

特徴点の局所的配置に基づくリアルタイム文書画像検索と その拡張現実への応用

中居 友弘[†] 黄瀬 浩一[†] 岩村 雅一[†]

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

あらまし 本稿で示すリアルタイム文書画像検索とは、Web カメラを用いて撮影された文書に対応する電子文書をデータベースからリアルタイムで検索する処理である。我々はすでに特徴点の局所的配置に基づく文書画像検索法を提案しており、リアルタイム文書画像検索はこの手法の高速性を利用したリアルタイム処理への応用である。また、この手法では検索の過程で文書の部分ごとの対応関係が得られるため、撮影範囲や紙面の傾きを推定することができる。そこで、リアルタイム文書画像検索の応用として、文書の関連情報を撮影画像に重畳表示する拡張現実も示す。

キーワード 文書画像検索, リアルタイム処理, Web カメラ, 拡張現実

Real-Time Document Image Retrieval Based on Local Arrangements of Feature Points and Its Application to Augmented Reality

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

[†] Graduate School of Engineering, Osaka Prefecture University

1-1 Gakuencho, Naka, Sakai, Osaka, 599-8531 Japan

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

Abstract In this report, we propose a method of real-time document image retrieval. It is a task of searching document images from the database based on query images captured by a web-camera. We have already proposed a method of document image retrieval based on local arrangements of feature points. Real-time document image retrieval is realized by applying the method to each frame captured by a web-camera. As an application of the real-time retrieval, we also propose a method of augmented reality for documents. This method is to superimpose relevant information onto the camera-captured image naturally by using the parameters of perspective transformation calculated as a subsidiary result of retrieval.

Key words Document image retrieval, Real-time processing, Web camera, Augmented reality

1. はじめに

文書画像検索とは、利用者の与えた検索質問に対応する文書画像をデータベースから検索する処理である。本稿では特に、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、データベースから同一の文書を検索する場合を考える。

この問題について、我々はすでに特徴点の局所的配置に基づく文書画像検索法 [1] を提案している。この手法は、特徴点の局所的な配置に基づいて幾何学的不変量を利用した特徴量を求め、これをインデックスとして特徴点をハッシュ表に格納することで高速検索を実現する。

この手法の特長として、特徴量計算の計算量が検索対象の特徴点数 N に対して $O(N)$ であるため、文書画像のように多くの特徴点をもつ画像についても高速に特

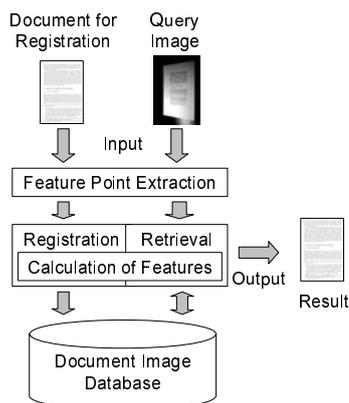


図 1: 文書画像検索システム

微量計算ができる点が挙げられる。Geometric Hashing (GH) [2] のような従来手法では、射影歪みに不変なものにするためには $O(N^5)$ の計算量が必要となるため、処理速度の面で大きな問題がある。提案手法はこの問題を解決するものである。また、ハッシュによる近似近傍探索を用いることで、検索処理についても高速化が実現されている。具体的には、特徴量ベクトルが 15 次元のとき、検索質問の約 1.7×10^4 個の特徴量ベクトルとデータベースの約 5.0×10^7 個の特徴量ベクトルとのマッチングを 80ms 程度の時間で実現している。

本稿では、この手法に基づくリアルタイム文書画像検索法を提案する。これは、入力機器として Web カメラを用いて、そこから得られるフレームの画像を検索質問とし、対応する文書画像を検索するという処理をフレームごとに繰り返し行うものである。

また、リアルタイム検索を用いたサービスとして、文書への拡張現実を提案する。これは、検索の過程で得られる検索質問と登録文書の特徴点の対応関係を用いて撮影された文書の位置と姿勢を推定し、関連情報を重畳表示するものである。

以下では本稿の構成について述べる。2 節では我々の既に提案した特徴点の局所的配置に基づく文書画像検索法について説明する。3 節では本稿で提案するリアルタイム文書画像検索法について述べる。4 節ではリアルタイム検索の拡張現実への応用について述べる。

2. 特徴点の局所的配置に基づく文書画像検索法

2.1 処理の流れ

文書画像検索システムの構成を図 1 に示す。登録文

書 (document for registration) および検索質問 (query image) は、特徴点抽出 (feature point extraction) により点の集合に変換された後、登録の場合は登録処理 (registration) へ、検索の場合は検索処理 (retrieval) に入力される。これらの処理では、特徴点から不変量を用いた特徴量を求める特徴量計算 (calculation of features) に同じ処理を用いる。登録処理では、特徴点から得られた特徴量をハッシュ表のインデックスに変換し、それを用いて文書画像データベース (document image database) へ登録する。一方、検索処理では、特徴量から同様にインデックスを計算し、投票する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。

2.2 特徴点抽出

本稿では [1] と同様に英文文書における単語領域の重心を特徴点として用いる。単語領域の重心は射影歪みやノイズの影響下でも比較的安定に抽出できるため、安定な特徴量の計算に有効である。

2.3 アフィン不変量

提案手法では、同一平面上の 4 点から計算されるアフィン不変量を用いて特徴量を計算する。アフィン不変量はアフィン変換における不変量である。アフィン変換とは、直線の平行性が維持される幾何変換であり、デジタルカメラで撮影された画像で生じる射影変換よりも自由度の低い変換である。射影変換を受けた平面においても、多くの場合、局所領域ではその変換がアフィン変換に近似されるため、アフィン不変量を利用することが可能となる。アフィン不変量は同一平面上の 4 点 $ABCD$ の座標から $P(A, C, D)/P(A, B, C)$ で計算される [3]。

さて、画像上の特徴点から得られるアフィン不変量の値は連続値だが、インデックスに用いる際には離散化する必要がある。本手法では、予備実験により特徴点から得られるアフィン不変量のヒストグラムを作成し、各離散値におけるアフィン不変量の頻度が等しくなるように離散値を割り当て、 k 階調に量子化する。

2.4 特徴量計算

2.4.1 特徴量の満たすべき条件

提案手法における特徴量とは、文書画像の特徴点を表現する量である。検索質問および登録文書のそれぞ

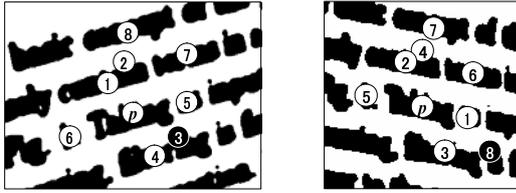


図 2: 射影歪みによる特徴点 p の近傍点の変化 (数字は点 p との距離の順位)

れについて、特徴点から得られる特徴量を計算し、それらの値を比較することで検索質問と登録文書が対応しているか否かを判断する。

特徴量の評価基準としては、検索精度や検索に要する計算量が挙げられる。検索質問に対応する登録文書を正確に、かつ高速に検索できる特徴量が優れた特徴量であるといえる。

本稿では、特徴量の精度に関する条件として以下の 2 条件を定義する。第一の条件は、各種歪みの影響を受けても同じ文書の同じ点からは同じ特徴量が得られなければならないということである。もし登録文書と検索質問から異なる特徴量が得られれば、検索によって正しく対応する特徴点を見つけることはできない。この条件を「特徴量の安定性」と呼ぶ。第二の条件とは、異なる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ことである。もし異なる文書から同じ特徴量が得られれば、検索の際に正しく対応する特徴点だけでなく対応しない特徴点まで見つかることになる。この条件を「特徴量の識別性」と呼ぶ。

また、いうまでもなく、安定性や識別性の高い特徴量であっても、計算量が膨大であれば利用は困難となる。従って、特徴量は精度に関する上記の 2 条件を満たすだけでなく、計算量が小さい必要もある。

これら 3 つの条件が高速かつ高精度な文書画像検索における特徴量の満たすべき条件である。

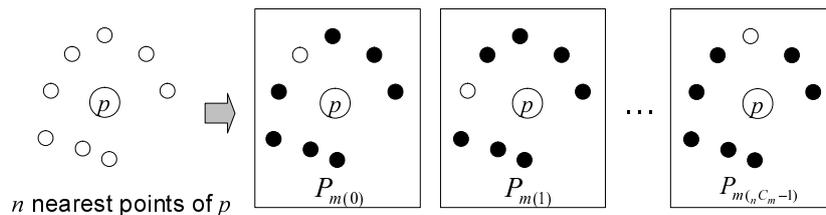


図 3: $n(=8)$ 点から $m(=7)$ 点を選ぶすべての組み合わせ

2.4.2 特徴量の安定性

まず特徴量の安定性について述べる。前述したように、提案手法では各特徴点の近傍点から不変量を用いて特徴量を計算する。特徴量が安定であるためには、射影歪みによって近傍点の座標が変化しても特徴量の計算に用いる点に同じ点が得られる必要がある。図 2 に示されるように、近傍点は射影歪みの影響で変化するため、特徴点 p の近傍 4 点から計算される不変量の特徴量とすると、同じ特徴点 p から同じ特徴量を得ることはできない。

そこで、提案手法ではより広い範囲の近傍点から複数の点の組み合わせを作成し、そこから複数の特徴量を計算する。これは、図 2 のように、射影歪みの影響があってもある程度広い範囲の近傍 n 点 (図 2 では 8 点) のうち m 点 (図 2 では 7 点) までは同じ点が含まれるという仮定に基づいている。近傍 n のうち m 点が同一であるならば、図 3 のように n 点からすべての m 点の組み合わせ $P_{m(0)}, P_{m(1)}, \dots, P_{m(nC_m-1)}$ を作成し、それぞれ特徴量を計算することで、少なくとも 1 つは同じ特徴量が得られると考えられる。

2.4.3 特徴量の識別性

次に、特徴量の識別性について述べる。提案手法では、1 つの特徴量の計算に用いる特徴点の数 m を増やすことで識別性を高める。 m 点の配置の表現方法としては、図 4 のように m 点から得られるすべての 4 点の組み合わせから計算されるアフィン不変量の列 $r_{(0)}, r_{(1)}, \dots, r_{(mC_4-1)}$ を用いる。 m が大きければ大きいほど、計算される不変量の数が多くなるため、同じ特徴量が偶然に現れる可能性は低くなる。

ただし、 m が大きすぎると特徴量の安定性が低下する。なぜなら、特徴量が一致するためには特徴量の不変量すべてが一致する必要があるが、 m が大きいと不変量の数 mC_4 が増加するので、誤差の影響で異なる不変量が計算される可能性が高くなるためである。

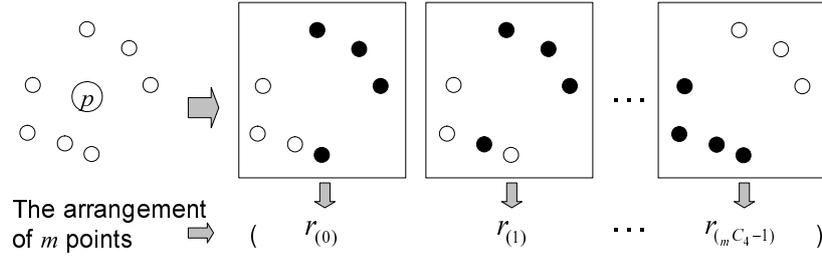


図 4: すべての 4 点の組み合わせから得られる不変量で表現される $m(=7)$ 点の配置

2.4.4 計算量と記憶容量

以上に述べたように、 n を大きくすると広い範囲から複数の特徴量を計算するようになり、特徴量の安定性が向上する。また、 m を大きくすると 1 つの特徴量の計算に用いる点の数が増加するため、特徴量の識別性が向上する。しかし、これらのパラメータを大きくし過ぎると計算量の面で問題が生じる。 n および m が大き過ぎると、不変量の計算回数が増加する。そのため、登録および検索に要する処理時間が大きくなる。また、登録時には計算される特徴量を記憶しておく必要があるため、大きな記憶容量が必要となる。

2.4.5 不変量の量子化レベル

安定性および識別性に影響するパラメータは n, m だけではない。不変量の量子化レベル k もこれらに影響する。 k の値が大きいく（不変量が細かく離散化される）と、誤差の影響で同じ点から計算される不変量が異なる値に離散化される可能性が高くなり、安定性が低下する。 k の値が小さい（不変量が粗く離散化される）と、異なる 4 点から計算されたアフィン不変量が同じ値に離散化される可能性が高くなり、識別性が低下する。

以上のことから、高い検索精度と同時に高速な処理や小さい記憶容量を実現するには、 n, m および k を適切に設定する必要があるといえる。

2.5 登録・検索

以上で説明した特徴量を用いた文書画像の登録および検索方法について述べる。本稿では詳細なアルゴリズムについては省略し、概略のみを説明する。詳細については [1] を参照されたい。

まず、登録について述べる。ここまでで述べたように、この手法では、各特徴点の近傍 n 点から m 点を取

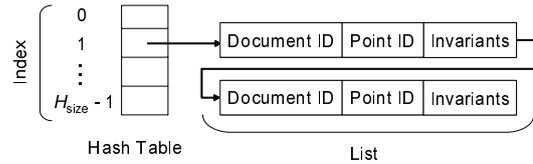


図 5: ハッシュ表の構造

り出し、 mC_4 次元のアフィン不変量のベクトルによって表現される m 点の配置を特徴量としている。この特徴量を、以下に示すハッシュ関数によってハッシュ表のインデックス H_{index} に変換する。

$$H_{\text{index}} = \left(\sum_{i=0}^{mC_4-1} r_{(i)} k^i \right) \bmod H_{\text{size}} \quad (1)$$

ここで、 k は不変量の量子化レベル、 H_{size} はハッシュ表のサイズである。そして、インデックスを用いて図 5 に示されるようなハッシュ表へ、登録文書の識別番号である文書 ID (Document ID) と点の識別番号である点 ID (Point ID)、不変量 (Invariants) $r_{(i)}$ ($i = 0, 1, \dots, mC_4 - 1$) を登録する。登録時に衝突が生じた場合、データは図 5 のようにリスト構造で付け加えられる。

次に検索について述べる。登録時と同様に、特徴点の局所的配置から特徴量を求め、式 (1) を用いてハッシュ表のインデックスを求める。インデックスを用いて登録処理で作成されたハッシュ表にアクセスし、登録されている文書 ID の文書に対して投票する。このような処理をすべての点について繰り返し、最終的に最も多くの得票数を得た文書を検索結果とする。

3. リアルタイム検索

3.1 処理の概要

文献 [1] の手法は、デジタルカメラで撮影された画像

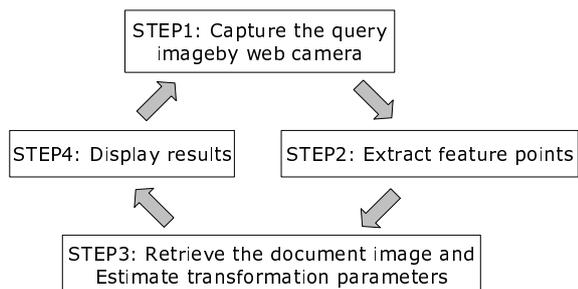


図 6: リアルタイム検索の処理の流れ

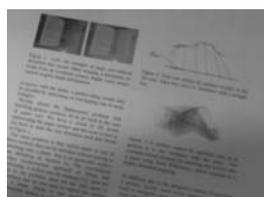


図 7: キャプチャ画像

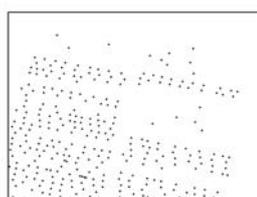


図 8: 特徴点抽出結果

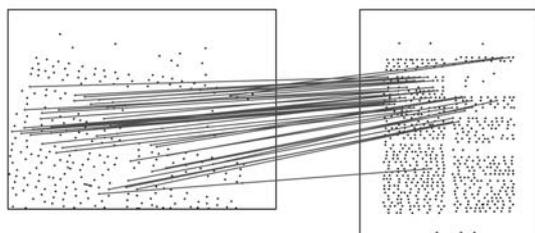


図 9: 検索結果と特徴点の対応関係

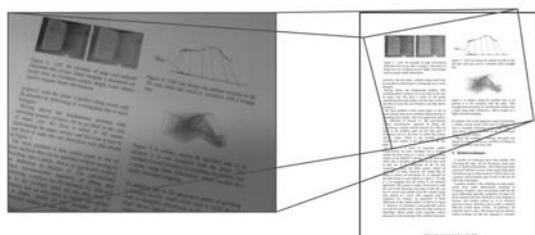


図 10: 射影変換パラメータの推定による撮影範囲の特定

を検索質問として、対応する文書画像をデータベースから検索するものである。本稿で提案するリアルタイム文書画像検索は、図 6 に示すように文献 [1] の手法を Web カメラで取得した画像の各フレームに適用し、繰り返し行うものである。

以下では繰り返し処理について述べる。まず Web カメラによって検索質問となる画像を取得し (STEP1) 、

- 1: for $i = 0$ to T do
- 2: ランダムに 4 つの対応関係を選ぶ
- 3: 4 つの対応関係から射影変換パラメータを求める
- 4: すべての対応関係について、実際の対応点とパラメータによって推定される対応点との距離を調べ、閾値以下のものの数を数えてスコアとする
- 5: スコアが最大であれば暫定的な推定値とする
- 6: end for
- 7: if 最大スコアが閾値以上 then
- 8: 推定値を射影変換パラメータとする
- 9: else
- 10: 推定失敗
- 11: end if

図 11: RANSAC による射影変換パラメータ推定アルゴリズム

特徴点を抽出する (STEP2) 。Web カメラによるキャプチャ画像を図 7 に、特徴点抽出結果を図 8 に示す。次に、特徴点に基づいて文書画像検索を行う。ここでは、対応する登録文書の特定だけでなく、図 9 のような特徴点単位の対応関係も取得する。そして、対応関係を用いて撮影画像の射影変換パラメータを推定する (STEP3) 。射影変換パラメータによって図 10 のように撮影範囲を特定することができる。最後にこれらの結果を表示する (STEP4) 。一連の処理が完了すると、再び Web カメラによる画像の取得に戻り、処理を繰り返す。

このように、提案手法では Web カメラから得られる各フレームを検索質問とし、文書画像検索を繰り返すことでリアルタイム処理を実現している。フレームごとに検索を行うことで、入力の変化に対する高い追従性を実現される。つまり、撮影される文書が変わったとき、以前のフレームの情報を蓄積して検索を行う手法では数フレーム反応が遅れるが、本手法ではフレームごとに処理を行うことでより迅速な反応が可能にする。

3.2 射影変換パラメータの推定

ここでは、検索質問と登録文書の間の特徴点の対応関係に基づく射影変換パラメータの推定方法について述べる。図 9 のようにして得られる特徴点の対応関係には、誤ったものも含まれている。そのため、誤った対応を除いて適切な変換パラメータを推定する必要がある。提案手法では、RANSAC [4] を用いてパラメータの推定を行う。

推定アルゴリズムを図 11 に示す。まず、対応関係の集合からランダムに 4 つの対応関係を選ぶ。そして、4 つの対応関係から射影変換パラメータを求める。次に、得られた射影変換パラメータの評価を行う。これは、射

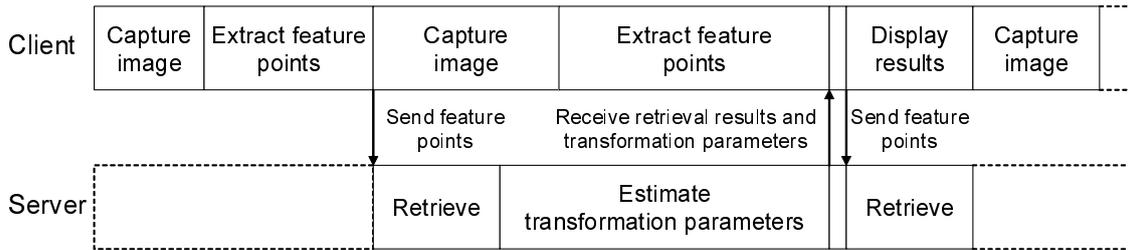


図 12: クライアントサーバシステムによるパイプライン処理

影変換パラメータを支持する対応関係の数をスコアとすることで行われる。具体的には、対応関係の検索質問側の点をパラメータによって変換し、得られた点の座標と対応関係の登録文書側の点の座標との距離を求めて、閾値以下であれば対応関係がパラメータを支持するとみなす。このような処理を T 回繰り返し、最大のスコアが閾値以上であれば最大のスコアを得たパラメータを推定結果とする。そうでなければ推定失敗とする。

3.3 パイプライン処理

上述したように、リアルタイム検索処理は画像の取得や特徴点抽出などのいくつかの処理に分かれている。これらの処理は独立に行うことが可能であるため、クライアントサーバシステムによってパイプライン処理を実現する。なお、現在の実装ではクライアントおよびサーバはそれぞれ 1 つの計算機でシステムを構成している。図 12 にクライアントおよびサーバの担当する処理とその流れを示す。まず、クライアントにおいて Web カメラを用いて画像を取得 (Capture image) する。次に、画像から特徴点を抽出 (Extract feature points) する。得られた特徴点のデータはサーバに送られ、サーバで特徴点に基づいて検索 (Retrieve) および変換パラメータの推定 (Estimate transformation parameters) が行われる。サーバでの処理と平行して、クライアントでは次のフレームの画像取得および特徴点抽出が行われる。クライアントとサーバ双方の処理が完了した時点でサーバからクライアントへ検索結果および射影変換パラメータが送られ、クライアントからサーバへ特徴点のデータが送られる。クライアントでは検索結果を表示 (Display results) した後、画像の取得と特徴点抽出が繰り返し行われる。サーバでも同様に検索と変換パラメータの推定が繰り返される。このように、パイプライン処理によって並行して処理を行うことで高いフレームレートを実現することが可能

表 1: 処理ごとの計算時間

	Processing time[ms]		
	100pages	1,000pages	10,000pages
Extract feature points	66.9		
Retrieve	60.7	63.3	81.2
Estimate transformation parameters	0.27		

になる。

3.4 実験結果

提案手法の性能を調べるために行った実験の結果について述べる。実験では、入力としてあらかじめ 130 万画素の Web カメラを用いて 7fps で撮影しておいた長さ 27 秒、194 フレームの動画を用いた。この動画は [5] に公開されている。動画のフレームの例を図 13 に示す。このように、動画には文書の部分のみが写っているフレーム、隠れのあるフレーム、紙面の曲げられたフレームが含まれている。この動画を入力としてデータベースにおける登録ページ数を 100, 1,000, 10,000 と変化させたときの特徴点抽出、検索、変換パラメータ推定の処理時間と精度を調べた。ここで、精度とはすべてのフレームの検索結果における、撮影された文書が正しく検索されたものの割合である。

なお、実験ではクライアントとして CPU Intel Pentium4 3.0GHz、メモリ 1.5GB の計算機、サーバとして CPU Opteron 2.8GHz、メモリ 16GB の計算機を用いた。検索手法におけるパラメータは $n = 7$, $m = 6$, $k = 15$, $H_{\text{size}} = 2^{27} - 1$, $T = 100$ とした。

以下では実験結果について述べる。特徴点抽出、検索、変換パラメータ推定に要した処理時間を表 1 に示す。特徴点抽出と検索に多くの処理時間を要した一方、変換パラメータの推定の処理時間は比較的小さいものであった。なお、全体のフレームレートは約 7fps であった。登録ページ数ごとに計測した検索の処理時間と精度を図 14 に示す。まず、検索の処理時間について述べる。

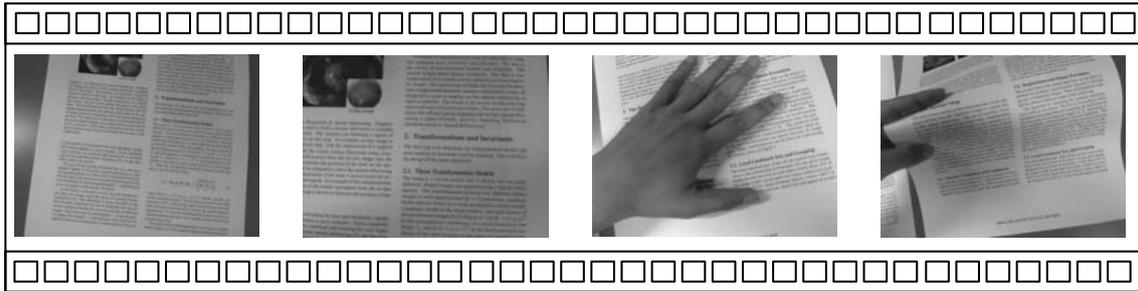


図 13: 入力として用いた動画のフレームの例

登録ページ数の増加にともなって処理時間も増加した。これは、ハッシュ表に登録されるデータの増大によって衝突が増加し、リストへのアクセス時間が増加したためと考えられる。ただし、登録ページ数 10,000 においても検索時間は約 80ms であり、リアルタイム処理の可能な処理時間は維持された。次に、検索精度について述べる。図 14 にはすべてのフレームを対象とした精度のグラフと、強いモーションブラーのかかったフレームを除いた精度のグラフを示している。図 15 のような強いモーションブラーがかかった文書画像の場合、本手法では原理的に特徴点の抽出が不可能となり、検索に失敗する。従って、検索処理の性能をより正確に測るためにはこのようなフレームを除く必要がある。なお、強いモーションブラーを受けたフレームは全 194 のうち 18 フレームであった。図 15 に示されるように、登録ページ数の増加にともなって精度もやや低下した。しかし、登録ページ数 10,000 の場合でも精度は約 88%、モーションブラーを受けたフレームを除くと約 93% であり、高い精度が実現された。なお、図 13 に示される 4 つのフレームは、すべての場合で検索に成功したものである。従って、部分のみが撮影された場合や、隠れ・紙面の湾曲のある場合でも、検索は可能であることが示された。

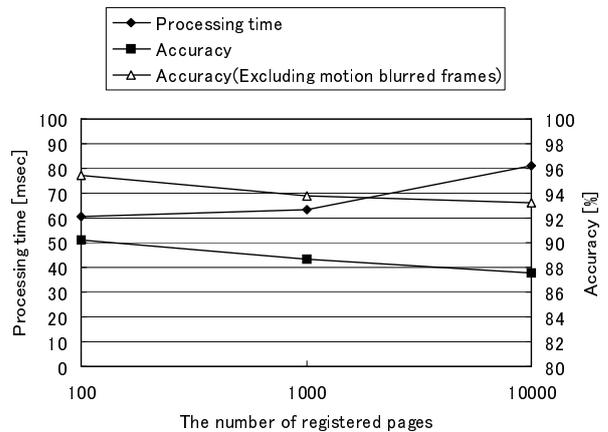


図 14: 実験結果

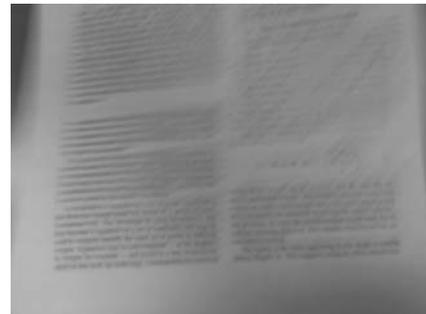


図 15: モーションブラーを受けたフレームの例

4. 拡張現実への応用

本稿では、リアルタイム文書画像検索の応用として拡張現実 [6] のサービスを提案する。これは、Web カメラで撮影された文書をリアルタイムで検索し、文書に関連付けられた情報を撮影画像に重畳して表示するものである。

関連付ける情報としては、テキスト、画像、下線、書き込みなどを想定している。これらの情報はリアルタイム検索の過程で得られた射影変換パラメータに基づいて図 16 のように紙面の位置・姿勢に適合した形で

重畳表示される。このような形で表示することにより、関連情報が文書のどの部分に関連付けられているかをリアルタイムで示すことが可能となる。なお、拡張現実処理のサンプル動画は [5] に公開されている。

リアルタイム検索も含めた全体のシステムを図 17 に示す。3 節で述べたように、まず Web カメラで得られた画像から特徴点を抽出する。次に、特徴点のデータを用いて検索を行い、データベースから対応する文書を見つける。また、検索の過程で得られた特徴点の対

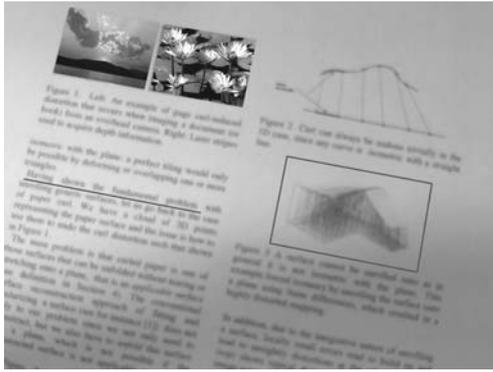


図 16: 関連情報の重畳表示

応関係を用いて撮影画像の射影変換パラメータを推定する．このようにして得られた検索結果と射影変換パラメータを用いて拡張現実による関連情報の提示を行う．これは，検索結果を用いて関連情報データベースから対応する関連情報を検索し，射影変換パラメータによって関連情報の形状を変換して撮影画像に重畳することで実現される．

関連情報の提示により，利用者は文書に記載されていない情報を得ることができる．また，データベース内の関連情報を更新することで常に最新の情報を提供することができる．利用者のプロフィールに基づいて提示する情報を変えるなど，利用者のニーズに適合したサービスを提供することも可能である．これらの性質から，リアルタイム検索に基づく拡張現実とは，文書のメディアとしての表現能力を拡張するものであるといえる．

5. まとめ

本稿では，我々のすでに提案している特徴点の局所

的配置に基づく文書画像検索法を用いたリアルタイム文書画像検索を提案した．提案手法の特徴は 7fps でのリアルタイム処理が可能である．本稿では，検索の過程で得られる特徴点の対応関係から推定される射影変換パラメータを用いた拡張現実への応用についても提案した．これは，文書のメディアとしての新たな利用方法を提示するものである．

今後の課題としては，拡張現実における重畳画像の安定化，様々な条件下でのロバスト性の向上などが挙げられる．写真やポスターなどの他の対象への提案手法の拡張も重要な課題である．

参考文献

- [1] 中居, 黄瀬, 岩村: “デジタルカメラを用いた高速文書画像検索におけるアフィン不変量および相似不変量の利用”, 信学技報, Vol. 105, No. 614, PRMU2005-188, pp.25-30 (2006).
- [2] H. J. Wolfson and I. Rigoutsos: “Geometric hashing: an overview”, IEEE Computational Science & Engineering, Vol. 4, No. 4, pp.10-21 (1997).
- [3] 佐藤: コンピュータビジョン—視覚の幾何学—, コロナ社 (1999).
- [4] M. A. Fischler and R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography”, Comm. ACM, Vol. 6, No. 24, pp.381-395 (1981).
- [5] <http://www.m.cs.osakafu-u.ac.jp/~nakai/drr/>.
- [6] R. T. Azuma: “A Server of Augmented Reality”, Presence, Vol. 6, No. 4, pp.355-385 (1997).

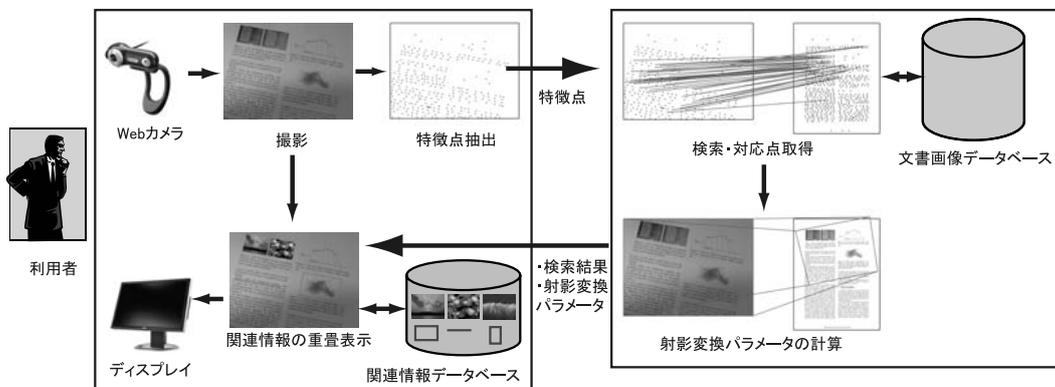


図 17: 全体のシステム