

研究速報

事例に基づく文書画像の回転角推定

内田 誠一[†](正員) 酒井 恵^{††}(学生員)
 岩村 雅一^{†††}(正員) 大町真一郎^{††††}(正員)
 黄瀬 浩一^{†††}(正員)

Document Skew Estimation by Instance-Based Learning

Seichi UCHIDA[†], Member,

Megumi SAKAI^{††}, Student Member, Masakazu IWAMURA^{†††},

Shinichiro OMACHI^{††††}, and Koichi KISE^{†††}, Members

[†]九州大学大学院システム情報科学研究所, 福岡市

Faculty of Information Science and Electrical Engineering,
 Kyushu University, Fukuoka-shi, 819-0395 Japan

^{††}九州大学大学院システム情報科学府, 福岡市

Graduate School of Information Science and Electrical En-
 gineering, Kyushu University, Fukuoka-shi, 819-0395 Japan

^{†††}大阪府立大学大学院工学研究科, 堺市

Graduate School of Engineering, Osaka Prefecture Univer-
 sity, Sakai-shi, 599-8531 Japan

^{††††}東北大学大学院工学研究科, 仙台市

Graduate School of Engineering, Tohoku University,
 Sendai-shi, 980-8579 Japan

あらまし 各文字の回転変形に対する変量と不変量を事例として学習しておき、それらを利用することで文書画像の回転角を推定する方法を提案する。本手法は、文字単位で回転角を効率的に推定するため、文字列が直線かつ平行にレイアウトされているという仮定が不要であり、したがって様々なレイアウトの文書に利用可能である。

キーワード 回転角推定, 文書画像, 事例

1. ま え が き

スキャナで取得された文書画像の回転角(いわゆるスキュー)の推定及び補正は、OCRに必須の処理である。多くの従来法では、文書中の文字列が直線的かつ平行に配列されているという仮定に基づいて回転角を推定する。文書画像の射影ヒストグラムに基づく手法は、その典型的な例である。

本論文では、事例に基づいた回転角推定という、従来法とは全く異なったアプローチによる方法を提案する。本手法は文字列の配列に関する上述の仮定に基づいていない。したがって非常に自由なレイアウトをもった文書、例えば短い文字列やポスターなど、カメラで取得される文書画像[1]には、特に有用といえる。実験を通して、十分な精度で回転角が推定可能であることを示す。

本手法の第1の特徴は、各連結成分(多くの場合、単文字に相当)ごとに独立に回転角を推定する点であり、これがレイアウトに対する自由度を高めている。

以前にも局所的な回転角推定を行う手法がいくつか提案されているが[2]~[4]、1単語程度の範囲が必要であり、また文字列配列の直線性の仮定に基づいていた。

第2の特徴は、前述のように事例に基づいている点である。具体的には、各文字カテゴリーについて回転変量と不変量を事例として学習しておき、それを利用することで回転角を推定する。この処理はテーブル参照により非常に効率的に実現できる。

第3の特徴は、用いる変量・不変量を変えることで、回転以外の幾何変形推定にも拡張できる点である。この特徴は、カメラを用いた文書認識において発生する射影変換ひずみの推定・補正においても、本手法が利用できることを意味する。なお、この拡張の実証実験については本論文の範囲外とし、稿を改めて論ずる。

2. 事例に基づく回転角推定

2.1 推定手順

事例に基づく回転角推定を説明するために、以下のような非常に単純な方法を考える。まず各文字カテゴリー $c \in [1, \dots, C]$ について標準文字画像 R_c を準備し、次にそれらを角度 θ で回転させた $R_{c,\theta}$ を事例として登録しておく。入力画像中の連結成分 X の回転角推定時には、 X とすべての $R_{c,\theta}$ を重ね合わせ、最も照合した $R_{c,\theta}$ についてその角度 θ を X の回転角とすればよい。すなわち X のカテゴリー c と回転角 θ を総当り的に探索することになる。文書全体の回転角はこれら連結成分ごとの局所推定角を多数決など適当な方法で統合すれば求まる。ただし、この単純な方法は明らかに非常に効率が悪く現実的ではない。

これに対し本手法では、各文字カテゴリー c の回転不変量 q_c 及び回転変量 $p_c(\theta)$ を事例として登録しておき、それらを用いて効率的な回転角推定を行う。具体的には以下及び図1に示すように、回転不変量は X のカテゴリー c の推定に用いられ、変量は X の回転角 θ の推定に用いられる。

(0) 事例の作成(学習)

各カテゴリー c の標準文字画像 R_c について回転不変量 q_c を測定し、更に R_c を回転させながら(すなわち $R_{c,\theta}$ を生成しながら)変量 $p_c(\theta)$ を測定する。これらの組 $(p_c(\theta), q_c)$ を事例として登録しておく。

(1) 不変量及び変量の計算

入力文書画像中の各連結成分 X について変量 p_X 及び不変量 q_X を求める。

(2) 不変量によるカテゴリー c の推定

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 の誤差の許容範囲である．文書画像の回転について 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．

(3) 変量による X の回転角の推定

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 c 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．

(4) 文書全体の回転角 $\bar{\theta}$ の推定

以上 (1)~(3) の処理を文書中のすべての連結成分 X について行った後，文書全体の回転角 $\bar{\theta}$ を投票により求める．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

事例生成のための回転変量 $p_c(\theta)$ 及び不変量 q_c としては何を用いてもよいが，本論文では以下を用いた．

$$p_c(\theta) = \frac{R_{c,\theta} \text{の外接方形の面積}}{R_c \text{の黒画素の面積}}$$

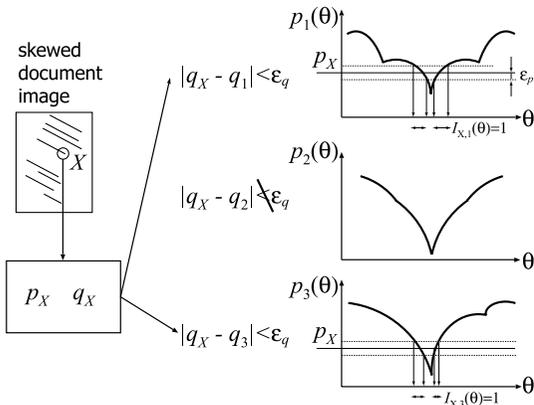


図 1 不変量 q_X と変量 p_X による回転角の推定
Fig. 1 Skew estimation by invariant and variant.

$$q_c = \frac{R_c \text{の凸包の面積}}{R_c \text{の黒画素の面積}}$$

各連結成分 X の変量 p_X 及び不変量 q_X についても同様に計算される．

画像 R_c の黒画素の面積で除していることから分かるように，これら変量及び不変量はいずれもスケール不変量である．位置不変量でもある．このため回転角推定はスケールや位置によらず， X の大きさが R_c と異なっても正しい回転角を推定できる．一方，変量としてスケール変量（かつ回転・位置不変量）を用いれば，スケールだけを推定できることになる．このように，1. で第 3 の特徴として述べたとおり，変量・不変量を変えることで多様な幾何変形を推定できる．

2.3 計算量

本手法の主要部分は，ステップ (2) 及び (3) において条件を満たすカテゴリ c や θ の区間を探索する部分である．いずれも逆引きテーブルを作っておくことで，1 連結成分当り $O(1)$ の計算量で済む．ここで p_X と q_X を求めるのに X の画素数程度の計算量がかかることを考えると，結局 1 文書画像の回転角推定に要する計算量はたかだか文書画像の総画素数のオーダーで済むことになる．射影ヒストグラムに基づく従来法では，この総画素数と回転角候補数の積の計算量が必要であることを考えると，本手法は非常に効率的であるといえる．

3. 実験

3.1 実験試料

評価実験のために，まず事例を作成した．具体的には，単一フォント (Times-Roman) の “A” から “Z”，“a” から “z” の 52 文字それぞれについて不変量 q_c を求め，更に -45° から 45° まで 0.1° 刻みで変量を測定して $p_c(\theta)$ を求めた．図 2 に実際に測定した変量 $p_c(\theta)$ の例を示す．

推定精度評価には，事例と同じ Times-Roman

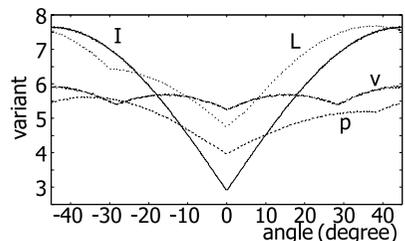


図 2 各カテゴリの変量 $p_c(\theta)$
Fig. 2 Variant $p_c(\theta)$.

(1) with the following l_2 and l_3 :

$$\begin{cases} l_2 = \frac{(L-l_1-2)(k-1)}{K-1} + \epsilon, \\ l_3 = L-l_1-l_2, \end{cases} \quad (2)$$

where ϵ is a positive constant specifying the minimum of l_2 and l_3 .

The above strategy is based on a simple linear quantization and may be weak against errors on the stripe widths l_1 , l_2 , and l_3 due to the insufficient camera resolution. In fact, larger k becomes, closer r_k and r_{k+1} become. Thus, a small error on the stripe widths may confuse these close cross ratios. Future work should focus on a more sophisticated strategy to avoid the confusion as possible. Recognition by cross ratio and shape similarity. In most cases, we cannot expect one-to-one assignment of K cross ratios to $||$ classes. Specifically, $||$ is often large (e.g., $|| > 1000$ for Chinese characters) whereas K is bounded by $L-l_1-2\epsilon$ (\approx character

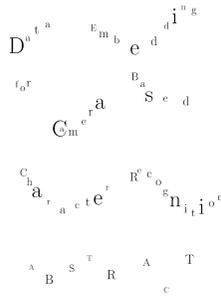


図 3 実験に使用した文書画像の例 (左より D4, D5)

Fig. 3 Subjected document images.

表 1 傾き推定誤差 (°)

Table 1 Estimation error (degree).

angle	D1	D2	D3	D4	D5	average (abs. error)
-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 (abs. error)	0.6	0.8	3.5	0.7	0.3	1.2

フォントで作成した 5 種類の文書画像 D1~D5 を $\pm 30^\circ, \pm 20^\circ, \pm 10^\circ, \pm 5^\circ, \pm 2^\circ, 0^\circ$ で回転させて生成した 55 個のテスト画像を用いた。画像 D4 と D5 を図 3 に示す。D3, D4 は数式を含む文書である。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得る。また D5 は文字が直線上に並んでおらず、従来法では回転角推定補正が困難な文書である。

3.2 本手法による傾き補正結果

55 枚のテスト画像に対する傾き推定誤差を表 1 に示す。1 例 (D3(-20°)) を除き、誤差 2.0° 以下で推定できている。また、D3, D4 のように事例をもたない数式部の文字を含む画像や、従来法では困難な D5 についても、本手法では正しく推定できている。図 4 は D4 (-5°) に対する投票結果 $h(\theta)$ であり、この場合真の回転角 (-5°) と $h(\theta)$ のピークが一致している。

表 1 にあるように、D3 (-20°) の推定誤差が -29.0°

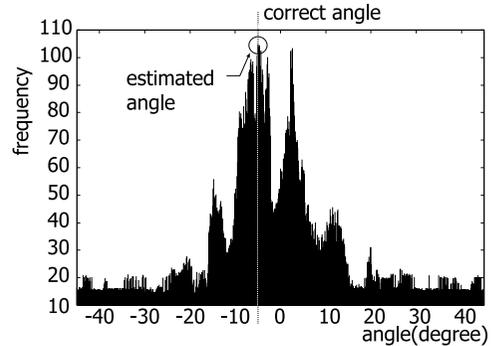


図 4 テスト画像 D4 (-5°) に対する投票結果 $h(\theta)$

Fig. 4 Voting result $h(\theta)$ of D4 (-5°).

と非常に大きくなっている。この原因には、不変量 q_X のノイズに対する不安定性が挙げられる。すなわち、いくつかのカテゴリーに属する X については、文字形状と解像度に起因して、推定されたカテゴリー集合 C_X に X の真のカテゴリーが含まれない場合があった。D3 には特にこれらのカテゴリーが多く含まれていたために、失敗となった。今後は他の不変量の利用や複数の不変量の利用によって安定化を図る必要がある。

4. むすび

変量と不変量を学習しておくことで、スキャナで取得された文書画像の回転角を効率的に推定する方法を提案した。また実験を通してその精度を確認した。特に従来法が苦手とする自由レイアウト文書についても正しく推定できることを示した。今後の課題としては、不変量の改良、及びカメラで取得した文書画像に起こる射影ひずみへの対応が挙げられる。

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,” 信学誌, vol.89, no.9, pp.836-841, Sept. 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.

(平成 19 年 7 月 30 日受付)