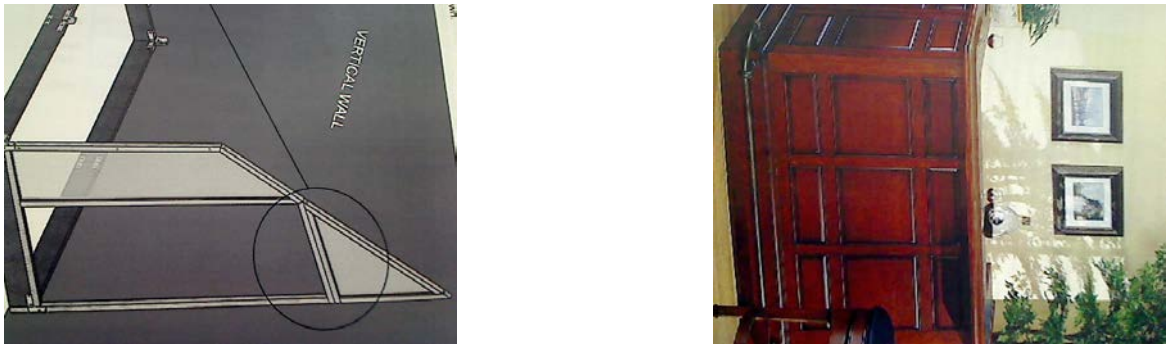


図 7　LLAH でのみ検索失敗した例

(a) 多数決の票の分散　　(b) LLAH の検索失敗

図 8　LLAH とハイブリッド方式で検索失敗した例

べた．精度は正しい文書を検索できた割合で，処理時間は特徴点を求めた後，検索を終えるまでの時間である．

7.2　結果と考察

表 1 に実験結果を示す．この結果より，提案手法は LLAH よりも精度が良く，画像検索よりも高速に検索可能であるということがわかった．図 6 に全手法で検索失敗した例，図 7 に LLAH でのみ検索失敗した例，図 8 に LLAH とハイブリッド方式で検索失敗した例を示す．図 6 より，ブレがひどい画像ではどの手法でも検索できないことがわかる．ブレが大きい部分からは特徴点がうまく取り出せないことが多い．よって，検索できなかった理由は抽出される特徴点数が少なかったからであると考えられる．図 7 のような文章が少ないクエリは LLAH では検索失敗している．けれども，画像検索とハイブリッド方式では検索成功している．これは，ハイブリッド方式により，適切に画像検索を適用できているからである．しかし，図 8 のクエリでは画像検索では検索成功しているが，LLAH だけでなくハイブリッド方式でも検索失敗している．これは，大まかに分けて二つの理由があった．まず一つ目の例を図 8(a) に示す．この例では最終結果を出す際の多数決に失敗している．得られた四つの結果により，二票ずつ別々の画像に投票していたため，結果が等しく分散し，検索失敗した．次に二つ目の例を 8(b) に示す．この例では，特徴点は数多く抽出できたものの，LLAH の検索時に誤投票により，他の文書に投票されていた．つまり，画像検索への切り替えができていないために，ハイブリッド方式で検索失敗していた．

8.　ま　と　め

本稿では，文書画像検索と画像検索とのハイブリッド方式について検討し，有用性を確かめるための比較実験をした．結果，LLAH と比べ精度が向上しており，画像検索と比べ処理時間が

短いということが分かった．また，ハイブリッド方式の最終結果を出す際の多数決のときに検索失敗があった．LLAH よりも時間がかかっており，画像検索よりも精度が下がっている．よって，実用性を求めるためには，改良を加え処理時間と精度の向上の必要性も感じさせる結果となった．

今後の課題として，図 8(a)(b) のような検索失敗を軽減することが挙げられる．図 8(a)(b) の例では，画像検索のみで検索すれば検索に成功したため，検索手法を切り替える領域の決定方法や，どちらの手法を適用するかの判断条件を改善することにより，対処可能であると考えられる．

文　　献

[1] ”http://84dialog.blogspot.jp/2010/03/layered-reading.html” 2011 年

[2] 近野恵，黄瀬浩一，岩村雅一，内田誠一，大町真一郎，“ カメラペンシステムにおける筆跡復元精度の向上，”電子情報通信学会技術研究報告，vol.111，no.317，PRMU2011-101，pp.13.18，(2011).

[3] 中居友弘，黄瀬浩一，岩村雅一. “ 処理速度とメモリ効率の改善された LLAH によるカメラベース文書画像検索，”画像の認識・理解シンポジウム (MIRU2008)，(2008).

[4] 黄瀬浩一，野口和人，岩村雅一. “ 参照特徴ベクトルの増加による低品質画像の高速・高精度認識，”電子情報通信学会論文誌 D，Vol. J93-D，No. 8，pp. 1353–1363，(2010).

[5] 黄瀬浩一，岩村雅一，中居友弘，野口和人. “ 局所特徴量のハッシングによる大規模画像検索，”日本データベース学会論文誌，No. 8，Vol. 1，pp. 119–124. (2009).

[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物体の高速認識 ”, 画像の認識・理解シンポジウム (MIRU2007) 論文集，OS-B2-02, pp. 111.118(2007).

[7] D. Doermann: “ The Indexing and Retrieval of Document Images: A Survey ”, Computer Vision and Image Understanding, 70, 3, pp.287.298 (1998).

[8] D. Doermann, H. Li and O. Kia: “ The Detection of Duplicates in Document Image Databases ”, Proc. ICDAR ’97, pp.314-318 (1997).

[9] K. Hannu : “ Document Image Retrieval with Improvements in Database Quality ”, Academic Dissertation of University of Oulu (1999).

[10] J. J. Hull : “ Document Image Matching and Retrieval with Multiple Distortion-Invariant Descriptors ”, Document Analysis Systems, pp.379.396 (1995).

[11] Lowe, D. G.: “ Distinctive image features from scale-invariant keypoints ”, International Journal of Computer Vision, Vol.60, No.2, pp.91-110 (2004).