IDIAP RESEARCH REPORT

IDIAP

Martigny - Valais - Suisse

A new normalization technique for cursive handwritten words

Alessandro Vinciarelli [a] Juergen Luettin [a]

IDIAP–RR 00-32

May 2001

PUBLISHED IN

Pattern Recognition Letters. Vol.22, no.9, pp.1043-1050, June 2001.

Dalle Molle Institute for Perceptual Artificial Intelligence • P.O.Box 592 • Martigny • Valais • Switzerland

phone +41 – 27 – 721 77 11
fax +41 – 27 – 721 77 12
e-mail secretariat@idiap.ch
internet http://www.idiap.ch

[a] Institut Dalle Molle d'Intelligence Artificielle Perceptive. Rue du Simplon 4, CP592 - 1920 Martigny, Switzerland. e-mail: {vincia,luettin}@idiap.ch

IDIAP Research Report 00-32

A new normalization technique for cursive handwritten words

Alessandro Vinciarelli Juergen Luettin

May 2001

published in
Pattern Recognition Letters. Vol.22, no.9, pp.1043-1050, June 2001.

Abstract. This paper presents new techniques for slant and slope removal in cursive handwritten words. Both methods require neither heuristics nor parameter tuning. This avoids the heavy experimental effort required to find the optimal configuration of a parameter set.
A comparison between the new deslanting technique and the method proposed by Bozinovic and Srihari was made by measuring the performance of both methods within a word recognition system tested on different databases. The proposed technique is shown to improve the recognition rate by 10.8% relative to traditional normalization methods. Moreover a long exploration of the parameter space is avoided.

1 Introduction

Off line Cursive Script Recognition (CSR) is an important technology for applications involving automated processing of handwritten data, such as bank check processing and postal address reading (Plamondon and Srihari, 2000). Almost every CSR system presented in the literature involves a normalization step, which consists in removing *slant* and *slope*. The slant is the angle between the vertical direction and the direction of the strokes that, in an ideal model of handwriting, are supposed to be vertical (see Fig. 1). The slope is the angle between the horizontal direction and the direction

Figure 1: Slant and slope in a raw word image.

of the line on which the writer aligned the word (see Fig. 1).
Slant and slope can be introduced by both handwriting style and acquisition process. Ideally, the removal results in a word image independent with respect to such factors. For this reason the process is called normalization.
The recognition techniques usually applied (Dynamic Programming and Hidden Markov Models) need a fragmentation of the word (Steinherz et al., 1999). In DP based systems, where the isolated subunits are expected to be characters, the normalization reduces the shape variability, which simplifies their classification by pattern recogniton techniques. In HMM based systems, where the signal is assumed to be piece-wise stationary, the normalization aims to transform the handwritten data into such segments.
In the literature, the description of the normalization technique is often neglected and the use of heuristic rules, frequently based on parameters that need to be set manually, is apparent (Morita et al., 1999). Parameter tuning can be a time consuming process and may only be optimal for a specific database.
This work presents new techniques to remove slope and slant that avoid such problems and that do not involve any heuristic parameters. The deslanting technique is compared with the widely applied "Bozinovic and Srihari Method" (BSM) (Bozinovic and Srihari, 1989) in terms of recognition rate.
Section 2 describes a common slope removal algorithm, section 3 illustrates the new desloping technique, sections 4 and 5 present the BSM and the new deslanting method respectively. Recognition results for the different methods are shown in section 6 and the conclusions are drawn in section 7.

2 The traditional slope removal algorithm

The traditional method for slope removal (Bozinovic and Srihari, 1989)(Senior and Robinson, 1998) starts by finding a first rough estimate of the *core region*, the region enclosing the character bodies (see figure 2). This estimate is biased by the fact that the word is not horizontal, so that the upper

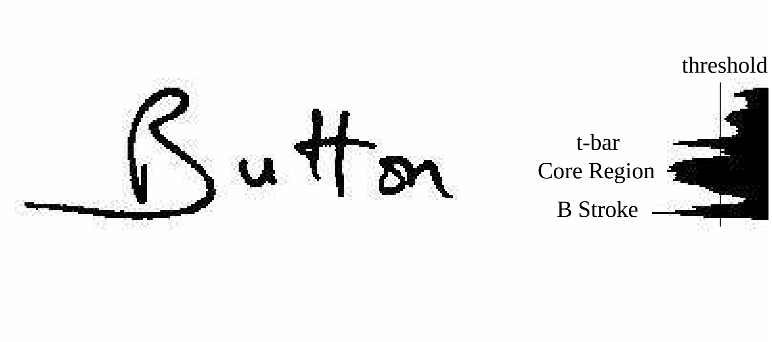
Figure 2: The proposed core region detection method: the density threshold gives a first rough estimate of the core region. A high density region is located in the ascender area, but can be discarded by simply selecting the high density region containing the highest number of foreground pixels.

and lower limit of the estimated core region do not fit, as they should, the extrema of the character bodies. To solve this problem, the stroke minima closer to the lower limit of the estimated core region are used to fit (e.g. with the Least Squares Method) the line connecting the bottom points of the character bodies. This is the line on which the word is aligned (called *lower base line*).
The image is desloped when this line is horizontal, which is a condition achieved by a rotational transform. After the deslanting step, the core region can be re-estimated under the assumption that the word image is now horizontal. It becomes evident that the first estimate of the core region is fundamental for removing the slope. The detection of the core region is based on the hypothesis that the density of the lines belonging to it is higher than the density of the other lines. This is evident in the horizontal density histogram (see figure 2), showing higher values in correspondence with the core region, and this property is used to perform the detection.
Core region lines are usually obtained as the ones surrounding the highest density peaks, but this technique is strongly affected by the presence of long horizontal strokes that can be confused with the actual core region and can lead to severe errors in normalising the word (see figure 3). Many heuristic rules are necessary to handle the problem and the resulting process is not robust.

3 Our slope removal algorithm

In order to avoid the problems described in the previous section we analyzed the distribution of the density values rather than the density histogram itself. The density distribution $P(h)$ is expected to be bimodal[1] : one mode corresponds to the high density lines of the core region, the other corresponds to the other lines, where the density is low. A thresholding algorithm can be applied to the distribution to separate the two modes. The threshold can be used to distinguish between core region lines (above the threshold) and remaining lines (below the threshold). Since the horizontal strokes have a negligible

[1]In the hypothesis that the data is distributed normally around the modes, these correspond to the averages of two gaussians.



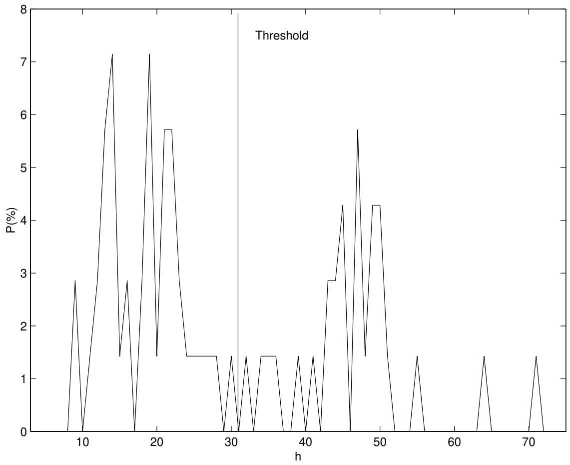


Figure 3: Peaks determined by horizontal strokes are evident in the density histogram (upper figure). In the density probability distribution (lower graphic), the density values of the peaks have small probability. Their influence is negligible and they do not influence the calculation of the threshold.

influence on $P(h)$, the threshold is robust with respect to their presence.
The density histogram $h(i)$ presents the density for each row i. The density distribution describes the probability $P(h)$ of having h foreground pixels in a line. Long horizontal strokes (e.g. t-bars) give rise to high peaks in the density histogram (see Fig. 3) but their influence in the density probability distribution is small because they only occur a few times.
We use the Otsu thresholding algorithm (Haralick and Shapiro, 1992) that is based on the minimization of the weighted sum of the variances of the following two sets: the group of the density elements less or equal to the threshold and the group of the density elements greater than the threshold. The weights are the probabilities of the respective groups, calculated as the percentage of elements belonging to each group. When several regions have a density higher than the threshold, the region with the highest number of foreground pixels is selected as the core region (see Figure 2).
Another approach focusing on the density distribution can be found in (Cote' et al., 1998). Here the entropy is calculated for distributions corresponding to several rotation transformed images of the word. The image giving the lowest value of entropy is assumed to be the desloped image.
After the core region is found, we select the points that are used to fit the lower base line. The lowest point l_i of each column i is found, then the set $L = \{l_j : [y(l_{j-1}) < y(l_j)] \wedge [y(l_j) > y(l_{j+1})]\}$, where $y(l_i)$ is the vertical coordinate of the points, is created[2]. The average distance of the L elements from the lower limit of the estimated core region is calculated. The elements with distance less than the

[2] Since the image vertical axis is directed downward, the condition $y(l_{j-1}) < y(l_j)$ means that point l_j is lower than point l_{j-1}

average are finally retained to fit the lower baseline. Fitting is done with the Least Mean Squares method. The slope of the fit is used as an estimate of the slope of the word.

4 The Bozinovic and Srihari deslanting technique

The BSM was proposed in (Bozinovic and Srihari, 1989) and is based on the idea that slant information is mostly contained in the almost vertical strokes of the word. These are selected as follows: first all lines containing a continuous stroke longer than a parameter MR (Max Run) are discarded. The remaining lines form stripes that are accepted if their height is higher than another parameter SH (Strip Height) as shown in figure 4. The strokes in the finally remaining stripes enclosed between

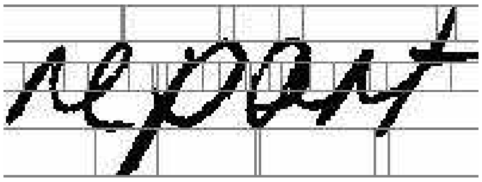


Figure 4: The Bozinovic and Srihari deslanting method: the strokes for the slant estimation are selected in the allowed stripes. These must be composed of lines where all of the continuous strokes are shorter than a value MR (Max Run) and must be higher than a second parameter SH (Strip Height). A stroke is enclosed between vertical lines connecting the borders of an allowed strip without intersecting foreground pixels. When one of the two halves of a stroke is empty, the stroke is discarded.

vertical lines connecting the strip borders without intersecting foreground pixels (see vertical segments in figure 4) are selected for the slant estimation.
For each selected stroke, the centroid of the upper and lower half are retained and the slope of the line connecting them is assumed as a local slant estimate. The average of all the local slant estimates is used as a global slant estimate. The image is deslanted with a shear transform based on the estimated slant.
The main disadvantage of the method is that parameter tuning is needed. This might result in a heavy experimental effort to find the optimal point in the parameter space (MR and SH), i.e. the point corresponding to the best performance of the recognition system. Furthermore, in the presence of several handwriting styles, pens, and handwritten word dimensions, the estimated parameters can be sub- optimal for many data.
Other methods based on averaging the slant of near vertical strokes are described in (Kim and Govindaraju, 1997) (Senior and Robinson, 1998) and (El-Yacoubi et al., 1999). They are similar in concept to the BSM, but use different criteria to select the near vertical strokes. In these methods, the slope of the selected strokes is estimated from the slope of their contours.
The contours of the near vertical strokes are used also in (Marti and Bunke, 2001). They are first approximated by a set of straight lines l_i. The so called angle histogram:

$$h_{sl}(\alpha) = \sum_{\{i:\lambda_i=\alpha\}} |l_i|^2 \tag{1}$$

is then obtained, and finally $\alpha_{sl} = argmax_\alpha(h_{sl}(\alpha))$ is assumed as slant estimate.

5 The new deslanting technique

The new deslanting technique is based on the hypothesis that the word is deslanted when the number of columns containing a continuous stroke is maximum. For each angle α in a reasonable interval, a shear transform is applied to the image and the following histogram is calculated:

$$H_\alpha(m) = \frac{h_\alpha(m)}{\Delta y_\alpha(m)} \tag{2}$$

where $h_\alpha(m)$ is the vertical density (number of foreground pixels per column) in column m, and $\Delta y_\alpha(m)$ the distance between the highest and lowest pixel in the same column. If the column m contains a continuous stroke, $H_\alpha(m) = 1$, otherwise $H_\alpha(m) \in [0, 1[$.
For each shear transformed image, the following function:

$$S(\alpha) = \sum_{\{i:H_\alpha(i)=1\}} h_\alpha(i)^2 \tag{3}$$

is calculated. The angle α giving the highest value of S is taken as the slant estimate (see fig. 5). By using in $S(\alpha)$ the square value of the density rather than the density itself, the contribution of the longest vertical strokes is enhanced with respect to that of short strokes, which are often due to stronlgy slanted vertical letters.
Histogram H_α and function $S(\alpha)$ are easier to calculate than the above mentioned angle histogram. The detection of near vertical strokes, their borders and their approximation with a set of straight lines strokes are in fact not required. A simple count of the foreground pixels of each column allows us to calculate H_α and $S(\alpha)$.
The need for multiple shear transforms makes the method computationally heavy, but the computa-

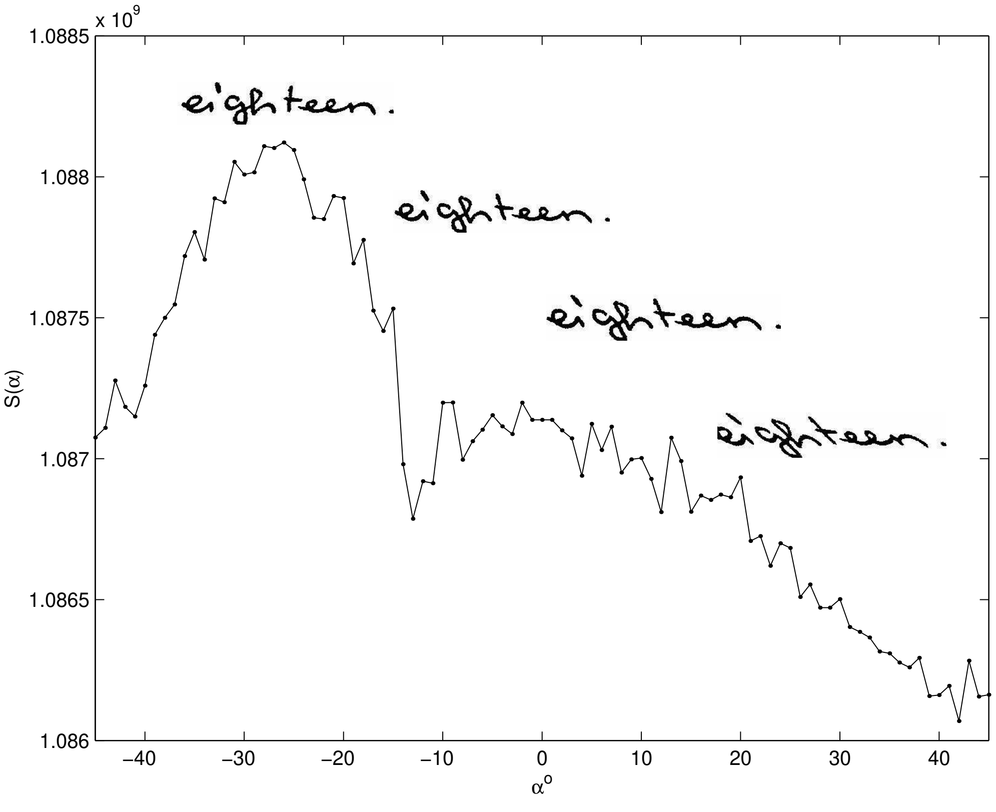


Figure 5: Deslanting algorithm. The plot of $S(\alpha)$ shows a high peak in correspondence of the deslanted image. For high values of α, when the image is strongly deformed, the value of $S(\alpha)$ becomes lower.

tion can be reduced using some recursive procedure to obtain the shear transformed image for angle

$\alpha + 1$ from the shear transformed image for angle α. On the other hand, the absence of parameters avoids any experimental effort in parameter optimisation.
A similar approach, based on slanted histograms, can be found in (Guillevic and Suen, 1994). Here, for each slanted image, the highest positive value $D(\alpha)$ of the first derivative of the vertical density histogram is calculated. The slant estimate is then found as $argmax_\alpha(D(\alpha))$. The main advantage of our algorithm with respect to such approach is that the function $S(\alpha)$ is not influenced by a single stroke but by the whole image. The derivative peak $D(\alpha)$ is, on the contrary, completely determined by a single vertical stroke which might be not representative of the slant across the word.
Another approach avoiding the selection of near vertical strokes and relying rather on a global measure over the word can be found in (Kavallieratou et al., 2000). In that work, the authors use the Wigner Ville Distribution to calculate the degree of *deslantedness* of the word. The measure is repeated over shear transformed word images corresponding to different angles in an interval. The angle corresponding to the image giving the optimal measure is taken as the slant estimate.

6 Experiments and results

The effectiveness of the proposed methods was evaluated as a function of the performance of a CSR system and compared to traditional normalization methods. The CSR system converts the handwritten data into a sequence of vectors with a sliding window shifting column by column from left to right. At each window position, a frame is isolated and a feature vector is extracted. The recognition is performed using continuous density Hidden Markov Models to calculate the likelihood of the observation sequence with each word in a lexicon. The most likely word is selected as the interpretation of the handwritten data. Words are modeled as concatenations of single letter HMMs. Training is based on Maximum Likelihood estimation, while recogniton is based on the estimation of Maximum A posteriori Probability. The number of states is the same for every letter model.
The *cross validaton* technique was used to find the best parameter values (Stone, 1974). Systems corresponding to different parameter sets were trained and tested on a separate training and validation set respectively. When using the new method, the only parameter to be changed is the number of states in the letter models. The range of this configuration parameter is delimited by the amount of training material available. When using the BSM, there are three parameters to be tuned: the number of states per letter model, MR and SH.
After having found the best parameter set for both deslanting techniques, the performance of the two corresponding systems was measured again over the test set, giving the final result.
Two different data sets were used in the experiments. The first is a database[3] of 4053 words produced by a single writer and described in (Senior and Robinson, 1998). The second is a set of 12198 words extracted from a handwritten page database collected at the University of Bern (Marti and Bunke, 1999). A page-into-words segmentation algorithm was applied to each. Only the words correctly segmented and labeled were retained. The words were produced by approximately 200 writers. The databases will be referred to in the following as the *Cambridge* and *Bern* database respectively.

6.1 Cambridge Database results

The database was partitioned into a training set (2360 words), validation set (673 words) and test set (1020 words). The lexicon size is 1334 and the number of states in the letter models tested spans from 1 to 11.
The best performance over the validation set obtained using the deslanting method described in section 5 is 84.23%. Exploration of the parameter space of the Bozinovic and Srihari algorithm was started at the point corresponding to the average value of the lengths of the continuous horizontal

[3]The database is publicly available on the web and can be downloaded at the following ftp address: ftp.eng.cam.ac.uk/pub/data.

		MR				
		8	9	10	11	12
SH	8	78.59	79.92	79.48	80.07	80.37
	9	77.99	79.33	79.48	80.22	80.67
	10	78.44	79.78	81.26	79.48	78.59
	11	79.18	79.63	79.03	79.33	78.44
	12	76.65	79.18	78.74	80.07	76.65

Table 1: Cambridge database results. Each column reports the accuracy results obtained changing SH for a given value of MR. Each line reports the results obtained changing MR for a given SH. The central point corresponds to the highest performance obtained. The results were measured over the validation set. Each reported score is the maximum among tests using different numbers of states per letter model, for a given pair (MR,SH).

strokes over the database (MR coordinate), and the average height of the allowed stripes given such value of MR (SH coordinate). We selected as optimal a point corresponding to the highest performance over a window centred on about thid point, and five points wide. Results over the validation set are shown in Table 1. The results over the test set are shown in Table 3. The recognition rate shows a 3.6% relative improvement using the new deslanting technique.

6.2 Bern database results

The database was partitioned into a training set (8106 words), validation set (1349 words) and test set (2744 words). The lexicon size is 100. Experiments involved letter models with a number of states between 9 and 20.
The best performance achieved with the deslanting described in Section 5 over the validation set is 57.52%. For the exploration of the parameter space of the Bozinovic and Srihari method the starting

		MR				
		8	9	10	11	12
SH	10	52.26	52.18	54.18	52.33	54.41
	11	51.96	53.66	53.39	52.77	52.70
	12	53.96	52.70	54.41	52.03	53.00
	13	51.22	51.81	53.15	54.41	53.52
	14	51.74	51.00	51.96	51.51	53.22

Table 2: Bern database results. Each reported value corresponds to different values of MR and SH (see caption for Table 1). The results are obtained over the validation set.

point was selected with the same method described in the previous subsection. The results obtained over the validation set are shown in Table 2. The two winning systems (one for each deslanting method) were tested over the samples of the test set. The use of the new method improves the recognition rate by 10.8% relative (see Table 3).

7 Conclusions

A new normalization technique for cursive handwritten words was presented. The algorithm is composed of two steps, deslope and deslant, and in both the use of any manually set parameters has been avoided. This is an advantage because parameter tuning often requires a heavy experimental effort, training and testing the system to find an optimal configuration. Furthermore, a parameter set can

DB	BSM (%)	new (%)
Cambridge	81.56	84.50
Bern	54.95	60.89

Table 3: Results over the test set. First column reports the database, The following columns report the performance obtained with BSM and the new method respectively. The results were obtained over the test set.

be optimal only for a given data set.
A comparison between the new deslanting technique and the well known BSM was performed.
Two benchmarks were used for the comparison, the first is composed of words produced by a single writer, the second of samples written by several persons with different pens (see Sections 6). In both cases, the system showed a significantly better performance when using the new deslanting technique. The effort needed to find the optimal system when using BSM is much higher and its optimal configuration is data dependent. Use of the new technique has therefore been shown effective in improving recogniton results, while also avoiding a lot of effort in parameter tuning.

Acknowledgements The authors wish to thank F.Camastra (University of Genoa, Italy), A.Morris (IDIAP) and D.Ugolini (Polo Nazionale Bioelettronica, Italy) for commenting on the draft. This work was done under the grant 21-55733.98 issued by the Swiss National Science Foundation.

References

Bozinovic, R. M., Srihari, S. N., January 1989. Off-line cursive script word recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1), 69–83.

Cote', M., Lecolinet, E., Cheriet, M., Suen, C. Y., january 1998. Automatic reading of cursive scripts using a reading model and perceptual concepts - the PERCEPTO system. International Journal of Document Analysis and Recognition 1 (1), 3–17.

El-Yacoubi, A., Gilloux, M., Sabourin, R., Suen, C., August 1999. An hmm-based approach for off-line unconstrained handwritten word modeling and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (8), 752–760.

Guillevic, D., Suen, C., 1994. Cursive script recognition: A sentence level recognition scheme. In: Proceedins of 4^{th} International Workshop on Frontiers in Handwriting Recognition.

Haralick, R. M., Shapiro, L. G., 1992. Computer and Robot Vision. Addison Wesley, USA.

Kavallieratou, E., Fakotakis, N., Kokkinakis, G., july 2000. A slant removal algorithm. Pattern Recognition 33 (7), 1261–1262.

Kim, G., Govindaraju, V., april 1997. A lexicon driven approach to handwritten word recognition for real time applications. IEEE Transactions on Pattern Analysis and Machine Intelligence 19 (4), 366–379.

Marti, U., Bunke, H., 2001. Using a statistical language model to improve the performance of an HMM-based cursive handwriting recognition system. International Journal of Pattern Recognition and Artificial Intelligence To appear.

Marti, U.-V., Bunke, H., 1999. A full english sentence database for off-line handwriting recognition. In: Proceedings of 5^{th} International Conference on Document Analysis and Recognition. Vol. 1. Bangalore.

Morita, M., Facon, J., Bortolozzi, F., Garnes, S., Sabourin, R., 1999. Mathematical morphology and weighted least squares to correct handwriting baseline skew. In: Proceedings of 5^{th} International Conference on Document Analysis and Recognition. Vol. 1. Bangalore.

Plamondon, R., Srihari, S., 2000. On-line and off-line handwriting recognition: A comprehensive survey. IEEE Transactions on Pattern Analysis and Machine Intelligence , 63–84.

Senior, A. W., Robinson, A. J., March 1998. An off-line cursive handwriting recognition system. IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (3), 309–321.

Steinherz, T., Rivlin, E., Intrator, N., February 1999. Off-line cursive script word recognition - a survey. International Journal of Document Analysis and Recognition 2 (2), 1–33.

Stone, M., 1974. Cross-validatory choice and assessment of statistical prediction. Journal of the Royal Statistical Society 36 (1), 111–147.