

レイアウトに依らない平面文書画像の射影歪み補正

丹羽 亮[†] 岩村 雅一[†] 黄瀬 浩一[†] 内田 誠一^{††} 大町真一郎^{†††}

[†] 大阪府立大学大学院工学研究科 〒 599-8531 堺市中区学園町 1-1

^{††} 九州大学大学院システム情報科学研究所 〒 819-0395 福岡市西区元岡 744

^{†††} 東北大学大学院工学研究科 〒 980-8579 仙台市青葉区荒巻字青葉 6-6-05

E-mail: [†]niwa@m.cs.osakafu-u.ac.jp, ^{††}{masa,kise}@cs.osakafu-u.ac.jp, ^{†††}uchida@is.kyushu-u.ac.jp,
^{††††}machi@ecei.tohoku.ac.jp

あらまし 近年、デジタルカメラをスキャナの代わりに用いて文書画像処理を行う試みが盛んに研究されている。しかし、既存の文書画像処理技術の多くはスキャナで取得した文書画像を対象としているため、デジタルカメラで斜めから撮影した文書画像には適用できない。このような場合、文書画像に生じた射影歪みを補正し、スキャナで取得したかのような真正面の文書画像を得る必要がある。既存の方法だと、文書の枠の有無や文字行の平行性あるいは、ステレオカメラによる撮影やビデオカメラの使用などの制約がある。本稿では、文字の面積や変形に着目した自由なレイアウトの文書へ適用できる歪み補正を提案し、実験により手法の有効性を確認する。

キーワード 文書画像, 正規化, 射影歪み, アフィン歪み, 文字認識

Dewarping of Planar Document Image without Layout Constraints

Ryo NIWA[†], Masakazu IWAMURA[†], Koichi KISE[†], Seiichi UCHIDA^{††}, and Shinichiro
OMACHI^{†††}

[†] Graduate School of Engineering, Osaka Prefecture University
1-1 Gakuencho, Naka, Sakai, 599-8531 Japan

^{††} Faculty of Information Science and Electrical Engineering, Kyushu University
744 Motoooka, Nishi, Fukuoka, 819-0395 Japan

^{†††} Graduate School of Engineering, Tohoku University
6-6-05 Aoba, Aramaki, Aoba, Sendai, 980-8579 Japan

E-mail: [†]niwa@m.cs.osakafu-u.ac.jp, ^{††}{masa,kise}@cs.osakafu-u.ac.jp, ^{†††}uchida@is.kyushu-u.ac.jp,
^{††††}machi@ecei.tohoku.ac.jp

Abstract For user convenience, processing of document images captured by a digital camera has been attracted much attention. However, most existing processing methods require an upright image such like captured by a scanner. Therefore, we have to cancel perspective distortion of a camera-captured image before processing. Although there are rectification methods of the distortion, most of them work under certain assumptions on the layout; the borders of a document are available, textlines are in parallel, a stereo camera or a video image is required and so on. In this paper, we propose a layout-free rectification method which requires none of the above assumptions. We confirm the effectiveness of the proposed method by experiments.

Key words document image, normalization, perspective distortion, affine distortion, character recognition

1. はじめに

近年、文字認識や文書画像処理にデジタルカメラを用いる試みが盛んに行われている [1], [2]. これは、文書画像の取得にスキャナを用いず、代わりにデジタルカメラを用いるものである。

デジタルカメラを使用する利点は文書画像取得の可搬性や簡便性である。スキャナとは違い、デジタルカメラは手軽に持ち歩くことができるため、思い立ったときに手軽に撮影することができる。また、ポスターや看板など、大きな物や移動できない物を撮影することができるなど、これまでになかった新しい使用形

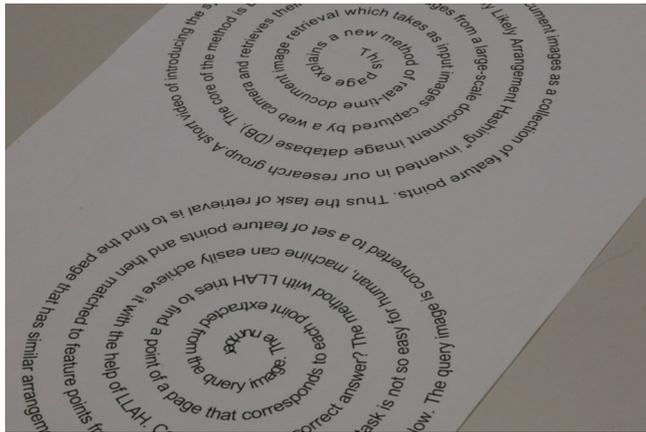


図1 既存の方法では補正が難しい平面文書画像

態に発展する可能性を秘めている。

このようにデジタルカメラを用いた文書画像処理は優れたアプリケーションを生み出す可能性がある反面、その実現は容易ではない。その理由は、既存の文書画像処理技術の多くがスキャナで取得した文書画像を対象としていることである。つまり、デジタルカメラで取得した文書画像に既存の技術を適用するためには、デジタルカメラで取得した文書画像を補正し、スキャナで取得したかのような画像を得る必要がある。本稿では、デジタルカメラを用いることによって生じる文書画像の劣化のうち、射影歪みに焦点を当て、文書画像の射影歪み補正を行う。

射影歪みを補正する既存の研究には、(1) 文書の枠が長方形になる変換を施す手法[3]、(2) 文字行の平行性を利用する手法[3]、(3) 複数のカメラ[4]もしくは動画[5]を用いる手法などがある。(1)と(2)の方法では、それぞれ画像中に文書の枠が含まれている必要や、文字行が平行である必要があり、図1のような複雑なレイアウトを補正できない。また、(3)の方法はステレオを用いるもので、1枚の静止画からは復元できない。

この問題に対して我々は、変量と不変量を用いて射影歪みからアフィン歪みに補正する手法を提案している[6]~[8]。この手法の特徴は、文書の傾きにに応じて値が変化する「変量」と変化しない「不変量」を組み合わせて用いることである。変量と不変量はレイアウトに関係なく求めることができるためレイアウトの制約を受けず、1枚の静止画のみを用いて補正することができる。

しかし、文献[6]はアフィン歪みが残りさらに精度も良くなかった。また、文献[7]は事前知識としてフォントの学習を行う方法で、アフィン歪みを補正しているが、図1のような複雑なレイアウトに対応していなかった。文献[8]と本稿では、大田らの手法[9]を応用して消失線を求め、文書画像の歪みをより精度良くアフィン歪みまで戻す方法を提案する。また、複雑なレイアウトでも正対する改良を加えている。文献[8]では実験データ数が少なかったため、本稿ではデータ数を増やし、提案手法の有効性を検証する。

2. 提案手法

提案手法では次の仮定に基づき、射影歪みを受けた平面文書画

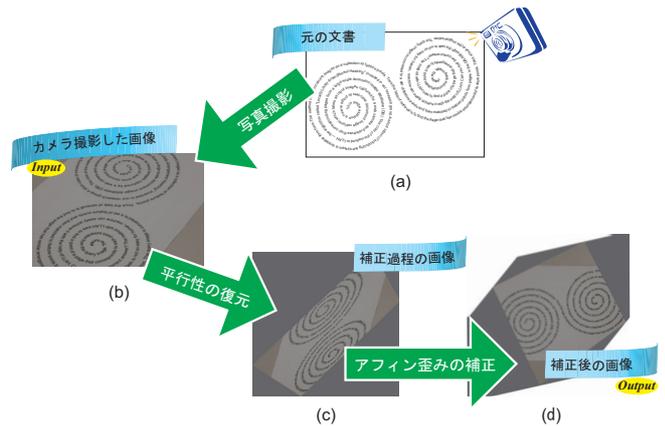


図2 提案手法の概略 (a) 元の文書 (b) 射影歪みをもつ画像 (c) アフィン歪みをもつ画像 (d) 正対する画像

像を補正する。

(1) 同じ字種は同じ大きさである

(2) 使用している文字が、2.4で説明する文献[10],[11]で認識できる

補正の流れを、図2で示す。図2(a)の文書を撮影した際に、図2(b)のように射影歪みをもつ画像を取得する。これを補正するため、提案手法では仮定1の下、2.1から2.3の手順で、同じ字種の面積を等しくする射影変換を行い画像を補正する。ただし、この仮定は、面積比を保存する変換であるアフィン変換に対して何の制約も持たない。従って、補正後の画像にはアフィン歪みが残されている(図2(c))。アフィン変換を受けても“平行線は平行線に変換される”ので、ここまでの手順を「平行性の復元」と呼ぶ。次に、画像に残されたアフィン歪みを補正する。そのため提案手法では仮定2の下、2.4の手順を用いる。すなわち幾つかのフォントを学習し、画像中の文字の変形から、画像全体のアフィン歪みを補正するアフィン変換を計算する。そして、変換によってアフィン歪みが無くなり、元の文書と同様に正対する画像を得る。

2.1 消失線の推定

デジタルカメラで撮像した文書画像中の各文字に注目すると、本来同じ字種であり同じ大きさで印刷されていても、位置によって大きさが変化している。例として、画像中からアルファベットのaを2つ選び注目する。すると、文字の黒画素の面積は射影歪みによって、カメラから近い方が大きく、遠い方が小さくなる。それぞれを面積 $S_1, S_2 (S_1 > S_2)$ で表し、図3(c)のように消失点を推定する。両文字の位置から消失点までの距離 f_1, f_2 は、大田らの手法[9]に従い、

$$\frac{f_2}{f_1} = \frac{S_2^{1/3}}{S_1^{1/3}} \quad (1)$$

とする。

2.2 クラスタリング

2.1の手法は同じ字種の文字から消失点の位置を計算している。他に、本来面積が同じである字種(フォントに依るが鏡像の文字bとdなど)の文字を利用することもできる。このような字種の間を以後“等価”と表現する。しかし、aとbのよ

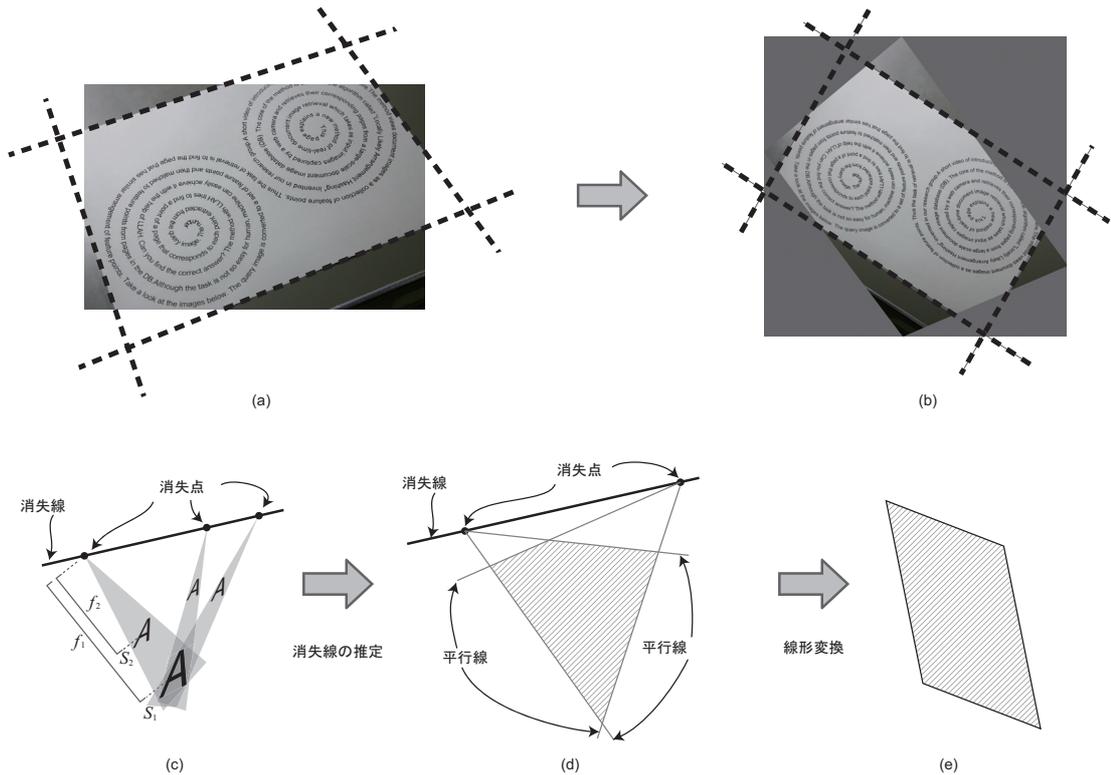


図3 アフィン歪みを持つ画像の取得までの概略 (a) 元画像 (b) アフィン歪みを持つ画像 (c) 文字の面積と消失線の推定 (d) 2組の平行線の交差と消失点 (e) 射影変換を用いたアフィン歪みを持つ画像の取得

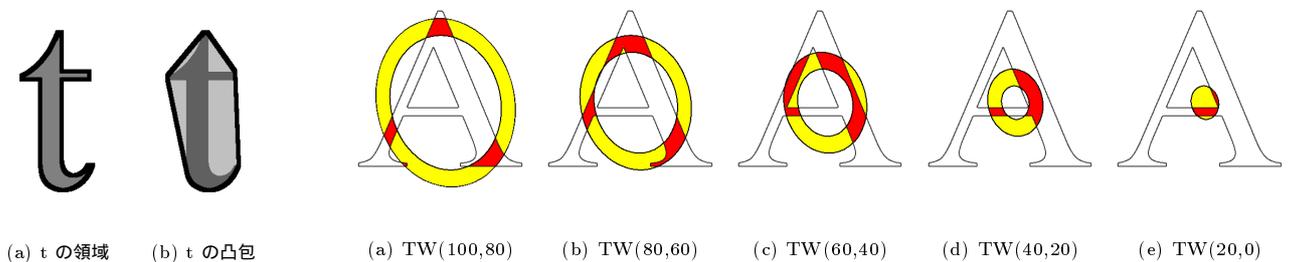


図4 文字と凸包

図5 TW値を計算する領域

うな等価でない字種間では消失点の位置を計算できない。文書中には様々な字種が混在しているため、等価な字種のグループに分ける必要がある。

字種を判別する方法として、文字認識が考えられるが、射影歪みを受けている場合は処理が難しい。また、ここでは字種を分類して奥行きを計算することが目的なので、文字認識のように文字にラベルを付ける必要はない。そこで、アフィン変換を受けても変化しない量(アフィン不変量)である面積比を用いた分類を行う。何故なら、アフィン不変量は射影変換に対して不変ではないが、文字領域のように面積の小さな領域において射影変換をアフィン変換に近似できることからである。提案手法では、アフィン不変量として「凸包との面積比」と「TW値」を用いる。

凸包との面積比は、対象図形である文字成分(図4(a))とそ

の凸包(図4(b))との面積比である。このとき凸包はアフィン変換に不変に求まり、この面積比はアフィン不変量となる。

TW値とは、図5のように2つの楕円に囲まれた領域から計算する値である。まず対象図形と同じ共分散行列を持つ楕円を描き、これをA-楕円と呼ぶ。そして、中心が同じA-楕円のm%の大きさの楕円とn%の大きさの楕円を用意する(m>n)。文字領域のうち、m%の楕円とn%の楕円で区切られる面積の割合をTW(m,n)と表す。TW値を式で表すと、

$$TW(m,n) = \frac{m\% \text{ と } n\% \text{ の楕円の間に含まれる領域の面積}}{\text{図形の面積}}$$

である。このように得られるTW値がアフィン不変量であることは文献[7]で示されている。

以上の不変量を用いてクラスタリングを行って、等価な文字ごとに分類する。本稿ではk-means法を用い、50クラスタに

分類する．

2.3 消失線のロバスト推定

2.1の方法を各クラスに適用して得られる消失点は直線の上に並ぶはずである．しかし，データの面積測定に誤差があることや，クラスタリングのときにノイズが混入している可能性があることから，実際にはばらつきを持つ．ばらつきの影響を受けないために，クラスタごとに消失線を推定する．推定にはRANSAC [12]を用い，ランダムに2点選んで引いた直線から，距離100pxの範囲により多く消失点を内包する線を選ぶ．50クラスタでそれぞれ1本ずつ推定しているので，消失線の候補が50本できる．その中から1本代表を選ぶため，角度の平均のロバスト推定 [13]を用いる．角度の平均 μ は以下のようにして求めることができる． x_1, \dots, x_n を n 個の消失線候補の傾きを角度で表したものとして，

$$\mu = \arg \min_{\mu} \sum_{i=1}^n \rho(t(x_i^T \mu; \kappa)), \quad (2)$$

ただし，

$$t(u; k) = \text{sign}(u) 2\kappa(1 - |u|)^{1/2} \quad (3)$$

$$\rho(t) = \begin{cases} t^2/2, & |t| \leq c \\ c|t| - c^2/2, & |t| \geq c \end{cases} \quad (4)$$

であり，パラメータは実験的に $\kappa = 20$ ， $c = 2$ と決めた．本稿では，式 (2) の μ を全消失線候補の傾きを代入し，最も総和が小さくなったとき，その角度の直線を消失線として選ぶ．

この消失線を用いて，文書画像を補正する．射影歪みを受ける前，平行であった2本の直線（以後単に平行線と呼ぶ）は，必ず消失点で交差し，逆に消失点で交差する2直線は本来平行である．消失線を得ることで，消失点を任意に選ぶことができるので，平行線を画像平面上に自由に引くことができる．2本ずつの平行線を2組引くと，図 3(d) のように4角形の領域ができる．この領域を図 3(e) のような平行四辺形に戻す変換を画像に与えることで，図 3(b) のように平行性の復元ができる．

2.4 アフィン変換パラメータと復元

アフィン歪みを持つ文書画像を補正して，正対する文書画像を得るため，本稿ではアフィン不変な文字認識手法 [10], [11] を用いる．この手法は，アフィン変換を受けた図形の正規化による認識手法を文字認識に応用したものであり，文字認識の副産物として，アフィン歪みを受けた文字を元の姿に戻すアフィン変換を得ることができる．これを図 6 に示す．図中の T_1, T_2 が正規化である．これは，本稿の 2.2 で用いた A-楕円を円に直すアフィン変換である．また， T_3 は回転を表しており， $T_2^{-1}T_3T_1$ が文字の姿を戻す変換である．

入力画像中の文字が全て同じ向きに揃っていれば，常に同じアフィン変換で正面を向くはずである．しかし，実際に変換を求めるとき，アフィン歪み以外の歪みの影響（平行性の復元誤差や紙面の反りなど）がある可能性や，計測に誤差が出ること，また，認識に失敗することもある．従って，画像中の全ての文字から共通のアフィン変換を得ることはできない．そこで，最

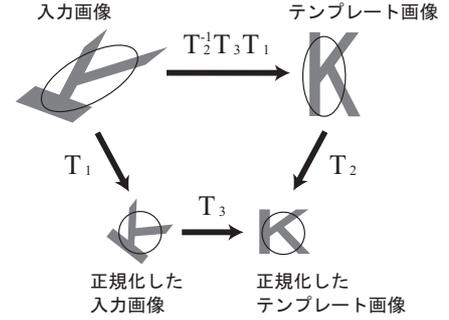


図 6 入力画像とテンプレート画像の正規化

も信頼できる変換を，全ての文字から得られた変換の中から選ぶ．すなわち，アフィン変換行列 $P = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ を，

$$P = L(\beta)R(\theta)S(\varphi)A(\alpha) \quad (5)$$

$$= \begin{pmatrix} \beta & 0 \\ 0 & \beta \end{pmatrix} \begin{pmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & \tan \varphi \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \alpha & 0 \\ 0 & 1/\alpha \end{pmatrix}$$

$$\alpha = \pm \sqrt{\frac{a^2 + c^2}{ad - bc}} \quad (6)$$

$$\varphi = \tan^{-1} \frac{ab + cd}{ad - bc} \quad (7)$$

$$\theta = \cos^{-1} \frac{\pm a}{\sqrt{a^2 + c^2}} \quad (8)$$

$$\beta = \pm \sqrt{ad - bc} \quad (9)$$

と分解して， α, φ の2パラメータだけを取り出し，投票処理を行う．特に誤認識が起きにくい字種をあらかじめ選んでおき，画像中の該当文字に対して2パラメータを求め，2次元空間に配置する．そして， k -近傍法による密度推定 [14] を利用して，密度の高い点を選ぶ．さらに，選んだ代表点から k 番目に近い点までの重心を選び，その点をより正確な代表点として利用する．本稿では $k = 20$ とした．

このようにして得た2パラメータから式を利用してアフィン変換行列を求めることができ，アフィン歪みを持つ文書画像を正対させることができる．

3. 実験

本稿で行った改良の有効性を実験により検証する．平行性の復元とアフィン歪みの補正の2段階に分けて実験し，評価を行う．実験に用いた文書は，(A) 文字の割合が多い英文雑誌の電子データ55ページと(B) 渦状にアルファベットが並んだ文書など複雑なレイアウトの文書3ページである．これらの文書に対し(1)コンピュータ上で射影変換した結果，(2)文書を印刷しカメラで撮影した結果を調べる．以上の組合せで4つのデータセットが用意でき，以後それぞれを(A1)，(A2)，(B1)，(B2)とよぶ．

(A2)，(B2)は用いたカメラやレンズ，さらに距離や角度を変えて，全ての文書に対し4パターンの撮影を行った．撮影に用いたカメラはCanon EOS Digital 5D と Fujifilm Finepix F710 である．また，一眼レフカメラのレンズはEF24-105mm

F4L IS USM と 100mm F2.8 マクロ USM を利用した .

pp.2055-2064 , Sept. 2006 .

(A1) と (B1) は文書の変形パラメータを 4 通り変えて実験する . 変形後の本来の文書の領域を計算できるため , これを利用して補正精度を評価する . (A2) は実験に用いる文書にあらかじめ , 21mm 間隔で縦横に 9×12 個 , 格子状に小さな印を印刷しておく . この印を用いて , 本来長方形であった領域のうち , 取り得る最大のものを利用して評価する . (B2) もほとんど同じで , あらかじめ等間隔に評価用の点を 4~6 個印刷してある .

具体的には , 平行性の復元とアフィン歪みの補正について実験を行った . 処理例を図 7 , 8 に示す . 結果の詳細は講演に譲るが , 現段階では , 平行性の復元については誤差が 2 度程度 , アフィン歪みの補正については , 誤差が 6 度程度にまで補正できることが分かっている .

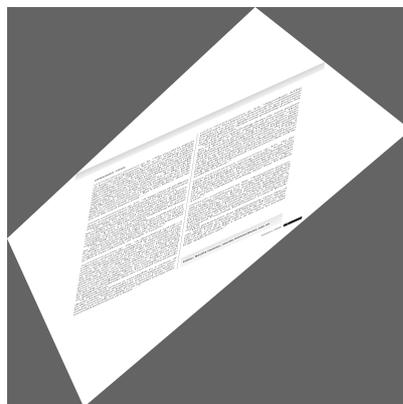
謝辞 本研究の一部は , 財団法人 大川情報通信基金の補助による .

文 献

- [1] J. Liang, D. Doermann and H. Li, "Camera-based analysis of text and documents: a survey," IJDAR, vol.7, pp.84-104, 2005.
- [2] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, "カメラを用いた文字認識・文書画像解析の現状と課題," 信学技報 PRMU2004-246, March 2005 .
- [3] P. Clark and M. Mirmehdi, "Recognising text in real scenes," IJDAR, vol.4, pp.243-257, 2002.
- [4] C. H. Lampert, T. Braun, A. Ulges, D. Keysers and T. M. Breuel, "Oblivious document capture and real-time retrieval," Proc. CBDAR2005, pp.79-86, Aug. 2005.
- [5] 池谷彰彦, 佐藤智和, 池田聖, 神原誠之, 中島昇, 横矢直和, "カメラパラメータ推定による紙面を対象とした超解像ビデオモザイクング," 信学論 (D), vol.J88-D, no.8, pp.1490-1498, Aug. 2005 .
- [6] 丹羽亮, 岩村雅一, 黄瀬浩一, 内田誠一, 大町真一郎, "変量と不変量に基づく射影歪みからアフィン歪みへの補正," MIRU2007 論文集, pp.720-725, July 2007 .
- [7] 丹羽亮, 堀松晃, 岩村雅一, 黄瀬浩一, 内田誠一, 大町真一郎, "変量と不変量と文字の正規化に基づく文書画像の射影歪み補正," MIRU2008 論文集, pp.1260-1265, July 2008 .
- [8] M. Iwamura, R. Niwa, A. Horimatsu, K. Kise, S. Uchida and S. Omachi, "Layout-free dewarping of planar document images," Proc. Document Recognition and Retrieval XVI (Proc. SPIE Volume 7247), Jan. 2009.
- [9] 大田友一, 前信潔, 坂井利之, "透視射影されたテクスチャから平面の傾きを得る一方法," 情報処理学会研究報告, 1981-CVIM-16, pp.1-8, Jan. 1982 .
- [10] 堀松晃, 丹羽亮, 岩村雅一, 黄瀬浩一, 内田誠一, 大町真一郎, "アフィン不変な文字認識手法とその高速化," MIRU2008, pp.1450-1455, July 2008 .
- [11] A. Horimatsu, R. Niwa, M. Iwamura, K. Kise, S. Uchida and S. Omachi, "Affine invariant recognition of characters by progressive pruning," Proc. DAS2008, pp.237-244, Sept. 2008.
- [12] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," Commun. ACM, vol.24, no.6, pp.381-395, 1981.
- [13] K. V. Mardia and P. E. Jupp, Directional Statistics, John Wiley & Sons, Inc., 1991.
- [14] K. Fukunaga, Introduction to statistical pattern recognition, Academic Press, 2nd edition, 1990.
- [15] 石田皓之, 高橋友和, 井手一郎, 目加田慶人, 村瀬洋, "携帯カメラ入力型文字認識におけるぼけやぶれに対処するための生成型学習法," 電子情報通信学会論文誌 D, vol.89, no.9,



(a) 入力画像

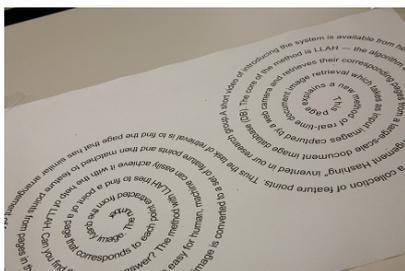


(b) 平行性を復元した画像



(c) 正対画像

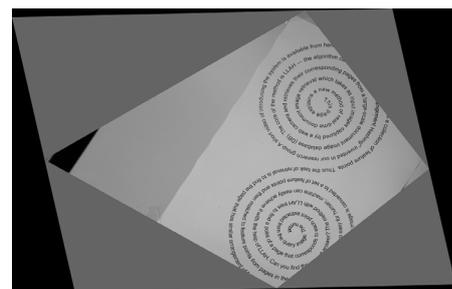
図 7 データ A1 の実験結果



(a) 入力画像



(b) 平行性を復元した画像



(c) 正対画像

図 8 データ B2 の実験結果