

特徴点の局所的配置に基づく位置合わせを用いた 文書からの書き込み抽出法

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 紙媒体の文書における書き込みには重要な情報が含まれているため、書き込みを抽出し、解析することでその情報を活用することが望まれている。本稿では、紙文書からの書き込み抽出の手法を提案する。これは、スキャナによって得られた書き込みの含まれる文書画像と元となった文書画像を比較することで書き込みを取り出すものである。提案手法の特徴は、高精度な位置合わせと柔軟な差分処理である。特徴点の局所的配置に基づく検索法によって得られる特徴点の対応関係に基づいて位置合わせを行い、撮像条件の変化によって生じる歪みに対してロバストな差分処理によってノイズの少ない書き込み画像を取得する。実験により、提案手法によってカラー文書の76%、モノクロ文書の94%から書き込みの抽出が可能であることが示された。

キーワード 文書解析, 文書画像, スキャナ, 書き込み

A Method of Annotation Extraction from Documents Using Alignment Based on Local Arrangements of Feature Points

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 Annotations on paper documents include important information, which can be exploited by their extraction and analysis. In this report, we propose a method of annotation extraction from paper documents. In the proposed method, annotations are extracted by comparing a document image of an annotated document captured using a scanner and its original document image. The proposed method is characterized by precise matching and flexible subtraction of images. Experimental results have shown that annotations can be extracted using the proposed method with 76% accuracy in the cases of color documents and 94% accuracy in the cases of monochrome documents.

Key words Document analysis, Document image, Scanner, Annotation

1. はじめに

紙媒体の文書（紙文書）は高い可読性および可搬性を持つため、多くの情報が紙文書を介して提供されている。また、紙文書に対して注目した部分に印をつけたり、メモを書き込むなどの書き込み行為は一般的である。従って、紙文書における書き込みにはユーザの興味や知識などの情報が豊富に蓄積されているといえ、書き込みを抽出・解析することで貴重な情報を得ることができると考えられる。

書き込み抽出の手法は、抽出後の書き込みの利用法、具体的

には抽出した書き込みを機械で認識するかどうかによって分類できる。利用法として、文書の校正や手書きフォームからの情報抽出を考えるならば、認識は必須となる。この場合、認識に耐え得るほど高い精度で書き込みを抽出する必要がある。多くの従来法 [1] ~ [4] はこのカテゴリに属するものである。これらの手法では、正確な抽出を実現するため、書き込みの色や種類に制限を課している。一方、書き込みを手がかりとした文書の検索やユーザの興味抽出、ユーザによる書き込みの閲覧などの利用法を考える際には、書き込みの有無の判定や位置の特定ができればよく、認識に利用するほどの抽出精度は必要とされな

い。しかしながら、検索や興味抽出という利用法において対象に制限を課すことは現実的ではないため、制限の少ない処理が望まれる。

本稿では、後者の立場、即ち、認識を施さない代わりに対象に制限を加えないという手法について考える。書き込みの色や形に前提を設けることなく抽出するための方策として、本研究では、書き込みのなされた紙文書の画像（書き込み画像）と、元となった書き込みのない電子文書の画像（元画像）の差分を取るというアプローチを採用する。近年では多くの紙文書がコンピュータを利用して作成されているため、紙文書の元となった電子文書が利用可能であると想定することはそれほど非現実的ではない。

このようなアプローチで書き込み抽出を行う場合、(1) 書き込み画像と元画像の位置合わせ、(2) 位置合わせされた書き込み画像と元画像の差分処理、の2つの問題を解決する必要がある。提案手法では、これらの問題点について次のように対処する。(1)の問題については、我々の既に提案している、特徴点の局所的配置に基づく文書画像検索手法 [5] を応用する。[5]の手法はモノクロの英文文書を対象としたものである。本手法ではこの制限を撤廃するため、色クラスタリングを導入する。(2)の問題については、局所領域で対応する画素を探索することで位置合わせの誤差を吸収するという柔軟な差分処理を用いる。実験により、提案手法を用いることで高い精度で位置合わせが可能であり、カラー文書でも任意の色と種類の書き込みが抽出できることが示された。

2. 関連研究

書き込みのなされた紙文書から書き込みを抽出する様々な手法が提案されている [1] ~ [4]。これらは以下の2つのタイプに分類される。

1つは書き込みの抽出ではなく、書き込みを用いた自動校正などの抽出された書き込みの利用に重点を置いたもの [1], [2] である。これらの手法では、抽出結果が機械で利用されるため、高精度な抽出が求められる。その実現のため、これらの手法では書き込みに制限を設けている。具体的には、書き込みに用いることのできるペンの色があらかじめ定められており、スキャンされた画像の画素の色によって書き込みであるか否かを判断する。そのため、これらの手法では書き込みの色に制限がある。

もう1つは、画像における連結成分を手書き文字と印刷された文字に分類するもの [3], [4] である。これらの手法では、書き込みのなされた文書の画像のみから書き込み抽出が可能であるという長所がある一方、抽出できる書き込みは文字に限られており、手書きの線や図形などを抽出することはできない。実際の書き込みでは下線や矢印などの図形は頻繁に用いられるため、そのような書き込みを抽出できる手法が必要である。

3. 提案手法

3.1 処理の概要

提案手法の目的は、元となった電子文書が利用可能であるという前提の下で、書き込みのなされた文書から書き込みだけを

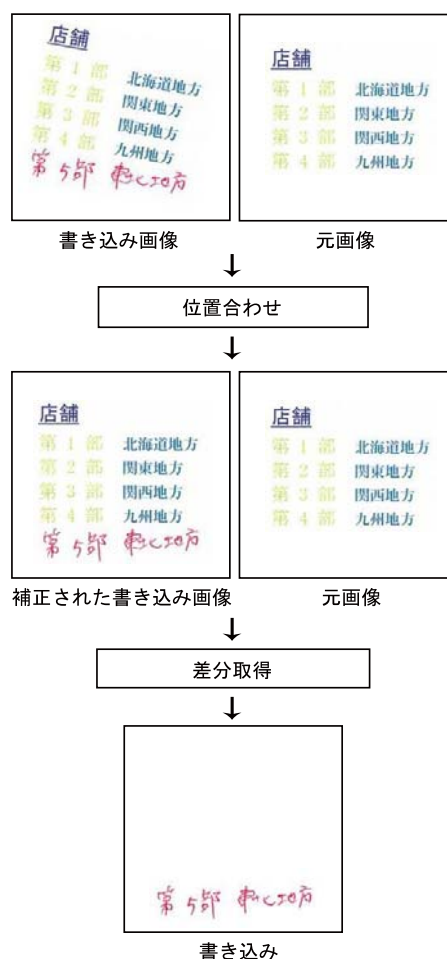


図1 処理の流れ

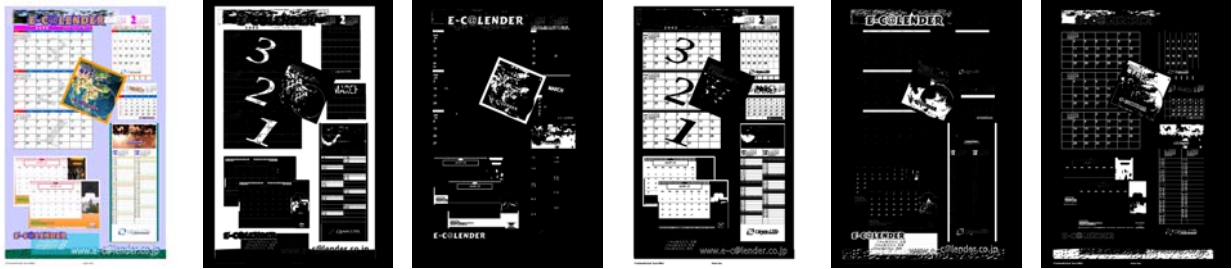
抽出することである。元となった電子文書が利用可能であれば、書き込みのなされた文書を画像化し、書き込みのない電子文書の画像との差分を得ることで書き込みを抽出することができると考えられる。ただし、そのためには書き込み画像と元画像の正確な位置合わせを行い、画像化の際のずれや回転を補正する必要がある。

図1に示すように、提案手法では、位置合わせと差分取得の2処理によって、以上の問題点を解決する。これらの処理について以下で説明する。

3.2 位置合わせ

この処理では、書き込み画像の元画像への位置合わせを行う。書き込み画像はスキャナによって取り込まれるため、位置のずれ、回転、サイズの変化が生じる場合がある。これらの変化は相似変換と呼ばれる。従って、この処理では書き込み画像と元画像の間の相似変換パラメータを推定し、パラメータを用いて書き込み画像を変換する。なお、書き込み画像には書き込みが存在するため、変換パラメータの推定は困難なものになる。

この問題を解決するため、特徴点の局所的配置に基づく検索法 [5] を応用する。なお [5] の手法での特徴量計算には相似不変量を用いる。処理の概要を以下に述べる。まず、書き込み画像から特徴点を抽出する。次に、得られた特徴点と、あらかじめ抽出しておいた元画像の特徴点を対応付ける。そして、対応



(a) 入力画像

(b) 色クラスタごとに分解した画像

図 2 色クラスタごとに画像を分解 ($k = 5$)

点を用いて相似変換パラメータを推定する．以上の処理で最も重要な部分は，特徴点の対応付けである．すべての可能な点対応について調べれば必ず正しい対応が見つかるものの，そのような処理を行うと組み合わせ爆発が生じるため，現実的ではない．そこで，提案手法では[5]の手法を用いて特徴点を対応付ける．この手法では，特徴点の数に比例した計算量しか必要ないため，効率的な点の対応付けが可能である．

ところで，[5]の手法は英語のモノクロ文書を対象としたものであり，特徴点抽出に英文特有の性質を利用している．そのため，任意のカラー画像に適用するためには特徴点抽出処理を変更する必要がある．以下で詳細について説明する．

3.2.1 特徴点抽出

ここでは書き込み画像および元画像から，変形やノイズに対してロバストな点を特徴点として抽出する．本手法では特徴点として，画像を RGB 色空間で色クラスタリングして得られた色ごとの連結成分の重心を用いる．これは，文字などの単色領域の位置を特徴点として用いることを意図したものである．文字は特徴的な配置をもち，また背景とのコントラストが大きいため安定した抽出が可能である．

特徴点抽出の手順を以下で述べる．まず， k -means 法を用いて画像を色クラスタリングする．ここで，色クラスタの数 k はあらかじめ定められた値である．そして，図 2 に示すように，色クラスタリングの結果に基づいて画像を k 枚の白黒画像に分解する．これらの画像から連結成分を抽出し，その重心を色クラスタごとの特徴点とする．

3.2.2 対応点探索

ここでは，書き込み画像の特徴点と元画像の特徴点を，文献[5]の文書画像検索法を応用して対応付ける．

この段階では，書き込み画像および元画像の特徴点は，それぞれ k 個の色クラスタごとに存在する．色クラスタによってはノイズの特徴点が多く含まれている場合があるため，全クラスタの特徴点をまとめてから対応付けるのではなく，色クラスタごとに対応付ける方が正しい対応点を得やすい．そのため，特徴点同士の対応付けに先立って，書き込み画像の色クラスタと元画像の色クラスタを対応付ける．本手法では，greedy 法を用いて色クラスタの RGB 色空間での重心の距離のより小さいものを優先的に対応付ける．

次に，対応付けられた色クラスタの特徴点単位の対応付けを行う．これは，文献[5]のように，元画像の特徴点をデータベース



図 3 対応点の例

に登録し，書き込み画像の各特徴点に対して対応するものを検索することで実現される．ただし，登録される画像が 1 つだけである点が文献[5]の場合と異なっている．特徴点の対応付け処理を行うと，図 3 に示すような特徴点同士の対応関係が得られる．

3.2.3 相似変換パラメータの取得

ここでは，特徴点の対応関係に基づいて相似変換パラメータを取得する．相似変換パラメータは 4 次元のベクトルであり，2 組の対応点から計算されるが，対応点には誤ったものが含まれている場合があるため，誤った対応点を除いて変換パラメータを計算する必要がある．提案手法では，RANSAC [6] を用いてパラメータを推定する．

得られた相似変換パラメータを用いて，書き込み画像に相似変換を適用し，元画像への位置合わせを行う．これにより，元画像と同じ大きさの補正された書き込み画像が得られる．

3.3 差分取得

この処理では，補正された書き込み画像と元画像を比較し，差分を取ることで書き込みを抽出する．以下で手順について詳しく説明する．

まず，補正された書き込み画像と元画像を画素ごとに比較し，差分および閾値処理を行う．ただし，スキャナによる画像取り込みの際の歪みや特徴点の誤差などによって，書き込み画像と元画像に微小なずれがある場合があるため，ずれに対処して差分を取る必要がある．

提案手法では，以下のようにして差分および閾値処理を行う．まず，書き込み画像の各画素と元画像の対応する画素を比較す

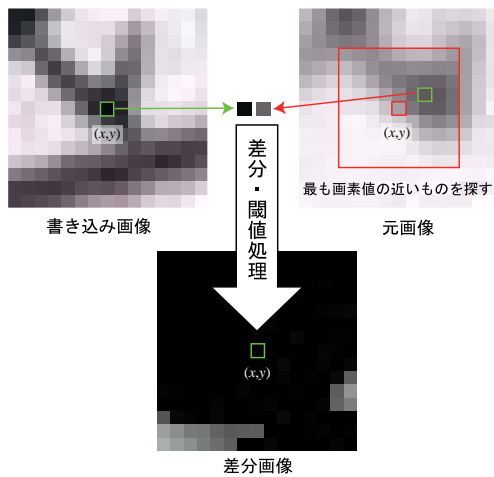


図 4 差分・閾値処理

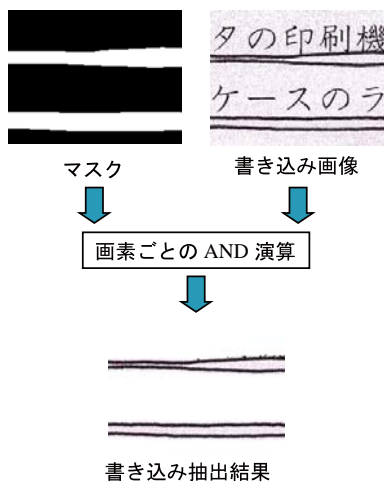


図 5 AND 処理

る．図 4 に示すように，書き込み画像の画素と比較する元画像の画素は，同一の座標のものではなく，同一の座標を中心とした正方形の領域における画素値の最も近いものとする．このように画素値の近いものを探すことで，書き込み画像と元画像に微小なずれがあっても対応する画素を見つけることができる．こうして得られた書き込み画像と元画像の画素値の差を求め，閾値処理によって 2 値化する．

ここまでの処理で得られた画像からノイズを除去し，モルフォロジ演算で連結成分を拡大してマスクを作成する．そして，マスクと書き込み画像の画素ごとの AND を取ることで書き込みを抽出する．処理の概要を図 5 に示す．このように，マスクが大きめに取られているため書き込みだけでなく背景も抽出される．しかし，人間が利用する場合は，書き込みの場所が特定されれば良く，書き込みと背景は容易に区別できるためあまり問題とはならない．

4. 実験結果

実験では，PDF ファイルを印刷したものに黒・赤・青のボールペンで文字や図形などを数箇所に書き込み，スキャナを用いて 600dpi で取り込んだ画像を書き込み画像とした．同様に，



(a) カラー文書



(b) モノクロ文書

図 6 実験に用いた文書の例

PDF ファイルを印刷し，書き込みを行わずにそのままスキャナで画像化したものを元画像とした．本来は印刷やスキャニングを経ずに，PDF ファイルをそのままラスタ形式の画像としたものを元画像として用いるべきだが，その場合は印刷とスキャニングの際に生じる画像の劣化を再現する処理を加えなければならない．今回の実験では，画像の劣化処理が適切に行われたと仮定して，実際に印刷およびスキャニングを行うことで劣化した元画像を作成した．

実験に用いた計算機は AMD Opteron 2.8GHz の CPU を搭載し，16GB のメモリを持つものである．実験では，文献 [5] の手法による対応点探索のパラメータを $n = 5$ ， $m = 5$ とし，不変量に相似不変量を用いた．

なお，提案手法を適用する上での問題点を明確にするため，図 6(a) に示されるようなポスターや Web ページなどの図形や写真を多く含むカラー文書と，図 6(b) に示されるような文字が紙面の大部分を占めるモノクロ文書を用いてそれぞれ実験を行った．

4.1 カラー文書を用いた実験

この実験では，109 枚の元画像を用いた．1 つの元画像に対して 3 通りの書き込み画像を用意したため，書き込み画像の数は 327 枚である．色クラスタリングの際のクラスタ数は $k = 5$ とした．

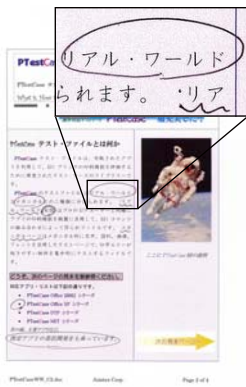
最終的な書き込みの抽出結果だけでなく，位置合わせ処理が終わった時点の補正された書き込み画像についても評価を行った．これは，位置合わせ処理と差分処理の，それぞれの性能を明確にするためである．位置合わせ処理および抽出処理の結果の判定は，目視によって行った．抽出結果の評価は，以下の 3 つに分類することで行った．

成功 書き込みが十分抽出され，ノイズがほとんどないもの (図 7)

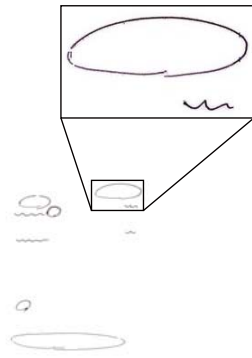
かすれ・ノイズあり 書き込みが部分的にかすれたり，目立つノイズがあるもの (図 8)

失敗 書き込みがほぼ完全に失われていたり，紙面の大半をノイズが占めているもの (図 9)

位置合わせの結果を表 1 に，抽出結果を表 2 に示す．これら



(a) 書き込み画像

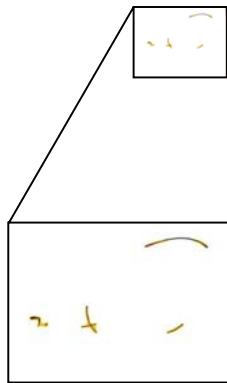


(b) 抽出結果

図 7 成功の例



(a) 書き込み画像



(b) 抽出結果

図 8 かすれ・ノイズありの例



(a) 書き込み画像



(b) 抽出結果

図 9 失敗の例

表 1 カラー文書の位置合わせ結果

	成功	失敗	合計
単色領域が多い	133(99%)	2(1%)	135(100%)
単色領域が少ない	167(87%)	25(13%)	192(100%)
合計	300(91%)	27(8%)	327(100%)

の表では、文字などの単色領域を多く含む画像と、単色領域の少ない画像とで分けて示してある。これは、単色の連結成分の重心を特徴点とする提案手法での入力画像の性質の影響を明ら

表 2 カラー文書の抽出結果

	成功	かすれ・ノイズあり	失敗	合計
単色領域が多い	119(88%)	13(10%)	3(2%)	135(100%)
単色領域が少ない	131(68%)	36(19%)	25(13%)	192(100%)
合計	250(76%)	49(15%)	28(9%)	327(100%)

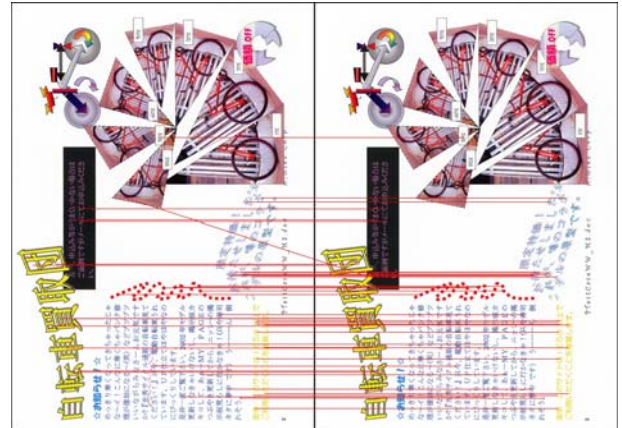


図 10 単色領域が多い場合の対応点

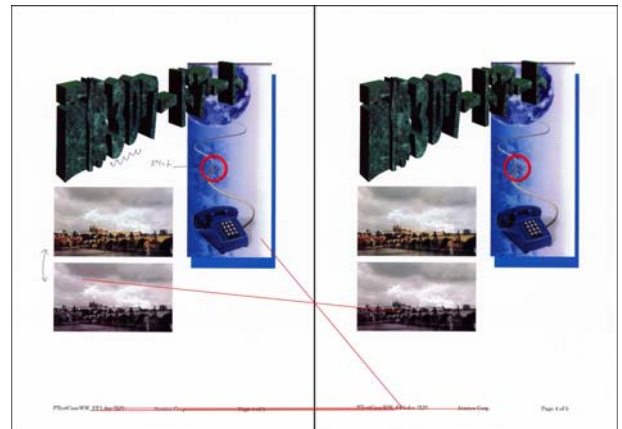


図 11 単色領域が少ない場合の対応点

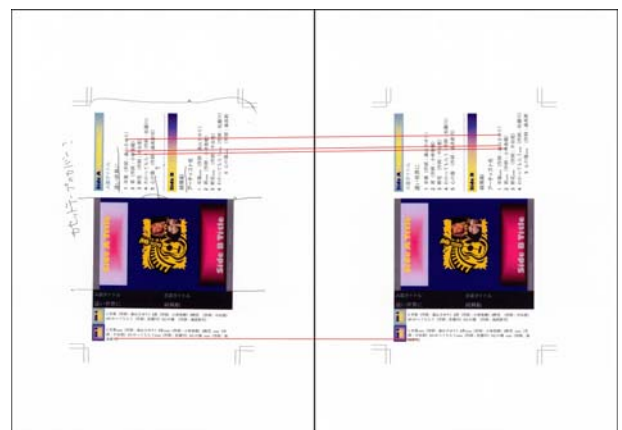


図 12 同じ特徴点の配置を複数の場所にもつ場合の対応点

かにするためである。なお、単色領域が多いかどうかの判定は目視によって行った。

以下では実験結果を考察する。まず、表 1 に示す位置合わせ

表 3 モノクロ文書の位置合わせ結果

成功	失敗	合計
34(100%)	0(0%)	34(100%)

表 4 モノクロ文書の抽出結果

成功	かすれ・ノイズあり	失敗	合計
32(94%)	2(6%)	0(0%)	34(100%)

の結果について考察する．全体では，91%が位置合わせに成功した．特に，単色領域の多いものでは99%が成功した．単色領域の多いもので成功が多く，少ないもので失敗が多いのは，以下の理由による．文字などの単色領域が多いと，それらの領域から書き込み画像と元画像で共通な，安定した特徴点を多く抽出することができ，図 10 に示すように多くの正しい対応点を得ることができる．一方，単色領域が少ないと，スキャンの際の微小な変動によって色クラスタリングの結果が変化するため，安定した特徴点を得ることが難しくなる．そのため，図 11 に示すように，正しい対応点があまり得られなくなる．単色領域が多いものでも 2 つの画像で失敗した．これらは，同一画像内で全く同じテキストが複数の場所に存在するなど，同じ特徴点の配置が複数あるものである．このような場合，図 12 のように誤った対応が生じるため位置合わせに失敗する．全体としては，ポスターなどのカラー文書でも 91%で位置合わせに成功したことから，手法のロバスト性が示された．

次に，表 2 に示す書き込み抽出の結果について考察する．全体では 76%，単色領域の多いものに限定すると 88%が書き込み抽出に成功した．また，単に閲覧するだけといった用途では，多少のかすれやノイズがあってもあまり問題はない．このような場合では，成功率は全体で 91%，単色領域の多いものでは 98%であるといえる．失敗に分類されたもののほとんどは，位置合わせの時点で失敗したものである．位置合わせが成功したにも関わらず，失敗あるいはノイズ・かすれありに分類されたものは，色付きの背景に書き込まれたために差分処理において書き込みが消失したものと，位置合わせに微小な誤差があったために差分処理後にノイズが生じたものである．

4.2 モノクロ文書を用いた実験

カラー文書を用いた実験では，文字などの単色領域の少ない画像や，色付きの背景に書き込みのなされた画像において書き込み抽出が困難であることが示された．本実験では，多数の文字を含み，白地に書き込みのなされるモノクロ文書に対して提案手法を適用し，提案手法に適した対象での性能を調べた．

実験には英語論文の PDF ファイルから作成した 34 枚の元画像と，それらに黒・赤・青のボールペンで書き込みを行った 34 枚の書き込み画像を用いた．色クラスタリングの際のクラスタ数は $k = 2$ とし，それ以外の条件はカラー文書の場合と同じものとした．

位置合わせおよび書き込み抽出の結果を，それぞれ表 3 と表 4 に示す．表 3 に示されるように，モノクロ文書の場合ではすべての画像で位置合わせに成功した．これは，多数の文字が含まれるため，多くの安定な特徴点を得ることが可能であり，

それによって多くの正しい対応点から適切な変換パラメータを推定できるためである．表 4 に示されるように，2 枚の文書において抽出された書き込みにかすれが生じていた．これらは，赤色のボールペンによる書き込みが薄くなっていたため，差分処理で部分的に消失したものである．ただし，94%のもので書き込み抽出に成功しており，失敗したものは存在しなかった．以上のことから，提案手法は文字中心のモノクロ文書に対しては極めて有効であることが示された．

5. ま と め

本稿では，文書に対する書き込みの抽出法を提案した．これは，元となった文書画像が利用可能であるという前提の下で，対応点探索によって相似変換パラメータを求めて位置合わせを行い，差分を取ることで書き込みを抽出するものである．提案手法の特徴は，文書画像検索法を応用した位置合わせと，局所的なずれに強い柔軟な差分処理である．

書き込み抽出の実験では，76%の文書で書き込み抽出に成功した．これは，図形や写真が大部分を占める処理が困難な画像も含んだ場合の結果であり，文字などの単色領域の多く含まれるものに限定すれば 88%の精度が得られた．また，特徴点が安定して得られる文字中心のモノクロ文書を対象とした実験では，94%の精度を得ることができた．以上より，本手法は適した対象からは高い精度で書き込みの抽出が可能であり，適さない対象についてもある程度の精度が得られることが示された．

今後の課題としては，特徴点抽出処理を改良して文字をあまり含まない対象についても安定な特徴点を得られるようにすることが挙げられる．これには，Harris オペレータ [7] などのコンピュータビジョンでの研究成果を導入することが有効であると考えられる．

文 献

- [1] D. Möri and H. Bunke, "Automatic Interpretation and Execution of Manual Corrections on Text Documents", in *Handbook of Character Recognition and Document Image Analysis*, ed. H. Bunke and P. S. P. Wang, pp.679-702, World Scientific, Singapore (1997).
- [2] J. Stevens, A. Gee, and C. Dance, "Automatic Processing of Document Annotations", In *Proc. 1998 British Machine Vision Conf.*, Vol. 2, pp.438-448 (1998).
- [3] J. K. Guo and M. Y. Ma, "Separating Handwritten Material from Machine Printed Text using Hidden Markov Models", In *Proc. 6th International Conf. on Document Analysis and Recognition*, pp.436-443 (2001).
- [4] Y. Zheng, H. Li, and D. Doermann, "The Segmentation and Identification of Handwriting in Noisy Document Images", In *Lecture Notes in Computer Science (5th International Workshop DAS2002)*, vol.2423, pp.95-105 (2002).
- [5] 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及び相似不変量の利用", 信学技報, PRMU2005-188 (2006).
- [6] 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).
- [7] C. Harris and M. Stephens, "A Combined Corner and Edge Detector", *Proc. 4th Alvey Vision Conf.*, pp.147-151 (1988).