Comparing Normalization and Adaptation Techniques for On-Line Handwriting Recognition

Anja Brakensiek, Andreas Kosmala
Dept. of Computer Science, Electrical Engineering
Gerhard-Mercator-University Duisburg
47057 Duisburg, Germany
{anja, kosmala}@fb9-ti.uni-duisburg.de

Gerhard Rigoll
Inst. for Human-Machine Communication
Technical University of Munich
80290 Munich, Germany
rigoll@ei.tum.de

Abstract

In this paper a writer-independent on-line handwriting recognition system is described comparing the influence of handwriting normalization and adaptation techniques on the recognition performance. Our Hidden Markov Model (HMM) -based recognition system for unconstrained German script can be adapted to the writing style of a new writer using different adaptation techniques whereas the impact of preprocessing to normalize the pen-trajectory is examined. The performance of the resulting writer-dependent system increases significantly, even if only a few words are available for adaptation. So this approach is also applicable for on-line systems in hand-held computers such as PDAs. In addition, the developed normalization techniques are helpful to improve completely writer independent systems. This paper presents the performance comparison of three different adaptation techniques either in a supervised or an unsupervised mode, in combination with appropriate normalization methods, with the availability of different amounts of adaptation data ranging from only 6 words up to 100 words per writer.

1. Introduction

Automatic recognition systems for unconstrained on-line handwritten words become more and more important, especially with respect to the use of pen based computers or electronic address books (compare also [2, 6, 8]). Although the performance of recognition systems increases, the error rate of writer independent recognizers is still too high – at least for some writer types – for a real application.

This scenario of writer-independent handwriting recognition systems, which are often used by specific single writers only, leads to two main requirements: a good preprocessing method to normalize different writing styles [3] and the implementation of writer adaptation techniques.

This paper describes an on-line handwriting recognition system, which operates in a writer independent mode and - if demanded - which can be adapted to a specific writer with a varying amount of adaptation data. In this field of research, HMM-based techniques (see [7]), which are well known in speech recognition, have been established because of their segmentation-free recognition approach and their automatic training capabilities. The individual writing style (see [10]) has to be normalized in a writer independent recognition system to improve the performance. However, the normalized pen trajectory still differs between various writers, such that an adaptation to a specific style will become necessary. We compare three different adaptation techniques: the maximum likelihood (ML) retraining, the maximum a posteriori (MAP) [5] and the maximum likelihood linear regression (MLLR, see [5, 9]) adaptation. One important aspect for a practical usage of adaptation methods is the amount of adaptation data, which is needed to reduce the error rate significantly. Another aspect is the availability of labeled adaptation data. To implement a user-friendly technique, our writer dependent recognition system can be developed by an unsupervised adaptation even with only a few words. The following sections describe our baseline recognition system, the normalization techniques and the adaptation framework. Results will be presented, which have been obtained by three investigated adaptation methods in combination with the normalization techniques.

2. System architecture

Our handwriting recognition system consists of about 90 different linear HMMs, one for each character (upper- and lower-case letters, numbers and punctuation marks). In general the continuous density HMMs consist of 12 states (with up to 15 Gaussian mixtures per state depending on the amount of training data per HMM) for characters and numbers and fewer states for some special characters depending on their width. To train the HMMs we use the Baum-Welch algorithm, whereas for recognition the Viterbi algorithm is used. The presented results refer to a single word recognition rate using a dictionary of about 2200 German words (no out of vocabulary).

For our experiments we use a large on-line handwriting database comprising several writers and a variety of different unconstrained writing styles (compare Fig.1). The database consists of script samples of 166 different writers, all writing several words or sentences on a digitizing surface. The training of the writer independent system is performed using about 24400 words of 145 writers. Testing is carried out with 4153 words of 21 different writers (about 200 words per writer). For each of these 21 test writ-

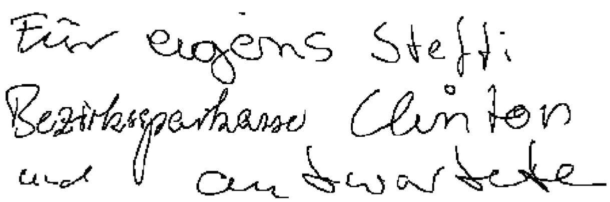

Figure 1. Examples of 7 different writers

ers up to 100 words are available for adaptation and nearly 100 words are used for evaluation of the developed writer dependent system (altogether 2071 test words).

3. Preprocessing and feature extraction

The first task in the processing pipeline is the resampling of the pen trajectory. The advantages of the resampling are twofold: 1.) the resampling with vectors of constant lengths compensates different writing speeds between different users. Obviously, this demands the first step towards writer independency. 2.) the resampling ensures a constant sample rate along the pen trajectory within words. As a result, the density of sampling points is constant within a single script sample, which is advantageous for the subsequent normalization methods.

3.1. Normalization

Normalization of the input-data implies basically the correction of slant and height. Standard methods for slant normalization rely on an analysis of the orientation histogram of the handwriting. Conducted investigations have shown that this method requires a relative huge amount of sampling vectors, in order to achieve reliable results.
As an alternative, a projection based approach has been investigated. The basic idea of the presented approach is to find a measurable parameter, which characterizes a slant free writing. Of course, this is the ideal vertical orientation of the up/down strokes. The challenge here is that it is obviously not feasible to measure this orientation directly on samples of very limited length. An auxiliary function might be the projection of the sample on the x-axes. Following this, the normalized x-projection $h(x)$ within a reasonable chosen interval $[x', x'']$ of the script sample can be computed as

$$h(x) = \frac{n_x}{\sum_{i=x'}^{x''} n_i} \tag{1}$$

with n_x, the absolute number of sample points in column x. Handwriting with (almost) no slant should yield in projections with clear peaks at positions with long up/down strokes, whereas any handwriting with very strong slant should yield nearly equal distributed normalized x-projections. Considering this, it is expected that this feature can be described well in terms of the entropy H of $h(x)$. Since $h(x)$ can be interpreted as the probability of a sample point being in column x, the entropy H of the whole sample is given as

$$H = -\sum_{i=x'}^{x''} h(i) \log h(i). \tag{2}$$

Finally the slant angle approximation is performed by shearing the resampled writing by small angles $\Delta\Phi$, while for each resulting shearing angle Φ, the entropy $H(\Phi)$ is computed. The optimal shearing angle $\Phi^\star$ for the slant minimization can than be obtained by searching for

$$\Phi^\star = \arg\min_{\Phi} H(\Phi). \tag{3}$$

The entropy $H(\Phi)$ is shown for the given sample in Fig. 2. After a final smoothing of the function $H(\Phi)$, an optimal shearing angle has been detected at $\Phi^\star = 5^o$ in the given example.

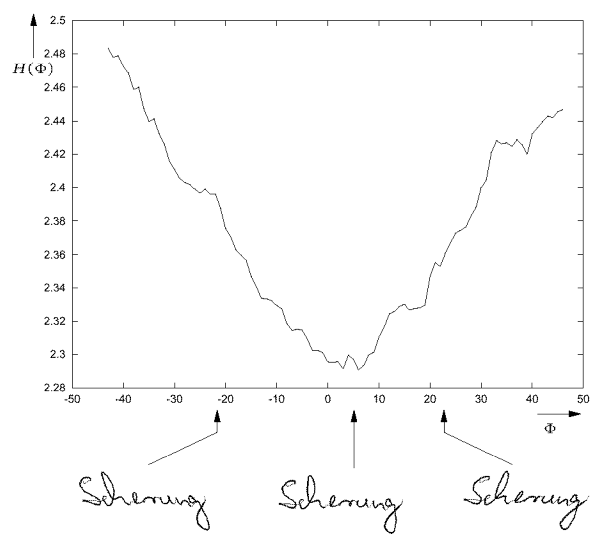

Figure 2. Entropy based slant normalization

For height normalization the baseline and the characters core height are estimated by means of an iterative region detection algorithm, which is independent of ascenders and descenders. The baseline and the coreheight are approximated by two horizontal lines determined by the local minima resp. maxima of the word[1]. The consideration of the minima and maxima for the region detection depends on their distance to the current approximated baseline/coreheight, removing disturbing outliers caused by ascenders and descenders.

[1]A horizontal line orientation can be achieved by applying the described slant minimization method in a similar manner to the y-projection.

3.2. Feature Extraction

In our on-line handwriting recognition system, the following features are derived from the trajectory of the pen input (compare also [8]):

- the angle of the spatially resampled strokes ($\sin\alpha$, $\cos\alpha$) and the difference angle ($\sin\Delta\alpha$, $\cos\Delta\alpha$)
- the pen pressure (binary)
- a sub-sampled bitmap slid along the pen trajectory (9-dimensional vector), containing the current image information in a 30×30 window

These 14-dimensional feature vectors x are presented to the HMMs in one single stream for training and testing the recognition system.

4. Writer adaptation

For handwriting adaptation we compare three different adaptation approaches, which are well known in speech recognition: a retraining according to the maximum likelihood (ML) criterion using the Baum-Welch algorithm (EM: expectation maximization, compare [7]), the maximum a posteriori (MAP, see [1]) adaptation and the maximum likelihood linear regression adaptation (MLLR, see [4]).

The goal is to maximize the matching between the Hidden Markov Models trained with the general database of 145 writers and a certain writer (wd: writer dependent) by considering different writing styles. When the amount of training or adaptation data is not sufficient for a robust training of all HMM parameters λ, it is useful to reestimate only some of these parameters, which consist of means μ, variances v, weights w and transitions t, and to leave the other parameters unchanged.

In the following the objective functions are given in principle. Applying the ML-retraining (Eq.4) to means adaptation, a similar transform can also be estimated for all other parameters.

$$\lambda_{ML} = \underset{\lambda}{\text{argmax}}\ P(X|\lambda) \quad \Rightarrow \quad \mu_{wi} \overset{EM}{\rightarrow} \mu_{ML} \quad (4)$$

The MAP (or Bayesian) approach (Eq.5) takes the prior probability $P(\lambda)$, which is estimated by the training of the writer independent model, into account. Here, a separate transformation for each Gaussian is performed and the objective function leads to an interpolation between the original mean and the estimated mean (analogical for variances) for the special writer.

$$\lambda_{MAP} = \underset{\lambda}{\text{argmax}}\ P(\lambda|X) \approx \underset{\lambda}{\text{argmax}}\ P(X|\lambda)P(\lambda) \quad (5)$$

Using the MLLR adaptation (Eq.6) only the means (or variances) of the Gaussians are reestimated by transforming them with a regression matrix M. To handle the sparse data problem, it is possible to cluster several similar Gaussians into regression classes or to use only one global regression matrix.

$$\lambda_{MLLR} = \underset{\lambda}{\text{argmax}} P(X|\lambda) \ \Rightarrow \ \mu_{MLLR} = M\mu_{wi} \quad (6)$$

The results in the following section have been achieved in a supervised and also in an unsupervised mode. Additionally, the influence of feature normalization is examined.

5. Experiments

In the presented experiments we examine the influence of writing style normalization (slant and height) as well as the performance of writer adaptation techniques using different amounts of adaptation data (6 or 100 words per writer).

Two separate recognition systems have been trained, one using the original (only resampled, but not normalized) data and the other using the normalized writing samples. Using these writer independent (wi) baseline recognition systems a word recognition rate of 85.8% for the original data and 86.7% for the normalized data is achieved testing the entire test-set of 4153 words.

This test-set is halved for adaptation experiments. Thus, using the same test-set as for the writer dependent (wd) systems, a word recognition rate of 85.7% resp. 87.0% can be obtained on the half test-set consisting of 2071 words (compare Tab.1: wi). This recognition accuracy results from the fact that the individual accuracies of the wi-baseline systems for each writer vary from 38.0% to 98.1% for the original and from 64.1% to 96.1% for the normalized system. So the advantage of a rational preprocessing is not only the higher recognition rate in general, but also the smaller range of recognition performance per writer (fewer outliers).

Tab.1 presents the dependency of the recognition accuracy on the amount of adaptation words (aw) and techniques (updating μ, w, v or t) in a supervised mode. To evaluate the recognition results obtained by a 6-word-adaptation, we repeat these tests with another disjoint adaptation set of (randomly chosen) 6 words (data-set (1.) and (2.)).

Table 1. Word recognition results (%) for 21 writers in a supervised mode

adaptation technique	original	normalized
baseline system (**wi**)	**85.7**	**87.0**
ML, μ, 100 adaptation words	93.7	93.3
ML, w, 100 aw	91.0	91.9
ML, v, 100 aw	90.8	91.7
ML, μw, 100 aw	**94.0**	**93.6**
ML, μv, 100 aw	89.1	88.1
ML, μwvt, 100 aw	89.5	87.6
ML, μ, 6 aw (1.) / (2.)	86.1 / 86.9	86.8 / 86.2
ML, w, 6 aw (1.) / (2.)	87.2 / 86.1	87.6 / 87.4
MAP, μ, 100 aw	91.1	92.2
MAP, μv, 100 aw	91.0	92.2
MAP, μ, 6 aw (1.) / (2.)	**87.0 / 87.2**	**88.4 / 88.2**
MLLR, μ, 100 aw, 1 cluster	86.1	87.5
MLLR, μ, 100 aw, 16 clu.	87.2	87.7
MLLR, μ, 100 aw, 128 clu.	88.1	87.1

Again, all experimental results refer to an average recognition rate of 21 different test writers. The recognition performance increases significantly, as it is expected, when the baseline system is adapted to a new writing style. Best recognition results of 94.0% resp. 93.6% on the wd-system can be obtained by the ML-method, reestimating means and weights, and using 100 words (about 6.5 characters per word) for adaptation. Characters, which do not appear, cannot be updated. After writer adaptation the influence of normalization decreases. Nevertheless, outliers of the writer-set can be handled better using data normalization (worst recognition rate of 58% for a single writer compared to 68% using the normalized wd-system).

If the amount of adaptation data is reduced to 6 words, the MAP-adaptation is most suitable. In this case the recognition rate increases to 87.1% resp. 88.3% in average.

Regarding the MLLR results, the error rate decreases only slightly compared to the other adaptation techniques. Here, it must be noticed, that this technique works very well for some writers depending on the parameter setting (e.g. number of clusters), but in average of 21 writers using identical adaptation parameters the performance enhancement is negligible. We assume, that it is problematic to cluster similar Gaussians, because of the variability in handwriting (depending on the kind of features) and the highly non-stationary feature sequence within a single grapheme model. The ML- and MAP-technique are much more robust against changes in parameter settings.

Tab.2 presents recognition results, which are obtained by an unsupervised adaptation with data labels, that are automatically recognized by the underlying wi-system.

Table 2. Word recognition results (%) for 21 writers in an unsupervised mode

adaptation technique	original	normalized
baseline system (**wi**)	**85.7**	**87.0**
ML, μ, 100 adaptation words	91.6	91.8
ML, w, 100 aw	89.9	90.8
ML, μw, 100 aw	**92.0**	**92.3**
MAP, μ, 100 aw	89.9	91.4
MLLR, μ, 100 aw, 16 cluster	87.1	88.1

As it is expected, the recognition improvement is a little bit smaller than in the supervised mode, whereas in principle the results are analogical. The important aspect is, that the error rate decreases although we use also incorrect labels for adaptation. From this it follows that an unsupervised adaptation during the use of the system (e.g. PDA) leads to higher recognition rates, and in this mode the amount of adaptation data is theoretically unlimited.

Future work will imply confidence measures to improve the quality of the adaptation data in the unsupervised mode. Other topics in the future will be the combination of different adaptation methods and an improvement of the baseline system by taking adjacent feature vectors (e.g. LDA) into account.

6. Summary and conclusions

In this paper we presented an HMM-based on-line handwriting recognition system for unconstrained German script samples. We investigated the performance of preprocessing methods consisting of height and slant normalization as well as some writer adaptation techniques (ML, MAP, MLLR) for a test-set of 21 writers. It has been shown that the recognition error of the writer-independent system can be reduced by about 7% relative using normalized data samples (compared to only resampled data). Performing a supervised writer adaptation with a large database of 100 words the error decreases by about 58% resp. 50% relative using a simple ML-reestimation of means and weights. This results in the best recognition accuracy of 94% using non-normalized data. Using only 6 words as adaptation-set, the MAP-approach leads to a significant error reduction of about 10% relative and 19% relative in combination with normalization. Regarding an unsupervised adaptation, the error can be reduced of up to 44% relative. After writer adaptation the gain of normalization methods decreases, except for a supervised adaptation with very few samples.

References

[1] J.-L. Gauvain and C.-H. Lee. Maximum a Posteriori Estimation for Multivariate Gaussian Mixture Observation of Markov Chains. *IEEE Transactions on Speech and Audio Processing*, 2(2):291–298, Apr. 1994.

[2] J. Hu, S. Lim, and M. Brown. HMM Based Writer Independent On-line Handwritten Character and Word Recognition. In *Proc. of IWFHR-6*, pages 143–155, Taejon, Korea, 1998.

[3] S. Jaeger, S. Manke, J. Reichert, and A. Waibel. On-Line Handwriting Recognition: The NPen++ Recognizer. *Int. Journal Doc. Analys. a. Recog. (IJDAR)*, 3:169–180, 2001.

[4] C. Leggetter and P. Woodland. Speaker Adaptation of Continuous Density HMMs using Multivariate Linear Regression. In *Int. Conference on Spoken Language Processing (ICSLP)*, Yokohama, Japan, Sept. 1994.

[5] P. Natarajan, I. Bazzi, Z. Lu, J. Makhoul, and R. Schwartz. Robust OCR of Degraded Documents. In *5th International Conference on Document Analysis and Recognition (ICDAR)*, pages 357–361, Bangalore, India, Sept. 1999.

[6] R. Plamondon and S. Srihari. On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 22(1):63–84, Jan. 2000.

[7] L. Rabiner and B. Juang. An Introduction to Hidden Markov Models. *IEEE ASSP Magazine*, pages 4–16, 1986.

[8] G. Rigoll, A. Kosmala, and D. Willett. A Systematic Comparison of Advanced Modeling Techniques for Very Large Vocabulary On-Line Cursive Handwriting Recognition. In *Advances in Handwriting Recognition*, chapter 2, pages 69–78. World Scientific, 1999.

[9] A. Senior and K. Nathan. Writer adaptation of a HMM handwriting recognition system. In *IEEE Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1447–1450, Munich, Apr. 1997.

[10] S. Srihari, S.-H. Cha, and S. Lee. Establishing Handwriting Individuality using Pattern Recognition Techniques. In *6th Int. Conference on Document Analysis and Recognition (ICDAR)*, pages 1195–1204, Seattle, Sept. 2001.