Reprint: 1996 Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP-96), Atlanta, USA, Vol. 6, pp. 3438-3441, May 1996

HMM-BASED HANDWRITTEN SYMBOL RECOGNITION USING ON-LINE AND OFF-LINE FEATURES

Hans-Jürgen Winkler
Institute for Human-Machine-Communication
Munich University of Technology
Arcisstr. 21, 80290 Munich, Germany
win@mmk.e-technik.tu-muenchen.de

ABSTRACT

This paper addresses the problem of recognizing on-line sampled handwritten symbols. Within the proposed symbol recognition system based on Hidden Markov Models different kinds of feature extraction algorithms are used analysing on-line features as well as off-line features and combining the classification results.
By conducting writer-dependent recognition experiments, it is demonstrated that the recognition rates as well as the reliability of the results is improved by using the proposed recognition system. Furthermore, by applying handwriting data not representing symbols out of the given alphabet, an increase of their rejection rate is obtained.

1. INTRODUCTION

This paper is concerned with the problem of recognizing on-line sampled handwritten symbols, which is one stage within the overall system presented at ICASSP'95 for understanding handwritten mathematical expressions [1][2].

Based on the soft-decision approach within the symbol segmentation and recognition stage of the overall system, next to the recognition rate two additional features are important:
- the reliability of the recognition results.
- the rejection of handwriting data not representing symbols out of the given alphabet.

Within the proposed symbol recognition system the classification is founded on single recognizers analysing different kinds of feature vectors. The obtained single recognizer results are finally combined. The improvements obtained by using the proposed recognition system are demonstrated by writer-dependent experiments.

2. SYMBOL RECOGNITION SYSTEM

2.1 System overview

Our system is based on the on-line sampled handwritten data. On-line means that the input data is a sequence of strokes captured during writing, each stroke itself is a sequence of (x,y)-coordinates corresponding to the pen positions. A stroke, in this connection, is the writing from pen down to pen up.
Based on the on-line data of each symbol, three different kinds of feature extraction algorithms are used. One of these algorithms is a typical on-line algorithm using the temporal information of the handwriting for generating a sequence of feature vectors $\{x_o\}$. The remaining two algorithms are off-line algorithms generating sequences of feature vectors $\{x_v\}$ and $\{x_h\}$ extracted by the image of the symbol.
Each sequence of feature vectors $\{x_o\}$, $\{x_v\}$ and $\{x_h\}$ represents the input of a symbol recognizer based on Hidden Markov Models (HMMs). The classification for each HMM is done by computing the maximum „a posteriori" probability that a symbol model generates the corresponding sequence of feature vectors.

As illustrated in fig. 1, the recognizer results are combined by multiplying their generation probabilities and taking the square root of the result.

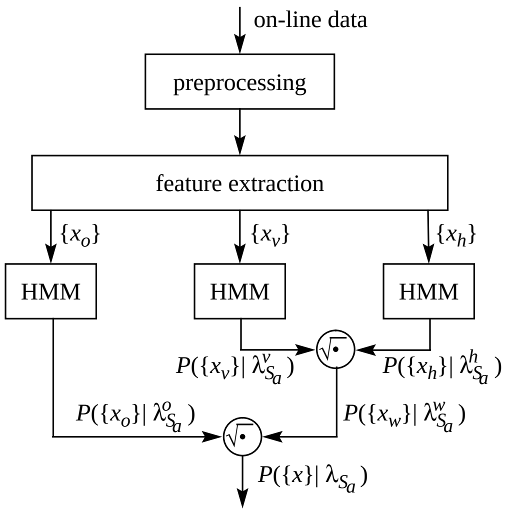

Figure 1: Overview to the symbol recognition system

2.2 Preprocessing

Preprocessing is subdivided into the following stages:

Reprint: 1996 Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP-96), Atlanta, USA, Vol. 6, pp. 3438-3441, May 1996

- The incoming temporally sequenced data of each stroke are smoothed by lowpass-filtering.
- The slant of the handwriting is estimated by detecting the near-vertical parts of each stroke. The correction is done by carrying out a shear.
- For size and position normalisation two pairs of (x,y)-coordinates, namely (x_{10},y_{10}) and (x_{90},y_{90}), are calculated. The thresholds x_{10} and x_{90} denote the coordinates where 10% of the overall stroke length of the symbol is on the left respectively on the right side of these thresholds. The thresholds y_{10} and y_{90} represents the analogous thresholds in the vertical direction.

2.3 „On-line“ Feature Extraction

Evaluating the temporal information is the most obvious kind for recognizing on-line sampled handwriting. Remarkable recognition results are concerned with recognizing handwritten characters [3] and words [4].

In our system so-called hidden strokes are integrated between each pair of temporal successive strokes within the symbol first. Each hidden stroke represents the pen movement from the final pen position of the actual stroke to the starting position of the successive stroke. This insertion is advantageous especially for the letters „i“ and „j“ containing small dots.

By using the coordinates (x_{10},y_{10}) calculated during the preprocessing stage position normalisation is carried out, size normalisation is done by means of the second pair (x_{90},y_{90}). Finally, the temporally sequenced strokes and the integrated hidden strokes are resampled at equispaced points along the trajectory retaining the temporal order. This resampling results in an elimination of the velocity in writing the symbol. For each point a feature vector x_{o_t} is constructed containing the local position, the sine and cosine value of the angle φ and the information whether the actual point belongs to a stroke or to an integrated hidden stroke. φ denotes the angle between the horizontal axis and the vector connecting the previous and the actual point. For the first feature vector x_{o_1} the sine and cosine value of φ is set to zero.

2.4 „Off-line“ Feature Extraction

In comparison with on-line symbol recognition off-line recognition is based on the image of the handwriting containing no temporal information.

In our system the image of the handwritten symbol is calculated by interpolating the on-line sampled data. Two sequences of feature vectors $\{x_v\}$ and $\{x_h\}$ are generated by mapping different grids upon this image. A detailed description and an illustration of this procedure is given in [1].

2.5 HMM-Topology

For each symbol S_a of our alphabet $\{S_1, \ldots, S_A\}$ three HMMs $\lambda^o_{S_a}$, $\lambda^v_{S_a}$ and $\lambda^h_{S_a}$ are generated corresponding to the different feature extraction algorithms.

The number of states N, the state transition probabilities $P(s_j|s_i)$ and the observation probabilities $P(x_t|s_i)$ completely specify each HMM [5].

Within our system semicontinuous first order left to right HMMs are used. This means that the state transition probabilities $P(s_j|s_i)$ are set to zero if $j < i$ or $j > i+2$. The observation probability $P(x_t|s_i)$ for a given feature vector x_t at state i is determined by a mixture of Gaussian distributions g_k and can be written as:

$$P(x_t|s_i) = \sum_k (P(x_t|g_k) \cdot P(g_k|s_i)) \; .$$

Furthermore, the first feature vector x_1 out of a sequence $\{x_t\}$ is fixed to the first state s_1, the last feature vector x_T to the last state s_N respectively.

Recognition by means of HMMs λ^f_S (f indicates any of the three feature extraction algorithms or a combination of them) is carried out by calculating the maximum „a posteriori“ probability $P(\{x_f\}|\,\lambda^f_{S_a})$ that a symbol model generates the observed sequence of feature vectors $\{x_f\}$:

$$\hat{S}^f = \underset{S_a}{\operatorname{argmax}}\,[\,P(\{x_f\}|\lambda^f_{S_a})].$$

As illustrated in fig. 1, the off-line results are combined by calculating the generation probability

$$P(\{x_w\}|\,\lambda^w_{S_a}) = \sqrt{P(\{x_v\}|\,\lambda^v_{S_a}) \cdot P(\{x_h\}|\,\lambda^h_{S_a})} \; .$$

Analogous, the final combination is done by

$$P(\{x\}|\,\lambda_{S_a}) = \sqrt{P(\{x_o\}|\,\lambda^o_{S_a}) \cdot P(\{x_w\}|\,\lambda^w_{S_a})} \; .$$

3. RESULTS AND DISCUSSION

3.1 Data Sets

Within the overall system symbols out of a 84-character alphabet are used. Three of these symbols, namely „Dot“, „Minus“ and „Fraction“, are classified by means of a prerecognition stage [6]. The remaining 81 symbols containing upper and lower case letters as well as digits, mathematical operators and other special symbols are used for HMM training and the recognition experiments. An illustration is given in fig. 2.

For each writer (currently nine) the data sampling was performed under two different environments: writing single symbols on the one hand and mathematical expressions on the other hand.

Within the single sampled data set each writer contributed 50 versions of each symbol within the alphabet. These symbols are used as part of the HMM training data set.

By means of the system presented in [6], 27 different math-

Reprint: 1996 Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP-96), Atlanta, USA, Vol. 6, pp. 3438-3441, May 1996

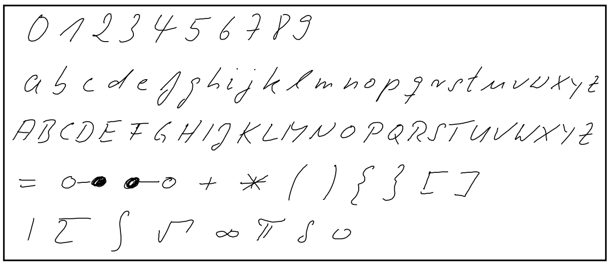

Figure 2: Handwritten symbols of the alphabet used for the recognition experiments, written by „wh“

ematical expressions written up to 10 times by each writer are sampled. A few samples are given in fig. 3.

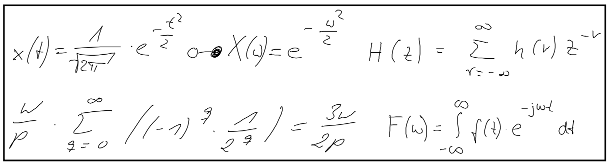

Figure 3: Handwritten expressions taken from the data set

Altogether, each writer contributed more than 5000 symbols by writing expressions, each symbol of the alphabet was represented. 60% of these symbols are used for HMM training too, the remaining symbols are applied to the recognition experiments.

3.2 Recognition experiments using symbols extracted from mathematical expressions

The results obtained with applying symbols extracted from mathematical expressions to the recognition system are summarized in tab. 1. In order to achieve independency of the symbol distribution within the recognition data set, for each writer averaging is done over the recognition rates for each symbol of the alphabet.
Additionally, the average Top-2 recognition rates (the „a posteriori“ probability $P(\{x_f\} | \lambda^f_{S_a})$ corresponding to the symbol S_a has to be one of the two most probable) are given.

By combining the three recognizer results a reduction of the average Top-1 error rate of about 13% in comparison with the best single recognizer is obtained.

Caused by the large number of nearly indistinguishable symbols such as „0“, „o“ and „O“ or „x“ and „X“ of the alphabet, a further improvement without using any contextual knowledge is not realistic. More than 80% of the final recognition errors are based on such confusions. This is indicated by the results achieved by the Top-2 recognition experiments. None of the three single recognizer is able to handle this problem, therefore the combination also fails in distinguishing these symbols.

writer	$\{x_v\}$	$\{x_h\}$	$\{x_o\}$	$\{x_w\}$	$\{x\}$
bw	86.2%	85.9%	91.9%	90.7%	92.3%
gv	88.0%	87.9%	91.8%	91.7%	94.4%
hs	87.4%	86.4%	91.9%	89.7%	92.2%
kb	90.5%	89.1%	93.1%	90.1%	92.5%
kl	89.6%	89.8%	90.9%	91.6%	91.8%
km	82.5%	82.6%	85.6%	86.7%	88.4%
ls	85.0%	83.1%	87.0%	87.9%	88.9%
sa	87.8%	87.4%	86.5%	88.9%	89.6%
wh	85.1%	83.8%	87.9%	87.3%	89.1%
AVG	86.9%	86.2%	89.6%	89.4%	91.0%
Top-2	97.0%	96.1%	98.5%	98.0%	98.7%

Table 1: Recognition rates using symbols extracted from mathematical expressions

As shown in tab. 1, the recognition rate is weak decreasing for some writers by combining the single recognizer results, which is also based on the mix-up of these symbols.

3.3 Experiments concerning the reliability of the recognition results

Using the results obtained by the symbol recognition system for further processing such as recognizing handwritten expressions [1] or words [4] using contextual knowledge, next to the Top-n recognition results the reliability of the results is important for reducing the perplexity of the following tasks. By using the reliability we want to determine how confident the recognizer system is in classifying the symbols.

By applying the feature sequence $\{x_f\}_{S_{a_i}}$ generated by the i-th handwritten version of the symbol S_a to the recognizer and analysing the generation probabilities obtained by the HMMs or their combination, a measurement for the reliability R of this classification result can be calculated by using the cost function C presented in [7][8]:

$$R = 1 - C = \frac{1}{A} \cdot \sum_a \frac{1}{I} \cdot \sum_i (1 + e^{-\gamma \cdot d_{a_i}})^{-1} ,$$

$$d_{a_i} = \log[P(\{x_f\}_{S_{a_i}} | \lambda^f_{S_a})] - \log[\max_{S_r \neq S_a}[P(\{x_f\}_{S_{a_i}} | \lambda^f_{S_r})]].$$

By distinguishing between correct and wrong classification results achieving independency from the Top-1 recognition rates, two reliabilities R_c and R_w are calculated by using $\gamma = 1$. Based on the feature extraction algorithms and their combination, the average results of the reliability calculation is given in tab. 2.
Additionally, the theoretical maximum values for R_c and R_w are given too.

Reprint: 1996 Int. Conference on Acoustics, Speech, and Signal Processing (ICASSP-96), Atlanta, USA, Vol. 6, pp. 3438-3441, May 1996

R	$\{x_v\}$	$\{x_h\}$	$\{x_o\}$	$\{x_w\}$	$\{x\}$	max.
R_c	0.78	0.77	0.77	0.81	0.83	1.00
R_w	0.37	0.36	0.39	0.37	0.40	0.50

Table 2: Reliabilities R_c and R_w calculated by the recognition experiments analogous to tab. 1

Comparing the results based on the final combination with any other of the given results, it is realized that the reliability R_c calculated by the correct classification results as well as R_w is always the highest. Therefore, the classification is more unequivocal if the result is correct and, if the result is wrong, the confusion error between the correct result and the classification result is minimized.
By calculating the reliability R without distinguishing between wrong and correct classification results, the improvement of the final combination will additionally increase caused by the higher recognition rate.

3.4 Experiments concerning the rejection of handwriting data not representing a symbol

In the preceding experiments handwriting data representing symbols S_a out of the alphabet $\{S_1, \ldots, S_A\}$ are used for demonstrating the performance of the proposed symbol recognition system.
Within the overall system data not representing symbols of a handwritten expression are also applied to the symbol recognition system. This is caused by the soft-decision approach within the symbol segmentation stage [6]. Therefore, the symbol recognizer has to reject these data by a poor generation probability in relation to the generation probability obtained by applying the symbols S_a out of the alphabet [1].

The writer-independent soft-decision segmentation results in a symbol hypotheses net composed of the elements:

- S_e containing the strokes belonging to a symbol used within the handwritten expression. These symbols are identical to the data set used within the preceding experiments.
- S_s containing a subgroup of strokes belonging to a symbol S_e.
- S_m containing strokes of two or more symbols.

In order to achieve independency of the number $N(S_{e,s,m})$ of these elements, writer-independent thresholds T_s and T_m are calculated maximizing the rejection rates J_s and J_m:

$$J_{s,m} = \frac{1}{2}\Big(N\big(S_{s,m}\big|_{(\max_{S_a}[P(\{x_f\}_{S_{s,m}}|\lambda^f_{S_a})] < T_{s,m})}\big) / N(S_{s,m}) \Big) + \frac{1}{2}\Big(N\big(S_e\big|_{(\max_{S_a}[P(\{x_f\}_{S_e}|\lambda^f_{S_a})] \geq T_{s,m})}\big) / N(S_e) \Big) .$$

Based on the three feature extraction algorithms and their combination, the optimized rejection rates J_s and J_m are given in tab. 3 concluding two different results.

J	$\{x_v\}$	$\{x_h\}$	$\{x_o\}$	$\{x_w\}$	$\{x\}$
J_s	66.5%	61.3%	66.3%	70.9%	71.6%
J_m	88.1%	90.0%	96.3%	94.7%	97.6%

Table 3: Rejection rates J_s and J_m

At first, the rejection rates J_s and J_m are increasing by combining the generation probabilities of the single recognizers. This is based on the combination by multiplication, just one poor generation probability of any single recognizer results in a significant devaluation of the final generation probability.
Furthermore, the rejection rate J_s is quite low caused by the fact that in many cases the elements S_s are symbols S_a of the alphabet even if they are not symbols S_e within the expression. For example, the letter „B“ given in fig. 2 consists of the letter „I“ and the digit „3“. Their devaluation has to be based on geometrical features [6].

4. REFERENCES

[1] M. Koschinski, H.-J. Winkler, M. Lang, *Segmentation and Recognition of Symbols within Handwritten Mathematical Expressions*, ICASSP 1995 Vol.4, pp. 2439-2442, 1995.

[2] H.-J. Winkler, H. Fahrner, M. Lang, *A Soft-Decision Approach for Structural Analysis of Handwritten Mathematical Expressions*, ICASSP 1995 Vol.4, pp. 2459-2462, 1995.

[3] K. S. Nathan, J. R. Bellegarda, D. Nahamoo, E. J. Bellegarda, *On-line Handwriting Recognition using Continuous Parameter Hidden Markov Models*, ICASSP 1993 Vol.5, pp. 121-124, 1993.

[4] E. J. Bellegarda, J. R. Bellegarda, D. Nahamoo, K. S. Nathan, *A Discrete Parameter HMM Approach to On-line Handwriting Recognition*, ICASSP 1995 Vol.4, pp. 2631-2634, 1995.

[5] L. R. Rabiner, *A Tutorial on Hidden Markov Models and selected Applications in Speech Recognition*, Proceedings of the IEEE Vol.77 No.2, pp. 257-286, Feb. 1989.

[6] S. Lehmberg, H.-J. Winkler, M. Lang, *A Soft-Decision Approach for Symbol Segmentation within Handwritten Mathematical Expressions*, to be published in ICASSP 1996.

[7] J.B. Hampshire II, A.H. Waibel, *A Novel Objective Function for Improved Phoneme Recognition Using Time-Delay Neural Networks*, IEEE Trans. on Neural Networks Vol.1 No.2, pp. 216-228, June 1990.

[8] W. Chou, B.H. Juang, C.H. Lee, *Segmental GPD training of HMM based speech recognizer*, ICASSP 1992 Vol. 1, pp. 473-476, 1992.