

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

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 利用者の利便性向上の観点から、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。本稿では、紙面の傾きに依りて値が変化する「変量」と変化しない「不変量」を組み合わせた文書画像の傾き補正手法を提案する。提案手法は、従来手法のように撮影方法やレイアウトに強い制約を用いないため、これまで補正不可能であった文書画像を復元することができる。実験により、射影歪みを受けた文書画像をアフィン歪み程度に復元できることを確認した。

キーワード 文書画像, 変量, 不変量, 射影歪み, アフィン歪み

Rectifying Perspective Distortion into Affine Distortion Using Variants and Invariants

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 Motooka, 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, document image processing (DIP) captured with a digital camera instead of a scanner has been researched. However, existing methods of DIP are not usable for a perspective document image captured by a digital camera since most of them are designed for the one captured by a scanner. Thus, we have to rectify the perspective of the document image and obtain the frontal image as if it was captured by a scanner. In this paper, we propose a new rectification method of a document image introducing *variants* which change according to the gradient of the paper and *invariants* which do not change against it. Since the proposed method does not use strong assumptions, it is widely applicable to many document images unlike other methods. We confirmed the proposed method rectifies a document image suffering from perspective distortion and acquires the one with affine distortion.

Key words document image, variants, invariants, perspective distortion, affine distortion

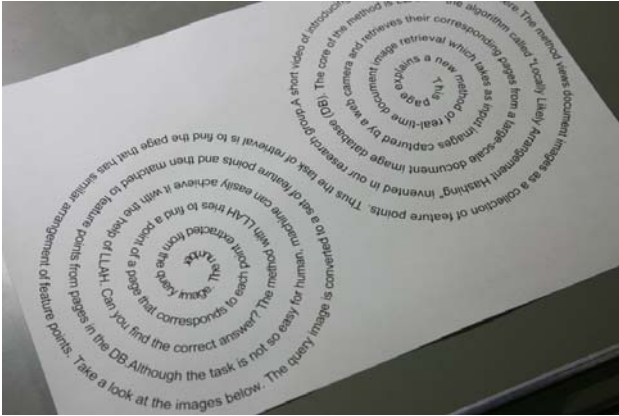


図 1 従来手法では補正困難な文書画像

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。デジタルカメラを用いる処理では、スキャナでは起こり得ない非一様な照明やフォーカスのずれ、対象を斜めから撮影したときに生じる射影歪みなどによって画像が劣化するため、一般にスキャナを用いた処理に比べて難しい。しかし、それにも関わらずスキャナに替えてデジタルカメラを用いる理由は、その可搬性や簡便性にある。例えば、スキャナの設置には時間と手間を要するため、手軽に移動したり、持ち歩いたりする用途には適さない。また、ポスターや看板など、大きな物や移動できない物には利用できない。一方、デジタルカメラであれば、思い立ったときに手軽に撮影することができ、これまでにない新しい使用形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にはデジタルカメラで取得した文書画像を変換・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法を提案する。すなわち、本稿で扱う問題は「デジタルカメラで斜めから撮影した文書画像から、スキャナで取得したような正対した文書画像をいかにして得るのか」である。提案手法の特徴は、文書の傾きに応じて値が変化する「変量」と変化しない「不変量」を組み合わせることであり、具体的には「変量」として連結成分の面積、「不変量」として面積の比を用いる。これらの値はどのような文書からも計算できる基本的な量であり、簡便に計算できるため、他の手法のように文書の長方形の枠が完全に写っていることや、文書中の文字が直線上に並んでいるなどのレイアウトに関する強い制約を課さない。そのため、図 1 に示す特異なレイアウトの文書を始め、幅広い対象に適用することが可能である。既存の手法との差異は 4. で述べる。

2. 提案手法

本節では、変量と不変量を用いた射影変換補正法について述

べる。提案手法の説明のために、最初にコンピュータビジョンにおける 3 次元座標と 2 次元座標の関係について述べ、その後提案手法の詳細について述べる。

2.1 カメラ座標系と画像座標系 [3], [4]

3 次元空間の物体をカメラで撮影したとき、どのように 2 次元画像が得られるかを考える。通常、カメラのモデルに図 2(a) のようなピンホールカメラを用いる。点 C がピンホールで、物体から来る光は全てピンホールを通り、画像平面 I に像を結ぶ。点 C を焦点と呼ぶ。また、画像平面 I と焦点 C の距離 f を焦点距離、焦点 C を通り画像平面に垂直な直線を光軸と呼び、光軸と画像平面の交点を画像中心 c とおく。このモデルでは平行な 2 線は必ずしも平行線に変換されない。このような変換を射影変換、射影変換による歪みを射影歪みと呼ぶ。また射影変換のうち、平行線を変換したとき平行線であるものをアフィン変換、アフィン変換を受けたときの歪みをアフィン歪みと呼ぶ。

コンピュータビジョンでは一般に、画像平面を図 2(b) のように配置し直して用いる。画像平面で画像中心 c を原点とし、図 2(b) のように x 軸と y 軸を持つ座標系を画像座標系と呼ぶ。また、焦点 C を原点、光軸を Z 軸とし、画像座標系の x 軸、 y 軸に対応する方向に X 軸、 Y 軸を持つ 3 次元座標系をカメラ座標系と呼ぶ。カメラ座標系の点 $(X, Y, Z)^T$ を画像平面に射影したとき、画像平面の座標系でこれに対応する点 $(x, y)^T$ は

$$\begin{pmatrix} x \\ y \end{pmatrix} = \frac{f}{Z} \begin{pmatrix} X \\ Y \end{pmatrix} \quad (1)$$

で求まる。

2.2 黒画素の面積と奥行き

文書画像中の各文字に注目すると、本来同じ字種であっても位置によって大きさが変化している。例えば、画像中からアルファベットの a のような特定の文字だけを取り出したときを考える。このとき、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。この面積の変化から文書の奥行き情報を求めることができるので、本研究ではこれを基に文書画像の補正を行う。

奥行きと文字の黒画素の面積の関係について求める。紙面上の同じ種類の文字を選んだ、中心の Z 座標がそれぞれ Z_1, Z_2 であるとする。問題を簡単にするため、傾いた紙面をカメラから見たときの模式図 (図 3(a)) を用いて、奥行きと文字の長さの関係を考える。取り出した文字の本来の長さを L とすると、このとき $Z_1 \neq Z_2$ なので、画像から得られる文字の射影の長さ l_1, l_2 は異なっている。それぞれの長さを簡単に求める図 3(b) のような近似を利用する。つまり、それぞれの中心に画像平面と平行な平面をおき、文字の正射影を得る。紙面と画像平面のなす角を α 、正射影の長さを L' とすると、

$$L' = L \cos \alpha \quad (2)$$

である。従って、式 (1) の x 座標についてのみ考えれば、

$$l_j = \frac{f}{Z_j} L' = \frac{f}{Z_j} L \cos \alpha \quad (j = 1, 2) \quad (3)$$

が得られる。

次に、これらの文字の面積に関しても同様に考える。それぞれの面積が S であり、正射影の面積を S' とすると、

$$S' = S \cos \alpha \quad (4)$$

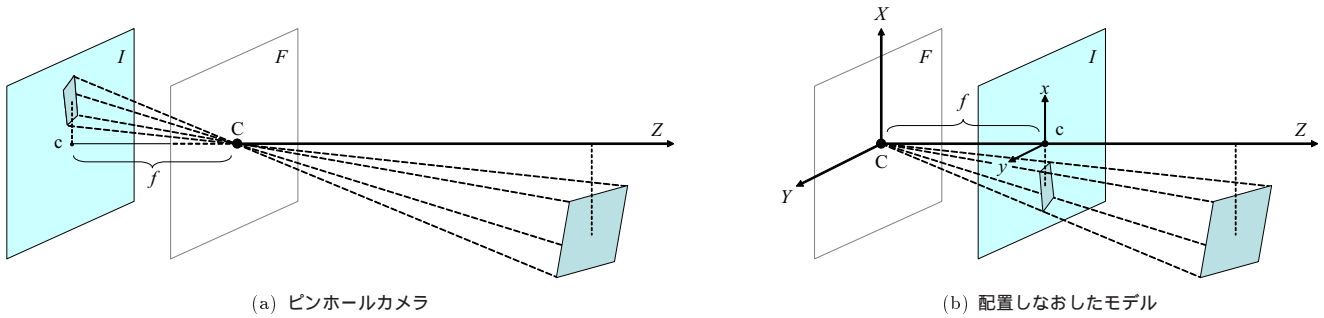


図 2 カメラによる結像

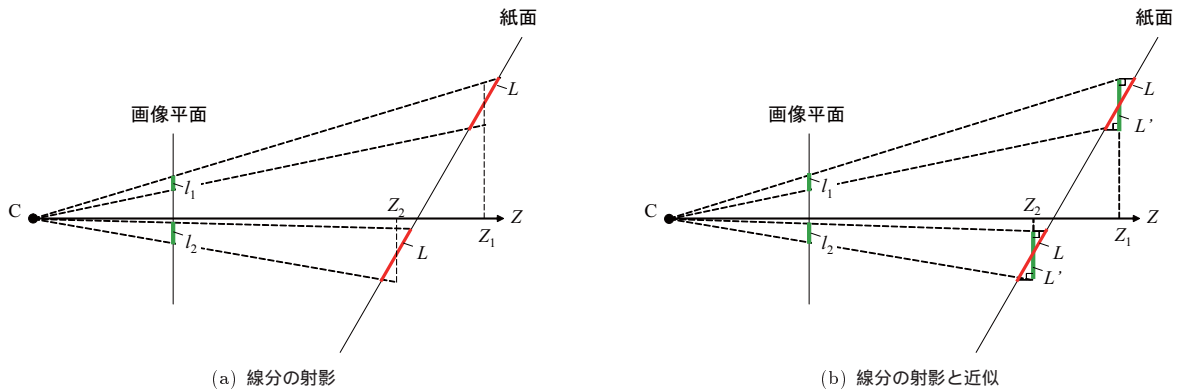


図 3 座標系の対応

となる．このとき，画像平面への写像の面積 s_j は，

$$s_j = \left(\frac{f}{Z_j}\right)^2 S' = \left(\frac{f}{Z_j}\right)^2 S \cos \alpha \quad (j = 1, 2) \quad (5)$$

である．

式 (5) より画像平面の面積は本来の面積に対して，奥行きの 2 乗に反比例していることがわかる．従って，ある字種の黒画素の面積に注目したとき， j 番目の文字の奥行き Z_j を黒画素の面積 s_j で表すと，

$$Z_j = \frac{f\sqrt{S \cos \alpha}}{\sqrt{s_j}} \quad (6)$$

となる．また，焦点距離 f ，紙面と画像平面のなす角 α は撮影時に決まっているので，それぞれ定数である．

次に各文字の奥行きからカメラ座標系における紙面の傾きを考える．式 (1) より j 番目の文字の座標 $(X_j, Y_j, Z_j)^T$ は，

$$\begin{pmatrix} X_j \\ Y_j \\ Z_j \end{pmatrix} = \begin{pmatrix} Z_j x_j / f \\ Z_j y_j / f \\ Z_j \end{pmatrix} \quad (7)$$

である．ここで， Z を $1/f$ 倍した Z' で置き換えた座標系で各座標を表現すると，

$$\begin{pmatrix} X_j \\ Y_j \\ Z'_j \end{pmatrix} = \begin{pmatrix} Z'_j x_j \\ Z'_j y_j \\ Z'_j \end{pmatrix} \quad (8)$$

となり，形の上では未知の定数 f を消去できる．画像中の文字は本来同一平面上にあるので，式 (7) を用いて各文字の 3 次元座標を計算し， $Z' = aX + bY + c$ で表される平面に当てはめることで紙面の傾きが推定できる．詳細は 2.4 で述べる．

2.3 面積比によるクラスタリング

2.2 で述べた，面積から奥行きを推定する方法は文書中に文字が一種類しか存在しない場合のみ使用できる．しかし，実際の文書中には複数の字種が混在しているため，あらかじめ文字をその種類ごとに分けておく必要がある．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できればよく，文字認識のように文字にラベルを付ける必要はない．

そこで提案手法では，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考える．すなわち，文字から 2 つの領域を得たとき，その 2 領域の面積比から不変量を得，これを用いて字種を分類する．アフィン不変量は射影変換に対して不変ではないが，局所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，文字領域のように面積の小さな領域の面積比を射影不変量のように扱うことができる．字種の判別に面積比が満たすべき条件は次の 2 つである．

(1) 面積比（面積）を計算する領域は射影変換を受けても同一でなければならない．

(2) 面積比は，字種を十分に判別できなくてはならない．(1) について，面積比は不変量であるため，同一の領域から（近似的に）同一の値を算出することができる．しかし，面積を計算する領域が異なれば同一の値を計算することができないため，射影変換に不変な領域抽出法が必要となる．本手法では，各領域の凸包が線形変換に対して不変であることを利用する．(2) については，異なる字種が偶然同一の面積比を持つ可能性がある．その場合，字種の混同が生じ，紙面の傾きを正しく推定できない．そこで複数の面積比を用い，面積比の判別性能を向上させる．異なる字種から計算した 1 種類の面積比が偶然近くなる確率に比べて，複数の面積比が偶然同時に近くなる確率は小

th

t

h



(a) t の黒画素の領域 (b) h の黒画素の領域 (c) t の凸包 (d) h の凸包 (e) t と h の凸包

図 4 画像から取り出した 2 文字の組

図 5 2 文字から取り出せる領域

さいからである．1 文字から計算できる面積の種類は限られているため，提案手法では最近傍の 2 文字を組み合わせ，2 文字から計算した面積を使用する．図 4 に示すように，2 文字を選んだとき（この場合は“t”と“h”），文字の黒画素領域と凸包領域から図 5 の 5 種類の領域が得られる．これらを組み合わせることで，面積比を複数作り出すことができる．得られた m 個の面積比は m 次元の不変量ベクトルとして用いる．

文書中から取り出された文字の集合（以降クラスタと呼ぶ）を，面積比の近い文字の部分集合に分けることを考える．本稿ではクラスタリングに k -means 法を用いる．クラスタリングにより得られたそれぞれのクラスタには同じ字種の組が含まれると期待される．以降の処理は 2 文字毎に行う．

ここでクラスタリングに関して，提案手法の文字の大きさに関する制約を述べておく．2.2 で述べた，面積と奥行きの関係は，本来の文字の大きさが同じであることを仮定している．しかし，面積比を用いたクラスタリングでは，字形が同じで大きさのみ異なる文字は区別できない．そのため，クラスタ内に異なる大きさの文字が存在すると，文字の大きさの違いが奥行きの変化によるものなのか，本来の大きさが異なることに起因するものなのかの区別ができず，奥行き情報の外乱要因となる．ただし，通常の文書のように，ほとんどの文字が同じ大きさで，見出し部分のような一部分のみが大きな文字である程度であれば，後述するノイズ除去処理で棄却することができるため，問題ないと考えられる．

2.4 平面への当てはめ

2.3 のようにクラスタリングを行うと，各クラスタにおいて 2.2 で述べた平面への当てはめを考えることができる．紙面の傾きを高精度に推定するためには同一字種が文書中で分散していることが望ましいが，このような状況は必ずしも期待できない．そこで，クラスタ（字種）毎に推定した紙面の傾き情報の統合を考える．その際に問題になるのは黒画素の面積 S である．2.2 の議論では S が既知であるとしていたが，実際には未知であり，文字毎に異なる．そこで，それぞれの平面の傾きが等しくなるように，クラスタ毎の黒画素の相対的な面積の比も同時に推定することにする．以下でその詳細を述べるが，以降の議論では 2.2 における S ， $(X_{ij}, Y_{ij}, Z_{ij})^T$ ， Z'_{ij} にクラスタ番号 i を付し， S_i ， $(X_{ij}, Y_{ij}, Z_{ij})^T$ ， Z'_{ij} としてそれぞれ用いる．式 (6)～式 (8) も同様である．

まず，式 (8) は式 (6) を利用して，

$$\begin{pmatrix} X_{ij} \\ Y_{ij} \\ Z'_{ij} \end{pmatrix} = \begin{pmatrix} Z_{ij}x_{ij}/f \\ Z_{ij}y_{ij}/f \\ Z_{ij}/f \end{pmatrix} = \sqrt{S_i \cos \alpha} \begin{pmatrix} x_{ij}/\sqrt{s_{ij}} \\ y_{ij}/\sqrt{s_{ij}} \\ 1/\sqrt{s_{ij}} \end{pmatrix} \quad (9)$$

と表すことができる．この式は，各文字の座標 $(X_{ij}, Y_{ij}, Z'_{ij})$ を画像から得られる x_{ij} ， y_{ij} ， s_{ij} を用いて計算できることを

意味している．しかし，式 (9) 中の文字の本来の面積を表す S_i と，紙面の傾き α が未知のため， $K_i = \sqrt{S_i \cos \alpha}$ とおき，平面に対する各文字の奥行きの誤差を

$$\begin{aligned} \varepsilon_{ij} &\equiv \left| \{aX_{ij} + bY_{ij} + c\} - Z'_{ij} \right| \\ &= \left| \left\{ a \frac{K_i x_{ij}}{\sqrt{s_{ij}}} + b \frac{K_i y_{ij}}{\sqrt{s_{ij}}} + c \right\} - \frac{K_i}{\sqrt{s_{ij}}} \right| \end{aligned} \quad (10)$$

と定義する．そして，全ての文字の誤差の和

$$E = \sum_i \sum_j \varepsilon_{ij} \quad (11)$$

を最小にする $\{K_i\}$ と平面のパラメータ a, b, c を求める．ここで， $\{K_i\}$ には定数倍の任意性があり，各パラメータが一意に定まらないため，本稿では $c = 1$ に固定して $\{K_i\}$ ， a, b を求める．

ただし，平面への当てはめを行う際にはノイズ（外れ値）の影響を考慮する必要がある．ノイズの原因としては，画像中から文字を抽出する際の画像処理の失敗やクラスタリングにおける誤分類，更には 2.3 で述べた，文書中に同一字種で複数の大きさの文字が存在する場合が考えられる．これらのノイズに対応するために，提案手法では 2 種類のノイズ除去を行う．

2 種類のノイズ除去は，(A) クラスタ内の外れ値を除去するもの，(B) クラスタ自体を除去するもの，である．(A) については，各クラスタごとに平面から各文字までの距離 ε_{ij} を計算し，距離が閾値 t_1 以上の文字を除去する．(B) については，要素数が少ない (t_2 以下である) クラスタを除去する．そのようなクラスタからは誤った平面が推定される可能性が高いと考えられるからである．

2.5 紙面の回転

最後に画像中の紙面を回転して，正面に向ける．これは紙面の正面に視点を移すことと等価であるので，傾いた紙面の法線ベクトルを求め，延長上に視点を移すことを考える．回転表現には，任意の回転を Z 軸周りの回転 ϕ ， Y 軸周りの回転 θ ， X 軸周りの回転 ψ の 3 段階の回転で表すロール・ピッチ・ヨー型の回転変換を利用する．ロール・ピッチ・ヨーを用いた回転行列 R は次式で表される．

$$\begin{aligned} R &= R(Z, \phi) R(Y, \theta) R(X, \psi) \\ &= \begin{pmatrix} \cos \phi & -\sin \phi & 0 \\ \sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \\ &\quad \begin{pmatrix} \cos \theta & 0 & -\sin \theta \\ 0 & 1 & 0 \\ \sin \theta & 0 & \cos \theta \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos \psi & -\sin \psi \\ 0 & \sin \psi & \cos \psi \end{pmatrix} \end{aligned} \quad (12)$$

紙面が $Z' = aX + bY + 1$ と表されているとき， $Z' = Z/f$

とおいたので、 $Z = afX + bfY + f$ である。その法線ベクトルは $(af \quad bf \quad -f)^T$ である。従って、この x 座標と y 座標を 0 にするように R を用いて回転変換すればよい。 Z 軸周りの回転を行わないことを考慮すると、 R の回転角は

$$\begin{pmatrix} \phi \\ \theta \\ \psi \end{pmatrix} = \begin{pmatrix} 0 \\ -\tan^{-1} \frac{af}{\sqrt{1+(bf)^2}} \\ -\tan^{-1}(bf) \end{pmatrix} \quad (13)$$

となる。式 (13) は角度の推定に未知パラメータ f が必要なことを示している。しかし、現段階では f を推定していないので、暫定的に $f = 1$ とした。この場合、射影歪み補正後に、本来は長方形である図形が平行四辺形になるアフィン歪みが残る可能性がある。

以下、回転後の 2 次元画像を求める。カメラ座標系の点 $(X_{ij}, Y_{ij}, Z_{ij})^T$ を回転して得られる点を $(\tilde{X}_{ij}, \tilde{Y}_{ij}, \tilde{Z}_{ij})^T$ とおけば、

$$\begin{pmatrix} \tilde{X}_{ij} \\ \tilde{Y}_{ij} \\ \tilde{Z}_{ij} \end{pmatrix} = R \begin{pmatrix} X_{ij} \\ Y_{ij} \\ Z_{ij} \end{pmatrix} \quad (14)$$

となる。さらにこれを画像平面に射影することによって、回転による 2 次元画像の座標

$$\begin{pmatrix} \tilde{x}_{ij} \\ \tilde{y}_{ij} \end{pmatrix} = \frac{f}{\tilde{Z}_{ij}} \begin{pmatrix} \tilde{X}_{ij} \\ \tilde{Y}_{ij} \end{pmatrix} \quad (15)$$

を得ることができる。

3. 実験

提案手法の有効性を検証するために実験を行った。実験データには Canon EOS 5D で撮影した、サイズが $4,368 \times 2,912$ の画像を使用した。提案手法は文書の枠を必要としないが、今回の実験では手法の効果がわかりやすいように、枠が写っている画像を選び、枠を利用して実験結果の定量評価を行った。

不変量の組み合わせを考える上で、2 文字のうち黒画素の面積が大きい方を「文字大」、面積が小さい方を「文字小」と呼ぶことにする。用いた 5 種類の不変量の組み合わせは、

- i. $\frac{\text{文字小の黒画素 (図 5(a))}}{\text{文字小の凸包 (図 5(c)) + 文字大の凸包 (図 5(d))}}$
- ii. $\frac{\text{文字大の黒画素 (図 5(b))}}{\text{文字小の黒画素 (図 5(a)) + 文字大の黒画素 (図 5(b))}}$
- iii. $\frac{\text{文字小の黒画素 (図 5(a))}}{\text{文字大の凸包 (図 5(d))}}$
- iv. $\frac{\text{文字小の凸包 (図 5(b))}}{\text{文字大の黒画素 (図 5(c))}}$
- v. $\frac{\text{文字小の黒画素 (図 5(a))}}{\text{文字大の黒画素 (図 5(b))}}$

である。不変量ベクトルの次元数が 5 次元の場合は 5 種類全て、3 次元の場合は i. ~ iii. のみを使用する。

提案手法による射影歪みの補正実験を画像 1~3 を対象に行い、その結果を補正に用いたパラメータとともに表 1 に示す。クラスタリングとノイズ除去の各パラメータは、クラスタリング結果が最適になるように調節した。表 1 の (A) が元画像である。

補正結果について、まず未知である焦点距離を $f = 1$ とした

ときの結果を、各画像の補正に用いたその他のパラメータとともに表の (B) に示す。理想的には文書の角の直角は復元されないものの、平行線は復元されるはずである。実験の結果、画像 1 の平行線はほぼ復元されるものの、画像 2 と画像 3 では若干の誤差が残った。この主な原因は、平面当てはめの際に外れ値が影響し、パラメータの推定に誤差が生じたことである。

次に焦点距離 f を手動で探索し、ほぼ最適となる値を定めたとときの結果を表の (C) に示す。理想的には、文書の枠の長方形が復元されるはずである。しかし、 $f = 1$ のときと同様の理由で、この場合も画像 1 の長方形はほぼ復元されたものの、画像 2 と画像 3 では誤差が残った。

平面当てはめにロバスト推定を用いて頑健にするとともに、自動的に f を取得して紙面の射影歪みを完成させることは今後の課題である。

4. 従来手法との比較

本稿で扱う文書画像の射影歪み補正は、カメラを用いた文書画像処理における基本的な課題であるため、既に様々な研究が行われている。それらは (1) 文書の枠を利用する方法、(2) 文書中の文字行を利用する方法、(3) ステレオ視を利用する方法、に大別できる。以下ではその概要を述べ、提案手法との違いを明らかにする。

(1) の方法は文書の枠が長方形であることを前提とする。枠の対辺の平行性は射影歪みの影響を受けると失われるが、これを長方形に戻すような変換を求めれば、射影歪みを受けた文書画像を正対した画像に復元することができる。このような手法は文献 [5] で提案されている他、市販のデジタルカメラ (Ricoh Caplio R6 等) に実装されている。この方法は、多くの文書の枠が長方形であるという合理的な仮定を用いている反面、枠を含めた文書画像全体の撮影を必要とする。

(2) の方法は文字行の平行性を仮定している。例えば文献 [5]^(注1) では消失点から傾き具合を推定する手法が提案されている。この手法の欠点は、文書のレイアウトに強い仮定を課しているため、適用範囲が限定されることである。特殊なレイアウトのページにはまず適用不可能であり、レイアウトが一般的であっても文書中に図や数式を多く含むページでは文字行の両端の推定は容易でないからである。

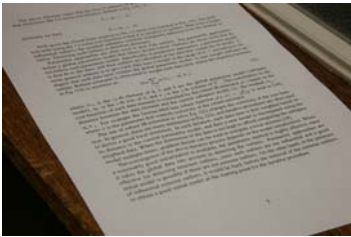




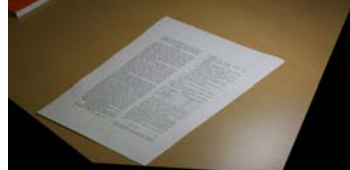
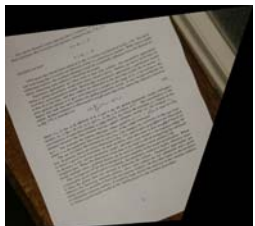


上記の (1) と (2) の方法は、図 1 に示すようなレイアウトが複雑 (文字行が平行でない) で、かつ枠が画像中に文書の枠が含まれていない画像に対して適用することができない。

(3) は複数のカメラ [6] もしくは動画 [7] を用いて 3 次元形状を復元する方法であり、提案手法と使用する装置が異なる。

また、提案手法のように変量と不変量を用いた手法としては、スキャナで取得した文書画像の傾き (回転角) を補正する手法 [8] がある。この手法では提案手法と同様に不変量で字種の判別、変量で回転角の推定を行っているため、他の手法に比べて類似点が多い。この手法と提案手法の一番の違いは事前に学習が必要かどうかである。文献 [8] の手法は字種毎に回転角と変量 (文字の外接矩形の面積) の対応をあらかじめ登録しておく必要があるため、登録していないフォントからなる文書には適用することができない。また、文書中の文字は全て同じ方向を向いている (回転していない) という仮定を用いている。一方提案手法ではそのような事前の処理やレイアウトに対する仮定を

(注1): 文献 [5] には (1) の手法と (2) の手法の両方が記載されている。

表 1 補正パラメータと補正結果 . (A) 補正前 . (B) 補正パラメータと補正結果 ($f=1$) . (C) 焦点距離 f に最適値を選んだときの補正結果 .

		画像 1	画像 2	画像 3
(A)	画像			
	対辺と成す角	(長辺) 8.98° / (短辺) 5.18°	(長辺) 1.73° / (短辺) 19.2°	(長辺) 6.64° / (短辺) 7.70°
(B)	不変量ベクトルの次元数	3	3	5
	クラス数	60	40	200
	閾値 t_1, t_2	0.05, 20	(閾値処理なし)	0.1, 20
	画像			
対辺と成す角	(長辺) 0.06° / (短辺) 0.73°	(長辺) 0.18° / (短辺) 2.81°	(長辺) 1.78° / (短辺) 5.43°	
(C)	焦点距離 f	$f=11000$	$f=2000$	$f=15000$
	画像			
	対辺と成す角	(長辺) 0.78° / (短辺) 0.02°	(長辺) 1.24° / (短辺) 1.09°	(長辺) 1.20° / (短辺) 3.88°
	四隅の角度の誤差平均	0.605°	1.35°	4.16°

一切用いていないため、多様なレイアウトの様々なフォントに (文字ではないマークの類いにさえ) 適用可能である .

5. ま と め

本稿では、紙面の傾きに対する「変量」と「不変量」を組み合わせることで文書画像の射影歪みを補正する方法を提案した . 提案手法では、文書画像中の各連結成分に対して、面積を「変量」として、面積比を「不変量」として用いることにより、各連結成分の相対的な奥行き情報を推定することができる . そして、紙面全体の情報を統合することにより、紙面の傾き角度の推定を行う . 提案手法は、連結成分が持つ相対的な情報のみを利用してあり、撮影方法やレイアウトに強い制約を用いていないため、多様な文書画像の射影歪みを補正することができる .

実験により、本来は長方形である物を平行四辺形に (射影歪みをアフィン歪み程度に) 復元できる潜在能力を確認したが、本来の長方形を復元するまでには至っていない . この原因は推定した奥行き情報に定数倍の不確実性が残っていることと、平面の当てはめに誤差が生じていることによる . 前者は本稿で用いたのは別の「変量」と「不変量」を利用することで解決できると考えられ、後者はロバスト推定の導入やノイズ除去の精度向上によって解決可能であると思われる .

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] 徐剛, 辻三郎, 3次元ビジョン, 共立出版, 1998 .
- [4] 出口光一郎, 画像と空間, 昭晃堂, 1991 .
- [5] P. Clark and M. Mirmehdi, "Recognising text in real scenes," IJDAR, vol.4, pp.243-257, 2002.
- [6] 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.
- [7] 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," 信学論 (D), vol.J88-D, no.8, pp.1490-1498, Aug. 2005 .
- [8] 酒井恵, 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, "変量と不変量の組合わせによる文書画像の傾き補正," 信学技報 PRMU2006-251, March 2007 .