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.

LLAHをノイズおよび距離変化に頑健にする手法の提案

文字とノイズの高速な推定と画像の正規化

宮本　康平†　　岩村　雅一†　　黄瀬　浩一†

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

あらまし　スマートフォンの普及により，カメラは誰しもが持つデバイスとなった．同時に，そのカメラで撮影された画像上にリアルタイムに情報を付加するサービスが普及し，それを実現する技術の一つにカメラベースの文書画像検索がある．これは，カメラで撮影した文書画像と同じ文書を，予め登録しておいた文書画像の中から検索するというものである．Locally Likely Arrangement Hashing (LLAH) は，主に単語領域の重心を特徴点とし，そこから計算された幾何学的不変量を用いることで，高速な検索を実現した．しかし，LLAH の特徴点抽出法はノイズに弱いという欠点があった．LLAH では，同じ単語内の文字だけが連結するように文字をぼかし，単語領域を得ていた．ここで，適切なぼかし具合は，文字の大きさによって変化するため，文字の面積からぼかし具合の推定を行っていた．しかし，ノイズがその推定に影響を与えるという問題があった．本稿では，画像から文字の成分を取り出し，その面積から文字とノイズを EM アルゴリズムによって動的に分離し，ノイズに頑健なぼかし具合の推定を行った．また，一定のぼかし具合であっても，画像を拡大縮小することによって，ぼかし具合を変化させたときと，同じ効果を得ることができる．これにより，推定されたぼかし具合から，画像の大きさを正規化することで，特徴点抽出の精度と速度が向上させた．
キーワード　Locally Likely Arrangement Hashing, EM アルゴリズム

1. はじめに

スマートフォンの普及によりカメラの付いた小型デバイスは世界中の人が持つようになった．それに伴い，その内蔵型のカメラを利用したサービスが数々提案されている．そういったサービスの基盤となる技術の一つにカメラベースの文書画像検索がある．カメラベースの文書画像検索とは，カメラによって撮影された文書画像と同じ文書画像を，予め登録しておいた文書画像の中から検索するというものである．Locally Likely Arrangement Hashing (LLAH) [1]〜[3] は，このようなカメラベースの文書画像検索手法の一つであり，高速に動作し，かつカメラと撮影された文書との撮影角度の変化や距離の変化に頑健という特徴を持つ．この LLAH のを使用したサービスに，株式会社ステークホルダーコムの AR-Sentence がある (注1)．これは，スマートフォンのアプリ内のカメラを文章にかざすと，撮影された文書がどの文書で，どの場所が撮影されているかを検索し，関連するデジタルコンテンツを表示させるというクラウドサービスである．

LLAH がこのような特徴を持つのは，文書の特徴点として，撮影条件の変化に対しても比較的安定して取り出せる単語領域の重心を用い，その重心を元に相似不変，アフィン不変または射影不変な特徴量を計算するためである．また，検索ではこの特徴量に基づいて投票を行い，得票数の多かったものを検索結果として選ぶ．そのため，登録した画像数に対して劣線形時間での検索が可能となっている．その結果，LLAH は特徴点が正しく抽出できれば，高速かつ距離変化や角度変化に対しても頑健に動作する．

しかし，LLAH は特徴点の抽出部分はノイズに弱いため改良の余地がある．LLAH は，主に単語領域を特徴点とするため，登録時とカメラで撮影された時で同じ単語領域を取り出す必要がある．単語領域を安定かつ高速に取り出すために，二値化された文字をぼかして，再度二値化することで各文字を連結させ，単語領域を得ている．したがって，ぼかし具合を適切に決めることができなければ，正しい単語領域を取り出せない．カメラで撮影された画像は，その文書との距離にに応じて見かけの文字の大きさが変化するため，従来はそれ合わせてぼかし具合を調整したり [1]，あるいは一定の距離での使用を仮定し，常に同じぼかしを適応していた [2], [3]．しかし，見かけの文字の大きさを推定する際，ノイズが現れると文字の大きさを正しく捉えることができず，常に同じぼかし具合にすると，カメラと文書との距離変化に対応できない，という問題があった．

また，ぼかし具合そのものを変えるのではなく，画像を拡大縮小することによって，同様の効果を得ることができる．カメラと文書との距離が近く，画像内において文字が大きく写っている場合，単語内の文字を連結させるためには，大きくぼかす必要がある．しかし，単語領域を抽出するという目的において

(注1)：https://ar-shcom.jp/

は，文字が大きい場合は画像を縮小すると，走査するピクセルが減少し，その後の処理が高速に行える．また，カメラと文書との距離が遠く，画像内において文字が小さく写っている場合，ぼかし具合の僅かな差が，単語領域の抽出に大きな影響を及ぼす．この場合は，一度画像を拡大してから適切なカーネルサイズのガウシアンフィルタを掛けることで，安定性を高めることができる．

本稿では，文字の面積とノイズの面積には差があり，またそれぞれの分布が異なることを仮定し，その分布を EM アルゴリズムによって推定することで，文字の面積とノイズの面積を自動で高速に分離し，撮影条件によってノイズが多く現れた場合でも，撮影距離の変化に左右されず単語領域を正しく取り出す手法を提案する．さらに，従来の特徴点抽出のステップを改良し，推定された文字の大きさを元に画像の正規化をすることで，精度と速度の向上を実現した．

2. LLAH

本設では，LLAH の概略について説明する．LLAH は特徴量抽出と登録，および検索の 3 つの処理で構成されている．特徴量抽出では，画像から特徴点を取り出し，その特徴点から特徴量を計算する．この処理は，登録時および検索時で共通である．登録処理は登録したい文書から得られた特徴量および文書 ID を文書画像データベースに登録する．検索処理では，検索対象の文書から得られた特徴量と同じ特徴量を持つ文書を，文書画像データベースから検索し，特徴量に紐づけされた文書 ID に投票し，最も得票数が多かった文書を対応する文書画像とする．

2.1 特徴点抽出

LLAH では特徴点の位置関係によって文書画像を検索する．したがって，撮影された画像に射影歪やノイズが付加されたり，あるいは解像度が異なった場合であっても，同じ特徴点が抽出できる必要がある．特に，英語やフランス語などの分かち書きする言語においては単語の重心が，日本語や中国語などの分かち書きしない言語においては，文字の連結成分から特徴点を抽出することが望ましい [4]．

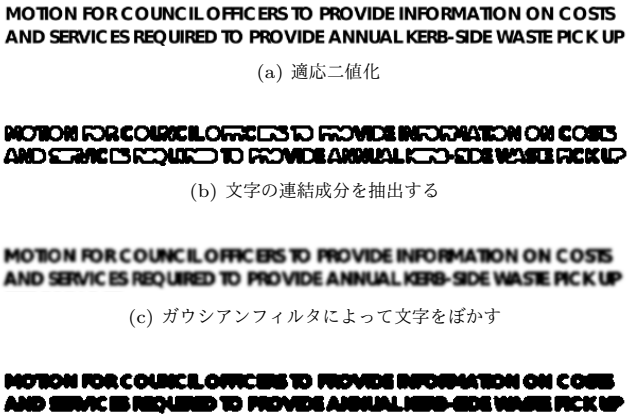
LLAH では特徴点は次のステップで計算される．また，それぞれのステップで得られる画像例を図 1 に示す．

STEP1　適応二値化を行い，ノイズ除去をする．（図 1(a)）適応二値化は，注目するピクセルの周辺のピクセルの値の平均から，定数を引いたものを閾値とする．

STEP2　文字やノイズを含む，すべての連結成分を抽出し，その面積の平方根の最頻値を求める．（図 1(b)）

STEP3　文字の面積の平方根の最頻値をもとに，ガウシアンフィルタのカーネルサイズを推定し，ガウシアンフィルタを適応する．これにより，同じ単語内の文字同士が連結する．（図 1(c)）

STEP4　もう一度適応二値化を行い，単語領域を得る．（図 1(d)）



(a) 適応二値化

(b) 文字の連結成分を抽出する

(c) ガウシアンフィルタによって文字をぼかす

(d) 適応二値化により連結された単語を得る

図 1: 単語領域を得るプロセス

STEP5　単語領域からその重心を計算し，特徴点とする．

撮影する距離が固定であると仮定して，固定値のカーネルサイズを用いる場合は，STEP1 では予め決めた固定値のカーネルサイズを用い，STEP2 の処理は省く．

最も重要な部分は，文字のぼかし具合を決めるガウシアンフィルタのカーネルサイズであり，これは文字の面積の平方根の最頻値によって求められる．カメラと文書との距離が変化すると，それにともなって画像内の文字の面積が変化するため，適切なぼかし具合は変化する．そこで，文字の面積の変化を捉えることにより，適切なぼかし具合，つまりガウシアンフィルタのカーネルサイズを推定する．しかし，STEP1 の適応二値化のパラメータによっては図 2 のように多くのノイズが現れる場合がある．こういった小さなノイズは，画像中の文書領域よりも，文書の外の背景部分に数多く現れる．カーネルサイズを推定する際は，このノイズが最頻値として選択される可能性が高まるため，予め決めておいた閾値によってノイズと文字の連結成分を分離していた．

2.2 特徴量計算

文書を判別するためには，特徴点からその文書らしさを表す特徴量が必要となる．特徴点は撮影距離や角度の変化により，その位置関係が変化するため，そういった状況であっても，同じ特徴量が得られる必要がある．そうでなければ，同じ文書であるにも関わらず，登録された文書と，撮影された文書から異なる特徴量が計算され，検索に失敗してしまう．特徴点はその位置を表す座標しか持っていないため，一点だけでは文書を表すことが出来ない．そこで，常に同じ特徴量を得るために，特徴量は各特徴点の位置関係から計算する．文書の一部が隠れるなどして，文書の一部だけが撮影された場合でも，正しく検索するために，特徴量は局所的な特徴点の配置から計算される必要がある．さらに，カメラで撮影された画像には射影変換が掛かるため，そういった変換に頑健な幾何学的不変量を用いる．そこで，局所領域においては射影変換なアフィン変換に近似できること，アフィン変換は射影変換よりも特徴点の位置変動に頑健であることから，特徴点から計算されるアフィン不変量を特徴量とする．

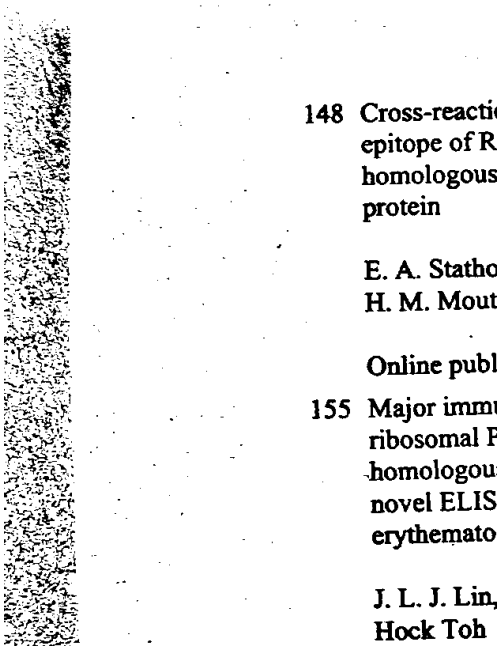


図 2: 撮影された画像で，適応二値化後多くのノイズが現れた部分を切り出したもの．紙面ではなく，背景の机の部分に集中してノイズが現れている．

具体的には，同一平面上の 4 点 A, B, C, D から以下の式で計算されるアフィン不変量を用いる．

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

ここで，$P(A,B,C)$ は 3 点 A,B,C を頂点とする三角形の面積を表す．そして，この特徴量はある点に最も近い点から計算される．ここで，この特徴量を計算するには最低でも注目する点と，その近傍の 3 点の合計 4 点あれば十分である．しかし，射影変換が掛かった画像においては，点同士の距離関係が変化し，近傍点が異なる場合が存在する．これを解決するために，近傍 n 点のうち $m(\leqq n)$ 点を選び，その選び方のすべての組み合わせを調べることにする．すると，n 点のうち m 点が共通であれば同じ m 点を必ず得ることができるため，特徴量が安定する．ここで，$m=4$ とすると最も単純な計算で済むが，これだけであると異なる文書で同じ特徴量が計算される場合がある．そこで，さらに識別性を高めるために $m \geqq 4$ とし，m 点のなかから 4 点を選ぶすべての組み合わせについて調べ，これを並べたものを特徴量とする．この並べ方は時計回りを考慮して一意に定める．したがって，特徴量の次元数は ${}_mC_4$ となる．

また，識別性の向上のために，単語領域の面積比を特徴量として追加することもできる [2], [3]．これは，選ばれた m 点の単語領域を時計回りに並び替え，隣り合う単語領域の面積比を計算し，その比の大きさの順番に並び替え，先程の特徴量の列に加える．したがって，最終的な特徴量の次元数は ${}_mC_4+m$ となる．この特徴量を実数値のまま扱うと，検索時にハッシュを使って高速化出来ないため，離散化する．

2.3 登録処理

データベースに登録する際，ハッシュテーブルのインデックスは以下に示すハッシュ関数で計算される．

$$\left(\sum_{i=0}^{{}_mC_4+m} r_i d^i\right) = \mathrm{Q}\mathrm{H}_{\mathrm{size}} + \mathrm{H}_{\mathrm{index}} \tag{2}$$

ここで，r_i は i 番目の特徴量を意味し，d^i は離散化レベル数である．$\mathrm{H}_{\mathrm{size}}$ はハッシュテーブルの大きさであり，このハッシュ関数により，ハッシュテーブルのインデックス $\mathrm{H}_{\mathrm{index}}$ および，商 Q は特徴量に対して一意に定まる．そして，文書の ID，特徴点の ID，Q の値を一組とし，ハッシュテーブルの $\mathrm{H}_{\mathrm{index}}$ 番目にリストとして繋ぐ．Q を同時に保存するのは，同じ $\mathrm{H}_{\mathrm{index}}$ を持つ特徴量であっても特徴量が同じだとは限らないため，同じ特徴量を持つかどうかを Q の一致で確認するためである．

2.4 効率化

上記のように，すべての文書のすべての特徴点を登録すると，その数は膨大になり，メモリ効率が悪くなる．また，同じハッシュテーブルのインデックスに大量のリストが繋がれば，そのインデックスは文書を判別する上で有効な指標とはならないばかりか，リスト構造は登録された要素数に比例する計算量が必要なため，文書検索に要する計算時間が大きくなる．そこで，一定数以上の特徴点が同一のインデックスに登録された際は，そのインデックスのリストを開放し，メモリ効率と計算効率を上げている．

さらに，文書から得られるすべての特徴点を使用するのではなく，特徴点数が多い場合は，特徴点のサンプリングを行う [2]．周囲の単語領域の中で，最も面積が小さいものを注目点としてその近傍点から，特徴量を計算する．これは，面積の小さい単語領域は，近傍点との距離が小さくなるため，それだけ射影歪みの影響を受けにくくなり，安定した特徴量が計算できるためである．しかし，このようにするとサンプリングされる特徴点が少なすぎるため，サンプリングされた特徴点の近傍の k 個の点からもサンプリングを行い，1 文書あたり 200 点が取り出されるようにしている．

また，アフィン不変量を求める上で必要となる 2 つの三角形は，m 点の中から選ばれた点によって定まる．この時，同じ三角形を用いて計算された 2 つの特徴量は冗長性を含んでいる．そこで，同じ三角形を用いないようにすることで，識別性を高めることができる．これにより，特徴量の冗長性が解消され，n および m の数を増やしたとしても，識別性および安定性を保つことができる．

特徴量の離散化の際，離散化誤差が大きくなるように離散化されると，カメラで撮影された特徴点から特徴量を計算する際に，異なる特徴量が計算される可能性が高まる．そこで，離散化誤差が小さい特徴量を優先してデータベースに登録し，メモリ量の削減と検索精度の低下の抑制を図る [3]．具体的には，それぞれの離散化領域に対し正規分布を適応し，その中心ほど高い点を，端に行くほど低い点を割り当てることで，特徴量に点数をつける．そして，最も合計点数の高かった特徴量から順番に N 個だけデータベースに登録する．このままでは，文書領域内で偏って特徴点が計算される可能性が高まるため，登録時には文書画像を複数の領域に分割し，各領域に含まれる特徴点

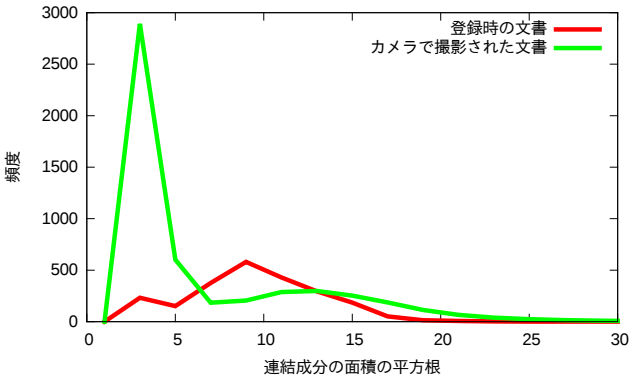


図 3: 1000 枚の文書画像のすべての連結成分の面積の平方根のヒストグラムの平均.

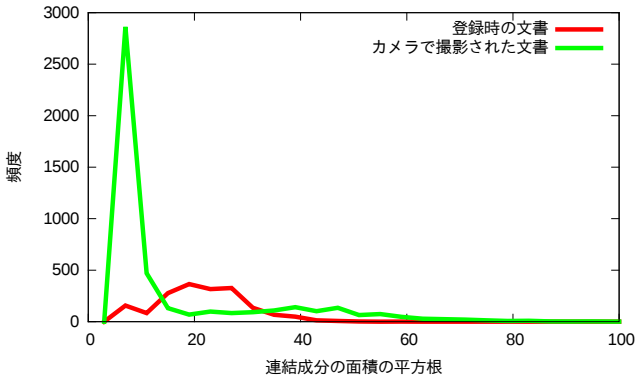


図 4: 文書画像の連結成分の面積の平方根のヒストグラム.

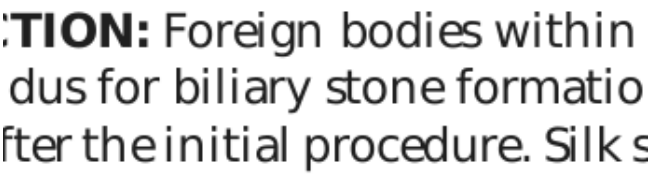


図 5: 図 4 のヒストグラムを得た文書の一部.

数に応じて登録する特徴量数を決定する.

2.5 検　　索

LLAH では，投票テーブルを用いて登録した文書へ投票処理によって文書を検索している．まず，撮影された画像から先程のステップと同様に特徴点を抽出し，特徴量を計算する．そして，特徴量のハッシュ値を計算し，ハッシュテーブルから同じハッシュ値を持つインデックスに紐づけされたリストを辿り，特徴量が一致するかを調べる．もし特徴量が，一致していればその文書 ID の投票テーブルに投票する．最後に，最大の得票数を得た文書を検索結果とする.

登録時に変動に強い特徴量を登録したとしても，検索時に異なる特徴量が計算されることは起こりうる．そこで，検索時においては，離散化の閾値に近い値を取った場合は，その閾値前後の 2 つの離散値を与えることもできる [3]．すると，少なくとも一方が登録時と同じ特徴量になる可能性が高まる．しかし，投票数が増え誤投票の可能性も高まり，処理時間が増加する.

3. 提 案 手 法

3.1 文字サイズの推定

従来は画像に現れたノイズについては，予め閾値を決めその閾値以下の面積を持つ連結成分を，ノイズとして除去していた．また，文字の連結成分の面積のヒストグラムを作成し，その最頻値からカーネルサイズの推定をしていた．このような仮定が成立するのは，画像に現れる連結成分の面積の平方根のヒストグラムを作成すると，図 3 に示すように，ノイズ部分と文字部分の二つに山を持つからである．またこの図 3 からわかるとおり，登録時の文書においても，カメラによって撮影された文書においても，連結成分の面積の平方根の値が 3 付近にピークを持っている．これは，i や j，ピリオドやコンマといった文字が持つ小さな点によるものである．カメラで撮影された文書については，おおよそ文字を構成する点と同じ大きさのノイズが大量に現れているため，大きなピークを持っている．したがって，カメラで撮影した文書において，連結成分の面積の最頻値はノイズによって決まることが多い．このままでは，適切なぼかし具合を推定することが出来ないため，従来は，閾値を予め決めることにより，あるいは，文書との距離が予め決まっていると仮定することで，事前に決めたカーネルサイズを使うことにより，この問題の発生を防いでいた．しかし，設定した閾値やパラメータが常に適切とは限らず，文字の面積がその閾値以下の部分になったり，あるいは文字の面積を表す部分にピークを持たなかった場合は，適切なカーネルサイズを選択できず，正しい単語領域を得ることができない．また，このヒストグラムは必ずしもノイズと文字領域の双峰を持つわけではなく，図 4 のように，複数の峰をもち，最頻値が安定しない図 5 のような文書も存在する．この文書は，文字の間隔が一定ではなく，近接する文字がるため，適応二値化時に文字が連結し，文字の大きさが正しく推定できない.

そこで，提案手法では，予め定めた閾値によってノイズを判別する代わりに，文字だけでなく，ノイズの大きさも推定することで，ノイズの閾値を適応的に定める．そのために，ノイズと文字の面積ははそれぞれ異なる確率分布によって生成されたと仮定し，その確率分布を EM アルゴリズムによって同時に推定する.

EM アルゴリズムとは，複数の分布に従った観測値を，E ステップ M ステップの 2 つのステップにより，推定したいパラメータを最適化するアルゴリズムである [5]．E ステップにより，現在推定されたパラメータによる分布のもとで，尤度関数の条件付き確率の期待値を計算し，M ステップにより，その期待値を最大化するパラメータを求める．このように，期待値とパラメータが収束するまで交互に計算することで，観測値が従う分布を推定することができる.

本稿では，ヒストグラムの形からノイズと文字がそれぞれ異なる正規分布に従うと仮定した．EM アルゴリズムは初期値によっては収束しない場合があるため，ノイズの初期平均値は 0 を，文字の連結成分の面積の平均値は全ての連結成分の面積の平均値とした．これは，文字の大きさが事前には不明であるた

め，暫定的な推定を行うためである．また，画像が含まれる場合など，極端な大きさの面積を持つ連結成分があった場合，EM アルゴリズムは収束しないため，全ての連結成分の平均値と分散から求めた，ある閾値以上の面積を持つものは除外した．この閾値は，平均値に分散の定数倍を足したものである．それでも EM アルゴリズムが収束しない場合があるため，その場合は，連結成分の面積の上下 10 パーセントを削除した上で，残った連結成分の平均値を推定された文字の面積とした．

3.2 画像の正規化

従来は，文字の面積やそれに準ずるものから，ガウシアンフィルタのカーネルサイズを推定し，適切なぼかし具合を推定していた．本稿では，同じぼかし具合に対し画像を拡大縮小することで同じ効果を得る．これにより，推定された文字の面積が大きい場合，画像を縮小することにより，走査すべきピクセル数が減り，高速化が望める．逆に，推定された文字の面積が小さい場合，カーネルサイズの僅かな違いが特徴点の抽出に大きな影響を与えることがあったが，画像を適切な大きさまで拡大することにより，僅かな違いに左右されることなく，安定的にぼかすことができる．また，画像の拡大縮小をすることで，ガウシアンフィルタのカーネルサイズは奇数の値しか取れないという制約を大幅に緩和し，擬似的に実数値のカーネルサイズであっても適応することが可能となる．

3.3 提案する特徴点抽出法

以上の処理を含めた提案する特徴点抽出のステップを次に示す．

STEP1 従来手法と同様に適応二値化を行い，ノイズ除去をする．

STEP2 すべての連結成分を抽出し，その面積に EM アルゴリズムを適応することで，文字の面積とノイズの面積の推定値を得る．

STEP3 推定された文字の面積を元に，ある一定の文字サイズになるように画像を拡大縮小し，正規化する．

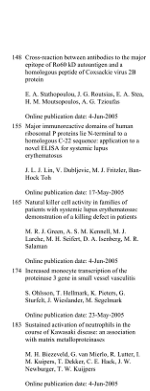
STEP4 予め決めておいたカーネルサイズのガウシアンフィルタを適応させ，文字をぼかす．その結果，同じ単語内の文字が連結する．

STEP5 適応二値化を行い，単語の連結成分を取り出す．

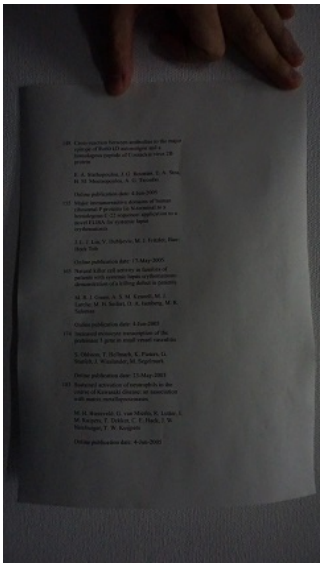
STEP6 単語の連結成分を元に連結した単語領域の重心を計算し，特徴点とする．

4. 実　　験

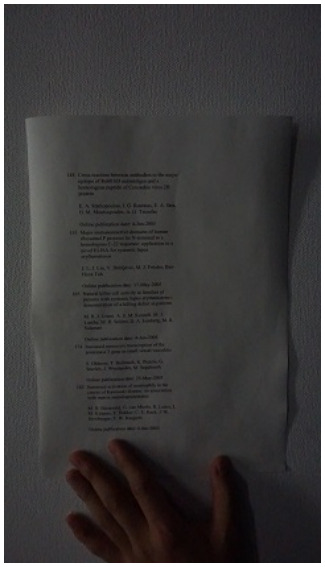
従来手法と提案手法を比較するため，実験を行った．今回は，登録用として 1700 x 2200 の画像を 1000 枚，検索用の画像として室内の壁に手で貼り付けた文書を一眼レフカメラで撮影した．画像の大きさは 2400x4240 であり，カメラと文書との距離を三段階に分け，それぞれ 100 枚ずつ，計 300 枚を使用した．実際に使用した画像は図 6 に示したようなものである．



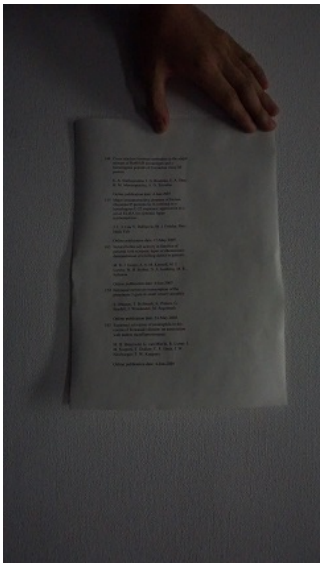
(a) 登録時に使用した文書画像



(b) 近距離から撮影された文書画像.



(c) 中距離から撮影された文書画像.



(d) 遠距離から撮影された文書画像.

図 6: 実験に使用した画像の例

LLAH のパラメータは $n = 8, m = 7$ で，ハッシュサイズは $2^{30} - 1$ である．実験に使用した計算機は，CPU が Opteron6238 2.80GHz，メモリが 512GB である．

従来手法として，カーネルサイズを予め固定する手法 [2], [3]，文字の連結成分の平方根の最頻値から求める手法 [**?**]，提案手法として，カーネルサイズを EM アルゴリズムによって求める手法，および，EM アルゴリズムを用いてカーネルサイズを推定し，それを元に画像を正規化する手法の 4 種類を比較した．

画像の登録は，すべて同じ固定値のカーネルサイズでぼかすことで得た特徴量で行い，すべての手法で共通とした．また，すべての手法のパラメータは中距離から撮影された文書画像に合わせて，最適化を行った．その結果を表**??**としてまとめる．検索正答率は，100 枚の画像のうち何枚が正しく検索できたかを表し，検索時間は画像が与えられてから，検索結果が返って

表 1: 検索正答率 (%)

カーネルサイズの推定法	近距離	中距離	遠距離
固定値	41	66	65
最頻値	36	42	37
EM アルゴリズムのみ	43	54	36
EM アルゴリズム+画像の正規化	66	66	56

表 2: 検索時間 [sec]

カーネルサイズの推定法	近距離	中距離	遠距離
固定値	0.338301	0.348164	0.325396
最頻値	0.432979	0.458000	0.458706
EM アルゴリズムのみ	0.524494	0.634122	0.805311
EM アルゴリズム+画像の正規化	0.341380	0.427528	0.591711

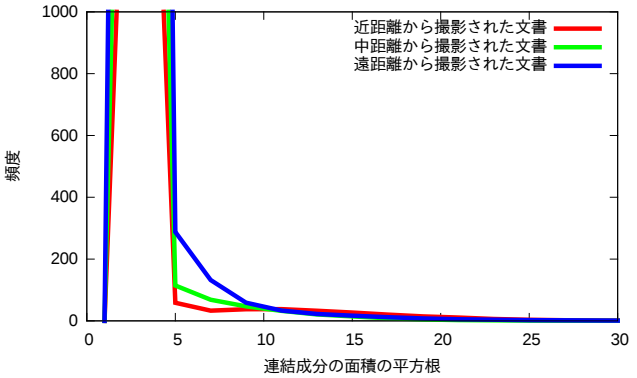


図 7: 使用した 100 枚の撮影画像の文字の連結成分の平方根のヒストグラムの平均.

くるまでの一枚あたりの平均時間を表している.

この結果より，検索正答率については，遠距離から撮影されたもの以外について，EM アルゴリズムを用い，画像を正規化した手法が，最も正答率が高かった．遠距離から撮影された画像の正答率が低かった原因として，背景がより広く写されたため，ノイズが背景に多く現れ，文字サイズの推定の妨げになったと考えられる．また，最頻値を用いた手法の精度が低かった原因は，実験に使用した画像はノイズが大量に現れ，文字の連結成分の面積の平方根のヒストグラムを作ると，図 7 のように，ノイズのみにピークを持つため，最頻値がノイズとなり，適切なぼかし具合の推定に失敗したためである．

検索時間に関しては，固定値のカーネルサイズを用いたものが最も速く，EM アルゴリズムのみを用いたものが最も遅かった．固定値のカーネルサイズが高速だった理由は，文字の面積の推定の処理が省かれているためである．EM アルゴリズムのみを用いたものが低速だった理由は，近距離の画像では正しく文字面積を推定できた結果，大きめのカーネルサイズが適応されたためであり，遠距離の画像で特に低速だったのは，文字が小さいため，カーネルサイズの僅かな違いによって，文字の連結に大きな影響を受けたためと考えられる．さらに，EM アルゴリズムを用い，画像の正規化も行った手法では近距離で撮影されたものは固定値のカーネルサイズを使用した手法と同程度の速度で検索が行えている．これは，大きなカーネルサイズが推定され，その結果画像が縮小されたため，その後の処理が高速化したからである．

また，遠距離の画像に EM アルゴリズムのみを適応した場合は精度が低く，画像の正規化も適応した場合には精度が高くなった理由は，画像の正規化を行うと，小さな文字サイズが推定されているため，画像が拡大される，するとぼかし具合の誤差が許容される範囲に収まったため，適切にぼかすことができたからだと考えられる．

5. ま と め

本稿では，LLAH の弱点であった，特徴点抽出がノイズに弱いという点を，文字とノイズを，その面積に確率分布を仮定し，EM アルゴリズムによって同時に推定することで，適応的に分離することで補った．さらに，ぼかし具合を変化させるのではなく，事前に定めたぼかし具合に合うように，画像を拡大縮小させることで，より柔軟にぼかし具合を変化させることができるようにした．これにより，文字が小さい場合はより精度を高く，文字が大きい場合はより高速に，特徴点を抽出することができるようになった．

しかし，撮影時にピントが合わないなどして，文字がぼやけた場合，適応二値化により，文字が太くなり，その面積が大きく推定される場合がある．この場合，文字が太くなったことにより，文字間の距離が短くなり，小さなぼかし具合を選択すべきであるが，文字が大きく推定された結果，必要以上に大きくぼかしてしまうという問題が発生する．したがって，ぼかし具合も同時に推定することが今後の課題として挙げられる．

謝辞　本研究は，JST マッチングプランナープログラム「企業ニーズ解決試験」MP28116808744, JSPS 基盤研究 (A)25240028 ならびに JST CREST JPMJCR16E1 の助成を受けたものである．

文　献

[1] 中居友弘，黄瀬浩一，岩村雅一，“特徴点の局所的配置に基づくデジタルカメラを用いた高速文書画像検索，” 電子情報通信学会論文誌 D，vol.J89-D，no.9，pp.2045–2054，Sept. 2006.

[2] 竹田一貴，黄瀬浩一，岩村雅一，“1,000 万ページのデータベースを対象とした実時間文書画像検索のためのメモリ削減と安定性向上，” 電子情報通信学会技術研究報告，第 110 巻，March 2011.

[3] 竹田一貴，黄瀬浩一，岩村雅一，“1 億ページのデータベースを対象とした大規模文書画像検索，” 電子情報通信学会技術研究報告，第 112 巻，pp.131–136，Feb. 2013.

[4] T. Nakai, K. Kise, and M. Iwamura, “Real-time retrieval for images of documents in various languages using a web camera,” Proceedings of the 10th International Conference on Document Analysis and Recognition (ICDAR2009), pp.146–150, July 2009.

[5] A.P. Dempster, N.M. Laird, and D.B. Rubin, “Maximum likelihood from incomplete data via the em algorithm,” Journal of the royal statistical society. Series B (methodological), pp.1–38, 1977.