

複比とハッシュに基づく文書画像検索

中居 友弘[†] 黄瀬 浩一[†] 岩村 雅一[†] 松本啓之亮[†]

[†] 大阪府立大学大学院工学研究科
〒 599-8531 大阪府堺市学園町 1-1

E-mail: †nakai@ss.cs.osakafu-u.ac.jp, ††{kise,masa,matsu}@cs.osakafu-u.ac.jp

あらまし 本稿では、デジタルカメラを用いた新しい文書画像検索手法を提案する。提案手法は、デジタルカメラで撮影した文書画像の一部あるいは全部を検索質問として、それを含む文書画像をデータベースから検索するものである。これを実現するためには、撮影時に生じる射影変換歪みに対処しつつ、部分的な手がかりでも柔軟に照合する必要がある。この問題に対して、提案手法では、(1) 射影変換の不変量である複比を用いたインデックス付け、(2) ハッシュを用いた投票処理の2特徴により対処する。高解像度デジタルカメラと携帯電話付属のデジタルカメラで撮影した画像を用いた実験により、提案手法の有効性を検証する。

キーワード 文書画像検索, デジタルカメラ, 複比, ハッシュ, 投票

Document Image Retrieval Based on Cross-Ratio and Hashing

Tomohiro NAKAI[†], Koichi KISE[†], Masakazu IWAMURA[†], and Keinosuke MATSUMOTO[†]

[†] Graduate School of Engineering, Osaka Prefecture University
Gakuen-cho 1-1, Sakai, Osaka, 599-8531 Japan

E-mail: †nakai@ss.cs.osakafu-u.ac.jp, ††{kise,masa,matsu}@cs.osakafu-u.ac.jp

Abstract In this report, we propose a new method of document image retrieval using digital cameras. The proposed method is to retrieve document images which include parts or the whole of document images acquired by a digital camera. For this purpose, it is required to solve the problem of “perspective distortion” of images, as well as to establish a way for matching parts of document images flexibly. These are achieved based on the following characteristic points of the proposed method: (1) Indexing of document images using the projective invariant called the “cross-ratio”, (2) Voting with hashes whose indices are defined by the cross-ratios. From experimental results using digital cameras with high resolution and a camera phone, we demonstrate the effectiveness of the proposed method.

Key words Document image retrieval, Digital camera, Cross-ratio, Hashing, Voting

1. はじめに

デジタルカメラの一般化、高機能化、小型化、低廉化は、パターン認識・メディア理解の分野に新しい可能性を与えつつある。これは一般的には、ユーザが取得する画像を様々なサービスと連携させることと考えられる。文字・文書の分野も例外ではなく、カメラに基づく文字認識、文書画像解析の研究が盛んに行われている [1], [2]。特に、携帯電話に付属のデジタルカメラを用いたインタフェースは重要であり、文字読み取り [3] や翻訳 [4] など、種々の処理が検討されている。

本稿では、このような認識処理とは異なる可能性の一つとして、検索処理の一手法を提案する。具体的には、デジタルカメラで撮影した文書の一部あるいは全部を「検索質問」とし、そ

の検索質問を含む文書画像をデータベースから検索する処理である。このような検索処理は、手元の印刷物から対応する電子文書を検索したり、印刷物に対する書き込みを抽出 [5] するための前段階の処理としても用いることができる。

さて、デジタルカメラを用いた文書画像検索が、従来の文書画像検索 [6] ~ [9] と大きく異なる点は、画像が受ける種々の歪み [2] にある。例えば、スキャナにより取得した従来の文書画像が受ける幾何歪みはほぼ回転のみであったのに対して、デジタルカメラによる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。これらの点は、従来と比べて検索処理をかなり困難な問題としている。

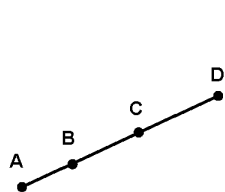


図 1 複 比

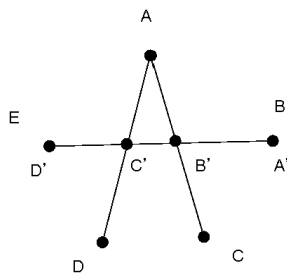


図 2 5点からの複比

この問題に対処するため、本研究では2つのアイデアを導入する。一つは射影変換に対する不変量である複比 (cross-ratio) [12] に着目し、文書画像を複比によってインデックス付けすることである。文書画像の様々な「部分」から複比を計算し、インデックスとすることにより、部分検索にも対処する。もう一つは、ハッシュを用いた投票処理により検索を実現することである。これにより、特徴点の対応を陽に求めることなく、柔軟かつ比較的高速な検索が可能となる。このような考え方は、Geometric Hashing [10] に代表されるように、物体認識の分野では広く用いられているものである [11]。提案手法は、複比をハッシュのインデックスに用いるという点で、従来のものと多少異なる。なお、処理の都合上、対象文書は現在のところ英文文書とする。

以下、2. で関連研究についてまとめ、3. で提案手法について述べる。4. では、通常のデジタルカメラに加えて携帯電話付属のカメラを用いた検索実験の結果について述べる。

2. 関連研究

2.1 文書画像検索

画像ベースの文書画像検索の従来手法としては、以下のようなものがある。Kauniskangas の方法 [8] では、文書画像を段落や図などの領域に分け、それぞれの領域を判別し、木構造で表現する。検索を行うときは検索質問とデータベースの文書画像の各領域をマッチングし、最もよくマッチするものを検索結果として出力する。Hull の方法 [9] では、テキストを単語ごとに分割し、各単語の文字数の並びを特徴量として文書を表現する。あらかじめデータベース中の文書の各部分の特徴量を計算してハッシュに登録しておき、検索時には入力画像に同じ処理をハッシュにアクセスし、投票により検索を行う。

上に挙げた手法は、フラットベッドスキャナなどで得られた高解像度の正対画像を対象としているため、本研究で対象とするようなデジタルカメラを利用した文書画像検索には利用できない。例えば、Hull の方法では入力画像が文字ごとに分離されていることが前提であり、これは本研究で対象とするような射影変換を受けた画像では困難なことがある。

2.2 複 比

コンピュータビジョンの分野では、種々の変換に対する不変量がしばしば用いられる。射影変換については、不変量として複比 [12] と呼ばれるものがある。これは、図 1 に示すように同一平面上で一直線状に並ぶ点 ABCD に対して、以下の式で求められる。

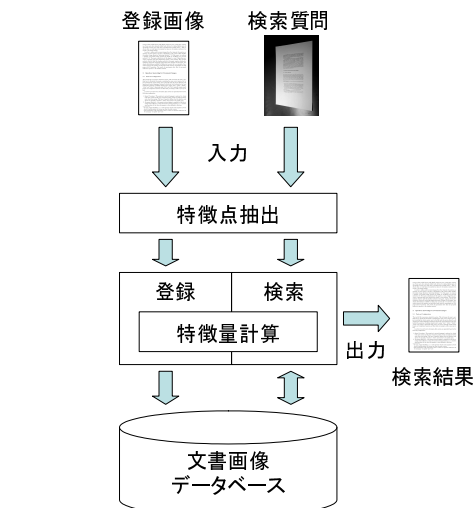


図 3 文書画像検索システム

$$\frac{AC}{BC} / \frac{AD}{BD} \quad (1)$$

また、複比は同一平面上の5点から、図2のようにして一直線上の4点を生成して計算することもできる。ここで、点ABCDEの複比はA'B'C'D'の複比として求められる。本手法では、5点からの複比を用いて文書画像固有の特徴量を計算し、文書画像の検索を行う。

3. 提案手法

3.1 基本方針

まず、本研究の目的とその達成のための方針について述べる。本研究の目的は、デジタルカメラで撮影された文書画像を検索質問として、文書画像データベースを検索する手法の考案である。ここで問題となるのは、検索質問の文書画像はデジタルカメラで撮影されるため射影変換の歪みを受け、また全体が撮像されているとは限らないことである。さらに、検索質問画像の解像度や照明条件がデータベース中の文書画像とは大きく異なることも問題をより困難にする要因である。

以上のような問題を克服するために、次の方針を立てる。

- 特徴量…射影変換歪みによらない文書画像の特徴量を得るため、複比を用いて特徴量を計算する。複比の計算に用いる特徴点には、射影変換やノイズに対する安定性を得るため、単語の重心を用いる。また、部分画像からの検索を可能にするため、文書画像の部分ごとに得られる複比を用いて特徴量を計算する。

- 登録・検索…特徴点の対応を取らずに検索を行うため、ハッシュを用いた投票を導入する。登録時には、まず文書画像から特徴点を求めて特徴量を計算し、そこからインデックスを求めてハッシュに登録する。検索時には、検索質問から同様の処理で特徴点、特徴量、インデックスを求めてハッシュにアクセスし、登録されている文書に対して投票を行うことで検索する。

3.2 処理の流れ

提案する文書画像検索システムの構成を図3に示す。文書画

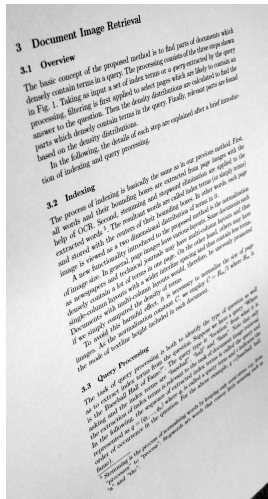


図 4 入力画像

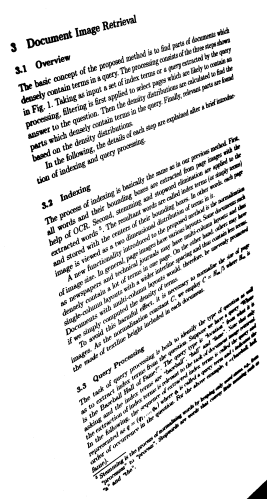


図 5 2 値化後

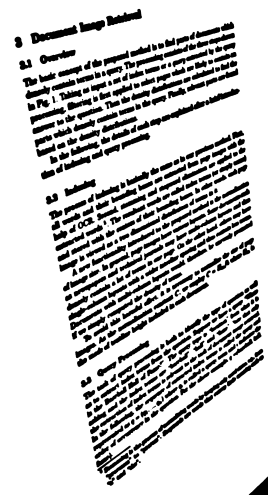


図 6 単語ごとに結合

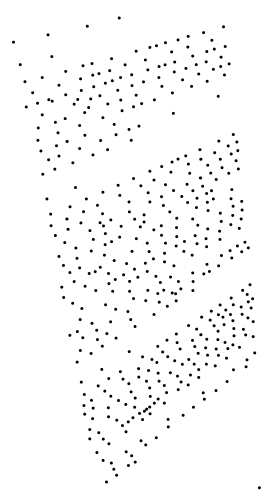


図 7 特徴点

像は、特徴点抽出により点の集合に変換された後、登録の場合は登録処理へ、検索の場合は検索処理に入力される。登録処理では、特徴点を用いて複比を計算し、それをインデックスに変換することによってハッシュに文書を登録する。一方、検索処理は、特徴点から同様にインデックスを計算し、投票を行うことで所望の文書画像を検索する。以下、各々について述べる。

3.3 特徴点抽出

特徴点抽出で重要なことは、特徴点の再現性、すなわち射影変換・ノイズ・低解像度の影響下でも同一の特徴点が見られることである。英文文書における単語の重心は、この条件を満たす特徴点の1つである。これは、英文文書では単語と単語の間に空白があり、分離が比較的容易なためである。

例を用いて手順の概略を説明する。入力画像(図4)は、まず適応2値化^(注1)により2値画像(図5)に変換される。次に、2値画像から以下のように単語領域を得る。まず、ガウシアンフィルタを適用して2値画像をにじませる。このときガウシアンフィルタのパラメータは文字サイズの推定値(連結成分の面積の最頻値の平方根)に基づいて適応的に定める。そして、

(注1): 適応2値化, ガウシアンフィルタには OpenCV [13] の関数を使用した。

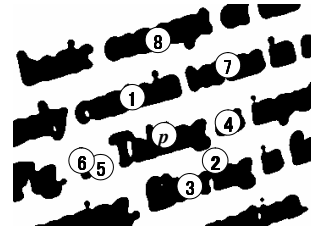


図 8 近傍点 1

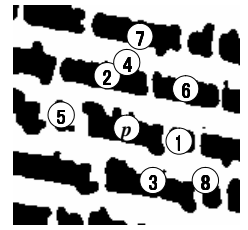


図 9 近傍点 2

にじませた画像に対して、再度適応2値化を適用し、2値画像(図6)を得る。この画像の連結成分を単語領域とみなし、その重心を特徴点とする。図6に対しては図7が得られる。

3.4 インデックスに用いる特徴量計算

登録、検索の鍵となるのは、複比を用いていかにハッシュのインデックスを計算するかである。ここでは、登録、検索の詳細について述べる前に、インデックスに用いる特徴量の計算方法についてまとめる。

画像上の特徴点から得られる複比の値は連続値だが、インデックスに用いるときは k 個に離散化する。射影変換や撮影条件の変動による誤差を吸収するために、 k の値はある程度小さくする必要がある。ここでは予備実験の結果として得られた $k = 9$ を用いる^(注2)。また、部分検索を可能とするため、各特徴点の近傍ごとに定まる特徴量を用いる。

各特徴点の近傍で特徴量を定義するとき、以下のようなものが考えられる。

- (1) 近傍5点の複比
- (2) 近傍 n 点からの5点の複比
- (3) 近傍 n 点からの $m (< n)$ 点の配置、および m 点からの5点の複比

本手法ではこのうち最も複雑な(3)を用いる。以下ではこれら3つの手法を単純なものから説明し、本手法における特徴量と、なぜ複雑な手法が必要になるかについて述べる。

3.4.1 近傍5点の複比

特徴点の近傍から複比を用いて特徴量を定義するとき、まず容易に思いつくのは最近傍5点から複比を計算する方法である。例えば、図8に示す特徴点 p に対して最も近い1~5の5点を選び出し、そこから複比を計算して点 p の特徴量とすることが考えられる。

しかし、図8、図9に示すように、異なる角度から撮影した文書画像では、最近傍の5点は変化する。そのため、同じ文書の同じ点から同一の特徴量が得られないという問題が生じる。

3.4.2 近傍 n 点からの5点の複比

次に考えられるのは、近傍 n 点からあらゆる5点の組み合わせを選び出し、そこから複比を計算する方法である。

図8、図9の例では、最近傍の5点こそ異なるが、近傍8点中の7点までは同一である。このように、近傍 n 点にはあまり変化しない $m (< n)$ 点が存在する。従って、 n 点からすべての5点の組み合わせを生成して複比を計算すれば、共通する m 点

(注2): 複比の誤差や離散化については、文献[14]のような理論的検討が必要と考えられるが、現状では単純に複比の取り得る値を k 等分して用いている。

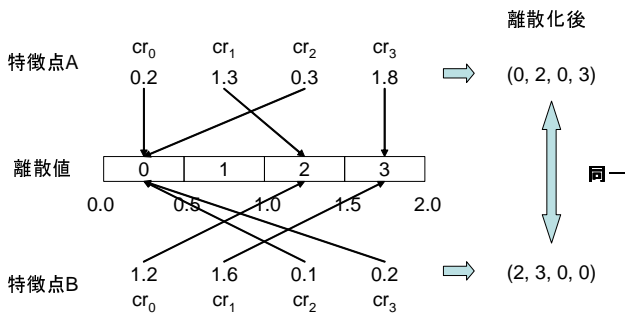


図 10 n 点からの 5 点の複比の問題点

から計算される複比は一致すると考えられる。このことから、近傍 n 点のすべての 5 点の組み合わせの複比を特徴量とし、比較する際にはそのうちの変化しない m 点から計算される複比が一定数以上一致する場合に特徴量が一致するとみなす方法が考えられる。

しかし、実際にこの特徴量を用いて検索を行うと、正しい結果が得られないことが多い。図 10 に示す例を用いて考える。いま簡単のため、近傍 n 点から得られるすべての複比を図 10 の 4 個とする^(注3)。ここでは、特徴点 A の近傍で定義される複比が (0.2, 1.3, 0.3, 1.8)、同様に特徴点 B から (1.2, 1.6, 0.1, 0.2) が得られている。いま、複比を 0.5 ステップで離散化するとすれば、離散化後は各々 (0, 2, 0, 3), (2, 3, 0, 0) となる。種類のみを考えると両者とも 0,2,3 の出現となり、同じものと認定される。実際の画像で計測するとこのような状況が頻出し、検索に失敗する。

3.4.3 近傍 n 点からの m 点の配置および m 点からの 5 点の複比

上記の問題を解決する一方策は、順序を含めて複比を考えることである。つまり、図 10 の例では、(0, 2, 0, 3) と (2, 3, 0, 0) を別のものとして区別することである。

具体的に、図 11 の例を用いて説明する。ここで点 ABCDEFGH と点 IJKLMNOP は、同一文書の異なる画像において、対応する点の近傍 8 点の組とする。8 点の組のうち異なるものは E, L の 1 点のみであり、残り 7 点は共通である。従って、それぞれの 8 点からすべての 7 点の組み合わせを取り出せば、そのうち ABCDFGH と IJKLMN の 1 組はまったく同じものになる。このように完全に共通な 7 点の組から、ある定まった順序ですべての 5 点の組み合わせを生成すれば、順序も含めて完全に一致する。つまり、最初は点 ABCDF から得られた複比と点 IJKLMN から得られた複比、次は点 ABCDFG から得られた複比と点 IJKLMO から得られた複比というように、同じ順序で共通な 7 点から複比を計算すれば、同じ順序で同じ複比が得られる。異なる 7 点の組から同じ複比がまったく同じ順序で現れるということはまれであるので、この順序付けられた複比を特徴量として用いれば、3.4.2 の問題の解決が期待できる。

特徴量の計算方法についてまとめる。まず、与えられた点の近傍 n 点を取り出す。次に、 n 点からすべての m 点の組み合

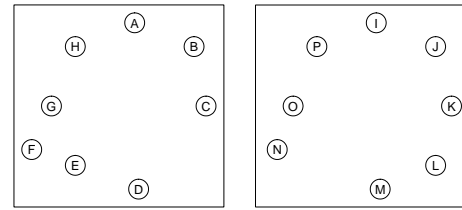


図 11 8 点

- 1: for all $p \in \{ \text{特徴点の集合} \}$ do
- 2: $P_n \leftarrow p$ の近傍 n 点 (時計回り)
- 3: for all $P_m \in \{ P_n \text{ からのすべての } m \text{ 点の組み合わせ} \}$ do
- 4: for all $P_5 \in \{ P_m \text{ からのすべての 5 点の組み合わせ} \}$ do
- 5: for $i = 0$ to 4 do
- 6: $cr_i \leftarrow P_5$ の i 番目を開始点とした複比
- 7: end for
- 8: $H_{index} \leftarrow cr_{0-4}, mC_5$ パターン ID から作成したハッシュのインデックス
- 9: H_{index} を用いて (文書 ID, 点 ID, nC_m パターン ID) の組をハッシュに登録
- 10: end for
- 11: end for
- 12: end for

図 12 登録アルゴリズム

わせを作成する。 m 点から定まった順序で 5 点の組み合わせを作成し、複比を計算する。ここで、複比の値が順序も含めて l 個以上一致したとき、特徴量は一致すると判定する。

3.5 登録

以上の準備に基づき、まず、登録アルゴリズムについて述べる。図 12 に登録アルゴリズムを示す。ここで、文書 ID は文書の識別番号であり、点 ID は文書ごとに各点に割り当てられる点の識別番号である。 nC_m パターン ID は n 点から m 点を取り出す際、組み合わせパターンに与えられる識別番号であり、 $0 \sim nC_m - 1$ の値をとる。同様に mC_5 パターン ID は m 点から 5 点を取り出す際の組み合わせパターンの識別番号であり、 $0 \sim mC_5 - 1$ の値をとる。

文書画像が登録されるハッシュの構造を図 13 に示す。登録処理では、5~8 行目でハッシュのインデックスを求め、9 行目でインデックスを用いてハッシュに登録するという処理を繰り返す。

アルゴリズムに沿って説明する。まず 5~7 行目では、5 点から 5 通りの複比を計算する。5 通りの複比は、得られた 5 点 ABCDE に対して、ABCDE, BCDEA, CDEAB, DEABC, EABCD というように先頭となる点を巡回させて得られるものである^(注4)。

次に、8 行目でハッシュテーブルのインデックスを計算する。ハッシュ関数を以下に示す。

$$H_{index} = \sum_{n=0}^4 cr_n (V_{max} + 1)^n + pat (V_{max} + 1)^5 \quad (2)$$

ここで、 $cr_n (n = 0 \sim 4)$ は離散化された 5 通りの複比、 V_{max} は

(注4): このような方法以外には、5 点の並び方に依存しない不変量である p^2 -invariant [10] の利用などが考えられる。

(注3): 後述するように、実際には 1 個の特徴点に対して多数の複比が得られる。

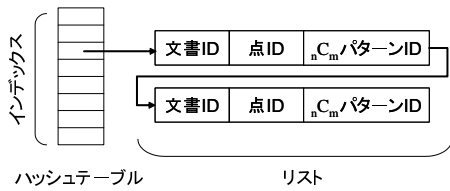


図 13 ハッシュの構造

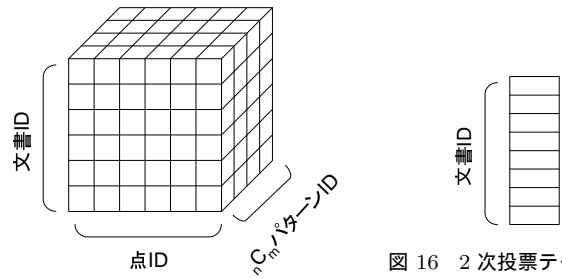


図 15 1次投票テーブル

図 16 2次投票テーブル

```

1: for all  $p \in \{ \text{特徴点の集合} \}$  do
2:    $P_n \leftarrow p$  の近傍  $n$  点 (時計回り)
3:   for all  $P_m \in \{ P_n \text{からのすべての } m \text{ 点の組み合わせ} \}$  do
4:     for all  $P'_m \in \{ P_m \text{の開始点を変えたもの} \}$  do
5:       for all  $P_5 \in \{ P'_m \text{からのすべての } 5 \text{ 点の組み合わせ} \}$  do
6:         for  $i = 0$  to  $4$  do
7:            $cr_i \leftarrow P_5$  の  $i$  番目を開始点とした複比
8:         end for
9:          $H_{index} \leftarrow cr_{0-4}, mC_5$  パターン ID から作成したハッシュのインデックス
10:         $H_{index}$  を用いてハッシュを読み, 1次投票テーブルに投票する
11:      end for
12:      1次投票テーブルで一定数以上の得票数をもっているセルがあったら, そのセルの属する文書IDで2次投票テーブルに投票する
13:      1次投票テーブルをクリア
14:    end for
15:  end for
16: end for
17: 2次投票テーブルで検索結果を判定する

```

図 14 検索アルゴリズム

離散化された複比の最大値, pat は mC_5 パターン ID である . 9 行目でインデックスを用いて (文書 ID, 点 ID, nC_m パターン ID) の組をハッシュに登録する . ハッシュで衝突が起きた場合, データは図 13 のようにリスト構造で付け加えられる . ここで, 文書 ID だけでなく, 点 ID および nC_m パターン ID もハッシュに登録するのは, 検索時に特徴量の比較を行う際, 順序付けられた複比の一致する個数を, 文書 ID, 点 ID, nC_m パターン ID ごとに集計するためである .

3.6 検索

次に検索について述べる . 図 14 に検索アルゴリズムを示す . 本手法ではまず図 15 に示す 1 次投票テーブルを用いて一定数 l 以上の複比が順序も含めて一致しているかどうかを調べ, 一致しているものについて図 16 に示す 2 次投票テーブルに対して投票を行い, 検索結果を決める .

同様にアルゴリズムに沿って説明する . まず 6~9 行目で登録時と同様の手法によりハッシュのインデックスを求め, 10 行目でハッシュを読んで図 13 にあるようなリストを得る . リストの各要素について, 1 次投票テーブルの該当するセルに投票を行う .

このような処理を繰り返し行い, m 点からのすべての 5 点の組み合わせについて投票が終わったら, 1 次投票テーブルの各セルを調べる . l 以上の得票数をもつセルが見つかったら, 文書の属する文書 ID で 2 次投票テーブルに投票する .

以上の処理により, 最終的に 2 次投票テーブルで最も得票数の多い文書を検索結果とする .

なお, 4 行目で P_m から開始点を変えた P'_m を作成し, すべての P'_m について処理を行っているのは, P_m として点 ABCDEFG が与られたとき, P'_m として BCDEFGA, CDEFGAB のように巡回させたものを m 通り作成することに相当する . これは, 画像が回転している場合に対応するためである .

4. 実験

4.1 実験概要

提案手法の有効性を検証するために, 通常のデジタルカメラと携帯電話付属のデジタルカメラそれぞれについて撮影された文書画像から対応するものを検索した . 通常のデジタルカメラには CANON EOS Kiss Digital (630 万画素) と付属のレンズ EF-S 18-55mm USM を, 携帯電話付属のデジタルカメラには京セラ TK31 付属のデジタルカメラ (18 万画素) を使用した .

文書画像データベースには, 1, 2 段組の英語論文の PDF ファイルから直接作成した文書画像を 50 ページ用意した . データベース中の文書画像の例を図 17, 図 18 に示す . なお, 実験は CPU Pentium 4 2.4GHz, メモリ 768MB の計算機で行った .

4.2 実験 1: 通常のデジタルカメラ

まず通常のデジタルカメラを用いた結果について述べる . 3. で述べた処理のパラメータは $n = 8, m = 7, k = 9, l = 10$ とした . 検索質問には, 異なる 10 ページに対して, 図 19~ 図 22 に示すような 4 段階の撮影範囲で撮影した画像を計 40 枚用意した . ここで撮影範囲 A は文書全体, B は文字領域全体, C は文字領域半分程度, D は文字領域 1/4 程度とした . 画像の撮影はやや傾いた角度から行った . 対応する文書画像の得票数が 1 位になったとき正答とし, 正答率と処理時間の平均を調べた .

表 1 に実験結果を示す . 撮影範囲に関わらず, すべての入力画像で正しい結果が得られた . なお, 撮影範囲が狭くなるにつれて処理時間が減少しているのは, 処理する特徴点の数が少なくなるためである .

表 1 実験結果

撮影範囲	A	B	C	D
正答率 [%]	100	100	100	100
処理時間 [sec]	231.6	173.1	157.6	118.1

ment in that article shows that digital camera use on existing document use. Some users have already utilized such devices for taking digital copies of documents. However, it is still rare to receive them [19]. This is because that not only legacy documents but also new documents continue to be stored in document images.

In order to address such document images from the viewpoint of providing a retrieval system, we have conducted a project of developing a QA-oriented "QIG" (Question Image Answering Engine). In this paper, we propose a method of document image retrieval (QIG) by matching the previous method [20]. The major contribution of this paper is a way of improving previous method which is called the baseline method in this paper, we give density distribution models for candidate response pages to images. In this paper, we describe distribution models by taking into account the occurrence of keywords in the question, the results of experiments show that the proposed method is superior to the baseline method.

2 Question Answering for Document Images

2.1 Task and Configuration

The task of QA is precision oriented in nature. This is because the user is satisfied not by having all documents containing the same correct answer, but by just receiving the correct answer once. In the QA task, the user is allowed to ask questions in natural language. Systems for electronic text developed so far have tackled the operation of online simple facts for using "what", "which", "when", and "where" questions using "why" and "how" generally require much longer and complicated answers than those that processing is still an open problem.

In order to locate facts in documents, QA systems are generally based on the following configuration.

1. Query Processing: The question in natural language is analyzed to obtain key-words and the type of question. Query terms are mapped to the user's intent of processing. The type of question defines what the question asks about. For example, "location", "time" and "person" are typical types.

2. Document Retrieval: A document retrieval engine is employed to find documents which is likely to contain the answer. Passage retrieval, i.e., to extract small portions of text from documents, is often utilized in this step.

3. Answer Output Displaying: It is to make pictures of books and magazines in book form by visible photos, but because no chart of public concern, the presentation of the text image is "black" in Japanese which has the two meanings "to explain" and "to say".

図 17 データベース中の画像 1

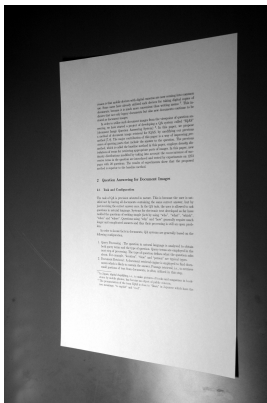


図 19 撮影範囲 A

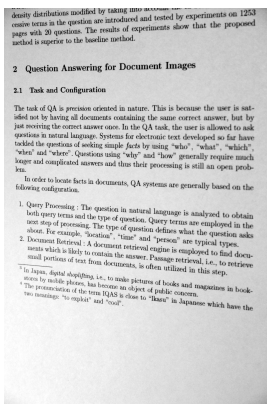


図 21 撮影範囲 C

Image	Resolution	Size	File Format	Color	Frame Rate	Shutter Speed	Aperture	ISO	White Balance	Focus Mode	Macro	Scene Mode	Self-Timer	Image Stabilization	Flash	Video Mode	Video Resolution	Video Frame Rate	Video File Format	Video File Size	Video File Format
Fig. 23	640x480	1.5MB	JPG	Color	30FPS	1/200	F5.6	100	Auto	Auto	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off
Fig. 24	640x480	1.5MB	JPG	Color	30FPS	1/200	F5.6	100	Auto	Auto	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off
Fig. 25	640x480	1.5MB	JPG	Color	30FPS	1/200	F5.6	100	Auto	Auto	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off
Fig. 26	640x480	1.5MB	JPG	Color	30FPS	1/200	F5.6	100	Auto	Auto	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off
Fig. 27	640x480	1.5MB	JPG	Color	30FPS	1/200	F5.6	100	Auto	Auto	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off	Off

Table 1. Camera specification of the digital camera used in this paper.

Table 2. The results of document images in the camera. The results of document images in the camera are shown in Table 2. The results of document images in the camera are shown in Table 2. The results of document images in the camera are shown in Table 2.

2.2 Document Retrieval

The task of QA is precision oriented in nature. This is because the user is satisfied not by having all documents containing the same correct answer, but by just receiving the correct answer once. In the QA task, the user is allowed to ask questions in natural language. Systems for electronic text developed so far have tackled the operation of online simple facts for using "what", "which", "when", and "where" questions using "why" and "how" generally require much longer and complicated answers than those that processing is still an open problem.

In order to locate facts in documents, QA systems are generally based on the following configuration.

1. Query Processing: The question in natural language is analyzed to obtain key-words and the type of question. Query terms are mapped to the user's intent of processing. The type of question defines what the question asks about. For example, "location", "time" and "person" are typical types.

2. Document Retrieval: A document retrieval engine is employed to find documents which is likely to contain the answer. Passage retrieval, i.e., to extract small portions of text from documents, is often utilized in this step.

3. Answer Output Displaying: It is to make pictures of books and magazines in book form by visible photos, but because no chart of public concern, the presentation of the text image is "black" in Japanese which has the two meanings "to explain" and "to say".

図 18 データベース中の画像 2

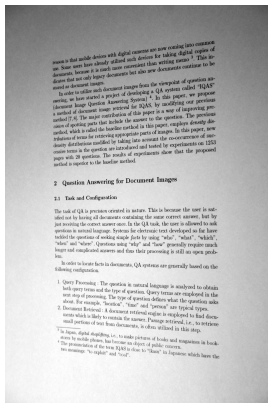


図 20 撮影範囲 B

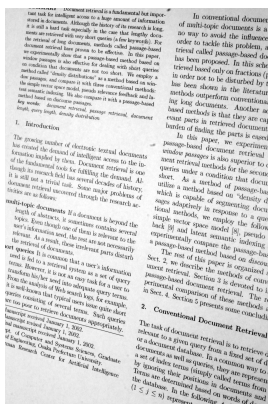


図 22 撮影範囲 D

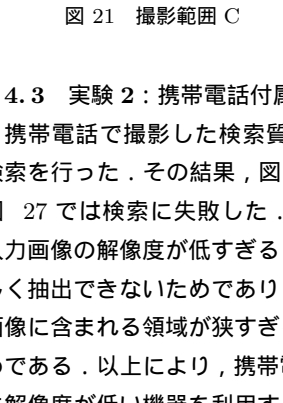


図 21 撮影範囲 C

図 22 撮影範囲 D

図 23 低解像度 1

図 24 低解像度 2

図 25 低解像度 3

図 26 低解像度 4

図 27 低解像度 5

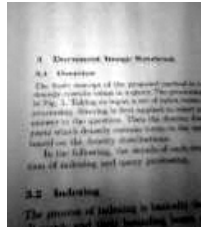


図 23 低解像度 1

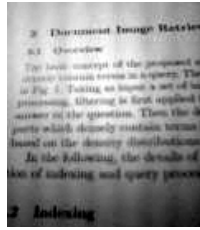


図 24 低解像度 2

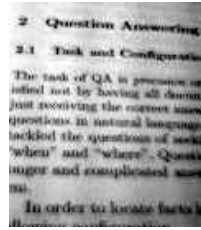


図 25 低解像度 3

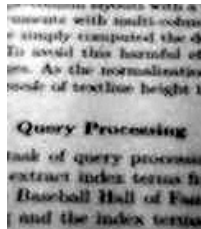


図 26 低解像度 4

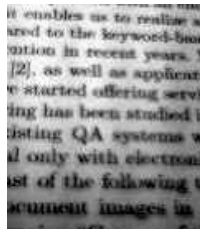


図 27 低解像度 5

メラで撮影した文書画像を検索質問とする文書画像検索法を提案した。実験により、提案手法が文書画像を高い精度で検索できることを確認した。また、撮影範囲を調整すれば、解像度の低い携帯電話付属のデジタルカメラでも検索が可能であることを示した。

今後の課題としては、速度の向上、日本語文書への対応、文書に限らないシーン中のオブジェクトの検索手法の考案が挙げられる。

文 献

- [1] D. Doermann, J. Liang and H. Li: "Progress in camera-based document image analysis", Proc. ICDAR'03, pp. 606-616 (2003).
- [2] 黄瀬, 大町, 内田, 岩村: "カメラを用いた文字認識・文書画像解析の現状と課題", 信学技報 (2005.3) [発表予定].
- [3] 山田, 仙田: "携帯カメラを用いたユビキタス情報インタフェース", 情報処理, 45, 9, pp. 923-927 (2004).
- [4] Y. Watanabe, Y. Okada, Y.-B. Kim, T. Takeda: "Translation camera", Proc. ICPR'98, pp.613-617 (1998).
- [5] 近藤, 中居, 黄瀬, 松本: "デジタルカメラを用いた文書画像からの書き込み抽出法", 2005 年信学総合大会 (2005) [発表予定].
- [6] D. Doermann: "The Indexing and Retrieval of Document Images: A Survey", Computer Vision and Image Understanding, 70, 3, pp.287-298 (1998).
- [7] D. Doermann, H. Li and O. Kia: "The Detection of Duplicates in Document Image Databases", Proc. ICDAR'97, pp.314-318 (1997).
- [8] K. Hannu: "Document Image Retrieval with Improvements in Database Quality", Academic Dissertation of University of Oulu (1999).
- [9] J. J. Hull: "Document Image Matching and Retrieval with Multiple Distortion-Invariant Descriptors", Document Analysis Systems, pp.379-396 (1995).
- [10] 松山, 久野, 井宮 (編): コンピュータビジョン: 技術評論と将来展望, 新技術コミュニケーションズ (1998).
- [11] 山口, 福井: "パターンハッシング: 部分画像と不変量索引を用いた分散アピアランスモデル", 情処学論: コンピュータビジョンとイメジメディア, 44, SIG5(CVIM6), pp.64-73 (2003).
- [12] 佐藤: コンピュータビジョン - 視覚の幾何学 -, コロナ社 (1999).
- [13] OpenCV: <http://sourceforge.net/projects/opencvlibrary/>
- [14] 金澤, 松永, 金谷: "複比による識別のための最良マーカーパタンの設計: 理論解析", 情処研報, 99-CVIM-115-13 (1999).

本稿では、複比とハッシュを用いた投票により、デジタルカ