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.

Webカメラを用いた多言語文書画像のリアルタイム検索システム

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　本稿では，さまざまな言語で書かれた文書画像のリアルタイム検索法を提案する．これは，Web カメラで撮影された文書画像を検索質問とし，データベースから元となった文書画像をリアルタイムで検索して提示するものである．我々はすでに英語文書を対象とした文書画像検索法を提案しており，これは従来手法をさまざまな言語の文書に適用できるよう拡張したものである．従来手法である英語文書画像検索法では，画像処理によって得られる単語領域の重心を特徴点としていた．しかし，日本語や中国語を含むいくつかの言語では，言語の特性上識別性の高い特徴点を安定的に得ることが難しい．提案手法では，記述子の追加によってさまざまな言語の文書における高精度なリアルタイム文書画像検索を実現する．

キーワード　文書画像検索，リアルタイム処理，カメラベース文書画像処理，LLAH

A System of Real-Time Retrieval for Images of Documents in Various Languages using a Web Camera

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 real-time image retrieval system using a web camera for documents in various languages. This system takes camera-captured document images as queries and find the corresponding document images from a database. This is an extension of our real-time image retrieval for English documents. In the English document image retrieval system, centroids of word regions are used as feature points to describe documents. These feature points are effective since English has clear between word gaps. On the other hand, in some languages including Japanese and Chinese, words are not separated. Therefore alternative feature points such as centroids of character regions are required. In the proposed method, this problem is solved by using the additional features extracted from character regions.

Key words　Document image retrieval, Real-time processing, Camera-based document image processing, LLAH

1. はじめに

近年，デジタルカメラの普及が急速に進んでいる．特に，携帯電話の分野においてデジタルカメラの付属が一般的になっており，またそれらの品質についても通常のデジタルカメラと比較して遜色のないものとなっている．これにより，一般の利用者が高解像度デジタルカメラを常に携帯するという状況が生じている．

そこで，デジタルカメラで撮影された画像を用いたサービスが注目されている．代表的なものに 2 次元バーコード [1] が挙げられる．これは，白黒のドットパターンで構成される特徴的な模様を対象に貼り付けておき，模様に埋め込まれた情報を撮影画像から抽出することでサービスを提供するものである．ドットパターンは認識が容易になるよう，非常に目立つ外見をしていることが一般的である．しかし，このようなアプローチは，認識対象の外観を損ねることが問題とされることがある．また，文書などに 2 次元バーコードを関連付けるためにはあらかじめバーコードを含んだ状態で印刷しておく必要があり，後からパターンの追加や変更を行うことは困難であるという問題もある．

そこで，物体の外観そのものをキーとして情報検索を行うア

プローチが望まれている．これは，物体の外観と，物体に関連付けられた情報をデータベースとして保持しておき，利用者から与えられた物体の撮影画像に基づいて情報の提供を行うものである．このような処理では，物体の外観から，それが何であるかを判断するという物体認識を行うことになる [2]．しかし，入力機器としてデジタルカメラを用いる場合，撮影画像が撮影方向の変化によって変化するため，物体の外観が一定ではないという問題がある．

特に対象を印刷文書に限定し，撮影画像に対応する文書画像をデータベースから検索する手法として，LLAH [3] が提案されている．これは，特徴点の配置を幾何学的不変量で表現することにより，平面物体に対する撮影方向の変化に伴う射影歪みに不変な特徴量を求め，撮影画像の認識を行うものである．同様の処理を実現する手法として，文書画像のレイアウト構造を利用するもの [4] も提案されているが，LLAH はこの分野における最も先駆的な研究の 1 つである．LLAH は，現実的な利用において生じうる撮影方向の変化にロバストであり，また単純な検索の繰り返しでリアルタイム検索が実現できるほどの高速性をもつ [5]．

しかし，現時点で LLAH は英語文書のみを対象としており，異なる言語の文書における有効性は示されていない．LLAH が有効に機能するためには，特徴点がある程度安定に得られる必要がある．単語領域の重心という形で安定な特徴点が得られる英語文書については LLAH がうまく働くことが示されているものの，他の言語では安定な特徴点を得ることが難しいため，未だ適用例が示されていない．特定の言語の文書でのみ機能する技術を，別の言語の文書に用いる際には，手法の改良を伴う調整が必要となり，困難な問題が生じることがある [6]．しかし，日本を始めとしたデジタルカメラの広く普及した国々の母国語文書に適用することができれば，情報検索の応用範囲が大きく広がるため，LLAH によるさまざまな言語の文書画像検索は困難であると同時に重要な問題といえる．

本稿では，LLAH のさまざまな言語の文書への適用を提案する．日本語や中国語などの，分かち書きのされない言語の文書においては，従来手法において用いられてきた単語領域の重心を抽出することは容易ではない．提案手法では，これらの言語から抽出される特徴点として，連結成分の重心を用いる．しかし，日本語や中国語においては文字が等間隔に配置されるため，特徴点の配置を記述子として用いる LLAH との相性はよくない．どの文書からも同じ特徴点の配置が得られると，文書の識別は困難となる．そこで，提案手法では従来の LLAH を拡張し，特徴点の配置のみではなく，特徴点の属する連結成分の面積も利用して記述子を求める．また，異なる言語の文書から，言語の判別を行うことなく検索するために，撮影された文書画像から常に 2 通りの特徴点を抽出し，それぞれに対する検索を行って，その結果を統合するという処理を加える．以上のような提案手法について，リアルタイム検索のシステムを実装し，実験を行った結果，性質の異なるさまざまな言語における有効性が確認された．

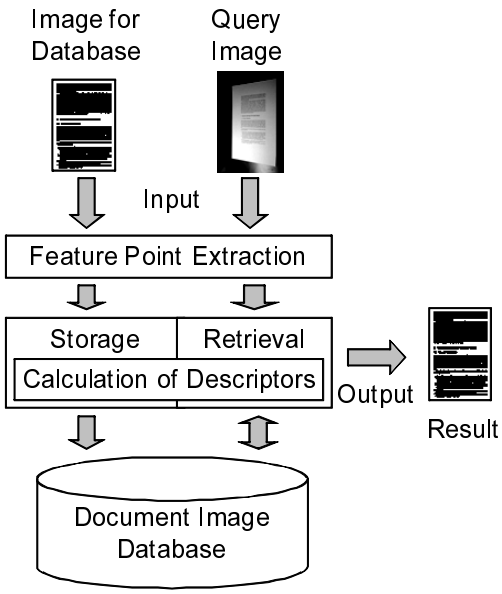


図 1　処理の概要

2. LLAH による英語文書画像検索

まず，オリジナルの LLAH と，それを用いた検索処理について説明する．

2.1 処理の概要

図 1 に処理の概要を示す．まず，特徴点抽出処理（Feature Point Extraction）で文書画像は特徴点の集合に変換される．次に，特徴点は登録処理（Storage）および検索処理（Retrieval）に入力される．これらの処理は記述子計算処理（Calculation of Descriptors）を共有している．登録処理では，各特徴点は独立に，その記述子に基づいて文書画像データベース（Document Image Database）に登録される．つまり，文書画像は特徴点を用いてインデキシングされる．検索処理では，検索質問の記述子を用いて文書画像データベースにアクセスし，投票処理で対応する文書画像を決定する．以下では各処理について説明する．

2.2 特徴点抽出処理

LLAH では特徴点の配置に基づいて文書画像のマッチングを行う．従って，特徴点抽出処理では，射影歪みやノイズが生じていたり，低解像度の場合でも同一の点を抽出する必要がある．そのため，単語領域の重心を特徴点として用いる．

特徴点抽出の手順を以下に示す．まず，入力画像（図 2(a)）を適応 2 値化し，2 値画像（図 2(b)）を得る．次に，2 値画像をガウシアンフィルタでぼかし，再度適応 2 値化を行うと，単語ごとに連結された画像（図 2(c)）が得られる．最後に連結成分の重心を計算して特徴点（図 2(d)）とする．

2.3 記述子計算

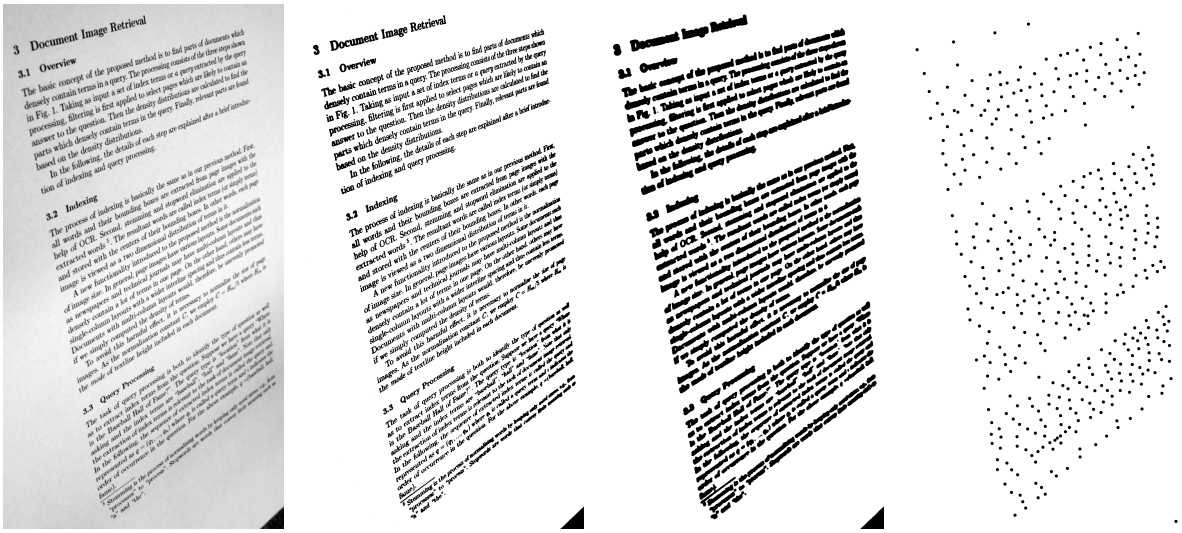
LLAH の記述子は，以下の特徴をもつ．

- 局所性

ロバスト性や隠れ耐性を実現するため，記述子は特徴点ごとに計算される．

- 幾何学的不変量の利用

デジタルカメラで撮影された画像において生じる射影歪みに



(a) 入力画像　(b) 2 値画像　(c) 連結成分　(d) 特徴点

図 2　特徴点抽出処理

対して不変となるため，幾何学的不変量が用いられる．具体的には，アフィン不変量が用いられる．アフィン不変量は同一平面上の 4 点 $ABCD$ から以下の式で定義される値である．

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

ここで，$P(A,B,C)$ は頂点 ABC で囲まれる三角形の面積である．

- 幾何学的不変量の組み合わせによる識別性の向上

記述子の識別性を増すため，複数の特徴点から計算される複数のアフィン不変量が用いられる．アフィン不変量は 4 点から計算されるため，5 個以上の点からは 2 個以上のアフィン不変量が得られる．具体的には，記述子は近傍する m 点から計算される ${}_mC_4$ 個のアフィン不変量 $(r_{(0)},\cdots,r_{({}_mC_4-1)})$ である．m 点から得られるすべての 4 点の組み合わせが用いられる．

- 複数の記述子の利用による安定性の向上

特徴点抽出の失敗や射影歪みによる近傍点の変化に対処するため，特徴点の近傍 $n(>m)$ 点から複数の記述子が計算される．具体的には，n 点から得られるすべての m 点の組み合わせが用いられ，${}_nC_m$ 個の記述子が計算される．

2.4　登録と検索

LLAH では，画像はハッシュ表を用いて検索される．まず，データベースの画像から得られた記述子を事前にハッシュ表に登録しておく．検索時には検索質問の画像から得られた記述子を用いてハッシュ表を調べ，同じ値の記述子をもつ画像を検索する．図 3 に示すように，画像の識別番号（Document ID）が記述子と共に登録されているため，識別番号を用いてデータベース画像に投票することで検索質問に対応する画像を検索することができる．また，点の識別番号（Point ID）も登録されているため，検索時に特徴点を対応付けることで図 4 に示すような特徴点の対応関係も得ることができる．

2.5　リアルタイム文書画像検索

LLAH による文書画像検索の応用として，Web カメラを用いたリアルタイム文書画像検索システム [5] が提案されている．

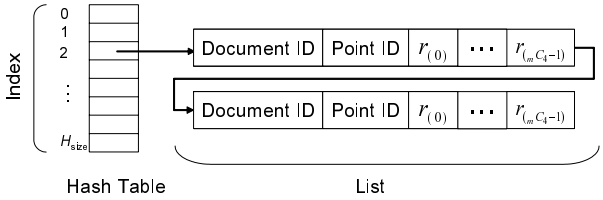


図 3　ハッシュ表の構成

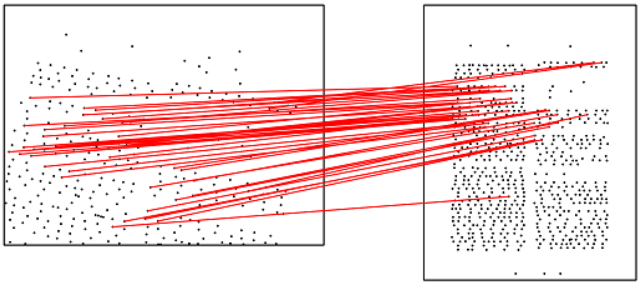


図 4　検索結果と特徴点の対応関係

これは，LLAH の高速性を利用して，Web カメラで撮影されたフレーム画像ごとに検索を行い，それを繰り返すことで検索結果をリアルタイムで提示するものである．クライアントサーバシステムによるパイプライン処理により，一般的な PC 程度の性能をもつ計算機において約 7fps のフレームレートが実現できることが示されている [5]．

3.　LLAH の多言語文書画像検索への拡張

3.1　特徴点抽出

従来手法で用いられている，ガウシアンフィルタで単語ごとの連結成分を作成し，それらの重心を特徴点とする手法は，単語ごとの分かち書きがなされる英語文書においては有効な特徴点抽出法である．従って，英語と同様にアルファベットが用いられ，分かち書きのなされるラテン言語（フランス語やスペイ

図 5　連結成分からの特徴点抽出

ン語など）においても同様に有効である．また，アルファベットは用いられないものの，分かち書きのなされる言語（アラビア語やヒンディー語など）においても，単語ごとの連結成分が得られるようガウシアンフィルタのパラメータを調整すれば安定な特徴点を得ることができる．ただし，英語ほど単語ごとの区切りが明確でない言語では，特徴点の安定性はやや低くなる．

分かち書きのされない言語（日本語や中国語）の場合は，単語という文書の構成単位を単純な画像処理で得ることが難しいため，異なる特徴点抽出法が必要となる．本稿では，文書画像における連結成分の重心を特徴点として用いる手法を提案する．連結成分は，図 5 に示されるように 1 つの文字を構成することもあれば，複数の連結成分が 1 つの文字を構成することもある．デジタルカメラで撮影された文書画像では，低い解像度やピントのずれ，あるいは印刷された段階でのインクのにじみなどにより，文字の細かいストロークが完全に分離した画像を得ることは困難である．そこで，ガウシアンフィルタを用いて入力画像をぼかし，細かいストロークは隣接する連結成分と結合させる．従って，この特徴点抽出法は，従来手法のガウシアンフィルタのマスクサイズを小さくしたものである．文字ごとの連結成分としないのは，密度の低い文字の場合，文字の中の連結成分の距離よりも隣接する文字との距離のほうが短い場合があり，文字だけを選択的に 1 つの連結成分にまとめることが困難なためである．

単語領域および連結成分の重心を抽出する際に用いられるガウシアンフィルタの適切なマスクサイズは，厳密には言語ごとに異なる．しかし，単語間および連結成分間の空白の存在により，適切なマスクサイズでなくても，ある程度安定した特徴点を抽出することが可能である．従って，言語間で共通したマスクサイズを設定することができる．文書画像の解像度が一定であれば，単語単位および連結成分単位の特徴点を抽出するための 2 通りのマスクサイズを用いることで，言語によらず特徴点を得ることができる．

ただし，紙面に対するカメラの距離が変化し，画像のスケーリングが変化すると，適切なマスクサイズも異なるものになる．提案手法では，カメラのフォーカスが合っていることを仮定し，フォーカスから紙面との距離を推定してマスクサイズを適応的に定める．

3.2　記述子の追加

日本語や中国の場合，多くの文字は 1 つの連結成分からなり，さらに文字はほぼ等間隔で配置される．従って，ほとんどの文書で特徴点の格子状の配置が支配的となり，特徴点の配置のみからでは識別性の高い記述子が得られにくい．そこで提案手法では，特徴点抽出の過程で得られる連結成分の面積から記述子を追加する．

連結成分の面積から記述子を計算する際，問題となるのはその安定性である．連結成分の面積は撮影条件によって変化するため，できるだけ変化しにくい値を用いる必要がある．例えば，カメラで撮影された文書画像においては，照明が弱ければ画像が暗くなり，すべての連結成分の面積が一様に大きくなる．そのため，連結成分の面積を正規化し，離散化するといった方法では安定な記述子を得ることが難しい．

そこで，提案手法では連結成分の面積の相対的な関係に着目する．具体的には，最も大きな面積をもつ連結成分は，ある程度の撮影条件の変化を受けても，やはり最大の面積をもつ．このように，面積の大小関係は変化しにくいため，連結成分の面積の順位を記述子とする．

記述子の例を図 6 に示す．この図では，近傍 7 点から選択された 6 点から記述子が計算されている．6 点には，中心点から見た角度の順に，時計回りで番号が与えられる．まず，従来手法と同様に，連結成分の重心から記述子が計算される．6 点から得られる 4 点の組み合わせの数は ${}_6C_4 = 15$ であるので，15 個のアフィン不変量が記述子として計算される．これは，図 6 において (Invariant 1, $\cdots$, Invariant ${}_6C_4 = 15$) と示されている．その右側に (3, 6, 1, 5, 2, 4) と示されているものが提案手法で追加された記述子である．この例では，6 点の中で最も大きな面積をもつ連結成分が 3 番の特徴点に対応している．そして，2 番目に大きな面積は 6 番の特徴点のものである．このように，面積の大きさの順番で特徴点の番号を並べると，(3, 6, 1, 5, 2, 4) となるため，この数字が記述子として追加される．

図 6 に示されるように，面積から計算される記述子は，従来の特徴点の配置から得られるものに追加される形で導入される．そのため，記述子の識別性が向上する一方で，安定性は低下することになる．そのため，特徴点の配置から得られるアフィン不変量については，従来手法より離散化レベル数を小さくし，変動を受けても一致しやすくする．これにより，安定性の低下を低減することができる．

3.3　検　　索

3.1 で述べたように，単語の分かち書きのされる言語とされない言語では，有効な特徴点が異なる．前者では単語領域の重心が，後者では連結成分の重心が安定な特徴点となる．検索の前に言語の判別が可能であれば，言語に応じて特徴点抽出処理を変えることができるが，判別のための処理を加える必要がある．また，判別に失敗した場合その後の処理がすべて失敗することになり，精度の面でも望ましくない．

提案手法では，あらかじめ登録時に単語領域の重心と連結成分の重心の 2 通りの特徴点を抽出し，それぞれを用いて 2 つのデータベースを作成する．検索時にも同様に 2 通りの特徴点を抽出し，それぞれ対応するデータベースに対して検索処理を行う．この結果，文書ごとに 2 つの得票数が得られることになる．これらを単純に足し合わせるだけでは，特徴点数の多い単語領域の重心から得られた得票数が支配的になるため，重み付けして合計し，文書ごとの得票数を求める．そして，合計された得

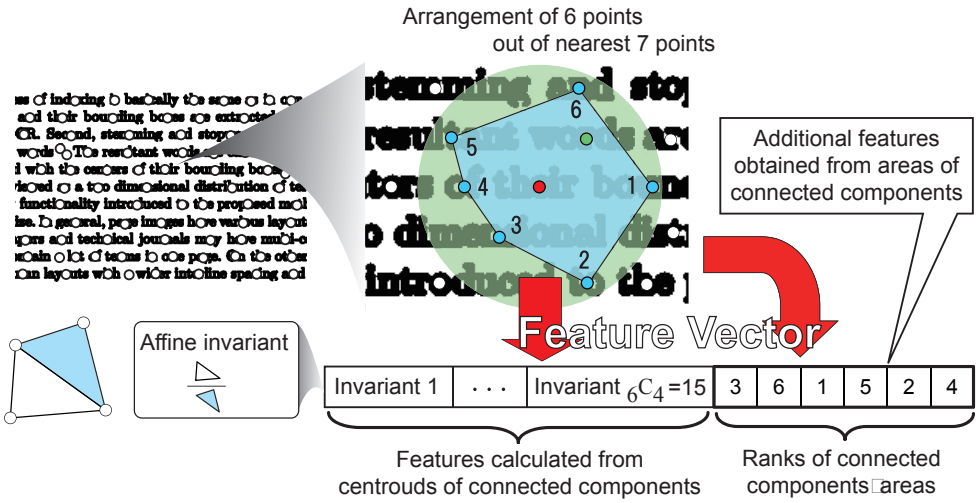


図 6　連結成分の面積の順位を追加した記述子

票数の最も大きいものを検索結果とする．

4. 実験結果

提案手法の有効性を確認するため，さまざまな言語の文書を用いたリアルタイム文書画像検索の実験を行った．実験に用いた文書の言語は，日本語，英語，アラビア語，中国語，フランス語，ヒンディー語，韓国語，ロシア語，スペイン語，タミル語の 10 言語である．これら 10 言語の文書画像をそれぞれ 100 枚，計 1,000 枚用意し，データベースを作成した．そして，言語ごとに 10 枚ずつ，計 100 枚の文書画像を抽出し，印刷して検索質問とした．実験の際には，市販の Web カメラを印刷文書から高さ約 6cm の場所に固定し，文書を次々と入れ替えて連続的に撮影した．撮影された画像の解像度は 1600×1200 である．撮影画像の例を図 7 に示す．各ページを 10 フレーム程度ずつ撮影し，合計 985 フレームについて検索を行った．印刷文書の写ったフレーム画像に対して，正しく対応する検索結果が得られたかどうかを調べた．あるページの写った約 10 フレームの画像において，正しい検索結果の得られたものが過半数であれば，そのページについては成功とみなした．なお，実験に用いた計算機は，クライアントが CPU 2.2GHz，メモリ 2GB のものであり，サーバが CPU 2.8GHz，メモリ 32GB のものである．LLAH のパラメータについては，$n = 7$，$m = 6$ とし，離散化レベル数は 7 とした．

言語ごとの検索精度を表 1 に示す．なお，平均処理時間は 359ms，平均フレームレートは 2.78fps であった．表 1 に示されるように，全体的に高い精度が得られた．特に，従来手法の英語文書における有効性から推測される通り，アルファベットを用い，分かち書きのされるラテン言語（英語，フランス語，スペイン語）では高い精度が得られた．

一方，中国語とロシア語の文書では比較的精度が低くなった．中国語文書での精度がやや低かったのは，図 8 に示されるような，文字数が少なく，写真が撮影範囲に入っていたケースがあったためと考えられる．写真からは安定した特徴点が得られないため，有効な特徴点の得られる文字が少なかったこともあり検索に失敗したと推測される．また，図 9 に示されるように，撮影範囲に英語と中国語が混在したケースでも検索に失敗した．英語と中国語では，それぞれ単語領域の重心と連結成分の重心が安定な特徴点であり，有効な特徴点の種類が異なる．連結成分の重心と単語領域の重心は，独立に登録および検索されるため，同一文書中に混在しているとうまく検索することが難しい．

ロシア語文書での精度がやや低かったのは，撮影範囲が狭かったためと考えられる．横幅の広いキリル文字で記述されるロシア語文書では，行あたりの単語数が少ない．そのため，撮影範囲が狭いと十分な数の特徴点が得られず，精度が低くなっ

表 1　検索精度

言語	精度
日本語	9/10(90%)
英語	9/10(90%)
アラビア語	10/10(100%)
中国語	7/10(70%)
フランス語	10/10(100%)
ヒンディー語	9/10(90%)
韓国語	10/10(100%)
ロシア語	7/10(70%)
スペイン語	10/10(100%)
タミル語	10/10(100%)
合計	91/100(91%)

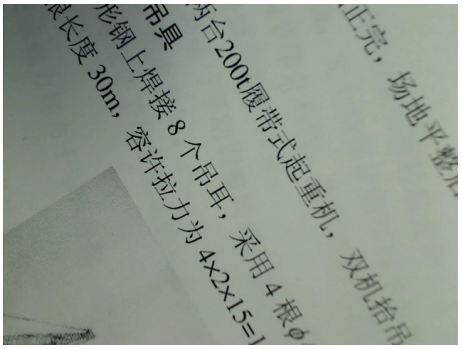
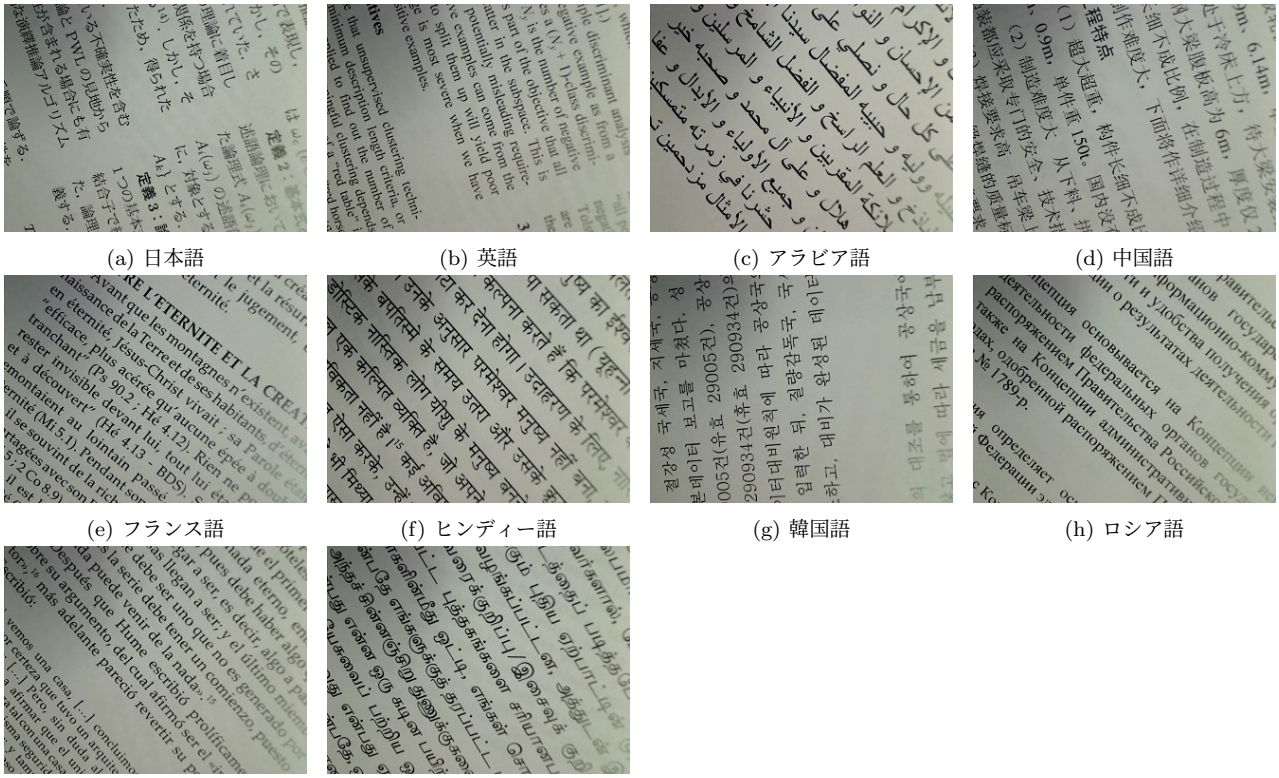


図 8　失敗例 1：文字が少なかった場合



(a) 日本語　(b) 英語　(c) アラビア語　(d) 中国語

(e) フランス語　(f) ヒンディー語　(g) 韓国語　(h) ロシア語

(i) スペイン語　(j) タミル語

図 7　実験に用いた文書画像の例

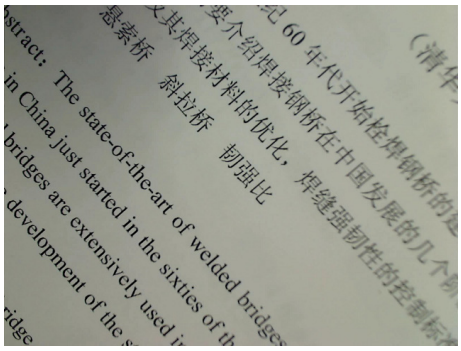


図 9　失敗例 2：英語と中国語が混在した場合

たと考えられる．

以上のように，いくつかの言語ではやや低い精度が得られたものの，性質の異なるさまざまな言語すべてにおいて 70%以上の精度が得られたことは提案手法の有効性を示しているといえる．

5. ま と め

本稿では，我々のすでに提案している画像検索手法 LLAH について，さまざまな言語で書かれた文書に適用するための改良手法を提案した．提案手法には，分かち書きのされない言語の文書からの特徴点抽出法と，連結成分の面積に基づく記述子の追加が含まれる．提案手法を用いてリアルタイム文書画像検索システムを実装し，実験を行った結果，性質の異なるさまざまな言語での有効性が確認された．今後の課題としては，比較的精度が低かったいくつかの言語のためのさらなる改良と，今回の実験で用いられなかった他の言語の文書における有効性の確認が挙げられる．

文　献

[1] “Qr code.com”, from http://www.denso-wave.com/qrcode/index-e.html.

[2] K. Kise, K. Noguchi and M. Iwamura: “Memory efficient recognition of specific objects with local features”, Proc. of the 19th International Conference of Pattern Recognition (ICPR2008), WeAT3.1 (2008).

[3] T. Nakai, K. Kise and M. Iwamura: “Camera based document image retrieval with more time and memory efficient llah”, Proceedings of Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2007), pp. 21–28 (2007).

[4] X. Liu and D. Doermann: “Mobile retriever - finding document with a snapshot”, Proceedings of Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2007), pp. 29–34 (2007).

[5] T. Nakai, K. Kise and M. Iwamura: “Real-time document image retrieval with more time and memory efficient llah”, Proceedings of Second International Workshop on Camera-Based Document Analysis and Recognition (CBDAR2007), pp. 168–169 (2007).

[6] F. Chang: “Retrieving information from document images: Problems and solutions”, International Journal on Document Analysis and Recognition, Special Issues on Document Analysis for Office Systems, **4**, pp. 46–55 (2000).