

特徴点の局所的配置に基づく文書画像モザイクングの実験的検討

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、デジタルカメラで撮影された、撮影範囲の異なる複数の文書画像をつなぎ合わせるモザイクング法を提案する。デジタルカメラで撮影された文書画像は任意の射影歪みを受けているため、その位置合わせは従来手法では困難なものであった。提案手法では、特徴点に基づく文書画像検索法である LLAH を利用することで、射影歪みを受けた文書画像の特徴点の対応関係を得て、それに基づいて画像間の変換パラメータを推定して位置合わせを行う。LLAH は射影歪みに対して不変であるため、射影歪みの補正なしに特徴点の対応付けが可能である。実験により、デジタルカメラで撮影された文書画像のつなぎ合わせが可能であることが示された。

キーワード 文書画像モザイクング、カメラベース文書画像処理、LLAH

Experimental Evaluation of Document Image Mosaicing 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 In this paper we propose a mosaicing method of camera-captured document images. Since document images captured using digital cameras suffer from perspective distortion, their alignment is a difficult task for previous methods. In the proposed method, correspondences of feature points are calculated using an image retrieval method LLAH. Document images are aligned using a perspective transformation parameter estimated from the correspondences. Since LLAH is invariant to perspective distortion, feature points can be matched without compensation of perspective distortion. Experimental results show that document images captured by a digital camera can be stitched using the proposed method.

Key words Document image mosaicing, Camera-based document image processing, LLAH

1. はじめに

近年の OA 化されたオフィス環境において、印刷文書を電子化することは頻繁に行われる。印刷文書から電子化された文書を得る際に、よく用いられるのはスキャナである。しかし、スキャナは持ち運びが困難であり、また貴重な文書を損なう可能性があるため適用できない場合もある。

そこで、デジタルカメラを用いて印刷文書を撮影し、画像として電子化することが考えられる。デジタルカメラを用いることで、持ち運びのできる機器によって文書を非接触で電子化することが可能になる。しかし、現時点ではスキャナに比べてデジタルカメラの解像度は低い。従って、デジタルカメラを用い

てスキャナと同等の解像度で文書を画像化するためには、何カ所かに分けて部分ごとに撮影し、後からつなぎ合わせる必要がある。このつなぎ合わせの処理が“モザイクング”と呼ばれるものである。

モザイクングを行うためには、重なり領域のある文書画像を適切に配置する必要がある。さらに、デジタルカメラを用いる場合は撮影角度が異なる場合があるので、画像間の射影歪みを補正する必要がある。また、現実的な処理時間で処理を完了できることが望ましい。画像のモザイクングを行うさまざまな手法が提案されているが、これらの問題をすべて解決したものは未だに存在しない。従来手法の多く [1], [2] はフラットベッドスキャナでスキャンされた画像をモザイクングすることを目的と

しているため、デジタルカメラで撮影された画像ではうまく動作しない。デジタルカメラを用いるもの [3] も提案されているが、射影歪みの除去のために重い処理を必要とするため、高速な処理は困難であると考えられる。

本研究では、デジタルカメラを用いて撮影された文書画像の高速なモザイクングを実現する手法を提案する。なお、提案手法は平面の文書を撮影した画像を対象としており、文書画像における湾曲などの非射影歪みについては想定していない。提案手法は 3 つのステップから構成される。1 番目のステップでは、2 枚の文書画像から特徴点が抽出され、LLAH [4] と呼ばれる画像検索法を用いて特徴点の対応付けが行われる。2 番目のステップでは、特徴点の対応関係に基づいて画像間の射影歪みを補正するための射影変換パラメータが計算される。変換パラメータを用いると、一方の画像がもう一方に合うように変換され、位置合わせが行われる。3 番目のステップでは、位置合わせされた画像をつなぎ合わせて、モザイク画像が生成される。LLAH は高速な検索手法であり、10,000 ページの文書を含むデータベースからの文書画像の検索がおよそ 30ms で可能である [4]。その高速性を利用したアプリケーションとして、リアルタイム文書画像検索も提案されている [5]。LLAH は高速かつ射影歪みに不変な手法であるため、前処理なしで射影歪みを受けた画像の特徴点の対応付けを高速に行うことができる。これにより、デジタルカメラで撮影された文書画像の高速なモザイクングが実現できる。

2. 関連研究

画像のモザイクング手法として、さまざまなものが提案されている。Whichello らの手法 [1] は、画像間の相関に基づいたものである。この手法では、2 枚の 2 値画像 $f(i, j)$ と $g(i, j)$ について、以下の式で定義される相関 $c(x, y)$ が最大となる x および y を見つけることで画像の位置ずれを計算する。

$$c(x, y) = \sum_i \sum_j f(i, j)g(i - x, j - y) \quad (1)$$

画像の位置ずれと (x, y) が等しいときは、重なり領域において $f(i, j)$ と $g(i - x, j - y)$ は同じ値となるため、 $c(x, y)$ は最大値をとる。その (x, y) を用いて画像 $f(i, j)$ と $g(i, j)$ をつなぎ合わせ、モザイクングを行う。この手法には、さまざまな高速化のための工夫も含まれているが、画像の相関を調べるという手法の性質上、位置ずれ以外の外乱には対処できない。従って、デジタルカメラを用いる際に生じる射影歪みはおろか、スキャナを用いた場合にも生じうる回転やスケール変化にすらロバストではない。

Isgro らの手法 [2] は、特徴量に基づいたものである。この手法では、まず片方の画像から特徴点を抽出し、もう一方の画像において対応する点の場所を得る。そして、対応付けられた特徴点から 2 枚の画像の間のユークリッド変換パラメータを推定する。得られたユークリッド変換パラメータに基づいて文書画像のつなぎ合わせを行う。この手法では、特徴点の対応付けを段階的に行うため、効率的な対応付けが実現されている。しか

し、ユークリッド変換は、位置ずれと回転のみを含むため、この手法ではスケールの変化や射影歪みに対処することはできない。さらに、特徴点の対応付けに画像の相関を用いているため、程度の大きい回転には対処することができない。従って、この手法もまたデジタルカメラで任意の角度から撮影された文書画像のつなぎ合わせを実現することはできない。

Lian らの手法 [3] は、カメラで撮影された文書画像のモザイクング手法である。前述の手法と異なり、この手法ではデジタルカメラで任意の角度から撮影された文書画像のつなぎ合わせを実現するものである。この手法では、まず文書の行の方向と、文字の縦方向ストロークの方向から文書の乗っている平面の消失点を推定し、そこから文書画像の射影歪みを補正して文書画像を正対したものにす。次に、PCA-SIFT [6] によって特徴点とディスクリプタを抽出し、対応付ける。限られた文字や単語から構成されるという文書画像の性質上、同一の文書から同じディスクリプタが複数得られて誤対応が生じることがある。そこで、対応点の一貫性を調べて誤った対応点を除去する。最後に、画像の相関を用いて正確な位置合わせを行い、文書画像をつなぎ合わせる。この手法では、デジタルカメラの姿勢に制限がないため高い柔軟性を持つ。しかし、特徴点の対応付けに先立って、射影歪みの補正が必要なことが問題となる。さまざまな外乱の下で安定して行や文字のストロークを抽出することは困難である。また、このような処理は膨大な計算量を必要とする。従って、この手法は事前の射影歪みの補正を必要とするという点において信頼性と効率に関する問題があるといえる。

文書画像のモザイクングを行う手法として、ビデオモザイクング [7] も提案されている。これは、ビデオカメラで撮影された印刷文書の動画について、フレーム画像をつなぎ合わせることでモザイク画像を生成するものである。この手法では、各フレーム画像から特徴点を抽出し、その対応関係から各フレームでのカメラパラメータを推定してフレーム画像の合成を行う。この手法は、湾曲した紙面を撮影した動画の平面画像へのモザイクングを実現している。しかし、類似したフレーム画像が連続するという撮影動画の性質を利用しているため、通常のデジタルカメラでランダムに撮影された静止画像のモザイクングにそのまま適用することは難しいと考えられる。

3. LLAH: Locally Likely Arrangement Hashing

LLAH は、特徴点に基づく画像検索法である。画像から特徴点を抽出し、特徴点の配置に基づいて記述子を計算する。そして、記述子の比較に基づいて検索質問に対応する画像をデータベースから検索する。記述子に幾何学的不変量を用いているため、検索質問とデータベースの画像の間に射影歪みがあっても動作する。また、検索の過程で特徴点の対応関係が得られる。従って、LLAH を用いることで、デジタルカメラで撮影された文書画像の特徴点の対応付けが可能になると考えられる。

3.1 特徴点抽出

特徴点抽出における要件は、射影歪み、ノイズ、解像度の変化を受けても、同じ文書画像からは同じ特徴点が得られなけれ

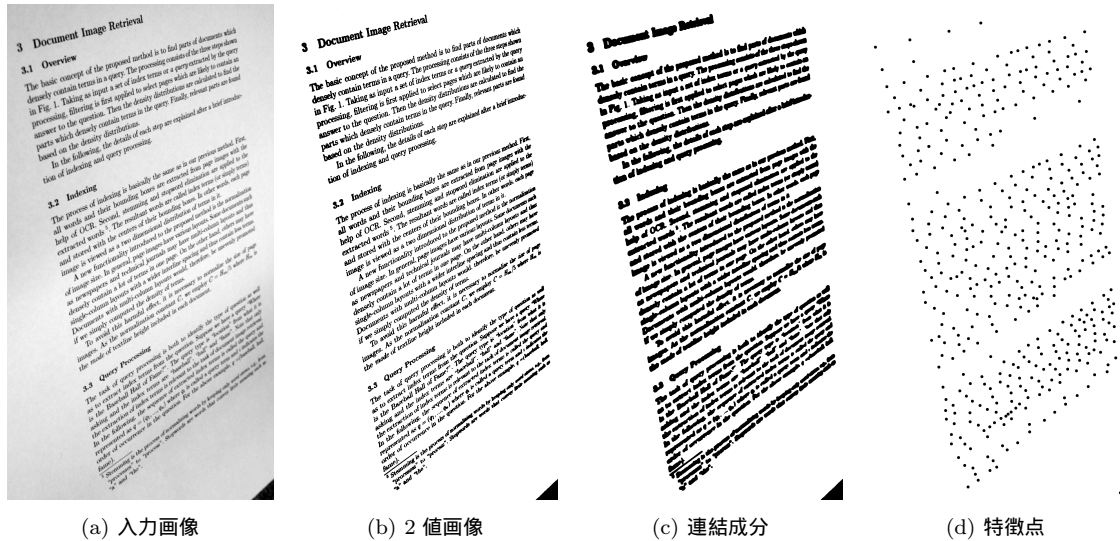


図 1 特徴点抽出

ばならないということである．この要件を満たすため，英語文書における単語領域の重心を特徴点として用いる．

特徴点抽出の処理を以下に示す．まず，入力画像（図 1(a)）は適応 2 値化され，2 値画像（図 1(b)）に変換される．次に，ガウシアンフィルタで画像をぼかし，再度適応 2 値化を施すことで，単語ごとに結合された画像（図 1(c)）が得られる．最後に，単語領域の重心（図 1(d)）が特徴点として抽出される．

3.2 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．

$$\frac{P(A, C, D)}{P(A, B, C)} \quad (2)$$

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 ${}_m C_4$ 個のアフィン不変量 $(r_{(0)}, \dots, r_{(mC_4-1)})$ 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， ${}_n C_m$ 個の記述子が計算される．

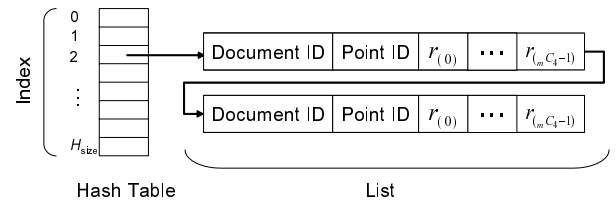


図 2 ハッシュ表の構成

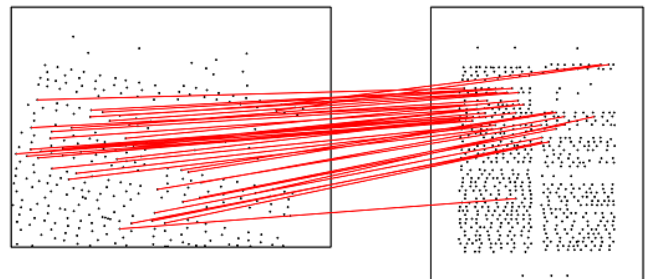


図 3 検索結果と特徴点の対応関係

3.3 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 2 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 3 に示すような特徴点の対応関係も得ることができる．

4. LLAH による文書画像モザイクング

提案手法は，以下の 3 つのステップで構成される．まず，LLAH を用いて 2 枚の文書画像の特徴点の対応関係を求める．

次に、対応関係に基づいて2枚の文書画像の一方をもう一方に合わせる射影変換のパラメータを推定する。最後に、2枚の画像をつなぎ合わせた画像を生成する。

LLAHを用いた文書画像検索では、検索の過程で検索質問の文書画像の各特徴点ごとに、データベースの文書画像の対応する特徴点が見られる。従って、2枚の文書画像の一方を検索質問とし、もう一方をデータベースの文書画像とすれば、図3のように2枚の文書画像の特徴点の対応関係が見られる。この場合、データベースに保存される画像は1枚だけである。

特徴点の対応関係を用いて、2枚の文書画像の位置合わせを行う。対応関係からの射影変換パラメータの推定には、RANSACアルゴリズム[8]を用いる。対応点にはいくつかの誤ったものが含まれる場合があるが、RANSACによりアウトライヤを除去した、正しい対応点のみに適合するパラメータを得ることができる。このパラメータを用い、文書画像をもう一方の文書画像に合うように変換する。

モザイク画像は、単純な合成処理で生成される。2枚の画像を $f(x, y)$ と $g(x, y)$ とし、それぞれの画像の範囲を F, G としたとき、モザイク画像 $h(x, y)$ は以下の式で定義される。

$$h(x, y) = \begin{cases} f(x, y) & (x, y) \in F \text{ and } (x, y) \notin G \\ g(x, y) & (x, y) \notin F \text{ and } (x, y) \in G \\ \frac{f(x, y) + g(x, y)}{2} & \text{otherwise} \end{cases} \quad (3)$$

なお、以上の処理で得られるものは、1枚目の文書画像に合うように2枚目の文書画像を変換し、合成したモザイク画像である。従って、1枚目の文書画像が斜め方向から撮影されていて正対していない場合、得られるモザイク画像も正対したものにはならない。文字認識などの処理に用いるために正対した文書画像が必要な場合は、モザイク画像を正対に戻す処理が必要となる。

この点から、正対した文書画像を入力として必要とする従来手法に対して、提案手法の優位性が存在しないと思われるかもしれない。しかし提案手法の場合、正対した文書画像を得るためには、モザイク画像に対して射影歪み除去の処理を一度加えるだけでよい。一方、従来手法の場合は入力画像すべてに対して射影歪み除去を行う必要があり、処理効率の面で提案手法に優位性がある。また提案手法の場合に行われるのは、多くの文字や行を含むモザイク画像への射影歪み除去であり、撮影範囲の狭い入力画像への処理よりも高精度な結果が期待できる。従って、処理精度の面でも提案手法の場合に優位性があるといえる。

5. 実験結果

提案手法を用いて、4組の文書画像の合成を行った。文書画像の撮影には Canon EOS 5D を用いた。文書画像のサイズは 4368×2912 である。画像のモザイクは 2.2GHz の CPU と 2GB のメモリをもつ PC で行った。3.2 で述べた LLAH のパラメータは、 $n = 7$ and $m = 6$ とした。撮影画像と特徴点の対応関係、得られたモザイク画像を図4-7に示す。

各ステップの平均処理時間は表1に示す。実験で用いられた

表1 平均処理時間

	処理時間 [ms]
特徴点の抽出	8,464
特徴点の対応付け	571
射影変換パラメータの推定	2
モザイク画像の生成	3,139

表2 特徴点数

	画像1	画像2	対応点
実験1	519	336	156
実験2	515	313	94
実験3	645	327	66
実験4	457	322	7

画像の解像度が高かったため、ガウシアンフィルタなどの画像処理を含む特徴点抽出処理に多くの処理時間を要した。この問題については、特徴点抽出の前に画像を縮小することで容易に対処できると考えられる。

図4-7に示される画像での実験における特徴点数を、表2に示す。表2において、画像1と画像2は、図4-7の左上の入力画像において、それぞれ左側と右側に示されるものである。表2には、対応関係の得られた特徴点の数も示されている。

図4と図5は、成功したケースである。LLAHによって多数の正しい対応点が見られ、そこから推定された射影変換パラメータによって正しく位置合わせが行われた。図5のケースでは、入力画像は射影歪みを受けているが、正しく位置合わせおよびモザイクができた。これらの結果により、提案手法はある程度の射影歪みに対するロバスト性をもつことが示された。

図6と図7は、失敗したケースである。おおまかな位置は合っているが、細部において微小なずれが生じている。これは、特徴点の位置の変化と、非射影歪みによって引き起こされたと考えられる。まず、特徴点の位置の変化について述べる。提案手法では、特徴点は単語領域の重心から得られる。具体的には、ガウシアンフィルタなどの画像処理によって単語ごとの連結成分を作成し、連結成分の重心を求めることで特徴点を得る。しかし、これらの処理は射影歪みを受けた状態で行われるため、異なる射影歪みを受けた文書画像では、単語領域の重心は微妙に異なることがある。特徴点の位置の変化は、強い射影歪みを受けると顕著となるため、正しく位置合わせができなくなると考えられる。次に、非射影歪みの問題について述べる。提案手法の位置合わせは、画像全体に補正のための射影歪みを施すことで行われる。しかし、レンズの歪みなど、文書画像の受ける歪みは射影歪みだけではない。従って、補正のための射影変換だけでは微小なずれが残る場合がある。

以上のような失敗は、画像の間の射影歪みが大きい場合に生じやすいと考えられる。従って、微小な位置ずれを補正するための追加の処理が必要である。例えば、従来手法で用いられているような画像の相関に基づく手法が有効であると考えられる。大まかに位置合わせされた画像の各ウィンドウについて、もう一方の画像での一致の最も大きい領域を探索することで正確な位置合わせが実現できると考えられる。

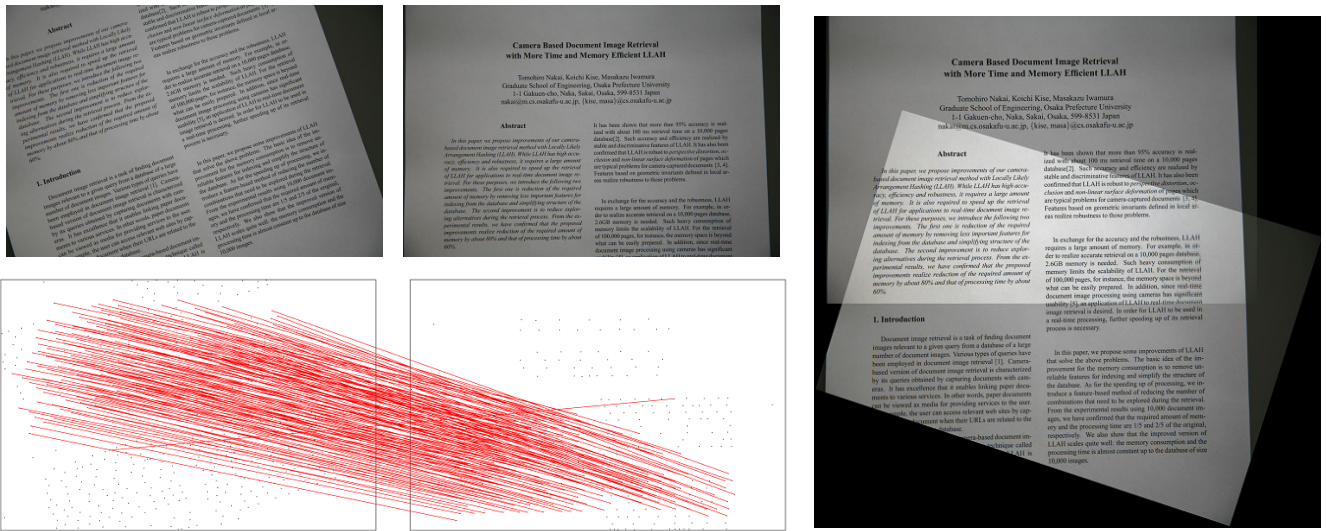


図4 実験結果 1 左上:入力画像 左下:対応点 右:モザイク画像

微小なずれの補正のために、追加の処理が必要であるとしても、提案手法の意義は大きい。画像の大きな位置が合っていれば、探索すべき領域は極めて限られるため、少しの処理で目的を達することができるからである。そもそも、図7のように過剰な射影歪みを受けた文書画像のモザイクングを行うというケースは、想定しにくいものである。利用者がモザイクングを行うためにデジタルカメラを用いて文書を撮影する場合、紙面に対する姿勢は可能な限り正対したものになると考えられる。図7のように過度の射影歪みを受けた画像では、大部分がピンぼけとなってモザイクングする意味があまりない。従って、多少の射影歪みに対するロバスト性でもモザイクング手法としては十分であるといえる。

6. まとめ

本稿では、デジタルカメラで撮影された文書画像のモザイクング手法を提案した。デジタルカメラで撮影された文書画像は任意の射影歪みを受けているため、その位置合わせは従来手法では困難なものであった。提案手法では、特徴点に基づく文書画像検索法である LLAH を利用することで射影歪みを受けた文書画像の特徴点の対応関係を得て、それに基づいて画像間の変換パラメータを推定して位置合わせを行う。LLAH は射影歪みに対して不変であるため、射影歪みの補正なしに特徴点の対応付けが可能である。

実験により、デジタルカメラで撮影された文書画像のつなぎ合わせが可能であることが示された。いくつかのケースでは微小な位置ずれが生じることも確認されたが、大きな位置は合っているため、追加の処理を加えることでより正確な位置合わせが可能であると考えられる。今後の課題は、より正確な位置合わせのための追加の処理の考案が挙げられる。

本稿での実験では提案手法が有効であることが示されたが、モザイク画像がどの程度正確か、といった詳細な評価は示されていない。実験結果の定量評価も今後の課題に含まれる。

文献

[1] A. P. Whichello and H. Yan: "Document image mosaicing",

Proceedings of ICPR 1998, Vol. 2, pp. 1081-1083 (1998).
 [2] F. Isgró and M. Pilu: "A fast and robust image registration method based on an early consensus paradigm", Pattern Recognition Letters, **25**, 8, pp. 943-954 (2004).
 [3] J. Lian, D. DeMenthon and D. Doermann: "Camera-based document image mosaicing", International Conference on Pattern Recognition (2006).
 [4] 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).
 [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] Y. Ke and R. Sukthankar: "Pca-sift: A more distinctive representation for local image descriptors", Proceedings of CVPR 2004, Vol. 2, pp. 506-513 (2004).
 [7] 佐藤, 池谷, 池田, 神原, 中島, 横矢: "動画像からの三次元復元による湾曲した紙面のビデオモザイクング", 電子情報通信学会論文誌 (D), **J90-D**, 8, pp. 1900-1911 (2007).
 [8] M. A. Fischler and C. Bolles: "Random sample consensus: A paradigm for model fitting with application to image analysis and automated cartography", Communications of ACM, **24**, 6, pp. 381-395 (1981).

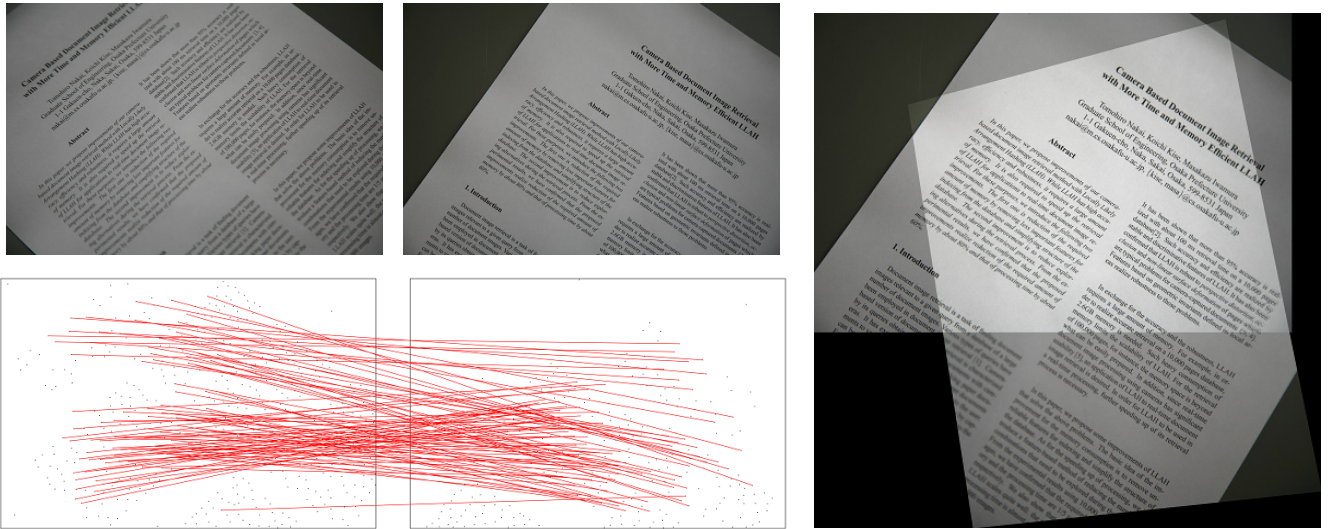


図5 実験結果2 左上:入力画像 左下:対応点 右:モザイク画像

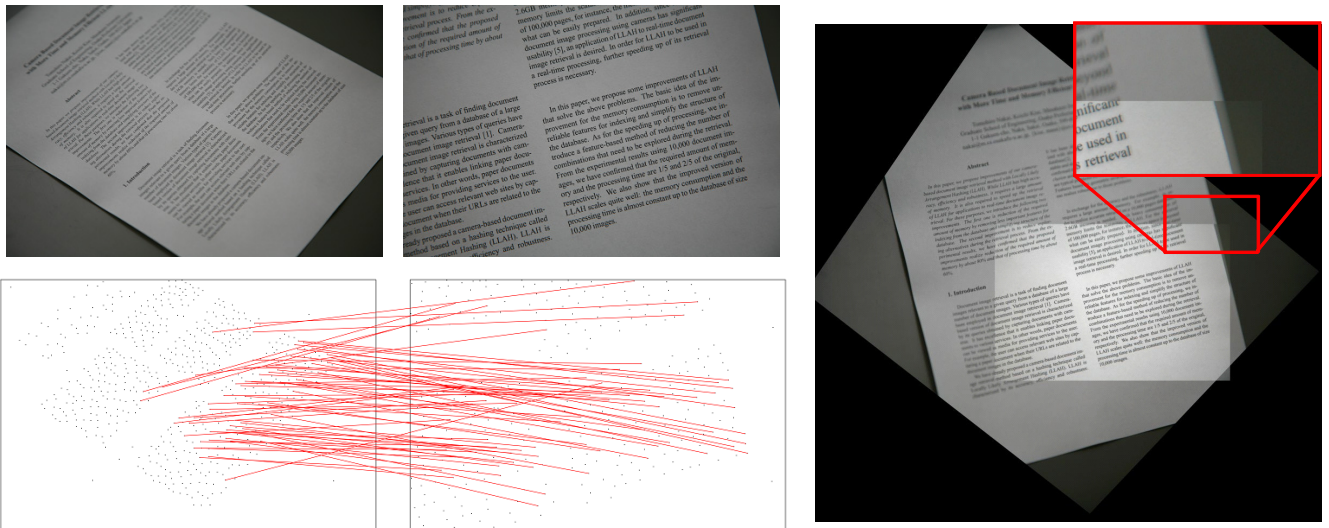


図6 実験結果3 左上:入力画像 左下:対応点 右:モザイク画像

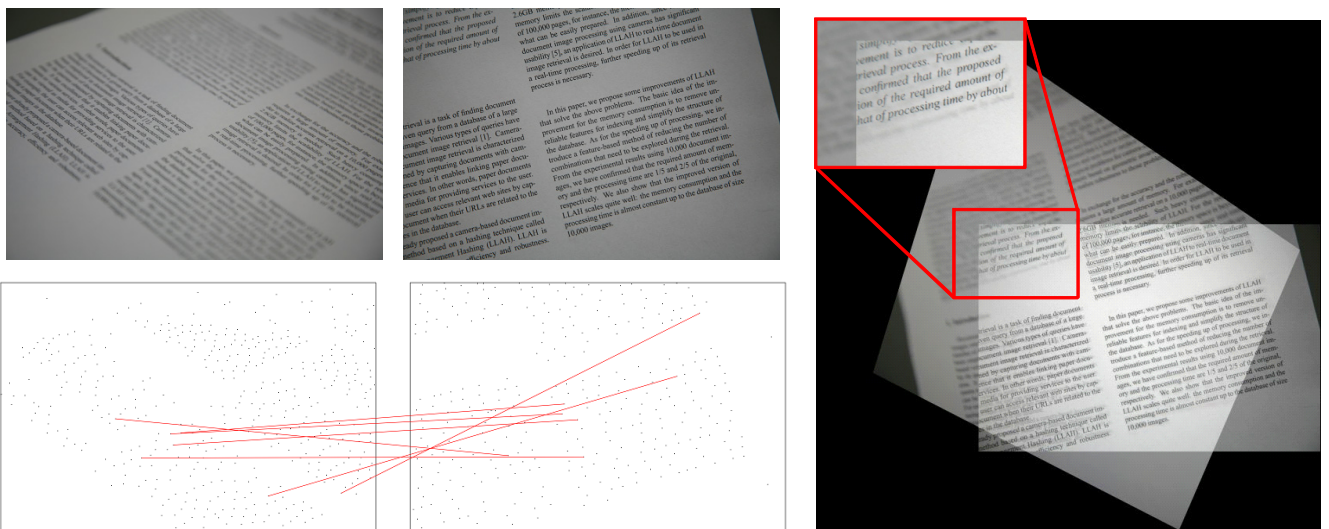


図7 実験結果4 左上:入力画像 左下:対応点 右:モザイク画像