PAPER

Nonuniform Slant Correction for Handwritten Word Recognition

Eiji TAIRA[†a)], *Student Member*, Seiichi UCHIDA[††], *Member*, *and* Hiroaki SAKOE[††], *Fellow*

SUMMARY Slant correction is a preprocessing technique to improve segmentation and recognition accuracy for handwritten word recognition. All conventional slant correction techniques were performed by the estimation of the average slant angle and the shear transformation. In this paper, a nonuniform slant correction technique for handwritten word recognition is proposed where the slant correction problem is formulated as a global optimal estimation problem of the sequence of local slant angles. The optimal estimation is performed by a dynamic programming based algorithm. From experimental results it was shown that the present technique outperforms conventional uniform slant correction techniques.

key words: *slant correction, dynamic programming, handwritten word recognition*

1. Introduction

In handwritten word recognition, slant correction is a widely used preprocessing technique to improve segmentation performance, which significantly affects recognition accuracy. In conventional slant correction techniques [1]–[9], the average slant angle of a word is estimated and then corrected *uniformly* by shear transformation. The average slant angle is estimated by averaging angles of near-vertical strokes [1], [2] and analyzing projection histograms [3]–[6] and statistics of chain-coded stroke contours [7]–[9].

These uniform slant correction techniques will perform successfully under the assumption that each word is written at a constant slant. The fact, however, is that the slant angle fluctuates at every component character in a word due to writer's habit, the inherent shape of each character, writing position, and so on. This makes it necessary to estimate local slant angles and to correct slants *nonuniformly*. The necessity of nonuniform slant correction is justified by Britto et al.'s result [9] that there are significant gaps between the average slant angle and the individual slant angles of component characters and the gaps degrade the performance of segmentation based word recognition systems.

It is noteworthy that Hase et al. [10] have proposed a realignment technique for inclined and curved texts, where the inclination and the curve of a text is approximated by some quadratic functions and then its component characters are realigned horizontally and slant-corrected using those functions. This technique, however, will not be appropriate for the slant correction of handwritten words whose component characters have slants regardless of the shape of their text line.

In this paper, a nonuniform slant correction technique is proposed where the slant correction problem is formulated as a global optimal estimation problem of local slant angles at all horizontal positions. The optimal estimation is governed by an objective function designed to globally evaluate the sequence of local slant angles and several constraints designed to impose the smoothness of the angle fluctuation. The optimal sequence maximizing the objective function subject to the constraints is efficiently searched for using a dynamic programming (DP) based algorithm.

2. Nonuniform Slant Correction

2.1 Problem Formulation

Let $\boldsymbol{A} = \{a(i, j)\}$ denote a binary-valued word image of size M (width) $\times N$ (height). The nonuniform slant correction technique proposed in this paper is based on optimal estimation of local slant angles at all horizontal positions of $\boldsymbol{A}$ (Fig. 1). Using the line segment between pixels $(i, 1)$ and (p_i, N) to represent the local slant angle at position i, the optimal correction of the local slants is equivalent to the optimal estimation of the sequence $p_1, \ldots, p_i, \ldots, p_M$. Hereafter, the line segment is called a *correction line* and denoted as $\boldsymbol{a}_i(p_i)$.

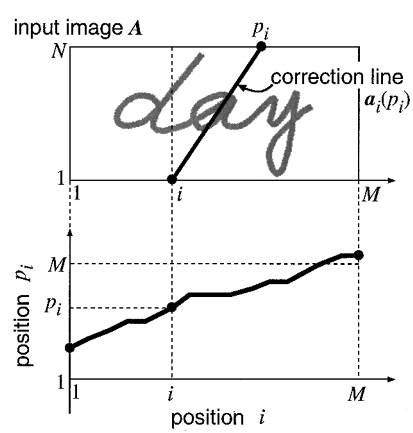

Fig. 1 Nonuniform slant.

Manuscript received April 14, 2003.
Manuscript revised October 24, 2003.

[†]The author is with the Graduate School of Information Science and Electrical Engineering, Kyushu University, Fukuoka-shi, 812–8581 Japan.
[††]The authors are with the Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka-shi, 812–8581 Japan.
a) E-mail: taira@human.is.kyushu-u.ac.jp

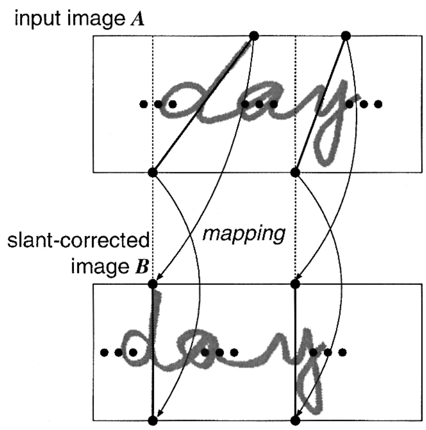

Fig. 2 Slant correction process.

The optimal estimation problem of the sequence $p_1, \ldots, p_i, \ldots, p_M$ is formulated as the maximization problem of the objective function

$$F(p_1, \ldots, p_i, \ldots, p_M) = \sum_{i=1}^{M} f_i(p_i \mid p_{i-1}), \tag{1}$$

where $f_i(p_i \mid p_{i-1})$ [†] is a function to evaluate p_i in the context of the value of p_{i-1}. The details of $f_i(p_i|p_{i-1})$ will be described in Sect. 2.2. For smoothing the estimating sequence, the sequence is subjected to a monotonicity and continuity constraint and a range limitation defined as

$$0 \leq |p_i - p_{i-1}| \leq 2, \tag{2}$$

$$|p_i - i| \leq W, \tag{3}$$

where W is a positive constant to specify the range of compensable slant angles as $[-\tan^{-1}(W/N), \tan^{-1}(W/N)]$.

Let $p_i^{\rm opt}$ denote the optimally estimated p_i. Slant-corrected images, denoted as $\boldsymbol{B} = \{b(i, j)\}$, are obtained by *mapping* the pixels on the correction line $\boldsymbol{a}_i(p_i^{\rm opt})$ of $\boldsymbol{A}$ onto the ith column of $\boldsymbol{B}$, for every position i (Fig. 2). This fact indicates that the nonuniform slant correction is closely related with nonlinear elastic image matching, such as piecewise linear two-dimensional warping [11]. It could be also noted that the above constraints are meaningful not only to smooth the optimal sequence $p_1^{\rm opt}, \ldots, p_i^{\rm opt}, \ldots, p_M^{\rm opt}$ but also to preserve the topological structure of the input word through the mapping.

2.2 Objective Function

The function $f_i(p_i \mid p_{i-1})$ is defined as a weighted sum of three functions, i.e.,

$$\begin{aligned} f_i(p_i \mid p_{i-1}) &= \omega_s s_i(p_i) + \omega_c c_i(p_i) \\ &\quad + (1 - \omega_s - \omega_c)\gamma_i(p_i \mid p_{i-1}), \end{aligned} \tag{4}$$

where ω_s, ω_c are nonnegative weights and $\omega_s + \omega_c \leq 1$.

The first function $s_i(p_i)$ is designed to measure a confidence level of the existence of a long stroke on the correction line $\boldsymbol{a}_i(p_i)$. This design policy is based on the observation that the long vertical strokes show the slant angle more clearly than other strokes. Here we use the maximum length of connected black pixels on the correction line as the confidence level. For suppressing the effects of short vertical strokes and horizontal strokes whose slant angles are less related with the character slant, $s_i(p_i)$ is set as 0 if the maximum length is shorter than $N/3$.

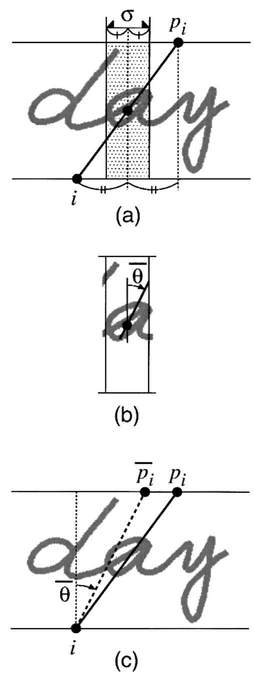

Fig. 3 Calculation of the horizontal position $\overline{p}_i$. Firstly, (a) the $\sigma \times N$ rectangular region is decided from p_i. Secondly, (b) the average slant angle $\overline{\theta}$ in this region is estimated. Lastly, (c) $\overline{p}_i$ is given by $i + N \tan \overline{\theta}$.

The second function $c_i(p_i)$ is designed to measure the difference between the slant of $\boldsymbol{a}_i(p_i)$ and the average slant around $\boldsymbol{a}_i(p_i)$. This is based on the consideration that the average slant angle in a small region is not far from the optimal local slant angle. In particular, using the chain code based method [7]–[9] as an average slant estimation technique, the slant angle of characters which inherently have slanted strokes, such as "X", can be estimated with reasonable accuracy. The function $c_i(p_i)$ is defined as

$$c_i(p_i) = -|p_i - \overline{p}_i|, \tag{5}$$

where $\overline{p}_i$ is the horizontal position which is calculated from the average slant angle in the $\sigma \times N$ rectangular region that centers the middle point of $\boldsymbol{a}_i(p_i)$ (Fig. 3). Local average slant angles are estimated by the chain code based method proposed by Ding et al. [8].

The third function $\gamma_i(p_i \mid p_{i-1})$ [††] is used to evaluate a smoothness between local slant of two consecutive position i and $i-1$, and defined as

[†] Although the function f_1 should be denoted as $f_1(p_1)$, the function f_i is denoted as $f_i(p_i|p_{i-1})$ regardless of i for notational simplicity.

[††] For $i = 1$, the value of the function γ_1 is set to 0 regardless of p_1.

$$\gamma_i(p_i \mid p_{i-1}) = \begin{cases} 0 & \text{if } p_i = p_{i-1} + 1 \\ -\Gamma_i(p_i) & \text{otherwise,} \end{cases} \tag{6}$$

where $\Gamma_i(p_i)$ is the number of black pixels on $\boldsymbol{a}_i(p_i)$. By maximizing $\gamma_i(p_i|p_{i-1})$, the local slant angles tend to be the same at i and $i-1$ (i.e., $p_i = p_{i-1} + 1$). Note that at ligatures and blanks, $\gamma_i(p_i|p_{i-1})$ becomes very small value regardless of p_i and p_{i-1}. Thus, this formulation allows quick changes in slant angles at ligatures and blanks.

2.3 Solution by DP

Considering a sequential optimization process of $p_1, \ldots, p_i, \ldots, p_M$, the process has the Markov property. This is because under the definition of the objective function (1) and the constraints (2) and (3), the value of p_{i-1} is necessary for determination of p_i, and the other past values (e.g., p_{i-2}) are not necessary. It is well known that the optimization problem with the Markov property can be efficiently solved using DP.

Figure 4 shows a DP algorithm for the optimization problem formalized in Sect. 2.1. The value $g_i(p_i)$ is the maximum (i.e., optimal) cumulated value of $f_k(p_k|p_{k-1})$ up to $k = i$. Step 5 is so-called DP recursion, and can be rewritten as

$$g_i(p_i) := \omega_s s_i(p_i) + \omega_c c_i(p_i) + \max_{p_{i-1}=p_i-\{0,1,2\}} \left[\begin{array}{l} g_{i-1}(p_{i-1})+ \\ (1-\omega_s-\omega_c)\gamma_i(p_i|p_{i-1}) \end{array} \right]. \tag{7}$$

Step 8 and 9 are the backtracking procedure to obtain the optimal sequence $p_1^{\mathrm{opt}}, \ldots, p_i^{\mathrm{opt}}, \ldots, p_M^{\mathrm{opt}}$.

The resulting DP algorithm requires $O(NMW)$ computations, since the calculation of $g_i(p_i)$ (i.e., the DP recursion) requires $O(N)$ computations and the number of the calculation of the DP recursion is $O(MW)$. When $M = 355, N = 64$, and $W = 90$ (i.e., the range of compensable slant angles is $[-54°, +54°]$), the DP algorithm required 25 ms on a PC (Xeon 1.7 GHz).

/* Initialization: */
1 **for all** $p_1 \in [1-W, 1+W]$ **do**
2 $\quad g_1(p_1) := \omega_s s_1(p_1) + \omega_c c_1(p_1)$
/* DP Recursion: */
3 **for** $i := 2$ **to** M **do**
4 $\quad$ **for all** $p_i \in [i-W, i+W]$ **do begin**
5 $\quad\quad g_i(p_i) := \max_{p_{i-1}=p_i-\{0,1,2\}} \left[g_{i-1}(p_{i-1}) + f_i(p_i|p_{i-1})\right]$
6 $\quad\quad b_i(p_i) := p_{i-1}$ which gives the maximum at Step 5
7 $\quad$ **end**
/* Backtracking: */
8 $p_M^{\mathrm{opt}} := \operatorname{argmax}_{p_M \in [M-W, M+W]} g_M(p_M)$
9 **for** $i := M$ **downto** 2 **do** $p_{i-1}^{\mathrm{opt}} := b_i(p_i^{\mathrm{opt}})$

Fig. 4 DP algorithm.

3. Experimental Results

Through several experiments, the effectiveness of the present nonuniform slant correction technique was qualitatively and quantitatively evaluated and compared with that of the conventional uniform slant correction technique proposed by Ding et al. [8].

3.1 Qualitative Evaluation

For qualitative evaluation of the present technique, several handwritten word images (city and state words) from the CEDAR CDROM database were subjected to a slant correction test. In advance to applying the present technique, each word image was binarized, linearly scaled to be $N = 64$, and then padded W columns of blanks at both left and right sides. The parameters of the present technique were set as $\sigma = 11$, $\omega_s = 0.56$, and $\omega_c = 0.26$. (While these parameters were experimentally determined, the investigation in Sect. 3.3 will show that this parameter setting is not a delicate task.)

Figure 5 shows original images and their slant-corrected images. Both of uniform and nonuniform slant correction techniques give good results for nearly uniformly slanted words (Fig. 5 (a)). When obvious nonuniform slant appears (Fig. 5 (b)), the present technique provides near-perfect correction while the conventional technique fails. Note that the present technique can correct the slants of component characters preserving their inherent shapes.

To clarify the necessity of the three functions $s_i(p_i)$, $c_i(p_i)$ and $\gamma_i(p_i|p_{i-1})$, slant correction was performed without one of these three functions. Figure 6 shows original images and their slant-corrected images. Firstly, when $s_i(p_i)$ was removed (Fig. 6 (a)), the slant correction was unsuccessful around long strokes. Secondly, when $c_i(p_i)$ was removed (Fig. 6 (b)), several characters which inherently have slanted strokes were excessively deformed. Finally, when $\gamma_i(p_i|p_{i-1})$ was removed (Fig. 6 (c)), most of slant-corrected images became unnatural due to the large fluctuation on estimated local slant angles. The reason why such degradations occur can be easily explained from the design policy of each function $s_i(p_i)$, $c_i(p_i)$ and $\gamma_i(p_i|p_{i-1})$, described in Sect. 2.2. The necessity of the three functions will be further confirmed through the quantitative investigation in Sect. 3.3.

3.2 Evaluation of Estimation Accuracy

For the quantitative evaluation of the present technique, an experiment to measure the mean error between given and estimated slant angles was conducted by using artificially slanted word images. Figures 7 (a), (b) and (c) show original machine-printed word images, slanted images at a constant angle of 30°, and nonuniformly slanted images at angles sinusoidally changing in $[-30°, 30°]$, respectively.

Table 1 shows the mean error between given and estimated slant angles, i.e.,

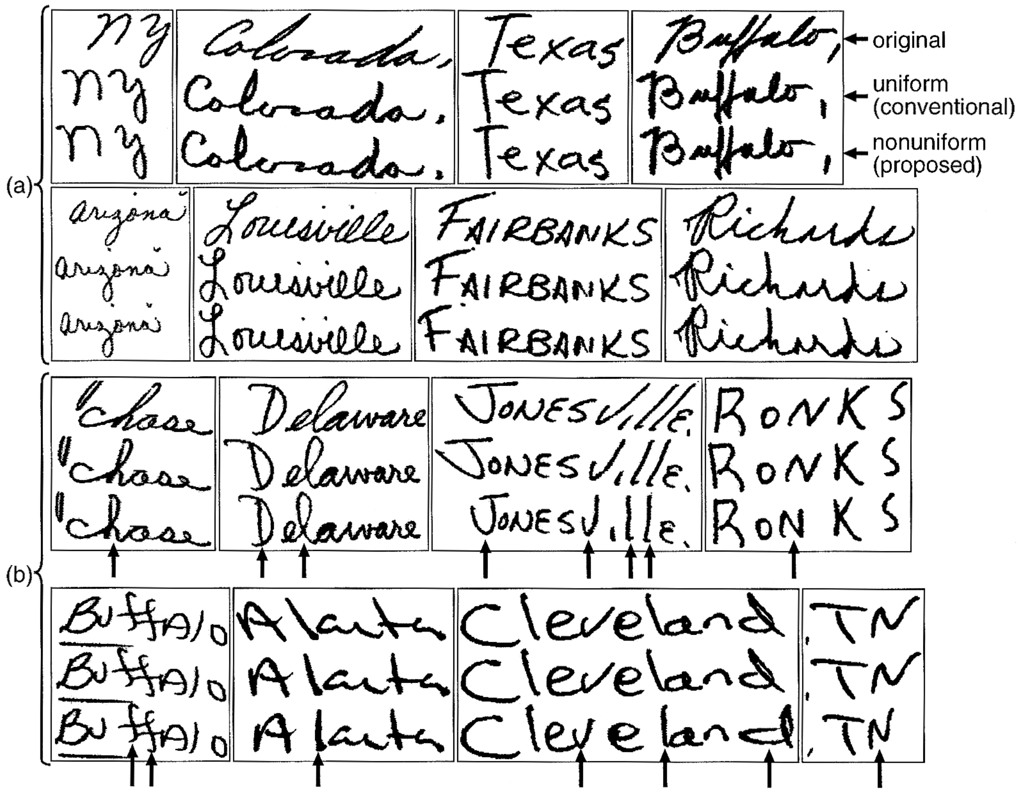

Fig. 5 Examples of slant correction. For each word, its original image (top), the result of the conventional uniform slant correction technique (middle), and the result of the nonuniform slant correction technique (bottom) are shown. The nonuniform slant correction results are similar (a) and superior (b) to the conventional uniform slant correction results. In (b), remarkably improved characters are indicated by arrows.

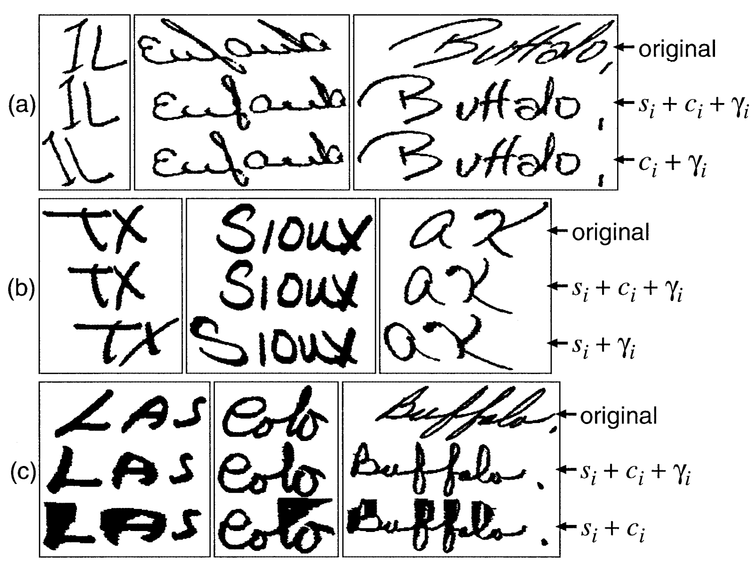

Fig. 6 The effect of the three functions $s_i(p_i)$, $c_i(p_i)$ and $\gamma_i(p_i|p_{i-1})$. For each word, the top image is the original image, and the middle image is the result of the present technique using all three functions. In (a), (b) and (c), the bottom image is the result of the nonuniform slant correction technique without the function $s_i(p_i)$, $c_i(p_i)$ and $\gamma_i(p_i|p_{i-1})$, respectively.

Fig. 7 Artificially slanted words. (a) Original image. (b) Slanted at the constant angle of 30°. (c) Nonuniformly slanted at angles sinusoidally changing in $[-30°, 30°]$.

Table 1 Mean error (in degree) between given and estimated slant angles for the ten images in Fig. 7 (b) and (c).

	constant angle					sinusoidally changing angle				
image no.	1	2	3	4	5	1	2	3	4	5
uniform correction	1.4	0.8	0.6	0.3	4.0	19.0	19.6	19.1	19.1	19.7
nonuniform correction	1.6	2.1	1.9	2.2	2.7	5.9	6.3	6.8	7.0	7.8

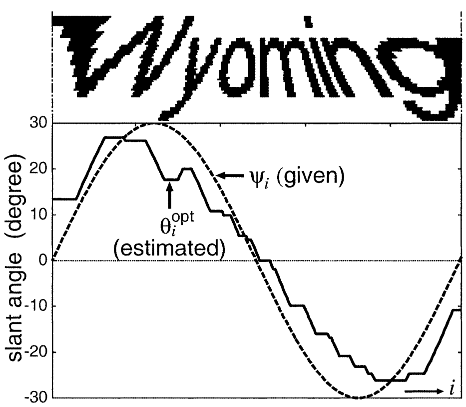

Fig. 8 Given and estimated slant angles for a nonuniformly slanted image of "Wyoming."

$$\frac{1}{M}\sum_{i=1}^{M}|\theta_i^{\mathrm{opt}} - \psi_i|, \tag{8}$$

where ψ_i and θ_i^{opt} are the given and the estimated slant angles at position i, respectively. The estimated slant angle θ_i^{opt} is calculated by $\tan^{-1}\{(p_i^{\mathrm{opt}} - i)/N\}$. This result shows that (i) the present technique provides almost the same estimation as the conventional technique when a constant slant appears, and (ii) the present technique provides more accurate estimation than the conventional technique when nonuniform slant appears. Figure 8 shows the estimated slant angle obtained by the present technique as a function of position i with the given slant angle for the case of word "Wyoming". It is shown that the estimated angle θ_i^{opt} follows the given slant angle ψ_i appropriately.

3.3 Improvement of Segmentation Performance

For better segmentation, the slant-corrected word image should be separable into its component characters by vertical boundaries. Thus, the effectiveness of a slant correction technique on the segmentation can be estimated by counting the number of vertical columns crossing two (or more) characters in its slant-corrected word images (Fig. 9). Hereafter,

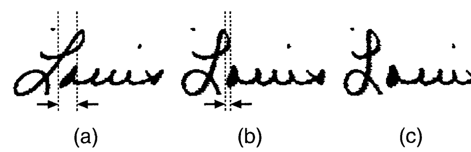

Fig. 9 The region of incomplete columns in (a) original image, (b) unsuccessfully slant-corrected image and (c) successfully slant-corrected image. If the slant correction is successfully performed, such columns become fewer. In fact, such column is disappeared in (c).

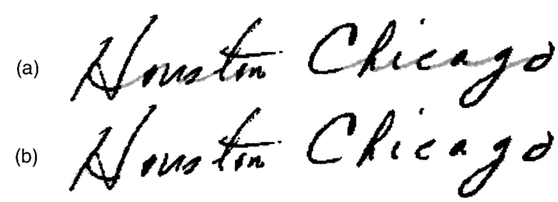

Fig. 10 (a) Division of all black pixels into component characters (thick parts) and ligatures (thin parts). (b) Word image without the ligatures.

such column is called an *incomplete column*. The fewer the number of incomplete columns becomes, the better segmentation performance can be expected.

This experiment was performed on 248 handwritten word images (city words composed of two or more characters) of BD0100 and BS0100 sets in the CEDAR CDROM database. Both of BD0100 and BS0100 sets contain hand-printed and cursive word images. After the preprocessing in the manner described in Sect. 3.1, we manually removed ruled lines, comma and periods from each image, and then manually divided all black pixels into component characters and ligatures. On counting the number of incomplete columns, the ligatures were neglected (i.e., considered as white pixels) because the belonging of ligatures is ambiguous. Figure 10 shows examples of division of all black pixels into component characters and ligatures.

Table 2 shows the number of incomplete columns in word images without slant correction, with the conventional uniform slant correction, and with the present nonuniform slant correction ($\omega_s = 0.56, \omega_c = 0.26$). The number of in-

Table 2 The number of incomplete columns.

dataset(#samples, #columns)	BD0100(137, 29076)	BS0100(111, 24102)	total(248, 53178)
without correction	2463	1589	4052
with uniform correction	684	554	1238
with nonuniform correction	582	478	1060

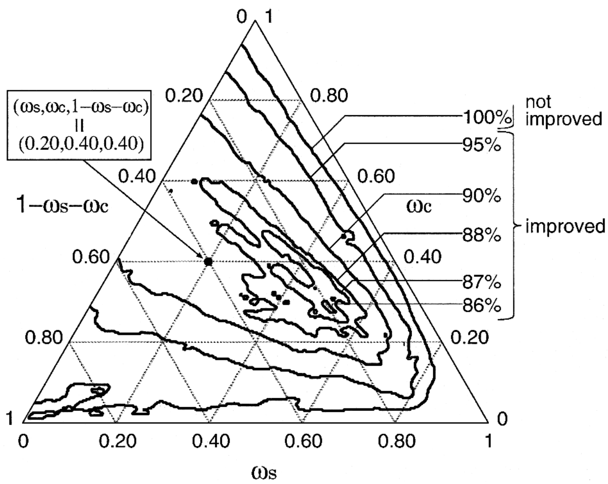

Fig. 11 The ratio of the number of incomplete columns by the present technique to that by the conventional uniform slant correction technique.

complete columns by the present technique was reduced to 86 percent (=1060/1238) of that by the conventional technique. Thus, it was quantitatively shown that the present technique will be more effective for improvement of segmentation performance than the conventional technique.

The necessity of the three functions $s_i(p_i)$, $c_i(p_i)$ and $\gamma_i(p_i|p_{i-1})$ was quantitatively clarified from the number of incomplete columns in slant-corrected images in addition to the qualitative investigation in Sect. 3.1. Figure 11 shows the reduction rate of the number of incomplete columns from the conventional technique at possible parameter settings. The improved results (i.e., lower rates) were obtained when all three functions are effective, that is, $\omega_s \neq 0$, $\omega_c \neq 0$, and $1 - \omega_s - \omega_c \neq 0$. Thus, it is shown that all three functions are imperative for the present technique.

It should be noted that the present technique consistently shows superiority over the conventional technique at extremely large area of Fig. 11, i.e., for most parameter settings.

4. Conclusion and Future Work

A nonuniform slant correction technique for handwritten word recognition was proposed. The nonuniform slant correction problem was formalized as a constrained optimization problem where local slant angles were variables to be optimized and then efficiently solved using a dynamic programming based algorithm. Experimental results indicated that the present technique can correct slants more appropriately than conventional uniform slant correction techniques.

Future works include the evaluation on actual word segmentation and recognition tasks. The embedment of the present technique into word recognition systems based on segmentation-by-recognition, such as Kim and Govindaraju [2], will realize a slant-correction-and-segmentation-by-recognition where segmentation boundaries are represented by the correction lines and optimized with the help of a character recognizer. This scheme will improve the slant correction performance since character shape information can be used in the estimation of local slant angles.

Acknowledgement

The authors thank the Center of Excellence for Document Analysis and Recognition (CEDAR), the State University of New York, for providing us CEDAR CDROM. This work was supported in part by the Ministry of Education, Culture, Sports, Science and Technology in Japan under a Grant-in-Aid for Scientific Research No.14780293.

References

[1] R.M. Bozinovic and S.N. Srihari, “Off-line cursive script word recognition,” IEEE Trans. Pattern Anal. Mach. Intell., 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. Pattern Anal. Mach. Intell., 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, Taipei, Taiwan, Dec. 1994.

[4] G. Nicchiotti and C. Scagliola, “Generalised projections: A tool for cursive handwriting normalisation,” Proc. 5th ICDAR, pp.729–732, Bangalore, India, Sept. 1999.

[5] E. Kavallieratou, N. Fakotakis, and G. Kokkinakis, “Slant estimation algorithm for OCR systems,” Pattern Recognit., vol.34, no.12, pp.2515–2522, Dec. 2001.

[6] A. Vinciarelli and J. Luettin, “A new normalization technique for cursive handwritten words,” Pattern Recognit. Lett., vol.22, no.9, pp.1043–1050, July 2001.

[7] L. Simoncini and Zs.M. Kovács-V, “A system for reading USA census’90 hand-written fields,” Proc. 3rd ICDAR, vol.2, pp.86–91, Montreal, Canada, Aug. 1995.

[8] 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, Barcelona, Spain, Sept. 2000.

[9] 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, Amsterdam, Sept. 2000.

[10] H. Hase, M. Yoneda, T. Shinokawa, and C.Y. Suen, “Alignment of free layout color texts for character recognition,” Proc. 6th ICDAR, pp.932–936, Seattle, Washington, Sept. 2001.

[11] S. Uchida and H. Sakoe, “Piecewise linear two-dimensional warping,” Proc. 15th ICPR, vol.3, pp.538–541, Barcelona, Spain, Sept. 2000.

Eiji Taira received B.E. and M.E. degrees from Kyushu University in 2000 and 2002, respectively. He is currently a doctoral course student in Graduate School of Information Science and Electrical Engineering, Kyushu University. His research interests include pattern recognition.

Seiichi Uchida received B.E., M.E., and Dr.Eng. degrees from Kyushu University in 1990, 1992 and 1999, respectively. From 1992 to 1996, he joined SECOM Co., Ltd., Tokyo, Japan where he worked on speech processing. Now, he is an associate professor at Faculty of Information Science and Electrical Engineering, Kyushu University. His research interests include pattern analysis and image processing. Dr. Uchida is a member of IEEE, IPSJ, ITE, and ASJ.

Hiroaki Sakoe received the B.E. degree from Kyushu Institute of Technology in 1966, and the M.E. and D.E. degrees from Kyushu University in 1968 and 1987, respectively. In 1968, he joined NEC Corporation and engaged in speech recognition research. In 1989, he left NEC Corporation to become a Professor of Kyushu University. His research interests include speech recognition and pictorial pattern analysis. He received 1979 IEEE ASSP Senior Award, 1980 IEICE Achievement Award and 1983 IECE Paper Award. He also received Kamura Memorial Prize from Kyushu Institute of Technology. Dr. Sakoe is a member of IEEE, IPSJ, ITE, and ASJ.