

変量と不変量を用いた文書画像の事例ベース傾き補正

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

[†] 九州大学大学院システム情報科学府 〒 819-0395 福岡市西区元岡 744

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

^{†††} 大阪府立大学大学院工学研究科 〒 599-8531 大阪府堺市学園町 1-1

^{††††} 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: sakai@human.is.kyushu-u.ac.jp

あらまし ある特定の幾何変形に対する「変量」と「不変量」のみを利用することで変形量を推定する方法を提案する．特定の幾何変形としてスキューすなわち回転を例に取り説明する．まずあらかじめ様々な文字フォントを回転させながら回転変量（例えば，文字の外接矩形の面積）の変化を測定し，これを回転不変量（例えば，文字凸包内の黒画素と白画素の面積）と共に事例として記憶しておく．そして補正したい画像の各連結成分（多くは単文字に相当）から算出した回転不変量により事例を呼び出し，同じく算出した回転変量から文字の回転角を得る．推定された回転角により投票を行い，最も投票結果の多いものを推定結果とする．本手法には，文字が直線的な部分形状と並びを持つという，従来一般に利用されてきた仮定が不要になる．文書画像 55 サンプルを用いて，簡単な傾き推定実験を行ったところ，45 サンプルにおいて誤差 1 度以下，54 サンプルにおいて誤差 2 度以下という精度を得た．

キーワード 幾何変形推定，変量，不変量，事例，投票，傾き補正

Instance-Based Skew Detection of Document Images by a Combination of Variant and Invariant

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

[†] Graduate School of Information Science and Electrical Engineering, Kyushu University

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

^{†††} Graduate School of Engineering, Osaka Prefecture University, Sakai-shi, Osaka, 599-8531 Japan

^{††††} Graduate School of Engineering, Tohoku University, Sendai-shi, 980-8579 Japan

E-mail: sakai@human.is.kyushu-u.ac.jp

Abstract A novel deformation estimation technique is proposed and applied to document skew estimation. The proposed method has two properties. First, it utilizes an invariant and a variant of a target deformation to be estimated. Second, it is an instance-based method where the deformation is estimated by referring stored instances which describe the relation among the deformation, the variant, and the invariant. The result of a skew estimation experiment on 55 document images has shown that the skew angles of 54 document images were successfully estimated with errors smaller than 2.0 degrees.

Key words geometric transformation, variant, invariant, instance, voting, skew detection

1. ま え が き

文字認識において，幾何変形の推定・補償は重要な問題の 1 つである．例えば，スキャナによって取得された文書画像中の

文字については，紙面の傾き（スキュー）による回転変形が重要な問題とされてきた．このスキュー補正については，後述のように従来より非常に多くの研究事例がある．一方，カメラによって取得された文書画像の場合，その画像中の文字が受け

ている幾何変形はより多様で複雑なものになる [1]。例えば、カメラと紙面が正対していないことによる射影変換歪み、紙面自体が平面になっていないことによる非線形な歪みが生じうる。これらの歪みの補正法は dewarping と呼ばれており、カメラベース文字認識の隆盛と共に、現在活発に研究されている。

本文献では幾何変形の推定手法を提案し、前述のスキュー補正問題に適用することでその定量的・定性的評価を行う。なお、本手法の考え方自体は一般的なものであり、原理的には射影変換などの幾何変形も補正するように拡張可能である。

本手法の第一の特徴は、ある特定の幾何変形に対する「変量」と「不変量」のみを利用することで変形量を推定する点にある。特定の幾何変形としてスキューすなわち回転を例にとると、回転により値が変わる特徴量と変わらない特徴量を組み合わせることで回転角度を推定することになる。したがって本手法では、文字行の傾きや各文字の縦横のストロークの傾きを手がかりとする従来のスキュー補正法とは全く異なるアプローチを取る。その結果、文書画像中の文字が直線状に並んでいなくても、またひらがなのように曲線の多い文字が支配的な場合でも、回転角度を推定することが可能となる。

本手法の第二の特徴は、事例に基づいて変形量を推定する点にある。この特徴をまず単純な方式を想定して以下に説明する。今、もし文書中の各文字パターンのカテゴリがわかっていたら、そのパターンに当該カテゴリの標準パターンを回転させながら重ね合わせ、最も重なりが大きかった回転角度がその文字の回転角度の推定値になる。この単純な方式の場合、様々な角度で回転されている標準パターンをすべて事例とし、文書中の文字パターンを各事例と比べることで、回転角度を推定していると言える。ただしこの単純方式には、回転角の範囲や拡大縮小まで考えると、事例の数が膨大になり、現実的ではないという問題がある。また、文書中の各文字パターンのカテゴリはそもそも不明であり、したがって事例の選出にも問題が残っている。そこで、第1の問題については、変量というスカラー量を事例として用いることで、変形量推定の効率化を図る。第2の問題については、前述の不変量を利用することで、カテゴリが不明であっても事例が参照できるような工夫をする。

2. 従来の傾き補正法

文書画像の傾き補正法については数多くの研究事例がある。その多くは、文書画像の大局的な特徴を利用している。例えば回転を変えながら求めた周辺分布を手がかりにした方法はその典型的な例である。一方、本手法のように、局所領域(例えば連結成分)毎に傾き角を推定し、それを画像全体で統合することで、文書画像全体の回転角を推定する方法も幾つか提案されている。以下にそれらを概観する。

- 文献[2]では、画像中にある一定の大きさの円の局所領域を取り出し、その円内に引いた直線上の文書の画素値の変化を測定する。角度を変えて直線を引き、その変化が最も大きかった直線の角度方向を局所領域の回転角推定結果とする。また、統合方法としては変化が最も大きかった直線の角度を文書全体の回転角とする。

- 文献[3]では、まずある局所領域内の連結成分から特徴点を検出する。各特徴点に対してその最近傍の特徴点3つの間の線分の比較によりおよその回転方向を表す直線を求め、それから微小距離以内にある点を一定数用いて最小2乗近似により最終的な局所回転角を表す直線を求める。この操作を各特徴点ごとに行い、その結果を投票し、全体としての回転角を求める。

- 文献[4]では、まずある連結成分について、一定の近傍条件を満たす連結成分を結合して拡大連結成分を形成する。次にその中で最遠点となる2画素間に直線を引き、その角度を局所回転角とする。ここでこの近傍条件とは、例えば文字行の直進性を考慮するようなものである。条件を満たす連結成分がない場合は、局所回転角も求められないことになる。以上の処理を各連結成分について行い、複数求めた局所回転角の中央値を全体の回転角とする。

この他にも、各文字の縦ストロークらしい部分の方向情報を利用した dewarping 手法 [5] なども存在する。

いずれも、文書の局所的な直線性を仮定した手法である。すなわち近傍の幾つかの連結成分が直線的な文字行を為すといった仮定や、文字には直線的な部分が存在するといった仮定である。これらに対し、本手法はこうした仮定を一切使用しない点で大きく異なっている。

3. 変量と不変量の組み合わせによる傾き推定

本章では、変量と不変量の組み合わせによる幾何変形推定の方法について述べる。本手法は、一般的な幾何変形の推定に拡張することもできるが、前述のように、幾何変形のうち回転を例にとって説明する。また対象についても文書画像に限定して論ずる。一般的な場合についても3.5において簡単に触れる。

3.1 傾き推定の原理

まず、文書画像内に存在する文字カテゴリについて、標準パターンを準備する。前述のように、この標準パターンのすべての回転画像を事例として登録しておけば、それらすべてと入力文書画像中の各文字(連結成分)を比較照合し、最も照合した回転標準パターン画像の角度から、文書画像の回転角を推定できる。この力づくの単純方式は、非常に明確ではあるが、明らかに効率が悪い。

そこで本手法は、回転変量を事例として用いることで効率的な傾き推定を行う。基本手順は以下の2ステップからなる。なお、理解を容易にするために、ここでの説明では、入力文書画像中のある文字のカテゴリが c とわかっているものと仮定する。(後に述べる手法によりこの仮定は不要になる。)

- 事例の収集(学習ステップ)

各カテゴリ c の標準パターンを少しずつ回転させながら変量 p を測定する。これを回転角 θ と変量 p の関係 $p = p_c(\theta)$ として保存しておく。一種の学習ステップであり、事例の収集段階でもある。カテゴリ集合については、英文書画像の場合、“A” ~ “Z”, “a” ~ “z” である。

- 傾きの推定(推定ステップ)

入力文書画像中の(傾いた)文字パターン x について変量 p_x を求める。そのカテゴリを c とすれば、 $p_x = p_c(\theta)$ を満たす θ

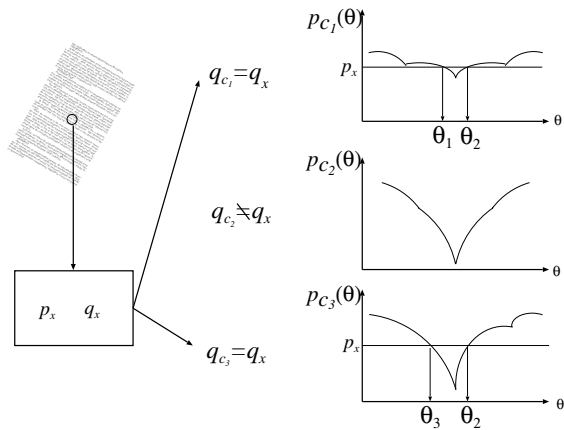


図 1 回転角の推定

が、入力文書画像の回転角の候補になる。この θ を求めるのは、1次元関数 $p_c(\theta)$ の逆引きテーブルを作っておけばよく、 $O(1)$ の計算量で済む。

このように回転変数を使えば、非常に容易に傾き推定が可能になる。以下では、カテゴリ c の推定法、推定ステップの詳細、ならびに実際に用いる変数と不変量について述べる。

3.2 不変量によるカテゴリ c の推定

前節では、各文字のカテゴリ c がわかっていることを仮定していた。しかし、傾き補正以前の段階でカテゴリ c を知ることができるならば、そもそも傾き補正の不要な状況であろう。このため、不確かながら、事例を参照するためのカテゴリが必要になる。

そこで、各文字（連結成分）のカテゴリ c を、画像情報から計算される回転不変量 q を用いて、回転変形に対してロバストに推定することを考える。具体的には、各カテゴリ c の標準パターンについて、その回転不変量 q_c を求めておく。そうして、入力文字 x のカテゴリを推定する際は、その文字から回転不変量 q_x を求め、 $q_x = q_c$ となる c を求める。文書画像の回転について q_x は不変であるから、原理的には正しい c が求められる。

しかし実際には回転不変量 q だけを頼りに文字カテゴリを唯一に絞り込むのは難しい。その難易度は用いる不変量やカテゴリ数、文書画像のノイズなどに因るが、要するに単一の特徴量による文字認識であり、厳密なカテゴリ推定を望むのには無理がある。このため、実際には小数 ϵ を用いて、 $|q_x - \epsilon_q| \leq q_c$ を満たす複数の c をカテゴリ候補として利用することになる。

3.3 推定ステップの詳細

以上の推定ステップで「候補」という言葉が用いられていることからわかるように、1文字だけでは θ を一意に決めることは困難である。これは次の理由による。

- 第一は、 p と θ が一対一対応でなく、同じ変数 p に複数の θ が対応することが多い点にある。したがって1文字では複数の回転角候補が与えられることになる
- 第二に、回転角 θ を変えても変数 p があまり変化しないような場合（関数 $p_c(\theta)$ が平坦になる場合）、変数 p の測定誤差に θ が鋭敏に反応し、推定値 θ の信頼性が低くなる。

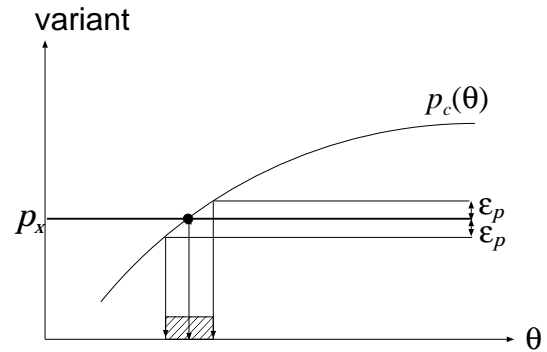
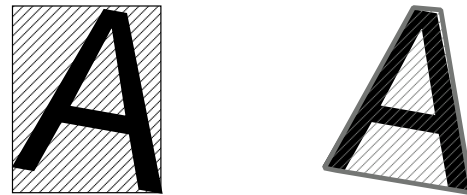


図 2 角度投票の実際



(a) (b)
図 3 文字 A の (a) 外接矩形, (b) 凸包

- 第三に、3.2 で述べたように推定されるカテゴリ c が複数存在する点である。結局、これら複数の c について、前節の推定ステップを駆動する他なく、結局複数の回転角の候補が与えられることになる。

そこで文書画像中の複数の文字から推定される角度を投票していき、最も投票結果の多いものを推定結果とする。文書画像中に異なるカテゴリの文字が複数存在すれば、それだけ異なった事例によって候補が与えられることになり、多くの誤った候補が混入したとしても、それらが誤った一点に集中する可能性は低いと期待される。

以上の推定ステップの処理を、図 1 を通して具体的に説明する。入力文書画像中の文字パターン x について p_x, q_x を計算する。 $|q_x - \epsilon_q| \leq q_c$ となるカテゴリを選択する。次に、 $p_x = p_c(\theta)$ となる角度、 $\theta_1, \theta_2, \theta_3$ にそれぞれ投票する。この場合、 θ_2 が 2 つあるので 2 回投票する。このような処理を文書画像中の全ての連結成分に対して行い、最も投票数の多かったものをその文書の回転角とする。

事例 $p_c(\theta)$ を基に投票する際、不変量の場合と同様に、 p_x の誤差に配慮し、 $|p_x - \epsilon_p| \leq p_c(\theta)$ を満たす範囲の θ についてすべて投票する図 2 にその様子を示す。変数関数 $p_c(\theta)$ の傾きによって投票範囲の大きさが異なることに注意されたい。

3.4 変数と不変量の具体例

本手法では、回転に対する任意の変数と不変量を用いてもよいが、本文献では最も単純なものを用いる。具体的には、変数 p として、文書画像中の 1 連結成分を回転させることにより変化する値、外接矩形の面積（図 3(a)）を用いる。また、不変量 q として、回転させても変化しない値、凸包内の面積（図 3(b)）を用いる。ここで、これらの面積と黒画素の面積の比をとること

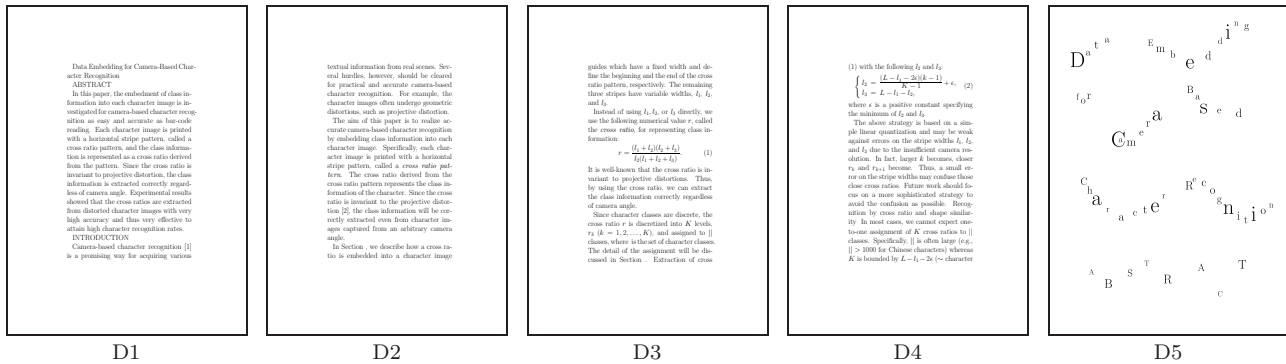


図 5 実験に使用した文書画像, D1,2 英字のみ, D3,4 英字と数式, D5 英字のみ

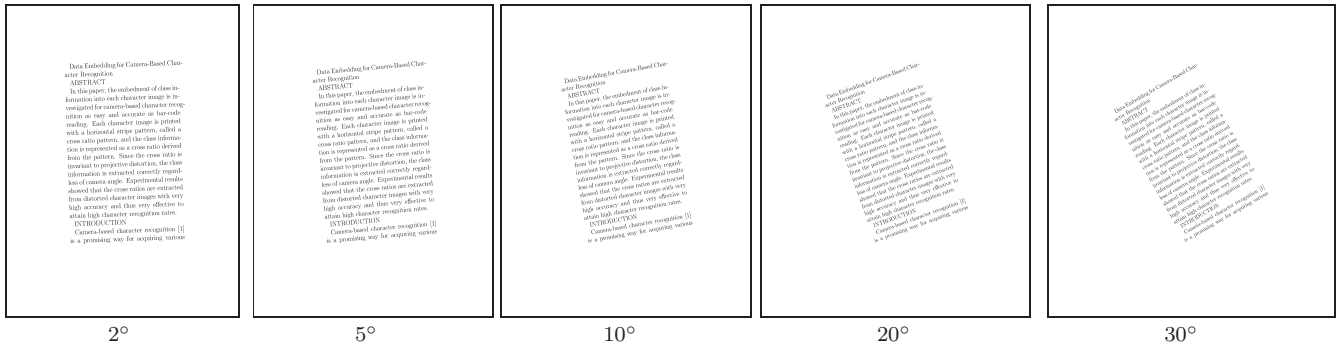


図 6 回転させた文書画像例 (D1)

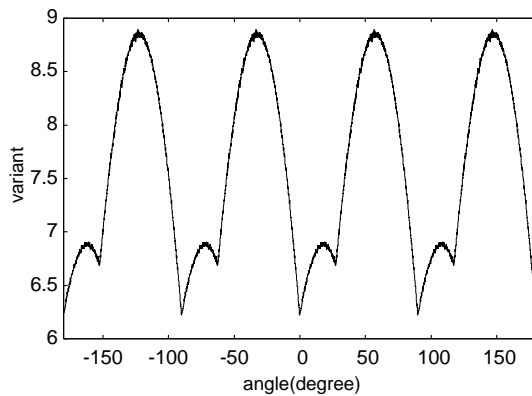


図 4 カテゴリ “y” の -180° から 180° の変量 p

とによりこの変量と不変量を, 画像中に占める連結成分の大きさに依らないようにする. すなわち, p, q 共にスケール不変量とする.

回転変量 p および不変量 q を改めて式で表すと以下のようになる.

$$p = \frac{\text{外接矩形の面積}}{\text{黒画素の面積}} \quad (1)$$

$$q = \frac{\text{凸包内の面積}}{\text{黒画素の面積}} \quad (2)$$

この変量 p は, 図 4 のように周期的に変化する. そのため, その 1 周期分, この場合 -45° から 45° だけが使用可能であり, 本手法による推定可能な回転角も同様の範囲に限定される.

3.5 より一般的な幾何変形の推定

以上, 回転を例にとって説明したが, これは他の幾何変形の推定にも応用できる. すなわち, 補正したい幾何学的変形の変

量と不変量を組み合わせることにより, 様々な幾何変形の程度を推定できる. 例えば, アフィン変形しうる対象について, アフィン変換不変量と相似変換不変量を事例として準備したとする. このとき, アフィン変換不変量で, アフィン変換によらないカテゴリ推定を行うことができ, 推定されたカテゴリ中の相似不変量を使用することによりアフィン変換のうちせん断成分を推定できる.

4. 実験

4.1 実験試料

スキュー推定の対象は, LaTeX で作成された 5 種類の文書画像である. それら文書の画像を図 5 に示す. 大部分は事例と同じフォントから構成されるが, 数式を含む文書もある. これら数式中のイタリックフォントや数学記号については対応する事例がなく, したがって誤推定の要因になりうる. それぞれの文書画像を $\pm 30^\circ, \pm 20^\circ, \pm 10^\circ, \pm 5^\circ, \pm 2^\circ, 0^\circ$ で回転させて 55 個のテスト画像を生成した. テスト画像の解像度は 600dpi とした. 図 6 はその例である. これらのテスト画像に対して左上から順に右下まで 1 連結成分 (多くの場合, 単文字) づつ投票を行うこととした.

単一フォント (Times-Roman) の “A” から “Z”, “a” から “z” の 52 文字それぞれを -45° から 45° まで 0.1° 刻みで変量と不変量を測定して事例を作成した. 事例を作成するために使用した画像の解像度は 1440dpi である. 図 7 に実際に測定した変量と不変量の例を示す. カテゴリごとに測定した不変量は誤差が生じるので, 事例として記憶しておく不変量 q_c はその平均をとった.

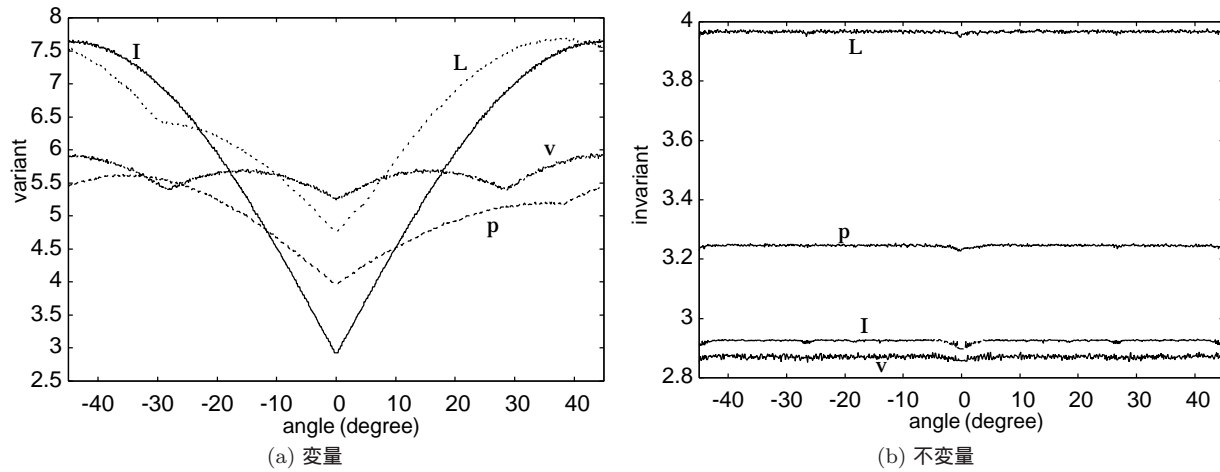


図 7 実際の事例

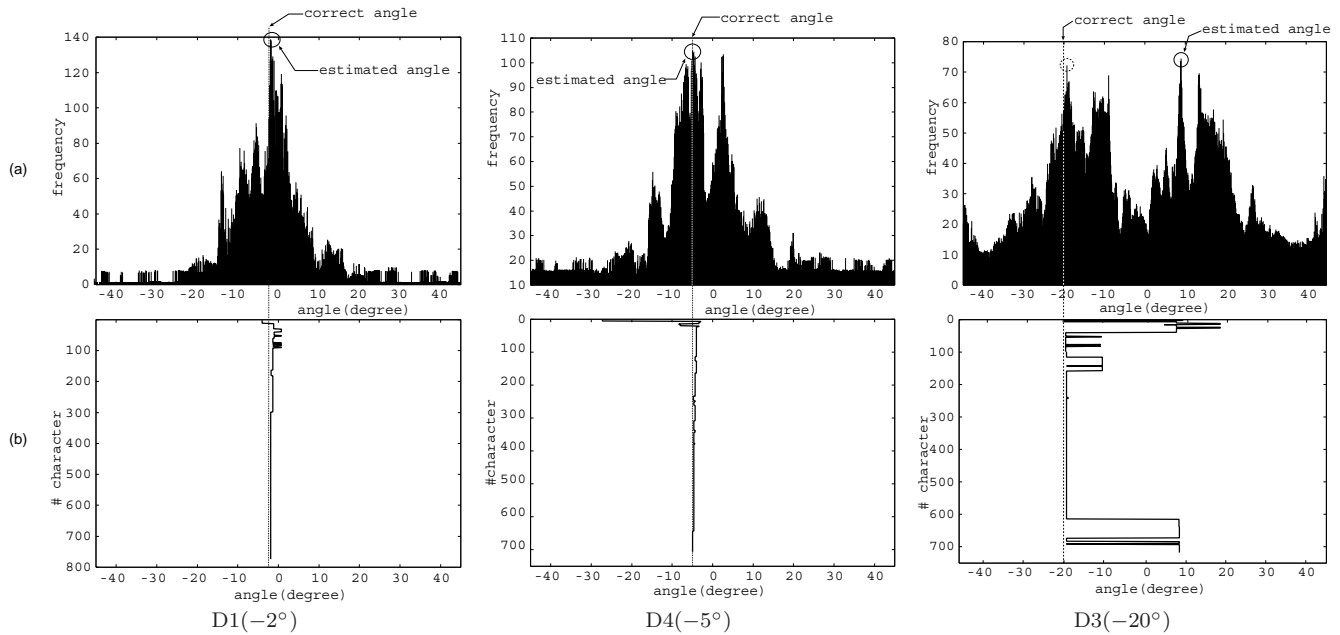


図 8 推定結果例 (a) 投票結果, (b) 使用した文字数による推定結果の推移

表 1 傾き推定結果. () 内は%

誤差 0.5° 以下	誤差 1.0° 以下	誤差 1.5° 以下	誤差 2.0° 以下
20/55(36)	45/55(81)	49/55(86)	54/55(98)

4.2 本手法による傾き補正結果

55 枚のテスト画像について傾きを推定した結果をまとめたものを表 1 に示す. テスト画像の 95% を誤差 2.0° 以下で推定した. この 2° の誤差がどの程度 OCR に影響するかは今後評価する必要があるが, 初期検討の結果としては許容範囲と考える.

図 8 の左及び中央の位置は D1(-2°), D4(-5°) の結果であり, いずれも傾き推定の成功例である. 同図 (a) の推定角度はそれぞれ 1.9°, 5.0° であり, 十分な精度で推定できたと言える. また, 同図 (b) の縦軸は推定に使用した文字数を表しており, 上から下に行く方向で投票が進む. すなわち, 推定に利用した文字数が増加する. 推定に使用した文字数が増えるに従い推定精度が向上していることが確認できる. ただし, D4(-5°) については正解角度付近でわずかながら振動が起きている. これは, 同図 (a) に示す通り, 正解角度付近にほぼ同数の投票

が行われているためである. 3.3 で述べたように, p_x の誤差を考慮するため, $|p_x - \epsilon_p| \leq p_c(\theta)$ を満たす θ に投票している. このため, p_x として常に同じ値が求めたととしても, $2\epsilon_p$ の幅の分だけ推定角度に曖昧さがでる. その結果, 正解角度付近に複数のピークが見られる.

各テスト画像に対する傾きの推定誤差を表 2 に示す. 傾き推定誤差の平均は 1.2° であった. どの角度においても大部分で誤差が小さいことがわかる. テスト画像 D3, D4 のように事例を持たない文字を含む画像でも, 大部分の連結成分が事例を持っていれば投票により正しく推定できている. また, D5 のような文字が直線上に並んでいないテスト画像の回転角は従来法では推定することが難しいが, 本手法では正しく推定できた.

テスト画像 D3(-20°) の推定誤差が -29.0° となり, 誤差が 2.0° 以上となった. 推定結果を図 8 の右列 D3(-20°) に示す. 途中まで正しい推定結果を示しているが, 使用連結成分数が 600 を超えた所で他の角度に推定している.

この誤った角度へ投票された主たる原因は, 不変量のノイズに対する不安定性にあった. すなわち, 解像度の問題により,

表 2 傾き推定誤差 (°)

angle	D1	D2	D3	D4	D5	average
-30	-0.7	-0.7	-0.7	-0.7	-0.4	0.6
-20	-0.7	-0.7	-29.0	-0.7	0.4	6.3
-10	-0.5	-0.9	-0.9	-0.2	-0.3	0.7
-5	-0.7	-0.7	1.5	0.0	-0.2	0.6
-2	-0.1	-0.9	0.2	-0.2	-0.2	0.3
0	1.7	1.7	1.7	1.7	0.4	1.4
2	-0.6	0.6	0.4	-0.3	0.2	0.4
5	0.3	0.3	0.9	0.9	0.3	0.5
10	1.0	1.4	2.0	1.4	0.2	1.2
20	0.7	0.7	0.9	0.7	0.6	0.7
30	0.7	0.7	1.0	1.2	0.4	0.8
average	0.6	0.8	3.5	0.7	0.3	1.2



図 9 不変量誤差が少ない文字とその凸包



図 10 不変量誤差の多い文字とその凸包

$|q_x - \epsilon_q| \leq q_c$ となるカテゴリ c において、真のカテゴリが存在しない文字が複数存在した。こうした傾向はカテゴリの種類によって異なっていた。図 9 に不変量誤差が少ない文字を、図 10 に不変量誤差が多い文字を、それらの凸包と共に示す。これらの比較により、不変量誤差の大小は黒画素と凸包の接し方により発生することが判明した。事例 (1440dpi) よりも入力画像 (600dpi) の解像度が低いため、文字線の細くなっている部分 (特にセリフ部分) において黒画素が削れる場合がある。こうした削れが凸包と接している部分に発生すると不変量誤差が大きくなる場合がある。図 11(a), (b) で円で囲った部分のような接し方の場合、黒画素の削れによる影響は殆どない。不変量誤差の少ない文字は全てそのような接し方をしている。しかし、同図 (c) で円で囲った部分のような接し方をしている場合、不変量誤差は無視できないものになる。黒画素の減少よりも、凸包の減少の方が大きくなってしまったためである。失敗した文書画像では、以上の原因による不変量の参照間違いが特に累積していた。

5. まとめ

ある特定の幾何変形に対する「変量」と「不変量」のみを利用することで変形量を推定する方法を提案した。特定の幾何変形としてスキューすなわち回転を例に取って、回転により値が変わる特徴量と変わらない特徴量を組み合わせることで回転角度を推定した。その結果、文書画像中の文字が直線状に並んでいなくても、回転角度を推定することが可能となった。また、

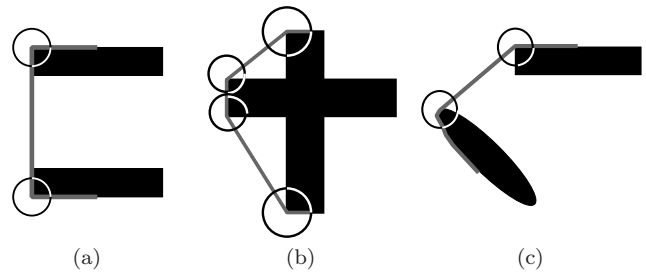


図 11 凸包の接し方

事例に基づいて変形量を推定した。標準パターンをそのまま事例として用いるのではなく、変量と回転角の関係を標準パターンから求め、それを事例として登録しておくことで、効率のよい変形量推定を可能にした。さらに前述の不変量を利用することで、カテゴリが不明であっても事例が参照できるような工夫を行った。

今後の課題としては、4つ挙げられる。

- 定量的な評価実験：今回は傾き角推定誤差を評価指標に用いたが、実際の OCR の許容範囲を含めて、補正後の文書画像の文字認識精度により評価する必要があると考える。
- 不変量の改良：考察により明らかになったように、誤推定の要因の一つに、不変量の曖昧性により誤った投票が多数行われた点が挙げられる。したがって、不変量の弁別性を上げるべく、隣り合う2つの文字の不変量を組み合わせるといった不変量のベクトル化を検討する。
- 事例の取捨選択：本論文の実験では、全てのカテゴリの事例を使用した。しかし、「e」のように悪影響を及ぼす事例が存在した。今後は、このような事例を適宜削除することにより精度向上を図る。なおこの事例の削除には計算量低減の効果があり、マルチフォント対応の際の事例の増加に対しても必要になろう。関連して投票の方式にも一考の余地がある。
- 回転以外の幾何変形への対応：本論文では回転に注目して補正を行った。今後は変換パラメータを増やすことによって、他の幾何変形へ応用していきたいと考える。

謝 辞

本研究の一部は、文部科学省科学研究費補助金 (17700198) によった。

文 献

- [1] 黄瀬浩一, 大町真一郎, 内田誠一, 岩村雅一, “デジタルカメラによる文字・文書の認識・理解,” 電子情報通信学会誌, vol. 89, no. 9, pp. 36–841, Sep. 2006.
- [2] Y. Ishitani, “Document Skew Detection Based on Local Region Complexity,” Proc. Int. Conf. Doc. Anal. Recog., pp. 49–52, 1993.
- [3] X. Jiang, H. Bunke, and D. Widmer-Kljajic, “Skew Detection of Document Images by Focused Nearest-Neighbor Clustering,” Proc. Int. Conf. Doc. Anal. Recog., pp. 629–632, 1999.
- [4] Y. Lu and C. L. Tan, “Improved Nearest Neighbor Based Approach to Accurate Document Skew Estimation,” Proc. Int. Conf. Doc. Anal. Recog., pp. 503–507, 2003.
- [5] S. Lu and C. L. Tan, “Camera Document Restoration for OCR,” Proc. Int. Workshop Camera-Based Doc. Anal. Recog., pp. 17–24, 2005.