手書き文字列の非一様な傾き補正法

平 英二 * 　石田敏之 * 　内田 誠一 † 　迫江 博昭 †

* 九州大学大学院システム情報科学府
† 九州大学大学院システム情報科学研究院
〒 812-8581 福岡市東区箱崎 6-10-1
TEL:092-642-3888
E-mail: {taira, tishida, uchida, sakoe}@human.is.kyushu-u.ac.jp

あらまし　文字列の認識において，各構成文字の傾き補正は認識精度向上を図る上で必須技術である．従来法では，文字列内の各構成文字が一定の角度で傾いているという仮定の下，一様な傾き補正を行なっていた．これに対し，本論文では非一様な傾きも補正可能な手法を提案する．本手法では傾き補正問題を局所傾き角の最適推定問題として定式化する．その最適推定を動的計画法に基づくアルゴリズムにより行なう．一様な傾き補正法との比較実験を行ない，本手法の傾き補正能力が優れていることを定性的および定量的に示す．

キーワード　傾き補正　動的計画法　手書き文字列認識

Nonuniform Slant Correction of Handwritten Word

Eiji TAIRA*, Toshiyuki ISHIDA*, Seiichi UCHIDA†, and Hiroaki SAKOE†

{*Graduate School, †Faculty} of Information Science and Electrical Engineering,
Kyushu University
6-10-1 Hakozaki, Higashi-ku, Fukuoka-shi, 812-8581 Japan
TEL:+81-92-642-3888
E-mail: {taira, tishida, uchida, sakoe}@human.is.kyushu-u.ac.jp

Abstract Slant correction is an indispensable technique to improve recognition accuracy for handwritten word recognition. In convensional slant correction techniques, uniform slant correction was performed under the assumption that each word is written with a constant slant. In this paper, a nonuniform slant correction technique is presented where the slant correction problem is formulated as an optimal estimaiton problem of local slant angles at all horizontal positions. The optimal estimation is performed by a dynamic programming-based algorithm. Experimental results show the slant correction ability of the present technique over the uniform slant correction technique.

key words slant correction, dynamic programming, handwritten word recognition

1 はじめに

手書き文字列の認識において，各構成文字の傾き補正は認識精度向上を図る上で必須技術である．例えば，文字列を全体論的 (holistic) に認識する場合，傾き補正処理は形の正規化として用いられる．また，各構成文字を切り出し，1 文字ずつ個別に認識する解析的認識の場合，傾き補正は認識精度に大きく関わる文字の切り出しの精度向上のために必要となる．

手書き文字列の傾き補正法は，これまでにもいくつか検討されている．Bozinovic ら [1] および Kim ら [2] は，長い縦ストロークの平均角度を文字列の傾きとする手法を提案した．Guillevic ら [3] や Nicchiotti ら [4] および Kavallieratou ら [5] は，投影ヒストグラムから文字列の平均傾きを推定する手法を提案した．Simoncini ら [6] や Ding ら [7] および Britto ら [8] は，文字の輪郭線のチェインコードから文字列の平均傾きを推定する手法を提案した．これらの従来法はいずれも，文字列全体が一様な傾きで書かれているという仮定の下，文字列全体にわたる平均的文字傾きを推定して補正している．しかし，実際には各構成文字毎に傾きが異なる文字列もあり，一様な補正の結果を部分的に見ると，依然として傾きが残ったり，補正過多のために逆方向に傾いてしまうことも起こりうる．そのため，部分的に異なった傾きも補正できる非一様な傾き補正法が必要となる．

本論文では，傾き補正問題を局所傾き角の最適推定問題として定式化した非一様な傾き補正法を提案する．最適推定の方針は，長い縦ストロークの傾き角を検出し，その傾き角を近傍に滑らかに伝播することとする．最適推定は推定した局所傾き角の大局的および局所的妥当性を評価する関数および制約条件に基づいて行なう．評価関数を最大化する局所傾き角の系列は動的計画法 (dynamic programming : DP) に基づいたアルゴリズムにより効率的に探索できる．

以下では，本手法の基本原理を述べ，英文字列を対象として一様な傾き補正法との比較実験を行ない，本手法の有効性を示す．

2 非一様な傾き補正法

2.1 傾き補正問題の定式化

画像 $\boldsymbol{A} = \{a(i,j)\}$ をサイズ $M \times N$ の 2 値文字列画像とする．本手法では，非一様な傾き補正を実現するために，局所傾き角の系列 $\theta_1, \ldots, \theta_i, \ldots, \theta_M$ の最適推定を行なう (図 1)．ここで，θ_i は第 i 列の局所傾き角を表す．最適推定の方針は，入力画像内から長い縦ストロークの傾き角を検出し，変形の際に文字の形を崩さないためにその傾き角を近傍に滑らかに伝播することとする．これは長い縦ストロークの傾きはそのストロークを含む文字の傾きを強く反映しているという推測による．この方針に従って，$\theta_1, \ldots, \theta_i, \ldots, \theta_M$ の最適推定問題を次の評価関数の最大化問題として定式化する．

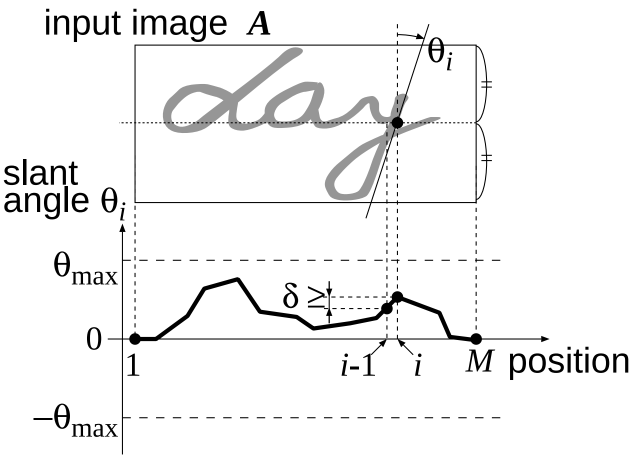

図 1: 非一様な局所傾き角の推定

$$F(\theta_1, \ldots, \theta_i, \ldots, \theta_M) = \sum_{i=1}^{M} f_i(\theta_i \mid \theta_{i-1}) \quad (1)$$

但し，θ_i は $\Theta = [-\theta_{\max}, \theta_{\max}]$ の範囲内の角度とし，最大化は次の単調連続性条件の下で行なう．

$$|\theta_i - \theta_{i-1}| \leq \delta \quad (2)$$

ここで，δ は正の定数とする．

関数 $f_i(\theta_i \mid \theta_{i-1})$ は θ_i の妥当性を評価する 2 つの関数の和として定義される．

$$f(\theta_i \mid \theta_{i-1}) = s_i(\theta_i) + \rho(\theta_i \mid \theta_{i-1}) \quad (3)$$

関数 $s_i(\theta_i)$ は第 i 列において θ_i 傾いた長い縦ストロークの存在を確認し，θ_i の妥当性を評価する．関数 $\rho(\theta_i \mid \theta_{i-1})$ は θ_{i-1} から θ_i への傾き角の変化の滑らかさを評価する．但し，$i = 1$ のとき $\rho(\theta_i \mid \theta_{i-1}) = 0$ とする．評価関数 (1) を最大化することで，関数 $s_i(\theta_i)$ によって長い縦ストロークの傾き角が検出され，関数 $\rho(\theta_i \mid \theta_{i-1})$ によってその傾き角は近傍に滑らかに伝播できる．

2.2 動的計画法による解法

評価関数 (1) を最大化する $\theta_1, \ldots, \theta_i, \ldots, \theta_M$ は図 2の DP アルゴリズムで探索できる．図 2において

```
    /* Initialization: */
1   for all θ_1 ∈ Θ do g_1(θ_1) := s_1(θ_1);
    /* DP Recursion: */
2   for i = 2 to M do
3     for all θ_i ∈ Θ do begin
4       g_i(θ_i) := max_{θ_{i-1}} [g_{i-1}(θ_{i-1}) + f_i(θ_i | θ_{i-1})];
5       b_i(θ_i) := θ_{i-1} which gives
                        the maximum at Step 4;
6     end;
    /* Backtracking: */
7   θ̄_M := argmax_{θ_M ∈ Θ} g_M(θ_M);
8   for i = M downto 2 do θ̄_{i-1} := b_i(θ̄_i);
```

図 2: DP アルゴリズム

$g_i(\theta_i)$ は第 i 列までの $f_k(\theta_k|\theta_{k-1})$ の最大累積値である．その最大累積値を求める Step 4 は DP 漸化式と呼ばれる．DP 漸化式における最大値選択は条件 (2) を満たす θ_{i-1} について行なう．θ_i の最適値を $\overline{\theta}_i$ と表すと，$\overline{\theta}_i$ は Step 7, 8 のバックトラック処理により求められる．

2.3 傾き角の離散化

図2の DP アルゴリズムは制御変数となる局所傾き角 θ_i が連続量であるためにディジタル計算機ではそのまま実行することはできない．そこで 2 点 $(p_i, 1)$, (q_i, N) を用いて，

$$\theta_i = \arctan \frac{q_i - p_i}{N - 1} \quad (4)$$

により θ_i を離散化する (図 3)．但し，p_i, q_i は整数値であり，

$$\left\lfloor \frac{p_i + q_i}{2} \right\rfloor = i \quad (5)$$

$$|p_i - i| \leq W \ , \ |q_i - i| \leq W \quad (6)$$

を満たすものとする (図 4)．ここで，$\lfloor x \rfloor$ は x を越えない最大の整数とする．また，条件 (6) は許容する傾きを規定する条件で，W は正の整数定数とする．以下，$(p_i, 1), (q_i, N)$ を結ぶ線分を補正線分と呼ぶ．

以上のような傾き角の離散化により，$\theta_1, \ldots, \theta_i, \ldots, \theta_M$ の最適推定問題は $p_1, q_1, \ldots, p_i, q_i, \ldots, p_M, q_M$ の最適推定問題として捉えられる．その場合，単調連続性条件 (2) を p_i, q_i についての条件として表すと次のようになる．

$$p_i \geq p_{i-1} \ , \ q_i \geq q_{i-1} \quad (7)$$

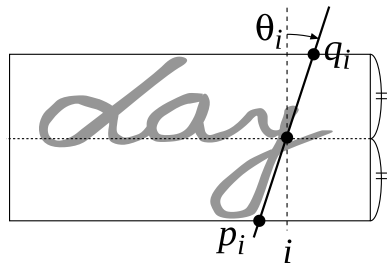

図 3: 局所傾き角 θ_i と p_i, q_i の関係

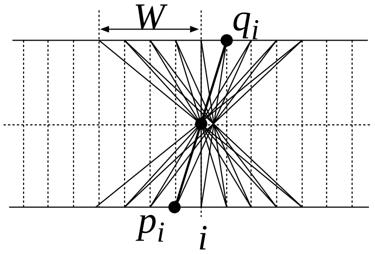

図 4: 第 i 列において可能な p_i, q_i の組合せ

この条件は単独では単調性条件であるが，条件 (5) と合わせて考えると p_i と p_{i-1} の間隔は高々3 となり，同時に連続性も保証している．また，評価関数は表現が変わり，関数 f_i, s_i, ρ はそれぞれ $f(p_i, q_i \mid p_{i-1}, q_{i-1})$, $s(p_i, q_i), \rho(p_i, q_i \mid p_{i-1}, q_{i-1})$ となる．p_i, q_i はそれ自体が位置情報であるため，f, s の添字 i は取り除いている．さらに，図 2の Step 3 の θ_i のループは条件 (5),(6) の下で可能な p_i, q_i の組合せのループとなり，Step 4 の DP 漸化式は次のように修正される．

$$g_i(p_i, q_i) := \max_{p_{i-1}, q_{i-1}} [g_{i-1}(p_{i-1}, q_{i-1}) + f(p_i, q_i \mid p_{i-1}, q_{i-1})]; \quad (8)$$

ここで，最大値選択は条件 (5)-(7) の下で可能な p_{i-1}, q_{i-1} の組合せについて行なう．

2.4 評価関数

関数 $s(p_i, q_i)$ は $(p_i, 1), (q_i, N)$ を結ぶ補正線分上において長い縦ストロークの存在を確認し，p_i, q_i の妥当性を評価する．具体的には補正線分に幅 λ を持たせ，その領域内での連結した黒画素の高さの最大値を計算する (図 5(a))．補正線分に幅を持たせたのは曲線ストロークにも対応できるようにするためである．ここで計算の簡単化のため，黒画素の連接性は縦方向のみで判断する (図 5(b))．ところで，垂直方向に短いストロークは横ストロークの一部もしくは縦ストロークであっても文字の傾きとは無関係であ

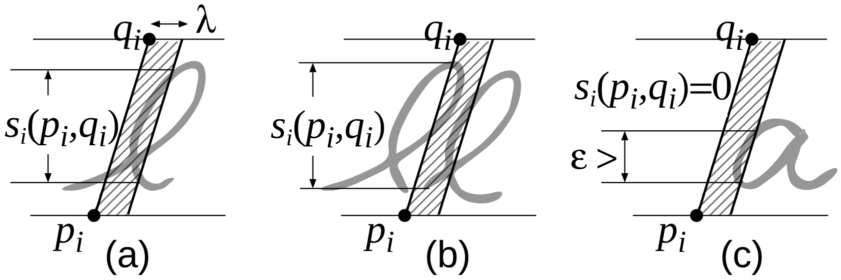

図 5: 関数 $s(p_i, q_i)$ の計算

ると推測される．またそのようなストロークは文字自体の形状が強く反映されており，評価対象としてしまうことはかえって不自然な補正結果を生じさせる．そこである閾値 ϵ を与えて，最大高さが ϵ 未満の場合 $s(p_i, q_i) = 0$ とする (図 5(c))．

関数 $\rho(p_i, q_i \mid p_{i-1}, q_{i-1})$ は傾き角の変化の滑らかさを評価するペナルティ関数である．具体的には次の式から求められる．

$$\rho(p_i, q_i \mid p_{i-1}, q_{i-1}) = -\alpha\Gamma_1 - \beta\Gamma_2 \tag{9}$$

ここで，α, β は正の重み係数である．Γ_1 は $\theta_i \neq \theta_{i-1}$，すなわち $p_i \neq p_{i-1} + 1$ または $q_i \neq q_{i-1} + 1$ のときに課せられるペナルティで，その値は補正線分上の黒画素の個数とする．文字と重なる補正線分ほど Γ_1 は大きい値を示す．よって，$-\alpha\Gamma_1$ を最大化，すなわち $\alpha\Gamma_1$ を最小化することで，1 文字毎の局所傾き角はできるだけ一定の角度となるように推定され，傾き角の変化は文字と文字の間で起きるようになる．Γ_2 は隣り合う列の補正線分の一端が交わるとき，すなわち $p_i = p_{i-1}$ または $q_i = q_{i-1}$ のときに課せられるペナルティで，その値は補正線分上の交わった一端の近傍にある黒画素の個数とする．本論文では近傍の範囲を $N/4$ としている．補正線分の一端が一点に集中した場合，傾き補正後の画像が不自然な形になることがある．Γ_2 はそれを防ぐためのペナルティである．

2.5　計算量

2.3節の離散化を考えた上での図 2の DP アルゴリズムに要する計算量を求める．この計算量は DP 漸化式の計算回数と DP 漸化式 1 回当たりの計算量の積で与えられる．

DP 漸化式の計算回数は Step 2,3 の繰り返し回数に等しい．Step 3 の繰り返し回数は条件 (5),(6) の下で可能な p_i, q_i の組合せの数であり，その数は $4W+1$ である．これと Step 2 の繰り返し回数 M の積より，DP 漸化式の計算回数は $O(MW)$ となる．

DP 漸化式 1 回当たりの計算量は関数 $f_i(\theta_i \mid \theta_{i-1})$ の計算量と最大値選択回数の積となる．前者は $O(N)$ であり，後者は条件 (5),(7) より高々7 となるため，DP 漸化式 1 回当たりの計算量は $O(N)$ となる．

以上より図 2の DP アルゴリズム全体の計算量は $O(MNW)$ となり，本手法は従来の一様な傾き補正法とほぼ同程度の計算量で済む．

2.6　傾き補正後の画像

画像 $\boldsymbol{B} = \{b(i,j)\}$ を画像 $\boldsymbol{A}$ の傾き補正処理後の画像とする．この $\boldsymbol{B}$ は，最適推定された $p_1, q_1, \ldots, p_i, q_i, \ldots, p_M, q_M$ より，$\boldsymbol{A}$ の $(p_i, 1), (q_i, N)$ を結ぶ補正線分上の画素を $\boldsymbol{B}$ の第 i 列上にマッピングすることで得られる．

3　実験と結果

本節では，一様な傾き補正法と本手法の定性的および定量的比較実験について述べ，その結果により本手法の傾き補正能力およびその有効性を示す．ここで比較対象とした一様な傾き補正法には，本手法に傾き角が画像全体で一定であるという制約を加えた方法を用いた．本実験において，W を 63，λ を 4，ϵ を 25，α を 1，β を 2 とした．なお，サイズ 256×64 の画像で傾き補正にかかる計算時間は 650 msec(PentiumIII,500 MHz) であった．

3.1　定性的評価実験

入力画像として CEDAR CDROM-1 (city and state words) に含まれる手書き英単語画像を用いて，本手法による傾き補正の効果を定性的に評価する実験を行なった．各画像には以下の手順で前処理を行なった．画像の縦サイズが 64 画素になるように正規

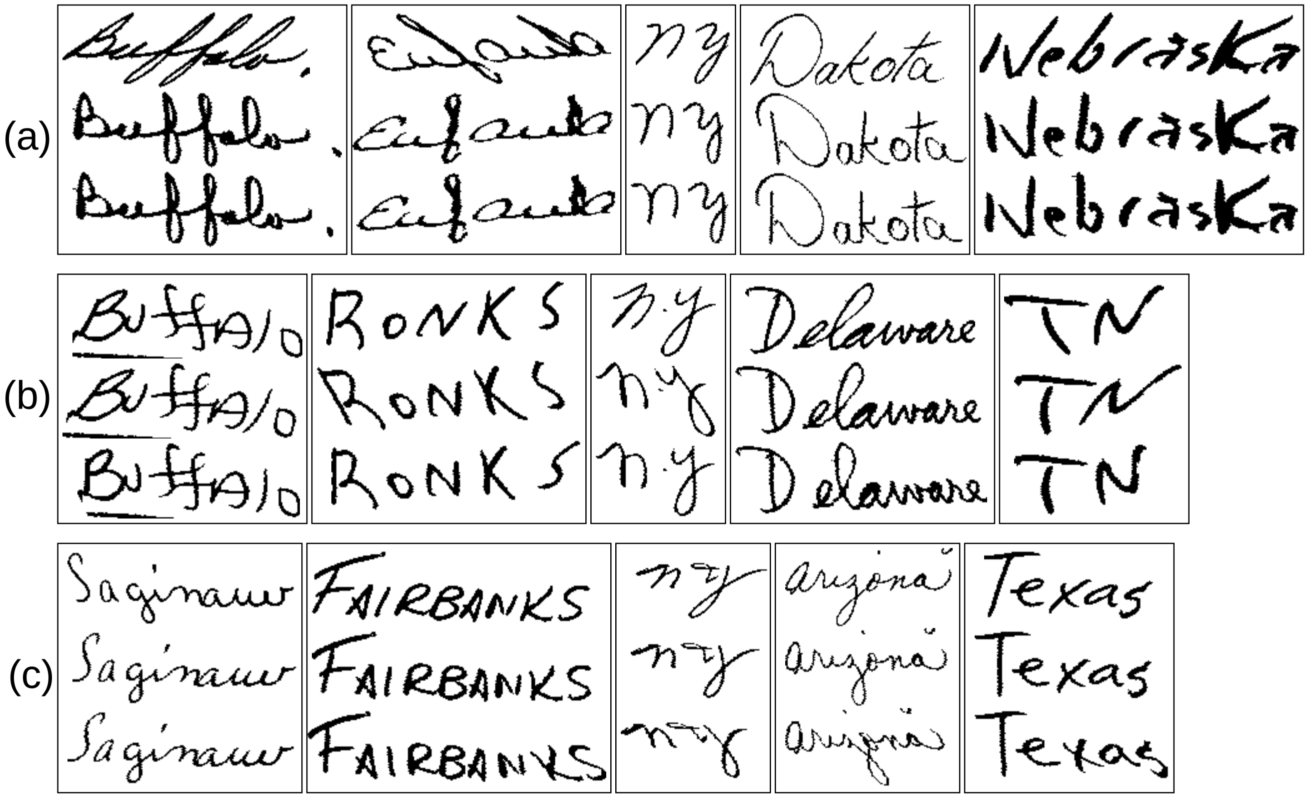

図 6: 傾き補正の例; 各単語について前処理後の入力画像 (上), 一様な傾き補正処理後の画像 (中), 非一様な傾き補正処理後の画像 (下) を示してある: (a) 両方とも良好な結果を示した例; (b) 非一様な傾き補正の方が良好な結果を示した例; (c) 非一様な傾き補正で補正過多が起きた例

化した．その際，画素の輝度値は共 1 次内挿法を用いて補間した．さらに，2 値化処理を施し，左右に W 画素分の空白列を加えた．

図 6に傾き補正の例を示す．図 6(a),(b) では，本手法によって，どの入力画像に対しても，傾いていた文字が形を崩すことなく立体に近い形に補正されているのが分かる．特に図 6(b) では，一様な傾き補正では補正が不十分もしくは過多となってしまう文字列に対し，本手法ではうまく補正できているのが分かる．しかし，図 6(c) のように，本手法では”X” のようにもともと斜線で構成される文字については過剰な補正が起きる．さらにその影響で近傍の文字まで傾いてしまうこともある．これは本手法における問題点である．

3.2 定量的評価実験

一様な傾き補正法と本手法の定量的な比較実験として，人為的に傾けた活字英単語画像を用いて，各列における局所傾き角の推定値と理論値の平均 2 乗誤差を測定する実験を行なった．図 7に原画像となる 5 種類の活字英単語画像およびそれらを一定の傾き角で一様に変形した画像と列毎に傾き角を正弦関数に従って変化させながら変形した画像を示す．また，表 1に測定した平均 2 乗誤差，すなわち $\sum_{i=1}^{M}(\overline{\theta}_i - \psi_i)^2/M$ の値を示す．ここで ψ_i は人為的に傾けた際の第 i 列の局所傾き角の理論値である．一様に傾けた場合，どちらの手法とも誤差は小さく，適切に対応できていることが分かる．image4,5 で本手法に多少の誤差が出ているのは，前述したように斜線で構成される文字の影響による．非一様に傾けた場合，本手法の傾き補正能力は一様な傾き補正法の傾き補正能力より明らかに優れていることが分かる．

さらに局所傾き角の理論値と本手法における推定値を列毎に比較したグラフを図 8に示す．ここでは用例として image1 (”Kentucky”) を用いた．図 8を見ると本手法が理論値を追従できているのが分かる．

4 まとめ

本論文では手書き文字列認識において問題となる各構成文字の傾きを非一様に補正する方法を提案した．本手法では傾き補正問題を局所傾き角の最適推定問題として定式化し，動的計画法 (dynamic programming : DP) に基づいたアルゴリズムにより効率

図 7: 人為的に傾けた英単語画像: (a) 原画像; (b) 一定の傾き角で一様に変形した画像; (c) 列毎に傾き角を正弦関数に従って変化させながら変形した画像

表 1: 図 7(b),(c) における局所傾き角の理論値と推定値の平均 2 乗誤差 ($\times 10^{-3}$)

	constant slant					sinusoidally changing slant				
image no.	1	2	3	4	5	1	2	3	4	5
uniform correction	0	0	0	0	0	427	537	286	185	186
nonuniform correction	0.188	0.007	0.337	8.47	6.66	18.3	169	41.5	110	152

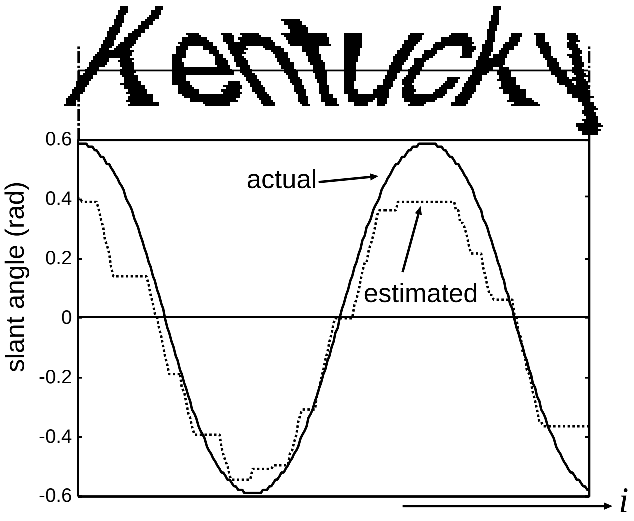

図 8: 図 7(c) の image1(“Kentucky”) における局所傾き角の理論値と非一様な傾き補正による推定値

的な探索を実現した．一様な傾き補正法と本手法との比較実験を行ない，本手法の傾き補正能力およびその有効性を示した．またその結果より本手法には補正過多の問題があることを指摘した．

今後の課題として，本手法を利用した文字列認識システムの作成が挙げられる．最近の文字列認識は Kim ら [2] のような認識処理の過程で文字の切り出し位置も同時に決定していく方法が 1 つの主流となっている．本手法をそのような認識システムに利用すれば，認識処理および文字の切り出しと同時に傾き補正処理も行なう認識システムが実現できると考えられる．その場合，文字種を仮定した上で傾き補正を行なうことになるため，本手法の補正過多の問題も解決すると期待される．

謝辞

本研究の一部は文部省科学研究費 (基盤研究 C(2), No.10680385) によった.

参考文献

[1] R.M. Bozinovic and S.N. Srihari. Off-Line cursive script word recognition. IEEE Trans. PAMI, Vol. 11, No. 1, pp. 68–83, Jan. 1989.

[2] G. Kim and V. Govindaraju. A lexicon driven approach to handwritten word recognition for real-time applications. IEEE Trans. PAMI, Vol. 19, No. 4, pp. 366-379, April 1997.

[3] D. Guillevic and C.Y. Suen. Cursive script recognition: a sentence level recognition scheme. Proc. 4th IWFHR, pp. 216–223, Dec. 1994.

[4] G. Nicchiotti and C. Scagliola. Generalised projections: a tool for cursive handwriting normalisation. Proc. 5th ICDAR, pp. 729-732, Sep. 1997.

[5] E. Kavallieratou, N. Fakotakis and G. Kokkinakis. A slant removal algorithm. Patt. Recog., Vol. 33, No. 7, pp. 1261–1262, July 2000.

[6] L. Simoncini and Zs.M. Kovács-V. A system for reading USA census'90 hand-written fields. Proc. 3th ICDAR, Vol. II, pp. 86–91, Aug. 1995.

[7] Y. Ding, F. Kimura, Y. Miyake and M. Shridhar. Accuracy improvement of slant estimation for handwritten words, Proc. 15th ICPR, Vol. 4, pp. 527–530, Sep. 2000.

[8] A. D. S. Britto JR., R. Sabourin, E. Lethelier, F. Bortolozzi and C. Y. Suen. Improvement in handwritten numeral string recognition by slant normalization and contextual information. Proc. 7th IWFHR., pp. 323–332, Sep. 2000.