

文書画像の事例ベース傾き補正の改良

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府 〒 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, uchida}@human.is.kyushu-u.ac.jp

あらまし 文書画像の回転(スキュー)を補正するため、事例ベース傾き補正法を提案している。事例を用いることにより、文書の直進性を仮定せずに回転角を推定できる。また、事例として回転変量と不変量を用いることにより効率的に回転角を推定できる。変量と不変量はどのようなものでも良いが、前回の報告で用いていた不変量はノイズに弱いという問題があった。そこで本報告では、回転角推定の精度向上を目的として、不変量の多次元化について述べる。計算機内実験において文書画像 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、48 サンプルにおいて誤差 1 度以下、全てのサンプルにおいて誤差 2 度以下という精度を得た。また、スキャナ取得画像に対しても評価実験を行った所、目視ではあるが入力 10 画像中 9 画像に対して正しく補正できた。

キーワード 幾何変形推定, 変量, 不変量, 事例, 投票, 傾き補正

An Improvement of Instance-Based Skew Estimation

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, uchida}@human.is.kyushu-u.ac.jp

Abstract The purpose of this report is to improve an instance-based deskewing technique which is free from the conventional assumption that text lines are straight and parallel. The instances describe the relation among the skew angle, a skew variant, and a skew invariant in an compact manner. A main idea of the improvement is to increase the number of the skew invariants for more stable estimation of the skew angle. An experimental result on 55 document images showed that their skew angles 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

紙面の文書をスキャナ等を使って画像として取得する際、文書画像に回転(いわゆるスキュー)が発生することがある。そこで文書の回転角を推定し、文書画像を補正する方法が研究されてきた。従来、スキュー補正は文字行の局所的もしくは大局的な直進性を仮定して回転角を推定している [1] ~ [3]。

本研究では従来法のような文書の直進性を仮定しない新しい

スキュー補正法を提案している [4], [5]。本手法ではあらかじめ登録された事例、具体的には回転変量と不変量を手がかりとして各連結成分(多くの場合、単文字に相当)ごとに独立して回転角を推定する。このように、文書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となっている。

原理的には事例として登録する変量と不変量として何をいっても構わない。前報告 [4], [5] では、扱う変量・不変量として連

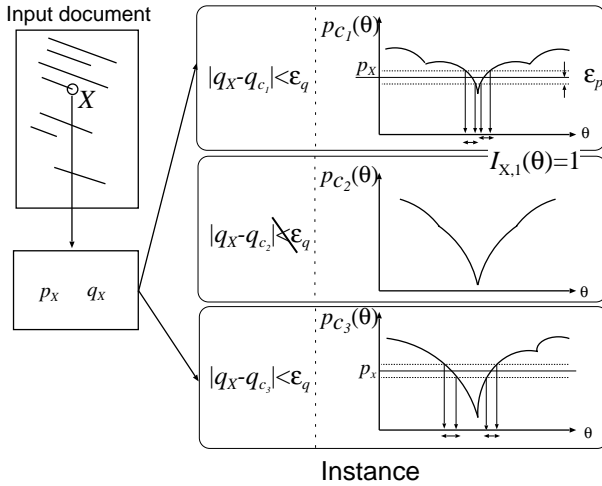


図 1 回転角の推定

結成分の凸包の面積と黒画素の面積の比と、連結成分の外接矩形の面積と黒画素の面積の比を用いた。しかし、この面積比に基づく不変量には、文字と構成する黒画素の小さな変化にも敏感であるという問題があり、結果として回転角推定に失敗する。文献 [4], [5] では計算機内で回転角を施した文書画像を対象としていたが、スキャナを通して実際に得られた画像を対象とすれば、この不変量に関する問題はより深刻になる。

そこで、本報告では、回転角推定精度の向上を目的として不変量によるカテゴリ推定の更なる安定化を目的として、不変量の多次元化を行う。そしてまずシミュレーション実験を通して精度評価を行う。その後、スキャナ取得画像に対しても実験を行い、同じく精度評価を行う。

以下、2. で、本手法の概要について説明し、3. では、シミュレーション実験とスキャナ取得画像に対する実験の結果を示す。最後に、4. で本報告の結果をまとめ、今後の課題について述べる。

2. 事例ベース傾き推定

2.1 推定手順

本節では、事例ベース傾き補正の基本的な手順 [4] について述べる。図 1 は推定手順をまとめたものである。具体的な変量と不変量については次節で説明する。

実際に回転角推定する以前に、事例を学習し登録しておく必要がある。あらかじめ標準画像から変量と不変量を測定し事例を作成する。具体的には、各文字カテゴリ $c \in [1, \dots, C]$ について標準文字画像 R_c から回転不変量 q_c を測定し、次にそれらを角度 θ で回転させた $R_{c,\theta}$ ごとに $p_c(\theta)$ を測定して事例として登録しておく。

作成した事例を基に回転角を推定する際には、まず入力文書画像中の連結成分 X に対して、その変量 p_X および不変量 q_X を計算する。そして、不変量を手がかりに X のカテゴリを絞りこむ。具体的には、条件 $|q_X - q_c| \leq \epsilon_q$ を満たす c の集合 C_X を求める。ここで $\epsilon_q > 0$ は q_X の誤差許容範囲である。集合 C_X には、不変量が近いカテゴリが複数選ばれる。しかし、文

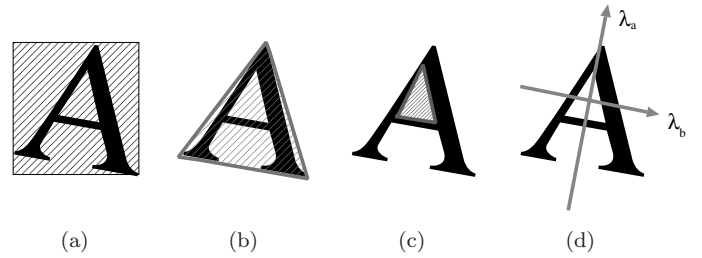


図 2 単文字に対する (a) 外接矩形, (b) 凸包, (c) ホール数, (d) 分散比

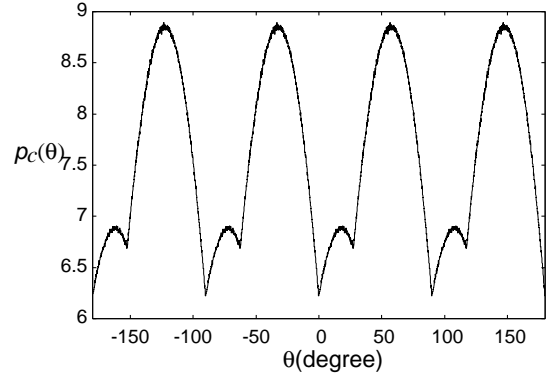


図 3 カテゴリ“y”の -180° から 180° の変量 p

書画像の回転については q_X は不変であるから、 C_X は X の真のカテゴリ c を含むと期待できる。図 1 の例では、 $C_X = \{1, 3\}$ である。

集合 C_X 中の各カテゴリについて変量を用いて回転角を推定する。具体的には各 $c \in C_X$ について、条件 $|p_X - p_c(\theta)| \leq \epsilon_p$ を満たす θ の区間、すなわち

$$I_{X,c}(\theta) = \begin{cases} 1, & \text{if } |p_X - p_c(\theta)| \leq \epsilon_p, \\ 0, & \text{otherwise} \end{cases}$$

について、 $I_{X,c}(\theta) = 1$ となる区間を求める。ここで $\epsilon_p > 0$ は p_X の誤差の許容範囲である。もし ϵ が X の真のカテゴリであれば、この区間は X の真の回転角を含むと期待できる。なお、この区間は図 1 のように不連続になることもある。

以上の処理を文書画像中の全ての連結成分 X に対して行う。複数の回転角が求められるが、これらを投票することで文書全体の回転角 $\bar{\theta}$ を求める。すなわち、前ステップで求めた θ の最頻区間として $\bar{\theta}$ が求まる。前出の $I_{X,c}(\theta)$ を用いれば、これは投票結果

$$h(\theta) = \sum_x \sum_{c \in C_X} I_{X,c}(\theta)$$

のピーク $\bar{\theta} = \operatorname{argmax}_\theta h(\theta)$ を求める処理に相当する。

2.2 変量と不変量

前述の通り、事例生成のための変量 $p_c(\theta)$ 及び不変量 q_c としてはどのようなものを使用してもよい。ここでは本報告で用いた変量と不変量について述べる。

変量 $p = p_c(\theta)$ として、外接矩形の面積 (図 2(a)) を正規化したものを用いる。すなわち

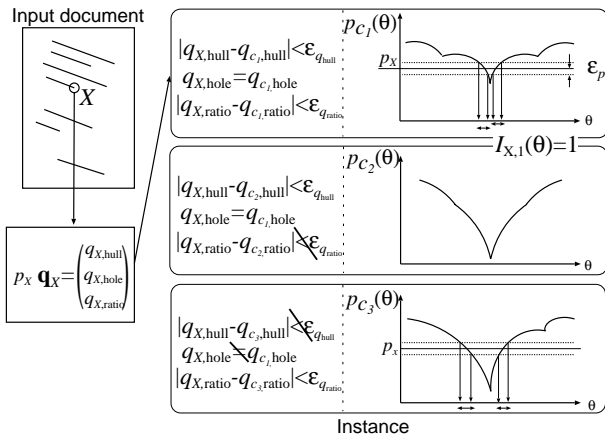


図4 不変量の多次元化による回転角の推定



図5 不変量 q_{hull} による誤差の多い文字とその凸包

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

この変量 p は、図3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．

前報[4],[5]では、不変量 q_{hull} として、凸包内の面積(図2(b))を用いた．すなわち、

$$q_{hull} = \frac{\text{凸包の面積}}{\text{黒画素の面積}}$$

黒画素の面積で正規化したことにより、画像中に占める連結成分の大きさに依らない、すなわちスケール不変量となっている．

2.3 不変量 q_{hull} の問題点

不変量 q_{hull} を用いて C_X に X の真のカテゴリが含まれないケースが散見される．こうした傾向は細い部分(セリフ部分)が多い文字(図5)において多く見られる．こうした、細い部分が欠けたり、膨らんだりすることにより黒画素の面積が解像度の影響を受け、得られる不変量が真値から大きく逸脱することになる．

また不変量 q_{hull} には、各カテゴリの値が非常に近いことが多く、結果的に誤ったカテゴリを C_X に多く含むという問題もあった．図6に不変量 q_{hull} の真値と測定値の最大値・最小値を示す．半数ほどのカテゴリが $q_{hull} = 2.8$ から 3.3 の間に集中しており、真のカテゴリ以外のものが参照されやすいことがわかる．

2.4 不変量の多次元化

本報告では新たに2つの回転不変量 q_{hole} , q_{ratio} を付加することにより、上記の問題を解決する事を試みる．不変量 q_{hole} とは、連結成分のホール数(図2(c))である．不変量 q_{ratio} とは、連結成分の長軸長 λ_a と短軸長 λ_b の比 λ_b/λ_a (図2(d))で

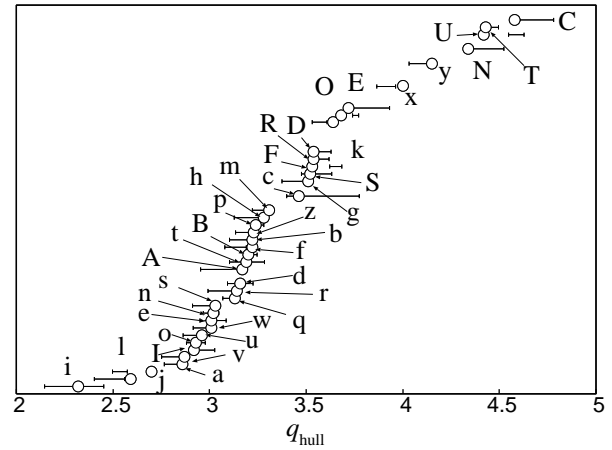


図6 不変量 q_{hull} の最大値と最小値．真値にしたがって昇順に並べて表示．丸印は真値．

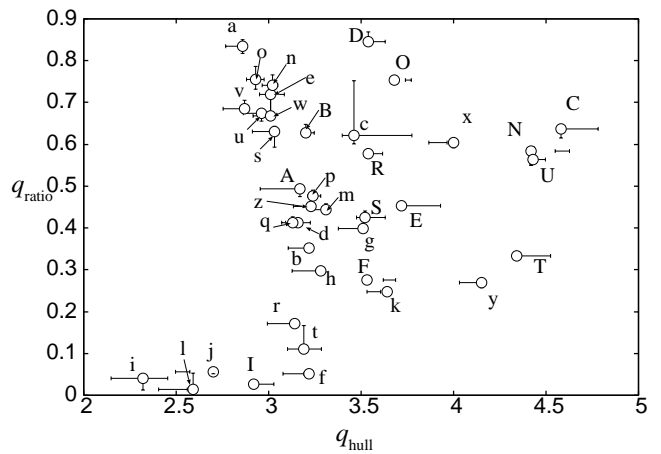


図7 不変量 q_{hull} と q_{ratio} の最大値と最小値．丸印は真値

ある．但し、 λ_a と λ_b は連結成分の黒画素座標の共分散行列の第1及び第2固有値である．

前出の q_{hull} を単独で用いる代わりに、以上3つの不変量 q_{hull} , q_{hole} , q_{ratio} を組み合わせて利用することを考える．こうした多次元化された不変量を用いた推定手順を図4に示す．不変量によるカテゴリ c の推定時に全ての条件 (ϵ_{hull} , ϵ_{hole} , ϵ_{ratio}) を満たしているものを参照していることに注意されたい．

図7に、不変量 q_{hull} と q_{ratio} の測定値の最大値の最小値を示す．なお、これらの値は後述するシミュレーション実験で用いた試料に基づいて測定したものである．同図の示す通り、多次元化することにより、カテゴリ間の分離度が上がっている．これにより、信頼度の低い連結成分はリジェクトされ回転角推定の精度向上が期待できる．

3. 実験

3.1 事例の作成

1440dpiの標準画像を単一フォント(Times-Roman)のアルファベット大文字“A”から“Z”，小文字“a”から“z”の52文字用意した．それぞれのカテゴリ c について不変量 $q_c = (q_{c,hull}, q_{c,hole}, q_{c,ratio})$ を求めた．更に -45° から 45°

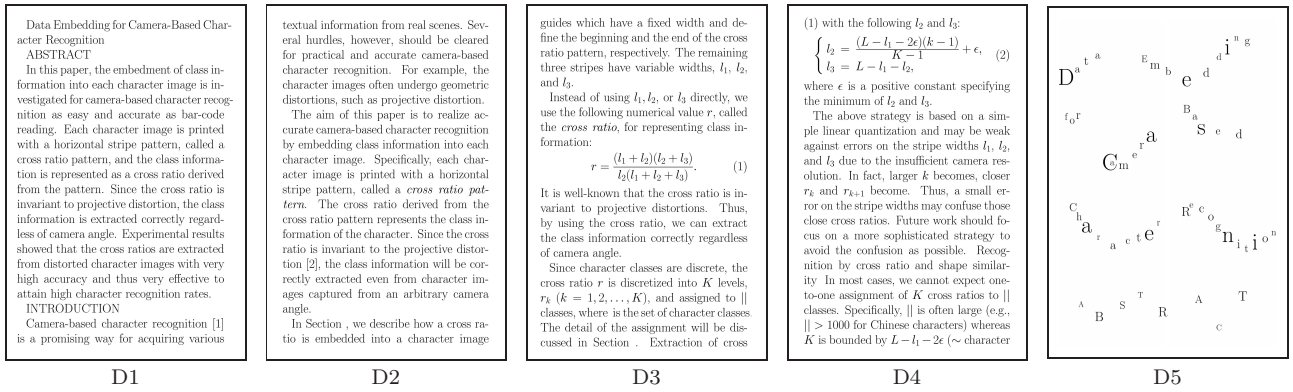


図 8 実験に使用した文書画像, D1,2 英字のみ, D3,4 英字と数式, D5 英字のみ

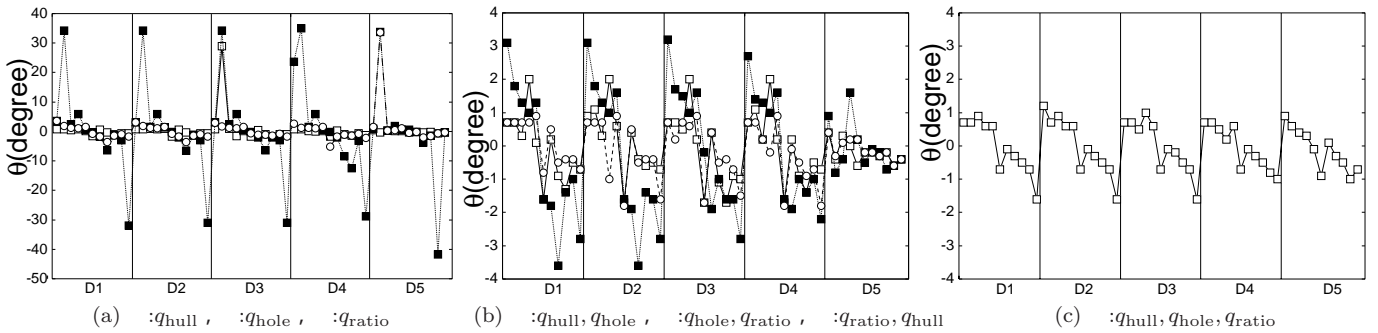


図 9 文書画像 D1~D5 の誤差評価. 各文書画像ごとに $-30^\circ, -20^\circ, -10^\circ, -5^\circ, -2^\circ, 0^\circ, 2^\circ, 5^\circ, 10^\circ, 20^\circ, 30^\circ$ の順番で推定誤差を示している.

まで 0.1° 刻みで θ を変化させて変量 $p_c(\theta)$ を測定した.

3.2 シミュレーション実験

3.2.1 実験試料

まず計算機内部での画像処理により回転変化を施した画像を用いたシミュレーション実験を行った. スキュー補正の対象は, 事例と同じ Times-Roman フォントで作成された 5 種類の 600dpi 文書画像である. この文書画像を図 8 に示す. 文書 D1, D2, D5 の大部分は事例と同じフォントから構成される. 文書 D3, D4 については数式も多く含まれる. 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. また D5 は文字が直線状に並んでおらず, 従来法では回転角推定補正が困難な文書である. また, この 5 画像を $\pm 30^\circ, \pm 20^\circ, \pm 10^\circ, \pm 5^\circ, \pm 2^\circ, 0^\circ$ で回転させて計算機内で生成した 55 個をテスト画像とした. 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.

3.2.2 実験結果

実験結果を図 9 に示す. 同図の横軸は画像の種類を表している. 横軸左から画像 D1, ..., D5 に大きく区別されており, 各画像内では左から $-30^\circ, \dots, 30^\circ$ となっている. また, 縦軸は正解角度からの誤差である.

同図 (a) は各不変量を単独で用いた場合の補正誤差である. 従来法に相当する不変量 q_{hull} 単独の場合にも成功例が多く見られるものの一部に大きな誤差が発生している. この傾向は q_{hole} や q_{ratio} を単独で用いた場合にも見られる.

同図 (b) および (c) は不変量を 2 つずつもしくは 3 つずつ組

み合わせた結果である. 全ての組み合わせで, 大幅な誤差がなくなっている. 同図 (c) では全ての画像において誤差が 2.0° 以下となり, そのうち 50 画像では 1.0° 以下となっている. 各不変量を統合して使用することにより精度が向上していることがわかる. また, (b) と (c) を比較すると不変量を 2 つ組で使用するよりも, 3 つ組で使用した方が精度が良くなっていることもわかる. この精度向上の要因となったのは, 不変量によるカテゴリ c 推定時の参照間違いの減少である. 図 10 に D1 (-5°) に対して不変量 q_{hull} のみで補正した時の事例の参照状況について示す. 同図 (a) は正しいカテゴリを参照し (すなわち C_X に正解カテゴリを含み), かつ変量により正しい回転角が引けた (すなわち $I_{x,c}(\theta)$ に正解が含まれた) 連結成分群である. これらは正しい回転角の推定に貢献している. 同図 (b) は事例中に一致するものがなかったためにリジェクトされたものである (すなわち $C_X = \emptyset$). これらは推定には用いられないため, 悪影響を及ぼすことはない. 同図 (c) は C_X に正解カテゴリを含みながら変量の誤差により $I_{x,c}(\theta)$ 中に真の回転角が含まれない場合であり, これらは推定に悪影響を及ぼす. 同図 (d) は間違ったカテゴリを参照し, かつ回転角も正しくない. これらも推定に悪影響を及ぼす. 同図 (e) は特殊なケースであり, C_X に真のカテゴリは含まれなかったものの, 偶然 $I_{c,x}(\theta)$ に正解が含まれた場合である. 補正に貢献しているものの非常に不安定である.

一方, 図 11 は同じテスト画像 D1 (-5°) に対して, 不変量 $q_{\text{hull}}, q_{\text{hole}}, q_{\text{ratio}}$ を全て用いて補正した時の事例の参照状況

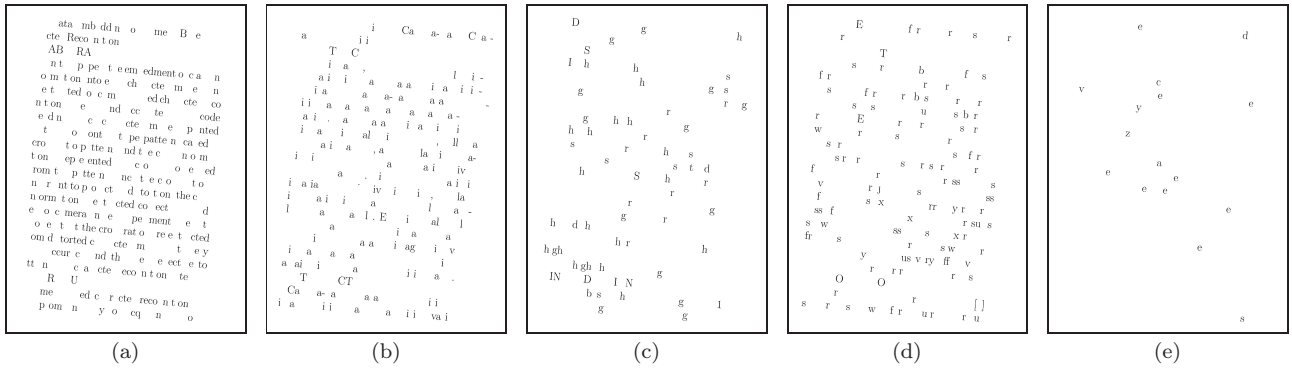


図 10 不変量 q_{hull} を使用した $D1(-5^\circ)$ の事例参照結果 . (a) 不変量, 変量共に正しい, (b) リジェクト, (c) 不変量間違い, 変量正しい, (d) 不変量, 変量共に間違い, (e) 不変量正しい, 変量間違い

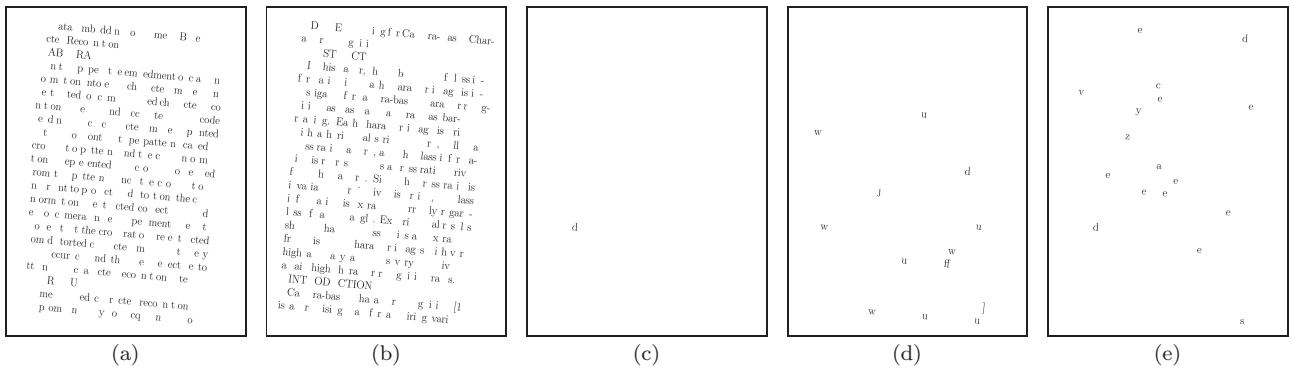


図 11 不変量 q_{hull} , q_{hole} , q_{ratio} を使用した $D1(-5^\circ)$ の事例参照結果 . (a) 不変量, 変量共に正しい, (b) リジェクト, (c) 不変量間違い, 変量正しい, (d) 不変量, 変量共に間違い, (e) 不変量正しい, 変量間違い

である . 図 10 と比較してみると, (c), (d) のような補正に影響を及ぼす連結成分の大部分をリジェクトすることに成功している . これは, 不変量 q_{hull} と q_{ratio} がお互いの得意なカテゴリを用いて信頼性の評価をしあい, より信頼度の高いものを選択した結果である .

不変量 q_{hull} , q_{ratio} は相補的な関係となっている . 実際不変量 q_{ratio} は q_{hull} が苦手とする 2.2 で述べた黒画素に対するノイズに強い . また, 不変量 q_{hull} は q_{ratio} が苦手とする未知のカテゴリ (コンマやノイズ, 数学記号等) に対して強い . 以上により, これら 2 種類を用いることで, 信頼度の低い連結成分をリジェクトし, 推定したカテゴリの信頼度を高めている . また, 不変量 q_{hull} は不変量 q_{hull} と q_{ratio} が判別しづらい, “e”, “n” のような正方形に近い文字を判別するのを助ける . シミュレーション実験では影響がわかりにくい, 後述する誤差やノイズが多量に発生するスキャナ取得画像に対する実験において, この効果が表れる .

3.3 スキャナ取得画像に対する実験

3.3.1 実験試料

スキャナで取得した画像を用いて, その回転角推定実験を行った . 用いた画像はシミュレーション実験で使用したものと同じ 5 種類である . この文書画像 $D1 \sim D5$ を A4 サイズに印刷し, 1 文書画像につき 2 種類の回転を発生させて 10 個の入力画像を取得した . 入力画像の解像度は 600dpi である .

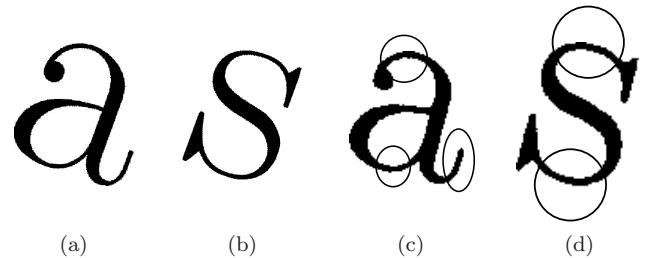


図 12 (a), (b) 標準画像 “a”, “s”, (c), (d) スキャナで取得した画像

スキャナで取得したため, 画像中には雑音が入っている . また, 連結成分の黒画素は細い部分が潰れて太くなっている . 例えば, 図 12 の標準画像とスキャン画像を比べると, 丸印で囲った部分が太くなっていることがわかる . また, $D5$ には大小様々な文字が存在しているが, そのうち小さいサイズの文字に関しては黒画素が全体的に太くなっている .

3.3.2 実験結果

不変量として $\mathbf{q} = (q_{hull}, q_{hole}, q_{ratio})$ を使用した場合, 入力 10 画像中 9 画像が補正に成功したことを確認した . なお, 本実験では真の回転角は不明なため目視により成否を判断した . また, 不変量として, $\mathbf{q} = (q_{hull}, q_{ratio})$ を使用した場合, 7 画像が補正に成功した . しかし, 各不変量を単独で, または不変量 2 つを組み合わせた (q_{hole}, q_{ratio}) , (q_{hole}, q_{ratio}) に関しては,

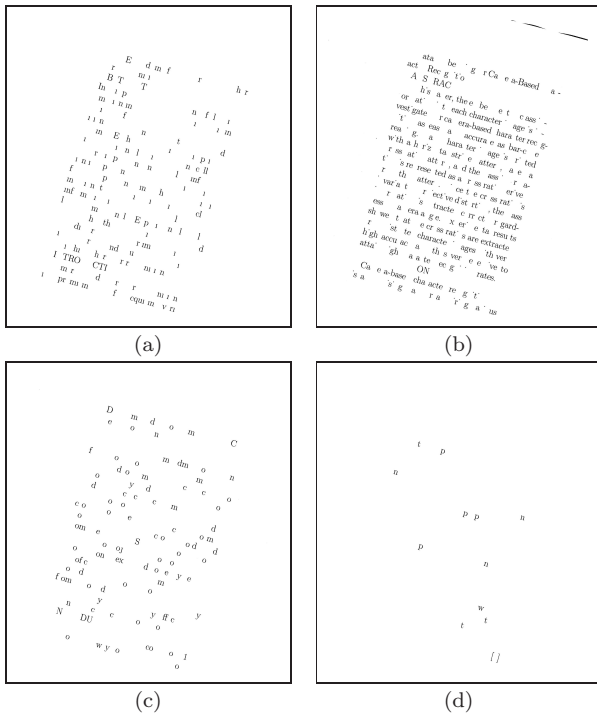


図 13 不変量 q_{hull} , q_{hole} , q_{ratio} を使用した D1 を回転させてスキャンした画像の事例参照状況 (a) 不変量, 変量共に正しい, (b) リジェクト, (c) 不変量失敗, 変量は正しい, (d) 不変量間違い, 変量失敗

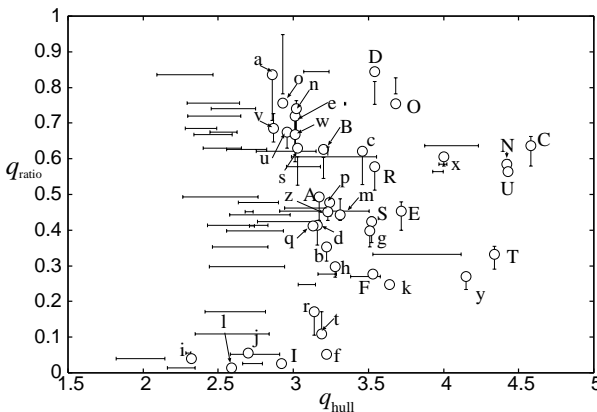


図 14 スキャナ取得画像における不変量 q_{hull} と q_{ratio} 測定値の最大値と最小値

入力画像全てにおいて補正は失敗した。

各不変量単独では十分なカテゴリ推定ができておらず、補正に失敗している。しかし、3つの不変量の組み合わせ方によっては、補正に成功した。これは、不変量 q_{hull} , q_{ratio} は前述の理由により誤差の大きい文字をリジェクトしあって、信頼度の高い文字を使用して回転角を推定しているためである。また、不変量 q_{hole} を用いることによって、この2種類の不変量では分別し難かったカテゴリを更にリジェクトし、より信頼度を高めている。例えば、図7で(“e”, “n”)や(“S”, “g”)は非常に近い値を有しているが、 q_{hole} を用いることで信頼度を高めている。図13に3つの不変量を用いて補正を行った際の事例参照結果を示す。同図の内不変量の参照を全て間違っている(C_X に真

のカテゴリが1つも含まれない)のは(c)であるが、3つの不変量の相互作用により、ほとんど参照されず、リジェクトが多くなっていることが確認できる。

以上より、不変量 q_{hull} , q_{hole} , q_{ratio} は相補的に作用している。不変量 q_{hole} のような参照されるカテゴリが多い場合も組み合わせることで誤差の大きい文字をリジェクトできるため、カテゴリ推定精度向上が期待できる。

図14に不変量 q_{hull} と q_{ratio} の測定値の最大値と最小値を示す。不変量 q_{ratio} は事例中の値と大幅に変わることは少ないが、不変量 q_{hull} は標準画像の値(図中の丸印)よりも小さく測定され、また、値の振れ幅が大きくなっている。これは図12で示すような、黒画素の潰れにより、黒画素の面積の比率が増し、標準画像よりも小さな値として計算されたと思われる。このことは図14を比較することでもわかる。なおこのため本実験では $\epsilon_p, \epsilon_{q_{hull}}$ を大きくして対応しておく必要があった。

4. まとめ

文書の回転角を推定するために、あらかじめ登録された事例、具体的には各文字カテゴリの回転変量と回転不変量を用いることで各連結成分(多くの場合、単文字に相当)ごとに独立して回転角を推定する方法を検討してきた。これまで不変量として、連結成分の外接矩形の面積と黒画素の面積の比を用いて実験を行ってきた。しかし、この不変量では、文字形状の小さな変化にも敏感であるという問題があった。本報告では、回転角推定精度の向上を目的としてを目的として、不変量の多次元化を提案した。また、計算機実験を通して精度評価を行い、不変量を多次元化することの有効性を示した。また、スキャナ取得画像に対しても実験を行い、本手法の有効性を確認した。

文献

- [1] Y. Ishitani, “Document Skew Detection Based on Local Region Complexity,” Proc. Int. Conf. Doc. Anal. Recog., pp. 49–52, 1993.
- [2] X. Jiang, H. Bunke, and D. Widmer-Klaja, “Skew Detection of Document Images by Focused Nearest-Neighbor Clustering,” Proc. Int. Conf. Doc. Anal. Recog., pp. 629–632, 1999.
- [3] 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.
- [4] 内田誠一, 酒井 恵, 岩村雅一, 大町真一郎, 黄瀬浩一, “事例に基づく文書画像の回転角推定,” 電子情報通信学会論文誌 (D), vol. J91-D, no.1, pp.136-138, Jan. 2008.
- [5] 酒井 恵, 内田誠一, 岩村雅一, 大町真一郎, 黄瀬浩一, “変量と不変量を用いた文書画像の事例ベース傾き補正,” 画像の認識理解シンポジウム (MIRU2007, 広島市立大学, 広島市) IS-4-01, 2007 7