

# 複比の投票に基づく文書画像検索と射影歪み補正

中居 友弘<sup>†</sup> 黄瀬 浩一<sup>†</sup> 岩村 雅一<sup>†</sup>

<sup>†</sup> 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@m.cs.osakafu-u.ac.jp, ††{kise,masa}@cs.osakafu-u.ac.jp

あらまし 本稿では、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という新しい文書画像検索手法を提案する。このような処理の実現には、撮影時に生じる射影歪みに対処しつつ、部分的な手がかりでも柔軟に照合する必要がある。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、この問題に対処する。また本稿では、文書画像検索の結果として得られる点対応を用いた射影歪みの補正手法についても提案する。提案手法は、我々が別途構築中のデジタルカメラを用いた書き込み抽出システムに用いられるものである。上記 2 手法について、高解像度デジタルカメラと携帯電話付属デジタルカメラを用いた実験により、有効性の検証を行う。

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, 射影歪み補正

## Document Image Retrieval and Removal of Perspective Distortion Based on Voting for Cross-Ratios

Tomohiro NAKAI<sup>†</sup>, Koichi KISE<sup>†</sup>, and Masakazu IWAMURA<sup>†</sup>

<sup>†</sup> Graduate School of Engineering, Osaka Prefecture University

Gakuen-cho 1-1, Sakai, Osaka, 599-8531 Japan

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

**Abstract** We propose a method of document image retrieval using digital cameras. The proposed method takes as input parts or the whole of documents acquired as queries by a digital camera, and retrieve document images that include the queries. 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. We also propose a method of removal of perspective distortion using point correspondence obtained through the retrieval. This method is utilized for extracting annotation from document images. 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, Removal of perspective distortion

### 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

ラを用いたインター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。

デジタルカメラを利用したサービスの一つに、文書画像検索がある。これは、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することなどに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 [5]~[8] と大きく異なる点は、画像が受ける種々の歪みにある。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（以後、射影歪みと呼ぶ）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。

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) [9] 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。文書画像を特徴点の集まりとして表現し、各特徴点の「近傍」で複比を計算して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。

ところで提案手法では、検索の過程において、デジタルカメラで撮影した画像とデータベース中の画像の間で特徴点の対応を求めることができる。本稿では、データベース中の正対画像に対してこの対応を求めることにより、デジタルカメラで撮影した画像の射影歪みを補正する手法についても述べる。この手法は、我々が検討している書き込み抽出 [12] において、書き込みのある文書画像から書き込み抽出を行うために用いられる。

以下、2. で関連研究についてまとめ、3. で複比の投票に基づく文書画像検索手法について、4. で射影歪み補正手法について述べる。5. 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

文書画像検索と射影歪み補正の従来法について順に述べる。

### 2.1 文書画像検索

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 [7] 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 [8] 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 射影歪み補正

射影歪みを受けた文書画像から正対画像を復元する手法には様々なものがある [2]。提案手法のように、1枚の画像から歪みを補正する手法としては、消失点と呼ばれる特徴点を利用するもの [13] がある。この手法では、文書画像における行の性質を利用して水平方向の消失点を、行端の並びから垂直方向の消失点を推定し、それらを用いて射影歪みの補正を行う。しかし、この手法では消失点の推定に時間がかかることや、また垂直消失点の推定精度が十分とはいえないなどの問題がある。

## 3. 複比の投票に基づく文書画像検索

本稿では、以上の問題点の解決を目指して、文書画像検索、射影歪み補正の両者について、共通の仕組みに基づく手法を提案する。以下では、まず文書画像検索について述べる。

### 3.1 基本方針

提案手法の目的は、デジタルカメラで撮影された文書画像を検索質問とした文書画像データベースの検索である。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することで検索を行う。

### 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1に示す。文書画像は、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・低解像度などの影響下でも同一の特徴点が得られることである。英文文書における単語の重心は、この条件を

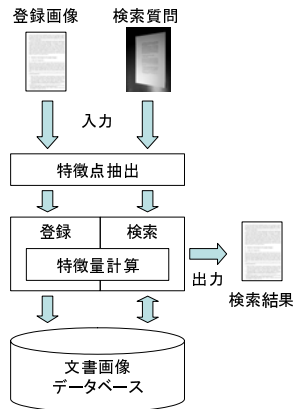


図 1 文書画像検索システム

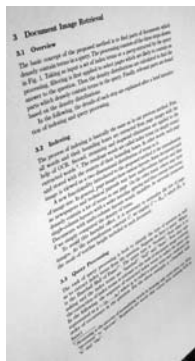


図 2 入力画像

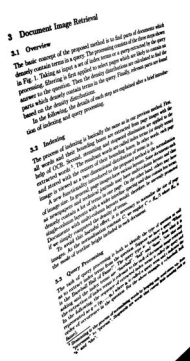


図 3 2 値化後

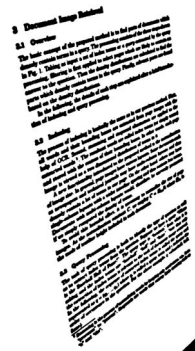


図 4 単語ごとに結合

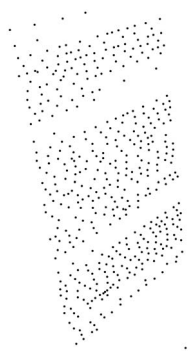


図 5 特徴点

満たす特徴点の一つである。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 2)は、まず適応 2 値化<sup>(注1)</sup>により 2 値画像(図 3)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、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 4)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(図 5)とする。

(注1): 適応 2 値化, ガウシアンフィルタには OpenCV [14] の関数を使用した。

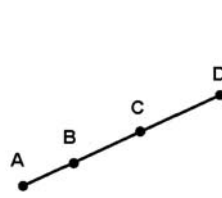


図 6 複比

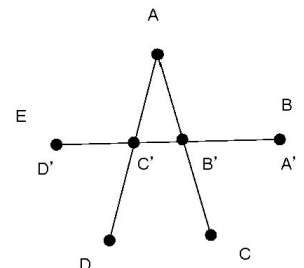


図 7 5 点からの複比

### 3.4 不変量を用いた特徴量

本研究では射影変換の不変量である複比を用いて特徴量を計算し、文書画像の登録・検索を行う。

#### 3.4.1 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 [9] と呼ばれるものがある。これは、図 6 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。

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

また、複比は同一平面上の 5 点から、図 7 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。

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<sup>(注2)</sup>。

#### 3.4.2 特徴量計算

本研究では、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。

#### 3.4.3 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

(注2): 複比の誤差や離散化については、文献 [15] 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。

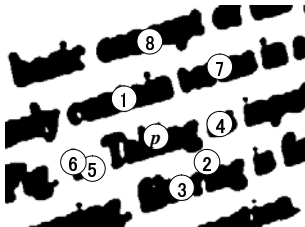


図 8 近傍点 1

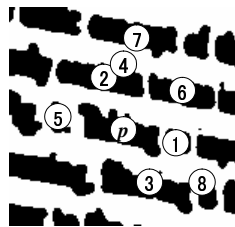


図 9 近傍点 2

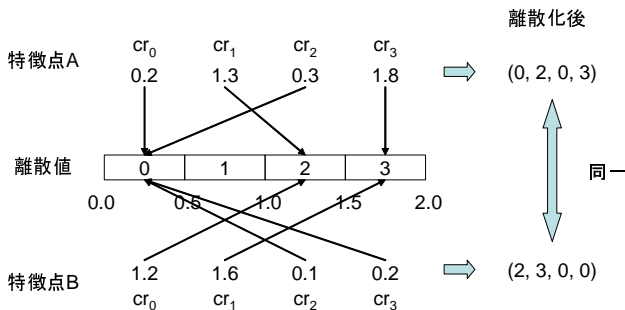


図 10  $n$  点からの 5 点の複比の問題点

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、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。

### 3.4.4 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、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<sup>(注3)</sup>。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。頻度のみを考えると両者とも 0 が 2 回、2, 3 が各々 1 回の出現となり、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。

### 3.4.5 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

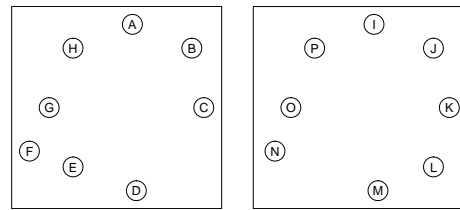


図 11 8 点

- 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:  $H_{index} \leftarrow cr_{0-4}, {}_m C_5$  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,  ${}_n C_m$  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

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 IJKLMNO P 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 ABCDEF から得られた複比と点 IJKLMN から得られた複比、次は点 ABCDGF から得られた複比と点 IJKMO 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、3.4.4 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。 ${}_n C_m$  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、

(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。

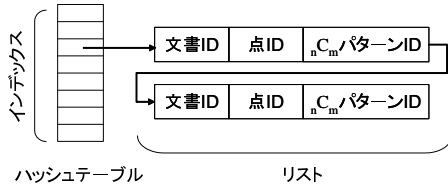


図 13 ハッシュの構造

$0 \sim {}_n C_m - 1$  の値をとる．同様に  ${}_m C_5$  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， $0 \sim {}_m C_5 - 1$  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らせて得られるものである<sup>(注4)</sup>．

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， $V_{max}$  は離散化された複比の最大値， $pat$  は  ${}_m C_5$  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,  ${}_n C_m$  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  ${}_n C_m$  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集計するためである．

なお，各点ごとのハッシュへの登録回数は  ${}_n C_m \times {}_m C_5$  となる．例えば， $n = 8, m = 7$  のときは 168 回である．

### 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，セ

- 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  $P_5 \in \{ P'_m \text{ からのすべての 5 点の組み合わせ} \}$  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:  $H_{index} \leftarrow cr_{0-4}, {}_m C_5$  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

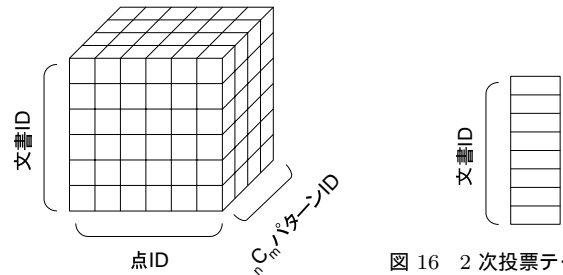


図 15 1 次投票テーブル

図 16 2 次投票テーブル

ルの属する文書 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 のように巡らせたものを  $m$  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．

なお，各点ごとのハッシュへのアクセス回数は  ${}_n C_m \times m \times {}_m C_5$  となる．同様に  $n = 8, m = 7$  のときは 1176 回である．

## 4. 書き込み抽出への応用

ここでは，3. で提案した文書画像検索手法の応用として，書き込み抽出 [12] を考える．以下，書き込み抽出システムの概要と，そこで用いられる射影歪みの補正方法について述べる．

### 4.1 書き込み抽出システム

書き込み抽出とは，書き込みのなされた印刷文書から書き込みだけを取り出すことである．提案手法では，書き込みのない文書 (元の電子文書) から得た画像にアクセス可能であるとの

(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．



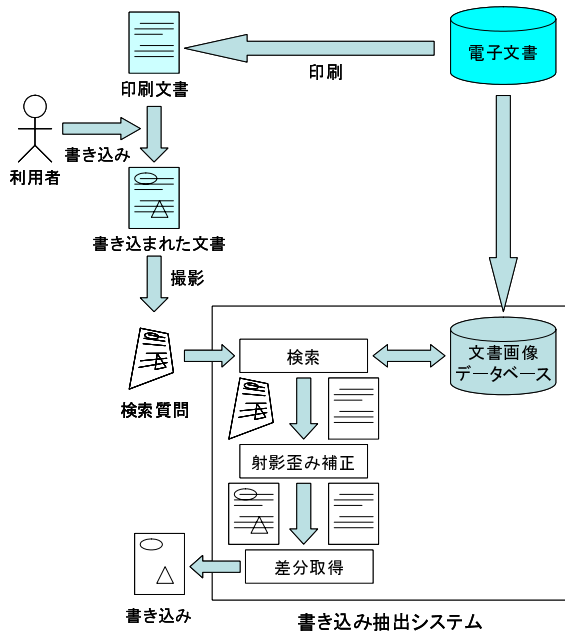


図 17 書き込み抽出の流れ

前提条件のもと、画像の差分をとることにより書き込みを抽出する。

書き込み抽出の流れを図 17 に示す。まず、コンピュータを用いて作成された電子文書は、印刷によって紙媒体の印刷文書となり、利用者のもとに届けられる。利用者によって書き込みのなされた文書は、デジタルカメラで撮影されて文書画像に変換され、検索質問として書き込み抽出システムに入力される。

書き込み抽出システムでは、まず前述した文書画像検索手法により、対応する文書画像を文書画像データベースから検索する。ここで、文書画像データベースには電子文書を収集・変換した文書画像を納めている。次に、得られた文書画像を用いて検索質問の射影歪みを補正する。このとき、検索の過程で得られた特徴点の対応関係を利用する。最後に、射影歪みの取り除かれた検索質問と、対応する文書画像の差分を取ることで書き込みを抽出し、出力する。

#### 4.2 射影歪みの補正

3. の検索処理では、その過程で入力画像の特徴点とデータベース中の文書画像の特徴点の対応関係を得ることが可能である。これを利用して、入力画像の射影変換パラメータを推定し、正対画像を得ることができる。

具体的には、図 14 の検索処理に以下の対応点の判別処理を付け加えることで入力画像とデータベース中の文書画像の特徴点の対応関係を得ることができる。まず 12 行目で 2 次投票テーブルに投票する際、同時に図 18 に示す対応点投票テーブルにも投票を行う。次に 16 行目で点  $p$  についての処理を終える際、対応点投票テーブルで最も多くの得票数を得た点を点  $p$  の対応点として記録する。

このようにして、図 19 に示すような特徴点の対応関係が得られる。射影変換のパラメータは 8 次元であるため、4 組の対応関係から求められるが、ノイズや誤った対応関係の影響を避けるため、以下のような手法で射影変換パラメータを求める。

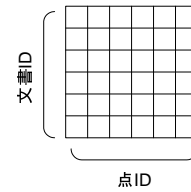


図 18 対応点投票テーブル

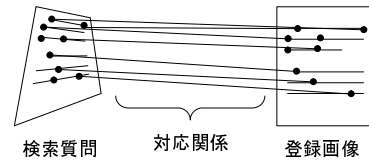


図 19 対応関係

- 1:  $a \leftarrow R$  の平均ベクトル
- 2:  $v \leftarrow R$  の分散ベクトル
- 3: do
- 4: for  $i = 1$  to  $s$  do
- 5:  $d_i \leftarrow r_i$  の重みつきユークリッド距離
- 6: end for
- 7:  $a, v \leftarrow d_i$  の小さいもの上位  $t$  個の平均および分散
- 8: while  $\sum d_i$  の減少率  $\geq e$
- 9:  $a$  を射影変換パラメータとする

図 20 射影変換パラメータ推定アルゴリズム

まず、対応関係を 4 組ランダムに取り出し、8 次元ベクトルで表される射影変換パラメータ  $r_i = (r_{i1}, r_{i2}, \dots, r_{i8})$  を求める。これを  $s$  回繰り返すことで、射影変換パラメータの集合  $R = \{r_1, r_2, \dots, r_s\}$  を得る。ここから図 20 に示すアルゴリズムを用いて射影変換パラメータを推定する。

アルゴリズムに沿って説明する。まず、1,2 行目で  $R$  の平均ベクトル  $a = (a_1, a_2, \dots, a_8)$  と分散ベクトル  $v = (v_1, v_2, \dots, v_8)$  を求める。次に、4~6 行目で  $R$  の各パラメータ  $r_i (i = 1 \sim s)$  について、以下の式で重み付きユークリッド距離  $d_i$  を求める。

$$d_i = \sqrt{\frac{\sum_{j=1}^8 \frac{(r_{ij} - a_j)^2}{v_j}}{\sum_{j=1}^8 \frac{1}{v_j}}} \quad (3)$$

次に、7 行目で  $R$  における  $d_i$  の小さいもの上位  $t$  個の平均および分散を求め、 $a, v$  とする。8 行目では、 $r_i$  の重みつきユークリッド距離の総和  $\sum_{i=1}^s d_i$  を求め、前回のループで求められた総和と比較して減少率を得る。減少率が  $e$  以上なら重みつきユークリッド距離の計算に戻って繰り返し処理を行い、そうでなければ  $a$  を入力画像の射影変換パラメータとする。

このようにして得られたパラメータを用いて入力画像を射影変換し、正対画像を得る。

## 5. 実験

### 5.1 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

There is that middle device with digital camera and color image data output...  
 In order to extract document images from the viewpoint of picture recognition, we have created a process of generating a QA system called "QAQ" (Question Answering Query) [1]. In this paper, we propose a method of document image recognition (QAQ) by modifying our previous method [1]. The main modification of this paper is the use of appropriate methods which is called the feature method in this paper. In this paper, we describe the details of the feature method and also present the experimental results of the feature method. The results of experiments show that the proposed method is superior to the feature 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 task is evaluated by the fact of all documents containing the same answer. In order to solve this task, the system must be able to extract the correct answer from the document. In this paper, we propose a method of document image recognition (QAQ) by modifying our previous method [1]. The main modification of this paper is the use of appropriate methods which is called the feature method in this paper. In this paper, we describe the details of the feature method and also present the experimental results of the feature method. The results of experiments show that the proposed method is superior to the feature method.

2.2. Query Processing

The query is processed in natural language. This is because the task is evaluated by the fact of all documents containing the same answer. In order to solve this task, the system must be able to extract the correct answer from the document. In this paper, we propose a method of document image recognition (QAQ) by modifying our previous method [1]. The main modification of this paper is the use of appropriate methods which is called the feature method in this paper. In this paper, we describe the details of the feature method and also present the experimental results of the feature method. The results of experiments show that the proposed method is superior to the feature method.

Table 1. Experimental results of the feature method.

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

図 21 データベース中の画像 1

図 22 データベース中の画像 2

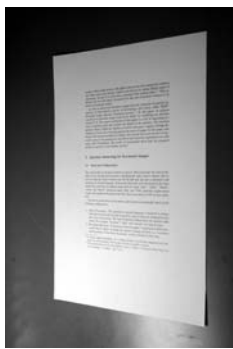


図 23 撮影範囲 A

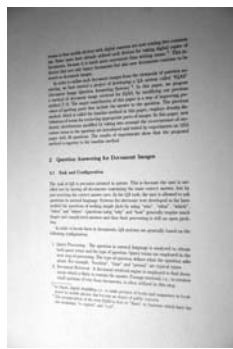


図 24 撮影範囲 B

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

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 段階の英語論文の 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 21, 図 22 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。

### 5.2 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 23~図 26 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。

### 5.3 実験 2: 携帯電話付属のデジタルカメラ

携帯電話で撮影した画像を検索質問として検索を行った。検索質問には、異なる 4 ページに対して、図 27~図 31 のように

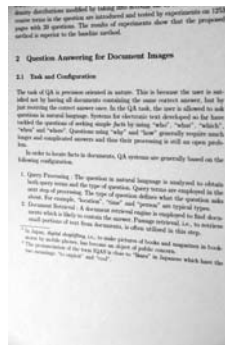


図 25 撮影範囲 C



図 26 撮影範囲 D



図 27 撮影範囲 A



図 28 撮影範囲 B

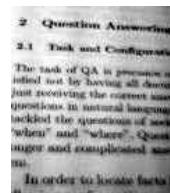


図 29 撮影範囲 C

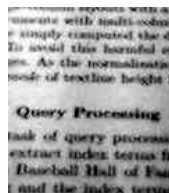


図 30 撮影範囲 D

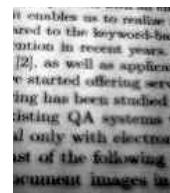


図 31 撮影範囲 E

表 2 実験結果

撮影範囲 (行数)	A (20)	B (16)	C (14)	D (12)	E (10)
正答率 [%]	50	100	50	75	0
処理時間 [sec]	28.2	25.4	22.5	20.4	18.2

5 段階の撮影範囲で撮影した画像を計 20 枚用意した。

なお、撮影範囲 A は 20 行、B は 16 行、C は 14 行、D は 12 行、E は 10 行程度のテキストが入るようにした。

表 2 に実験結果を示す。このように、高解像度デジタルカメラを利用した場合と比べて検索精度は低下した。これは、解像度が低いと単語の分離が困難になるためである。また、表 2 の撮影範囲 E で正答率が 0%なのは、画像に含まれる領域が狭すぎると近傍点が正しく得られないためである。

以上により、携帯電話付属のデジタルカメラのように解像度が低い機器を利用する場合でも、撮影範囲の調整が必要になるものの検索は可能であることが示された。

### 5.4 実験 3: 射影歪みの補正

入力画像とデータベース中の画像の特徴点の対応から射影変換のパラメータを推定し、入力画像の正対画像への復元を行った。対象は実験 1 で用いた入力画像 40 枚である。目視により復元結果の評価を行った結果、38 枚については正対画像が得られたと判定されたが、残り 2 枚では正対画像が得られず、成功率は 95%となった。

成功例における入力画像を図 32 に、復元された正対画像を

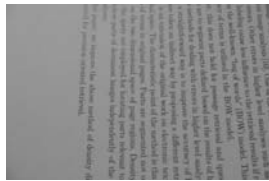


図 32 入力画像

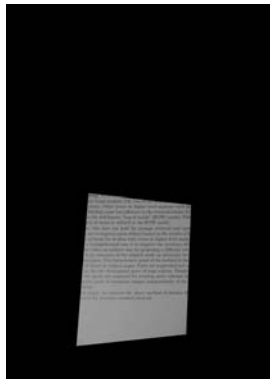


図 33 復元結果

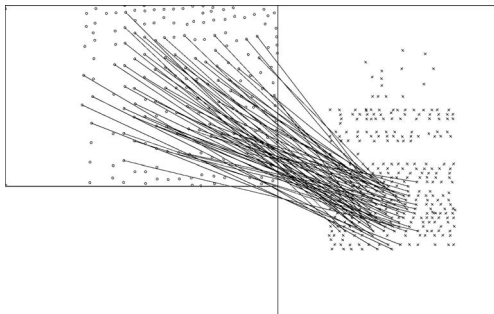


図 34 特徴点の対応



図 35 入力画像



図 36 復元結果

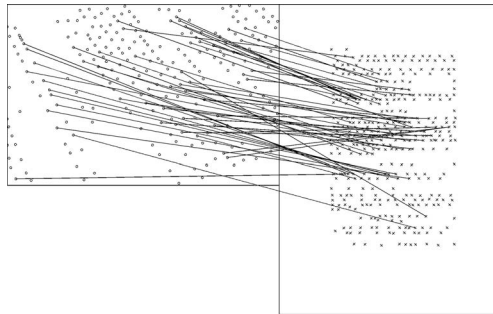


図 37 特徴点の対応

図 33 に示す。また、特徴点の対応関係を図 34 に示す。図 34 で左側に入力画像の特徴点の集合を、右側にデータベース中の対応画像の特徴点の集合を示し、対応関係の得られたものを線で結んでいる。このように十分な数の対応点が得られるときは正しい結果が得られた。

次に、失敗例における入力画像、復元結果、特徴点の対応関係を図 35～図 37 に示す。この例では、正しく対応する点の数が少ないため、検索は正しく行えるものの射影変換のパラメータを正しく推定することができなかった。

なお、実験 2 の携帯電話で撮影した画像についても復元を試みたところ、すべての画像で正対画像を得ることができなかった。これは、画像中の特徴点が少ないことと、ノイズの影響による特徴点の座標の変動が大きいためと考えられる。

## 6. ま と め

本稿では、複比の投票により、デジタルカ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。解像度の低い携帯電話付属のデジタルカメラについても、撮影範囲を調整すれば検索が可能であることを示した。また、本手法では検索過程で点の対応を知ることが可能であり、それにより入力画像の射影歪みの補正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。また、ゆがみなどにより入力画像が平面の条件を満たさない場合での提案手法の有効性の検討や、射影歪みの補正が可能になる条件についての定量的な調査も必要である。射影歪みの補正を利用した、本手法の書き込み抽出への応用についても継続して取り組んでいきたい。

## 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, PRMU2004-246 (2005).
- [3] 山田, 仙田: "携帯カメラを用いたコピ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] D. Doermann: "The Indexing and Retrieval of Document Images: A Survey", Computer Vision and Image Understanding, 70, 3, pp.287–298 (1998).
- [6] D. Doermann, H. Li and O. Kia: "The Detection of Duplicates in Document Image Databases", Proc. ICDAR'97, pp.314-318 (1997).
- [7] K. Hannu: "Document Image Retrieval with Improvements in Database Quality", Academic Dissertation of University of Oulu (1999).
- [8] J. J. Hull: "Document Image Matching and Retrieval with Multiple Distortion-Invariant Descriptors", Document Analysis Systems, pp.379–396 (1995).
- [9] 佐藤: コンピュータビジョン - 視覚の幾何学 -, コロナ社 (1999).
- [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ージメディア, 44, SIG5(CVIM6), pp.64–73 (2003).
- [12] 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, D-12-62 (2005).
- [13] P. Clark and M. Mirmehdi: "Recognising text in real scenes", IJDAR, 4, pp. 243–257 (2002).
- [14] OpenCV: <http://sourceforge.net/projects/opencvlibrary/>
- [15] 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).