board and to the recognition of sentences using language models to achieve further improvements of the word recognition rate.

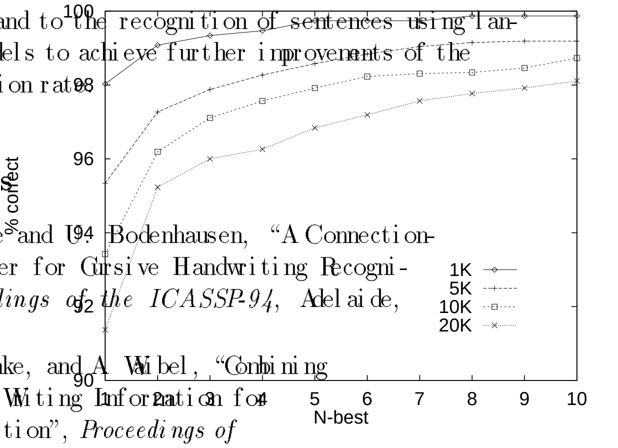

Figure 5: Recognition results with respect to the dictionary size if the $N = 1 \ldots 10$ best words are counted as correct.

incorrect output of the recognizer shows that we can expect further improvements of the word recognition rate by using language models for the recognition of sentences.

5 Conclusions

In this paper we have presented the **NPen**$^{++}$ system, a connectionist recognizer for writer independent on-line cursive handwriting recognition. This system combines a robust input representation, which preserves the dynamic writing information, with a neural network integrating recognition and segmentation in a single framework. This architecture has been shown to be well suited for handling temporal sequences as provided by this kind of input.

Testing the system on different dictionary sizes showed recognition rates from 98.0% for a ... to 82.9% for the 100,000 word ... largest dictionary used in the ... recognition time for a pattern ... seconds. These results are especially ... they were achieved with a ... other systems (e.g. [4]). ... system has proved to ... dictionary. Though ... dictionaries, ... and much ... depend ... length of ... full ... y-

References

[1] S. Manke and U. Bodenhausen, "A Connectionist Recognizer for Cursive Handwriting Recognition", *Proceedings of the ICASSP-94*, Adelaide, April 1994.

... Manke, M. Finke, and A. Waibel, "Combining ... with Dynamic Writing Information for ... writing Recognition", *Proceedings of* ... Jerusalem, October 1994.

... ke, and A. Waibel, "The Use ... mation in a Connection... ng Recognition Sys... *ation Processing*

... on, "On- ... De-

3.3 Training algorithm

During training the goal is to determine a set of parameters θ that will maximize the posterior probability $p(w|\boldsymbol{x}_0^T, \theta)$ for all training input sequences. But in order to make that maximization computationally feasible even for a large dictionary system we had to simplify that maximum a posteriori approach to
maximum likelihood training procedure that maxi-
$\boldsymbol{x}_0^T|w\theta)$ for all words instead.
rst step of our maximum likelihood training
rap the recognizer using a subset of ap-
500 words of the training set that were
the database: cur-
with the character boundaries to ad-
d a mixture of
word layer correctly. After train-
led data, the recognizer is used
set of unlabeled training data.
ining set is processed by the
its were se-
s determined automatically
urial
e target word unit serve
Then, in the second
l on both data sets to
recognizer.

lts

erent writer inde-
ging from 1,000
l in the dic-
r case let-
y 5,700
80

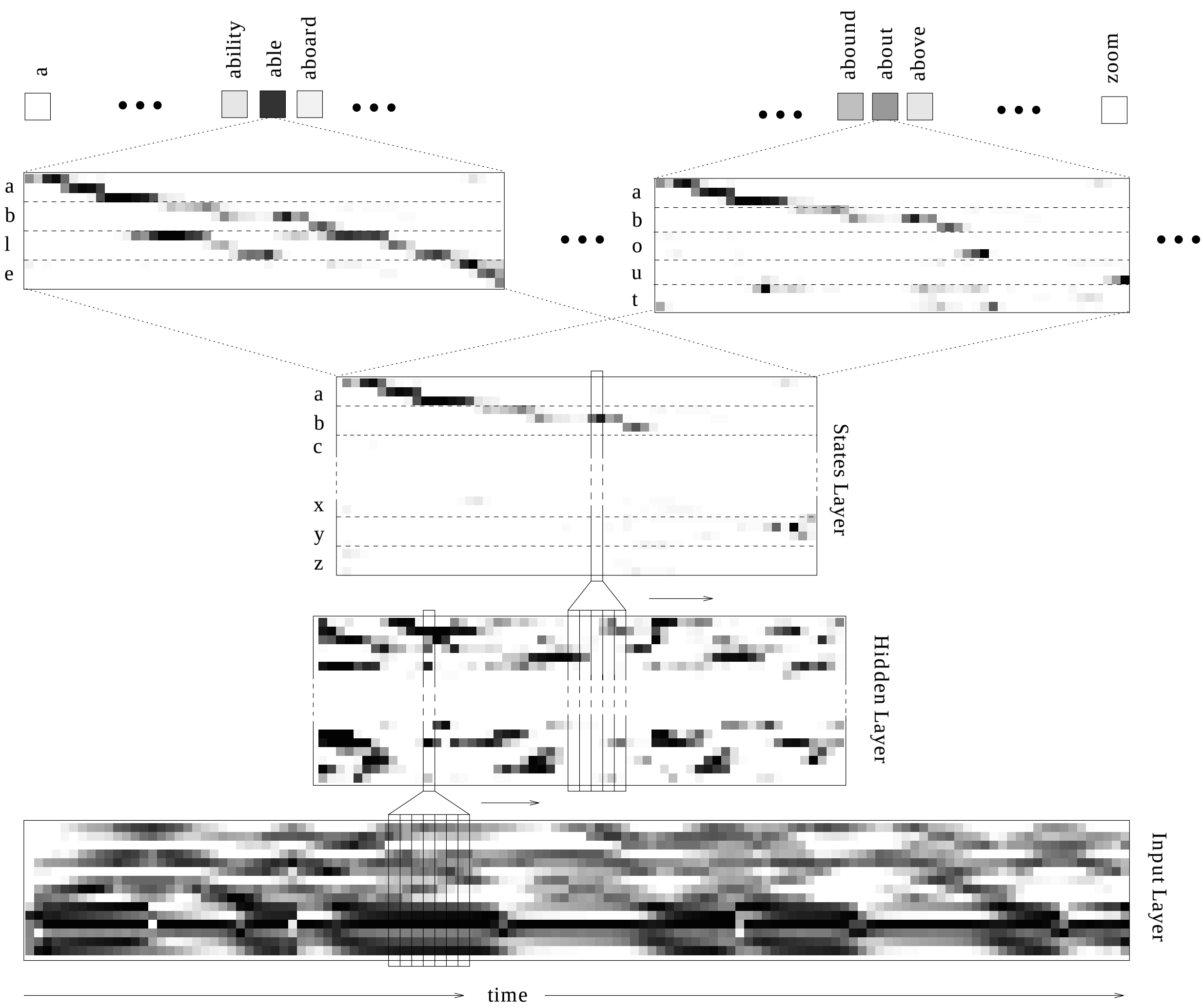

Figure 3: The Multi-State TDNN architecture, consisting of a 3-layer TDNN to estimate the a posteriori probabilities of the character states combined with word units, whose scores are derived from the word models by a Viterbi approximation of the likelihoods.

ndows in each layer. In the current implementation of the
system, a TDNN with 15 input units, 40 units in the hidden
en layer, and 78 state output units is used.
ime delays in the input layer and 5 time
den layer.
rmalized output of the states layer
stimate of the probabilities of the
put window $x_{t-d}^{t+d} = x_{t-d} \ldots x_{t+d}$,
.e.

$$\approx \frac{\exp(\eta_j(t))}{\sum_k \exp(\eta_k(t))} \quad (1)$$

the weighted sum of inputs to the output
sed on these estimates, the
efined to be a Viterbi approximation

approximation of the log likelihoods of the feature vector sequence, given the word model w, i.e. $\log p(x_0^T|w)$ is approximated by

$$\max_{q_0^T} \sum_{t=1}^{T} \log p(x_{t-d}^{t+d}|q_t, w) + \log p(q_t|q_{t-1}, w)$$

$$\approx \max_{q_0^T} \sum_{t=1}^{T} \log \frac{p(q_t|x_{t-d}^{t+d})}{p(q_t)} + \log p(q_t|q_{t-1}, w).$$

Here, the maximum is over all possible sequences of states $q_0^T = q_0 \ldots q_T$ given a word model, $p(q_t|x_{t-d}^{t+d})$ is the output of the states layer as defined in (1) and $p(q_t)$ is the prior probability of observing a state q_t estimated on the training data.

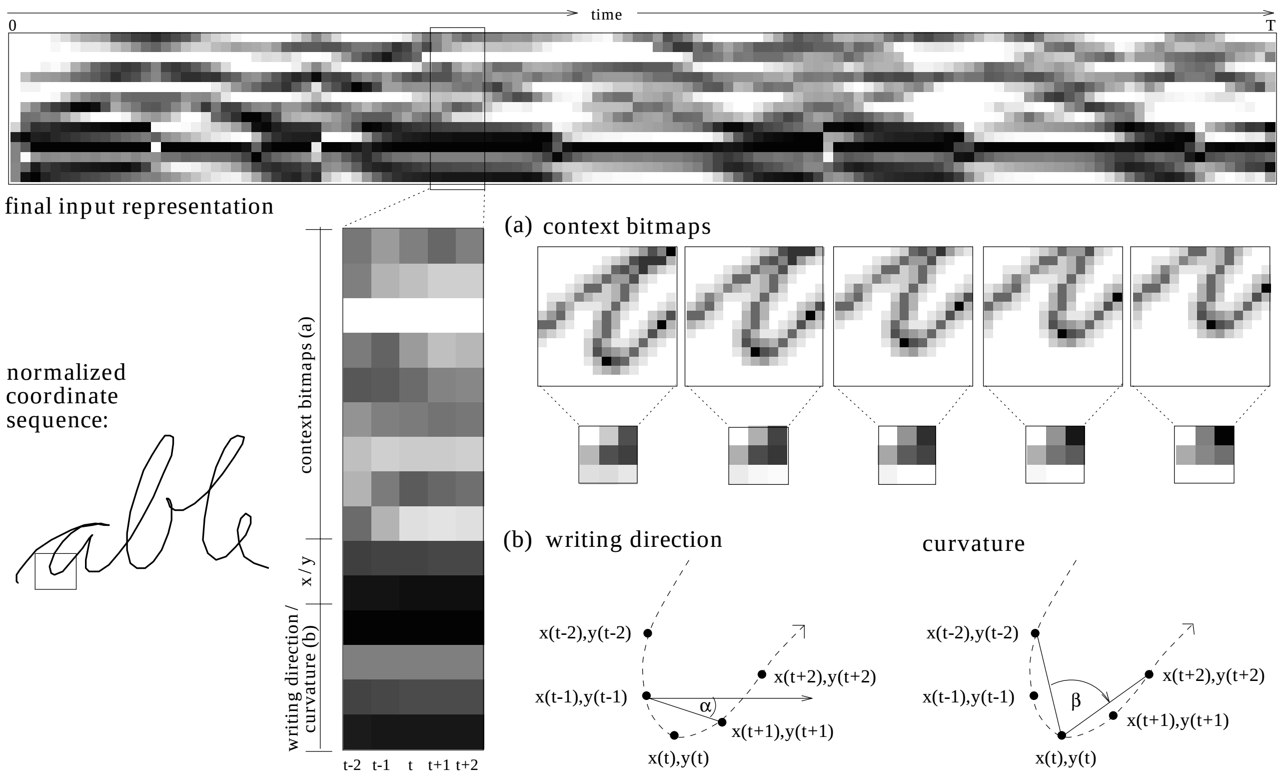

Figure 2: Feature extraction for the normalized word "able". The final input representation is derived by calculating a 15-dimensional feature vector for each data point, which consists of a context bitmap (a) and information about the curvature and writing direction (b).

recognition tasks [7, 8]; ... posterior probability $p(w|x_0^T, \theta)$ given a fixed set of ...ttern recognition ... parameters θ and a feature vector sequence ... -linear time alignment algorithm ... The handwritten word will then be recognized such ... for finding stroke and character boundaries ... handwritten words. ... that

$$w = \operatorname{argmax}_{w_i \in W} p(w_i|x_0^T, \theta).$$

...ling assumptions

In our Multi-State Time Delay Neural Network approach the problem of modeling the word posterior ... probability $p(w|x_0^T, \theta)$ is simplified by using Bayes' ... expresses that probability as ... each of these words w is represented as a ... characters $w \equiv c_{i_1} c_{i_2} \ldots c_{i_n}$, where each ... is modeled by a three state hidden ...

$$p(w|x_0^T, \theta) = \frac{p(x_0^T|w, \theta)P(w|\theta)}{p(x_0^T|\theta)}$$

... q_j^2. The idea of using three ... to model explicitly the initial ... of the characters. Thus, w is ... Instead of estimating $p(w|x_0^T, \theta)$ directly, we de- ... section a network that is supposed ... $w \equiv q_{j_1} q_{j_2} \ldots q_{j_n}$. In ... end of the feature vector sequence ... probabilities $p(q_j|q_j)$ ... $p(q_j|q_{j-1})$ are both ... transition probabilities

...architecture

... MS-TDNN architecture for ... The first three lay- ... with sliding input

ainst rota- on the LCD tablet or digitizer [10]. The system is de-
determine signed to make heavy use of this temporal information.
pecta- **NPen**$^{++}$ (Figure 1) combines a neural network rec-
- ognizer, which was originally proposed for continuous
speech recognition tasks [7, 8], with robust preprocess-
ing techniques, which transform the original sequence
of data points into a still temporal sequence of N
mensional feature vectors.
We have tested the system on the writer indepen-
cognition of isolated words with dictionary sizes
000 up to 100,000 words. Word recognition
rom 98.0% for the 1,000 word dictionary
100,000 word dictionary without us-
l. Even for the largest dictio-
ents the average recognition
less than 1.5 seconds.
escribes the preprocessing
em. The architecture and
nizer are presented in
experiments to eval-
have achieved on
4.

CR) input usu-
without any
s writ-
gh a
ese

Proceedings of the Internatio
on Document Analysis and Reco
Montreal (Canada), August 199
IEEE Computer Society

NPen^{++}: A Writer Independent, Large Vocabulary On-Line Cursive Handwriting Recognition System

Stefan Manke, Michael Finke, and Alex Waibel

University of Karlsruhe   Carnegie Mellon Universi
Computer Science Department   School
D-76128 Karlsruhe, Germany

Abstract

In this paper we describe t
for writer independent on-lin
tion. This recognizer needs no
ular writer and can recogniz
style (cursive, hand-printed,
The neural network architectur
proposed for continuous speech
the preprocessing techniques o
to make heavy use of the d
tion, i.e. the temporal sequen
on a LCD tablet or digitiz
the writer independent re
Tested on different dicti
100,000 words, recogniti
the 1,000 word dictionar
dictionary and 82.9% for th
No language models are used

1 Introduction

The success and user acce
or multi-modal systems highly de
ity of the on-line handwriting recog
these systems. To achieve acceptable
formance currently available handwriti
are often either writer dependent or
training for a particular writer to
writing. Additionally people usually have to
a particular writing style, e.g. hand-printed,
to use special character shapes defined in the
instead of the usual shapes to get the high perf
mance. All this together makes it very hard for
people using these systems to write as natural as they
usually would do on paper. Too small recognition
an additional restriction in some of the systems.