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

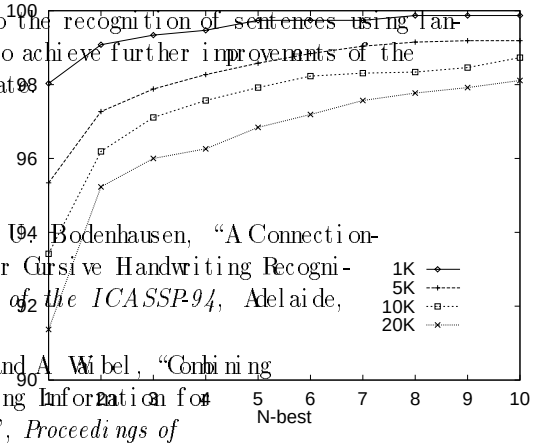

Figure 5: Recognition results with respect to the dictionary size if the $N = 1 \ldots 10$ best words are counted as correct.

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 framework. This architecture has been shown suited for handling temporal sequences as the kind of input.

the system on different dictionary cognition rates from 98.0% for a to 82.9% for the 100,000 word largest dictionary used in the cognition time for a pat- conds. These results are es- hey were achieved with a her systems (e.g. [4]). ystem has proved to ionary. Though dictionaries, t and much t depend gth of

full y-

References

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

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

e, and A. Waibel, "The use nation in a Connection- g 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-
ed data, the recognizer is used
set of unlabeled training data.
ining set is processed by the
ds were se-
s determined automatically
ursial
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

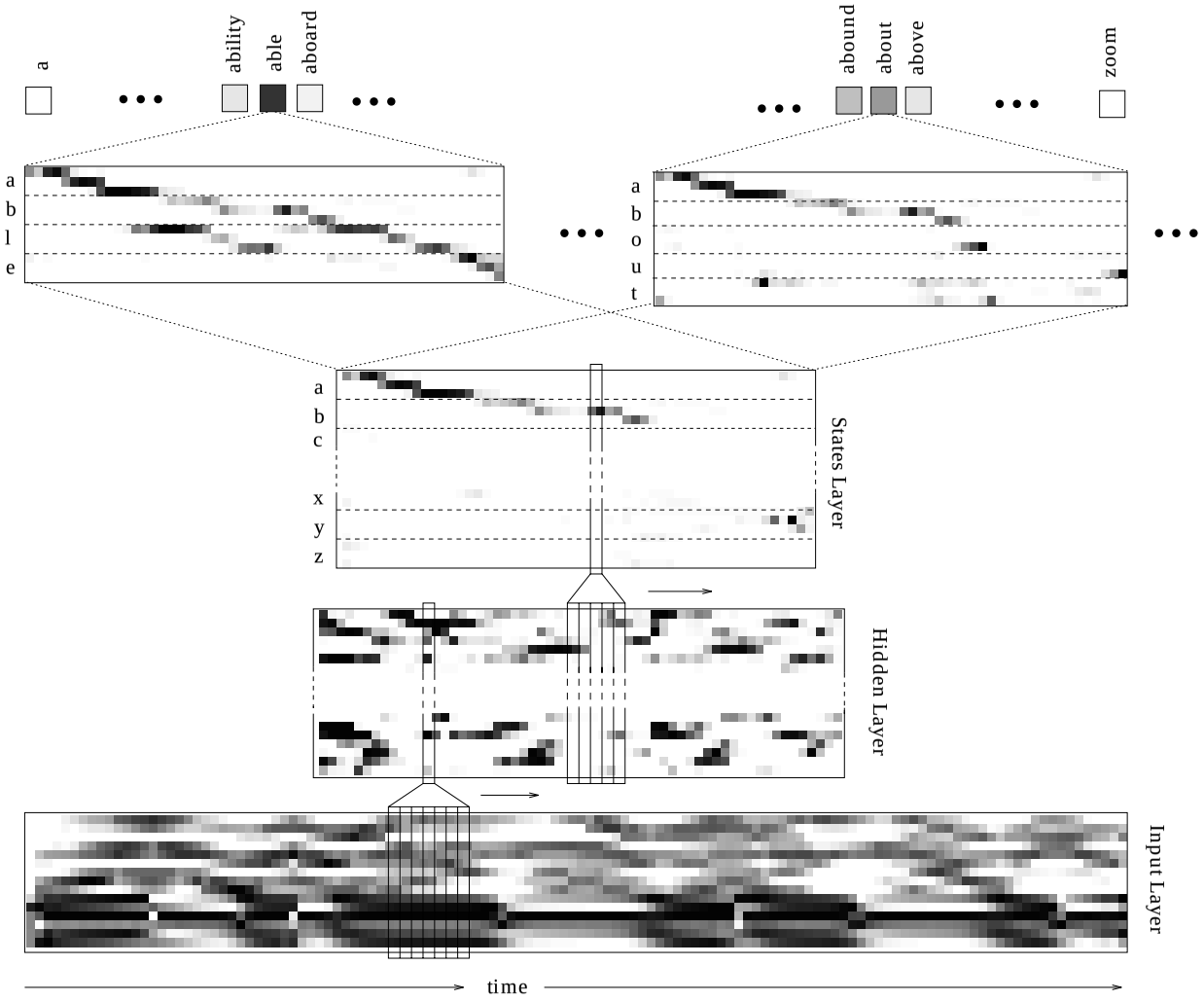

Figure 3: The Multi-State TDNN architecture, consisting of a 3-layer TDNN to estimate the a posteriori probabilities of the character states combined with word units, whose scores are derived from the word models by a Viterbi approximation of the likelihoods.

ndows in each layer. In the current implementation of this approximation the log likelihoods of the feature vector
system, a TDNN with 15 input units, 40 hidden units, the word model ψ, i.e. $\log p(\boldsymbol{x}_0^T|\psi)$ is
en layer, and 78 state output units is used. approximated by
ime delays in the input layer and 5 time
dden layer.

$$\max_{q_0^T} \sum_{t=1}^{T} \log p(\boldsymbol{x}_{t-d}^{t+d}|q_t,\psi) + \log p(q_t|q_{t-1},\psi)$$

rmalized output of the states layer
stimate of the probabilities of the
put window $\boldsymbol{x}_{t-d}^{t+d} = \boldsymbol{x}_{t-d} \ldots \boldsymbol{x}_{t+d}$
.e.

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

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

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

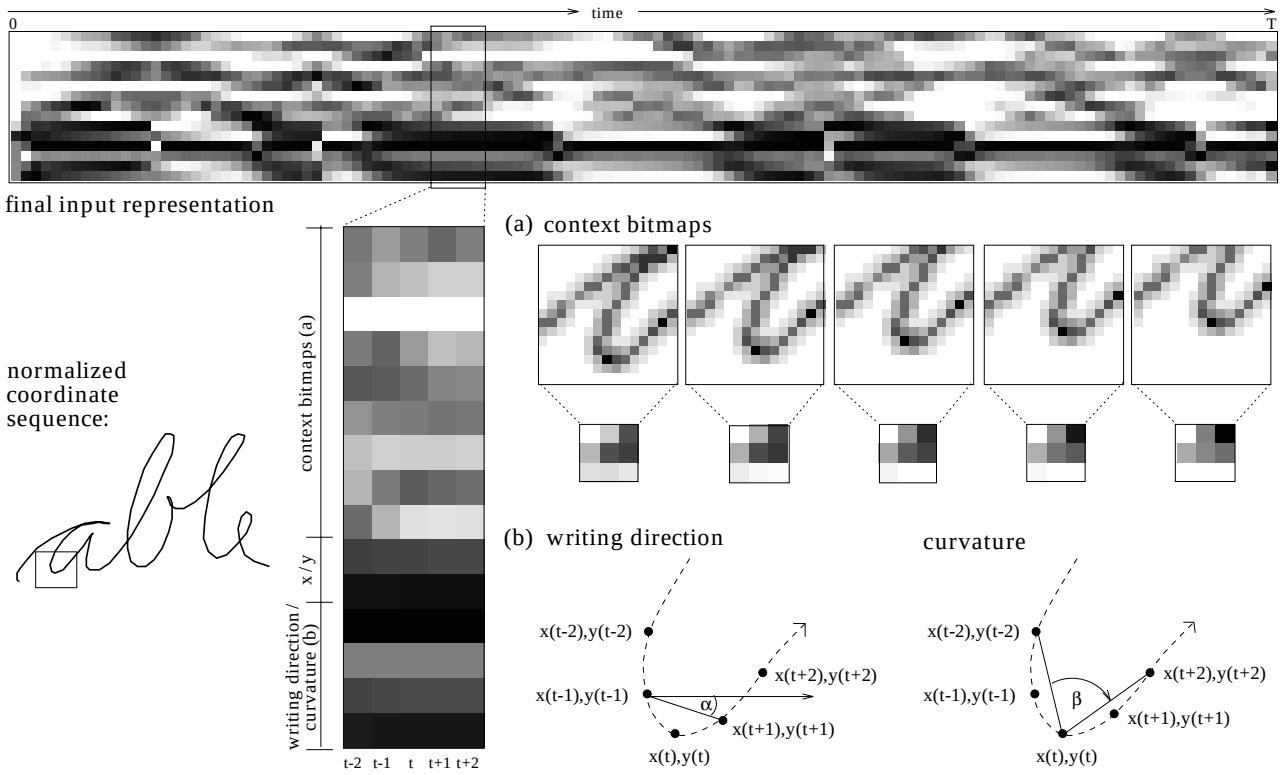

Figure 2: Feature extraction for the normalized word “able”. The final input represent
calculating a 15-dimensional feature vector for each data p
information about the curvature

recognition tas
p

on the LCD tablet or digitizer [10]. The system is designed to make heavy use of this temporal information.

NPen++ (Figure 1) combines a neural network recognizer, which was originally proposed for continu
speech recognition tasks [7, 8], wit
ing techniques, wh
of

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

Universi ty of Kar ls r uhe
Computer Sci ence Department
D-76128 Karlsruhe, Ger many

Carnegie Mellon Universi
School

Abstract

In this paper we describe the **N**
for writer independent on-line handwri
tion. This recognizer needs no training for
ular writer and can recognize any common writi
style (cursive, hand-printed, or a mixture of both).
The neural network architecture, which was originally
proposed for continuous speech recognition tasks, and
the preprocessing techniques of **NPen**$^{++}$ *are designed*
to make heavy use of the dynamic writing informa-
tion, i.e. the temporal sequence of data points recorded
on a LCD tablet or digitizer. We present results
the writer independent recognition of isola
Tested on different dictionary sizes fr
100,000 words, recognition rates ra
the 1,000 word dictionary to 9
dictionary and 82.9% for the
No language models are u

1 Introducti

The succes