A Handwriting Recognition System Based on Visual Input

Markus Wienecke, Gernot A. Fink, Gerhard Sagerer

Bielefeld University, Faculty of Technology
33594 Bielefeld
email: {mwieneck,gernot,sagerer}@techfak.uni-bielefeld.de

Abstract. One of the most promising methods of interacting with small portable computing devices such as personal digital assistants is the use of handwriting. However, for data acquisition touch sensitive pads, which are limited in size, and special pens are required. In order to render this communication method more natural Munich & Perona [11] proposed to visually observe the writing process on ordinary paper and to automatically recover the pen trajectory from video image sequences. On the basis of this work we developed a complete handwriting recognition system based on visual input. In this paper we will describe the methods employed for pen tracking, feature extraction, and statistical handwriting recognition. The differences compared to classical on-line recognition systems and the modifications in the visual tracking process will be discussed. In order to demonstrate the feasibility of the proposed approach evaluation results on a small writer independent unconstrained handwriting recognition task will be presented.

1 Introduction

In the field of human computer interaction the ultimate goal is to make interaction with computing machinery possible without the need for special input devices such as keyboard or mouse. Therefore, more natural modalities of communication such as speech, gesture, and handwriting are investigated. However, in order to capture a person's writing, again a combination of special input devices is necessary, namely touch sensitive pads and special pens. As handwriting is currently mostly used for the interaction with small personal digital assistants (PDAs) also the size of the input field is severely restricted. Therefore, a promising alternative might be to retain ordinary pen and paper to produce handwritten text and capture the writing process by observing it using a video camera — an idea first envisioned by Munich & Perona in 1996 [11].

In contrast to traditional methods to obtain so called on-line handwriting data — i.e. the dynamically sampled trajectory of the pen — an approach based on visual observation will be faced with severe difficulties in obtaining the trajectory data. As a consequence the input data of the recognition process itself

[1] This work was partially supported by the German Research Foundation (DFG) within project **Fi799/1**.

In 2nd International Workshop on Computer Vision Systems, pages 63-72, Vancouver, Canada, 2001. IEEE.

might be erroneous or noisy. Therefore, a subsequent recognition model has to be robust against the inherent limitations of the data acquisition method.

In this paper we present a complete handwriting recognition system based on visual input. A short review of related work in on-line handwriting recognition will be given in the following section. The architecture of our system will be described in section 3. Additionally, it will be shown in which respects the capabilities of our system go beyond the original work by Munich & Perona [11], which did not yet include a recognition module. The methods employed for tracking a pen in video data, extracting robust features, and for statistical recognition will be discussed in subsequent sections.

2 Related Work

In handwriting recognition usually two different research areas are distinguished[1]. When the handwritten text, which has been produced by ordinary writing on paper, is subsequently digitized for automatic processing using some sort of image acquisition device, so called off-line recognition is performed. Handwriting data is thus viewed as some special sort of images to be analyzed. On-line handwriting recognition, however, deals with dynamic data describing the motion of the pen as observed during the writing process. For recording such pen trajectories mostly special digitizing tablets are used. In contrast to off-line data the dynamics of the writing process can be exploited for the analysis of on-line pen trajectories.

Earlier approaches to recognition of on-line handwriting applied segmentation-based methods in combination with sophisticated classification techniques (cf. e.g. [16]). In more recent work, however, so called segmentation-free approaches based on Hidden-Markov-Models (HMMs) were successfully applied to the task and gained growing interest in the research community (e.g. [12, 9, 5]). This is certainly to some extent due to the fact that the HMM technology reached a state of maturity in the context of speech recognition (cf. e.g. [7]), where hardly any fundamentally different approaches are pursued today. As pen-trajectory data can also be viewed as a time series of samples transferring the HMM approach from speech recognition to the domain of on-line handwriting recognition is almost straight forward — a fact that has also been explicitly pointed out by some authors (e.g. [15]).

As for the acquisition of trajectory data special devices are necessary the investigation of alternative methods for obtaining on-line handwriting data seems promising. Munich & Perona [11] first proposed to use a video-based approach. Handwriting can thus be produced with an ordinary pen on paper. The writing process is observed by a video camera and the pen trajectory is automatically extracted from the recorded image sequence. In the original work the feasibility of the video-based tracking was demonstrated. However, even in more recent work [10] handwriting recognition based on video input was not investigated.

[1] For an overview of the field of handwriting and character recognition see e.g. [2].

A comparable approach using video-based data acquisition was recently proposed in [1]. Difference images are used to obtain the ink trace left on the paper during writing. Using an existing recognition system results on a small writer dependent task are presented.

3 System Architecture

The proposed complete system for handwriting recognition based on the visual observation of the writing process consists of four major building blocks depicted in Fig. 1. It shares modules for feature extraction from pen trajectory data as well as for recognition of handwritten characters or words with most current on-line handwriting recognition systems. However, the trajectory data can't be obtained directly from a special device such as a digitizing tablet but are the results of a method for tracking the pen tip in image sequences during the process of writing. We use a slightly modified version of the approach developed by Munich & Perona [11]. The most important modification is to include a completely automatic procedure for finding the pen initially before the actual writing starts. This feature is of fundamental importance in order to make the interaction with such a system as natural as possible.

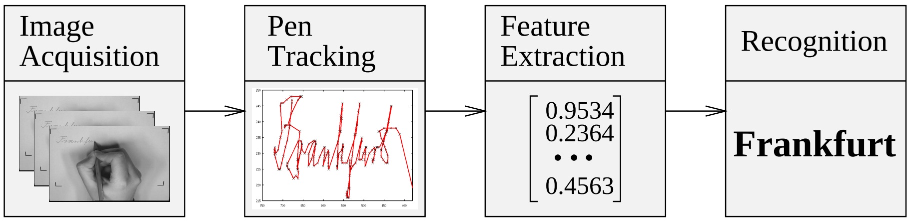

Fig. 1. Architecture of the video based handwriting recognition system.

The image sequences from which the pen trajectory is obtained during writing are captured using a standard video camera working in interlaced PAL mode. We obtain 50 frames per second with a resolution of 768×288 pixels. As the pen tip needs to be clearly visible in every single image the positioning of the camera has to be considered carefully. We found the setup show in Fig. 2 to provide a good compromise between the different requirements. It allows the user to write in a rather natural position. The writing space visible in the video images covers an area of approximately 18×13 cm so that the writing of multiple word instances or even short portions of text can be observed. As the camera can't be positioned directly above the writing area the scene is observed from an angle of approximately 65 degrees. An illumination that proved to be sufficient for our purposes is provided by two ordinary table lamps. During video based handwriting recognition the digitizing tablet shown in Fig. 2 is not used. However,

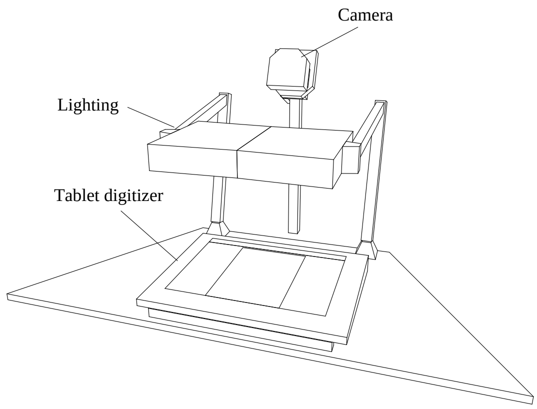

Fig. 2. Hardware setup of the video based handwriting recognition system: The digitizing tablet is only used to collect supplementary data.

it serves as a means to record precise pen trajectories as a sort of ground truth data for comparison with the results of video based tracking. Larger amounts of training material are also captured much more easily using the digitizing tablet.

4 Pen Tracking

The essential information for online handwriting recognition is the trajectory of the pen tip, hence the position of the pen tip has to be determined in every single image of the video sequence. The architecture of our system for video-based pen-tracking is illustrated in Fig. 3. In most parts it is based on the system developed by Munich & Perona [11], except for changes in the initial pen finding module and the Kalman filter. Additionally, a different approach for the pen up/down classification is used.

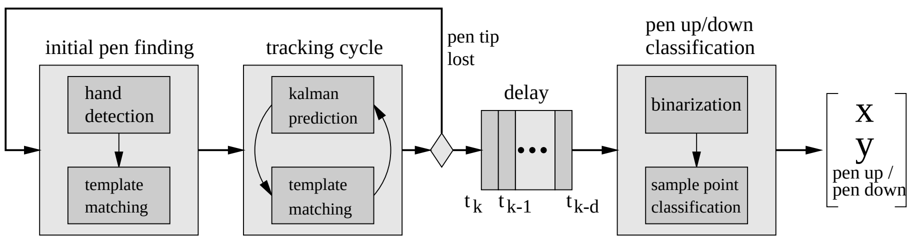

Fig. 3. Architecture of tracking system: The time-consuming template matching in the initial finding step works on the entire image whereas during the tracking process only a small cut will be used. In order to prevent occlusions pen-up/down classification is performed with a certain delay. The output vector consists of the position of the pen tip and pen-up/down information.

Fig. 4. Tracking: The search window for the template matching is illustrated for the time steps t_0, t_1 and the current frame t_k during the writing of the word *"Frankfurt"*.

An important step towards system usability is to automate the initial pen finding and the recovery of the tracking process when the pen is lost. Therefore, the initial pen finding procedure was designed to require very little user cooperation. In order to reduce the response time as much as possible, the initialization is subdivided into two parts: First, a fast hand detection module notices if the scene changes significantly, e.g. a hand moves into the scene. This is achieved by calculating the difference image of two consecutive frames. If the difference image yields regions with high activity the initial template matching will be performed as the second part of the initial pen finding. This step is time-consuming and cannot be performed in real time. A speed-up is attained by using a pyramidal scheme for comparing a set of template images of the pen tip with the current image. However, the user must not move the pen for about three seconds during this initial template matching phase.

The pen tip is assumed to be found, if the difference between the template and the image falls below a certain threshold. In order to increase the robustness of the template matching, a contrast enhancement is carried out by normalizing the intensity values of the templates and the image. This approach compensates well for variations in the lighting conditions.

After the initial pen position is determined, the pen tip has to be tracked during the whole writing process. In order to achieve real time performance the search space for the template matching is reduced by predicting the expected position of the pen tip in the succeeding video image using a Kalman filter. The position of the pen tip is then estimated by performing a template matching in a small region around the predicted position (Fig. 4). If the pen tip cannot be found, the processing returns to the initial finding procedure.

The Kalman filter represents a model of the kinematic motion of the pen tip, assuming that it may be described as a superposition of a constant velocity term and white noise, which compensates for the time-varying acceleration. For efficiency reasons we use a realization of the Kalman filter a proposed by Kohler

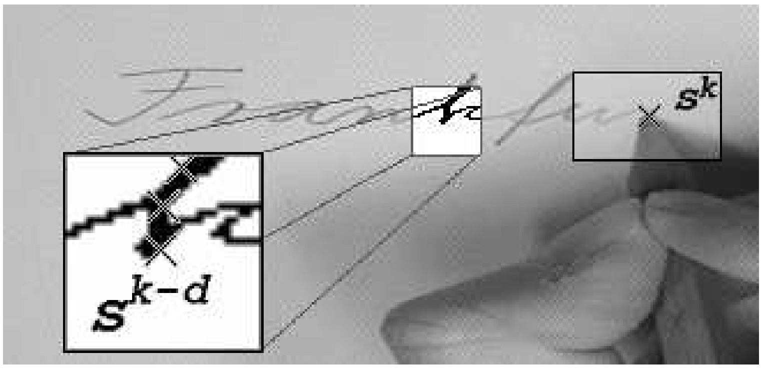

Fig. 5. Pen-up/down classification: The gray crosses mark the positions of the pen-tip found by the template matcher. The image is binarized in a small surrounding of the point s^{k-d}, extracted d frames ago, assuming that the ink trace is usually dark against the background.

[8]. In contrast to the method originally used in [11] the internal model is split up into two independent components for horizontal and vertical motion.

The method described so far is capable of tracking the pen tip in image sequences regardless of whether it is actually writing on paper or moved to another position during a so called pen-up phase. However, for the recognition process the distinction between the two writing modes is extremely important. Therefore, we use an approximate method for deciding between pen-up and pen-down strokes, assuming that the ink trace is usually dark against the background.

A small region surrounding the calculated pen position is binarized using a threshold that is automatically determined from the bimodal intensity histogram of that region [13]. If intensity values falling below that threshold can be found in the vicinity of the line connecting adjacent pen positions, this segment of the pen trajectory will be classified as pen-down and as pen-up otherwise. Additionally, if the intensity histogram is not bimodal, we assume that there is no ink trace visible in the region and classify the segment of the trajectory as pen-up.

In order to prevent occlusions by the pen tip or the writer's hand we perform the pen-up/down classification with a certain time delay, so that the current position of writing is sufficiently far away from the part of the image that has to be analyzed. Experiments indicated that after approximately one second the distance between the current writing position and the image area previously occupied by the pen tip is sufficiently large (Fig. 5).

5 Feature Extraction

Before feature vectors are calculated some preprocessing steps are applied to the trajectory data, mainly in order to make the features to a certain degree independent of the device or method used for pen-tracking. Therefore, not only

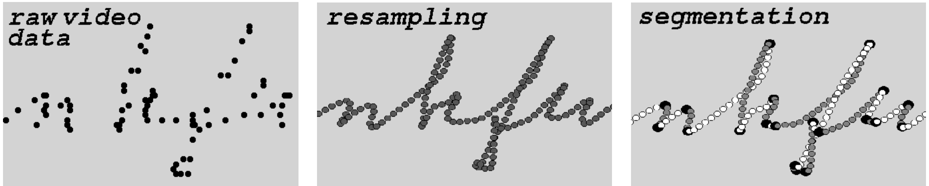

Fig. 6. Preprocessing: From left to right the raw samples obtained from the pen tracker, the resampled data, and the segmented trajectory are shown.

data obtained from the pen-tracking process described in the previous section can be used in the development of the video-based handwriting recognition system but also also traditional trajectory data obtained from a tablet digitizer.

After smoothing the trajectory by replacing each sample point with the mean value of its neighbors, the trajectory is spatially resampled using a linear interpolation scheme in order to compensate for device specific sampling frequencies and variations in the velocity of handwriting. This approach works well if trajectory data are sampled at higher frequencies of approximately 100 Hz and above which is common for tablet digitizers. However, when applied to the video-based tracking results sampled at only 50 Hz the method does not produce sufficiently smooth curves. In order to increase the smoothness we apply a momentum term, that considers the previous orientation of the vector connecting adjacent sample points.

The second step consists of partitioning the trajectory into stroke-like segments. The segmentation criterion is based on the local curvature and the vertical velocity of the pen tip, assuming that a high curvature or a low velocity signalize start and end points of handwritten strokes (Fig. 6). This idea goes back to the work of Plamondon & Maarse [14].

The main reason for segmenting the trajectory into stroke like entities is to allow the computation of writing-size invariant features — an important aspect for writer-independent systems. According to the work of Dolfing & Haeb-Umbach [3] an overlap of 50% of the segments was used for calculating 17 dimensional feature vectors, which include features like the sine and cosine of the mean writing direction, sine and cosine of the mean curvature, pen-up/down information, aspect ratio, and curliness of the segment as defined in [6]. In order to take the dynamics of the signal into account we also compute approximations of the derivatives of the above mentioned features. Inspired by [3] we additionally calculate so called delayed features, namely the sine and cosine of the change of writing direction and the relative change of writing size between the feature vectors at time t and $t-2$.

6 Handwriting Recognition

For the design of statistical recognition systems the availability of a sufficiently large database of training samples is an important prerequisite. Ideally, for a video-based recognition system it would be desirable to obtain a large amount of image sequences recorded while observing a subject's writing of words or sentences. However, the recording of such video data requires much more effort and storage than the direct recording of trajectory data with a suitable device.

Therefore, we decided to use the digitizing tablet (Wacom UltraPad A3) shown in Fig. 2 for the recording of training and cross-validation data. Furthermore, for this first data acquisition effort we decided to limit recording to isolated words. The lexicon was somewhat arbitrarily chosen to comprise the names of large German cities[2].

Using the same hardware setup as during video-based tracking we recorded data from 13 subjects who were asked to write the complete list of 201 city names in random order on ordinary paper using a special ball-point pen that can be tracked by the digitizing tablet. For collecting a test set of video-based data this procedure was repeated with four different writers using visual pen tracking. As we wanted to obtain unconstrained handwritten data no instructions concerning the writing style were given.

For configuration, training, and decoding of Hidden-Markov-Models (HMMs) for the video-based handwriting recognition task we used the methods and tools provided by the ESMERALDA development environment [4]. From the recorded trajectories we selected the data of 7 writers for setting up a semi-continuous HMM-system with a shared codebook containing 256 Gaussian mixtures with diagonal covariance matrices. Models for 52 characters and 2 diacritical marks – namely dot for *"i"* and German umlauts and dash for *"t"* – were trained using standard Baum-Welch reestimation. As we do not apply stroke reordering after a word has been completely written we need to allow for different temporal relations between base characters and associated diacritic marks. Therefore, for every word including diacritic marks a set of different realizations is contained in the lexicon depending on whether diacritics immediately follow the base character or are added after the word has been completed. The overall recognition lexicon thus contains 335 entries.

On the cross-validation set of unconstrained handwritten data from 6 different writers using the digitizing tablet we achieved a recognition rate of 92.5%. For the evaluation of the video-based approach the training set was enlarged to comprise the data of 11 writers. On the test set of the 4 writers using visual pen tracking a recognition rate of 75.5% was achieved. 3.4% of the errors were introduced by the video-based pen tracking.

Fig. 7 shows recognition results obtained on typical video data. After the initial pen finding the tracking process performs quite well also on faster pen movements between words. However, during the writing of the word *"Hamburg"* the pen tip is lost, but the tracking process recovers quickly and proceeds normally afterwards.

[2] From a list of German city names and associated zip codes we selected those cities that were assigned more than one code resulting in a list of 201 different names.

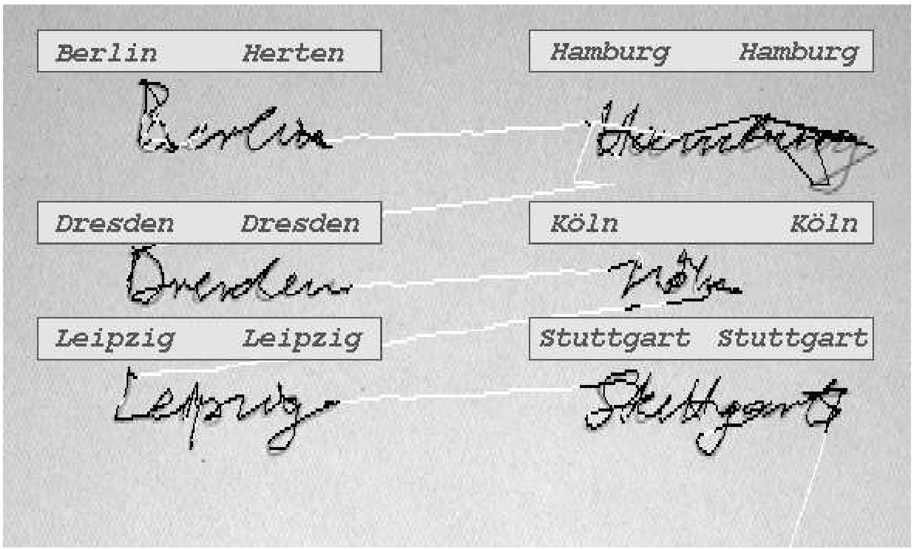

Fig. 7. Recognition results: A 6-word trajectory obtained from the pen tracker is superimposed to the captured video image. The small boxes illustrate the words that were written on the left and the recognition results on the right hand side. Furthermore, this sequence demonstrates the robustness of the pen tracking and recovery, because during writing the letter *"g"* of the word *"Hamburg"*, the pen tip is lost, but after a few frames automatically recovered.

7 Conclusion

We proposed a complete on-line handwriting recognition system based on visual input. Pen trajectories are automatically extracted from video image sequences recorded during the observation of the writing process. The pre-processing and feature extraction methods applied generate a data representation which is to a certain extent robust against the limitations of the video-based pen tracking and facilitates writer independent recognition. Promising recognition results could be achieved for experiments on a writer independent unconstrained handwriting recognition task.

8 Acknowledgments

We would like to gratefully acknowledge the support of Mario Munich, formerly California Institute of Technology, and Pietro Perona, California Institute of Technology, who allowed us to use their software for visual pen-tracking in the development of our visual handwriting recognition system.

Additionally, we would like to thank Marc Hanheide, Birgit Möller, Thomas Plötz, and Frank Seifert, students of Computer Science, Bielefeld University,

for enthusiastically developing a visual pen-tracking system during a project seminar.

References

1. H. Bunke, T. von Siebenthal, T. Yamasaki, and M. Schenkel. Online handwriting data acquisition using a video camera. In *Proc. Int. Conf. on Document Analysis and Recognition*, pages 573–576, 1999.
2. H. Bunke and P. S. P. Wang, editors. *Handbook of Character Recognition and Document Image Analysis*. World Scientific Publishing Company, Singapore, 1997.
3. J. G. A. Dolfing and R. Haeb-Umbach. Signal representations for hidden markov model based on-line handwriting recognition. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, volume IV, pages 3385–3388, München, 1997.
4. Gernot A. Fink. Developing HMM-based recognizers with ESMERALDA. In Václav Matoušek, Pavel Mautner, Jana Ocelíková, and Petr Sojka, editors, *Lecture Notes in Artificial Intelligence*, volume 1692, pages 229–234, Berlin Heidelberg, 1999. Springer.
5. Jianying Hu, Michael K. Brown, and William Turin. HMM based on-line handwriting recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 18(10):1039–1044, 1996.
6. S. Jaeger, S. Manke, and A. Waibel. Npen++: An on-line handwriting recognition system. pages 249–260, Amsterdam, September 2000.
7. Frederick Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, MA, 1997.
8. Markus Kohler. Using the kalman filter to track human interactive motion - modelling and initialization of the kalman filter for translational motion -. Technical Report 629, Informatik VII, University of Dortmund, January 1997.
9. John Makhoul, Thad Starner, Richard Schwartz, and George Chou. On-line cursive handwriting recognition using hidden markov models and statistical grammars. In *Human Language Technology*, pages 432–436. Morgan Kaufmann, 1994.
10. Mario E. Munich. *Visual Input for Pen-based Computers*. PhD thesis, California Institute of Technology, Pasadena, California, January 2000.
11. Mario E. Munich and Pietro Perona. Visual input for pen-based computers. In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 33–37, Vienna, Austria, 1996.
12. Krishna A. Nathan, Jerome R. Bellegarda, David Nahamoo, and Eveline J. Bellegarda. On-line handwriting recognition using continuous parameter Hidden Markov Models. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, volume 5, pages 121–124, Minneapolis, 1993.
13. N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man, and Cybernetics*, 9:62–66, 1979.
14. Réjean Plamondon and Frans J. Maarse. An evaluation of motor models of handwriting. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(5):1060–1072, 1989.
15. T. Starner, J. Makhoul, R. Schwartz, and G. Chou. On-line cursive handwriting recognition using speech recognition methods. In *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, volume 5, pages 125–128, Adelaide, 1994.
16. Charles C. Tappert, Ching Y. Suen, and Toru Wakahara. The state of the art in on-line handwriting recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(8):787–808, 1990.