Video-Based On-line Handwriting Recognition

Gernot A. Fink Markus Wienecke Gerhard Sagerer
Bielefeld University, Faculty of Technology
33594 Bielefeld, Germany
E-mail: gernot@techfak.uni-bielefeld.de

Abstract

The use of handwriting provides a natural way of interacting with small portable computers. However, in order to capture handwritten text on-line special input devices are necessary. Therefore, Munich & Perona [13] proposed to use visual input for pen-based computers. Writing can then be performed on ordinary paper and pen trajectories are automatically extracted from image sequences recorded during the writing process. On the basis of this work we developed a complete video-based on-line handwriting recognition system. We will present the techniques applied for pen tracking, pre-processing, feature extraction, and statistical modeling and recognition. Evaluation results on a writer independent unconstrained handwriting recognition task demonstrate that the inherent limitations of the video-based approach can be compensated using robust modeling combined with adaptation techniques.

1. Introduction

One of the most promising methods of interacting with small portable computing devices such as personal digital assistants is the use of handwriting. In contrast to the automatic processing of forms or postal addresses, where an image of the handwritten text is acquired and processed off-line, the movement of the pen is captured directly during the writing process. In order to obtain this so-called "on-line" handwriting data, however, special input devices such as touch sensitive pads together with special pens are necessary. Though these technical requirements restrict the use of handwriting in human computer interaction there has been little research in alternative methods. One idea first envisioned by Munich & Perona in 1996 is to produce handwritten text with an ordinary pen on paper and observe the writing process using a video camera [13]. In principle, the trajectory of the pen can then be recovered automatically by analyzing the recorded image sequence.

In contrast to traditional approaches dealing with on-line handwriting data the video-based acquisition of pen trajectories will never be able to produce perfect results. First, the tracking of an object in image sequences is a very challenging task. Therefore, errors can be introduced by the automatic pen-tracking method. Secondly, when using an ordinary video camera as sensor the data rate that can be reached by the tracking process will be limited by the video rate, which lies typically between 25 to 60 Hz. A digitizing tablet, however, can provide pen trajectory samples at 100 to 200 Hz or above. Finally, the spatial resolution of video-based pen-tracking is limited by the resolution of the video images (e.g. 768×576 for PAL) and the size of the desired field of view, which should be sufficiently large to facilitate the processing of not only single words but also smaller portions of text. In contrast, the spatial resolution of several thousand lines per inch that can be achieved by a digitizing tablet will be substantially higher. Therefore, the development of a video-based on-line handwriting recognition system will be a challenging task mainly due to the poor quality of the the input pen-trajectory data.

In the following section we will give a short review of relevant related work in on-line handwriting recognition. The combination of robust processing methods and statistical modeling we apply in our video-based approach will be described in section 3. Evaluation results on a writer independent unconstrained handwriting recognition task will be presented in section 4. Finally, the strengths and weaknesses of the approach will be discussed and possible method's for future improvements will be outlined.

2. Related Work

On-line handwriting recognition deals with dynamic data describing the motion of the pen during the process of writing. For recording such pen trajectories special devices such as digitizing tablets or touch sensitive displays are necessary. As long as the pen is touching the device's surface

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

In Proc. International Conference on Document Analysis and Recognition, pages 226-230. IEEE, 2001.

© IEEE. Personal use of this material is permitted.However, permission to reprint/republish this material for advertising or promotional purposes or for creating new collective works for resale or redistribution to servers or lists, or to reuse any copyrighted component of this work in other works, must be obtained from the IEEE. Contact: IEEE Intellectual Property Rights Office / IEEE Service Center / 445 Hoes Lane / Piscataway, NJ 08855-1331 / phone: (732) 562-3966 / fax: (732) 981-8062

it is assumed to be writing and leaving virtual ink traces forming the handwritten text. If the pen is lifted, however, so called "pen-up" strokes are generated.

In contrast to off-line data the dynamics of the writing process can be exploited for the analysis of on-line pen trajectories. However, also additional difficulties arise from the necessary distinction between pen-up and pen-down strokes. Furthermore, strokes may be arranged in many different temporal orders still producing the same handwritten word. For German this problem arises especially from the umlauts *"ä"*, *"ö"*, and *"ü"*. But also dots for *"i"* or the vertical bar of a *"t"* can be realized as a so called "delayed strokes" that do not immediately follow the production of the base character.

Earlier approaches to recognition of on-line handwriting applied segmentation-based methods in combination with sophisticated classification techniques (cf. e.g. [18]). 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. [14, 11, 5, 9]). 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. [17]).

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 [13] 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. Pen-tracking is accomplished by a combination of template matching for the pen tip and motion prediction using a Kalman filter. In order to make a distinction between pen-up and pen-down strokes possible, a rather complicated method is used to search for ink traces in the vicinity of the computed pen trajectory. As our approach to video-based on-line handwriting recognition is based on the work by Munich & Perona we use the same techniques for pen tracking with slight modifications. In the original study the feasibility of the video-based tracking was demonstrated. However, even in more recent work [12] handwriting recognition based on video input was not investigated.

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.

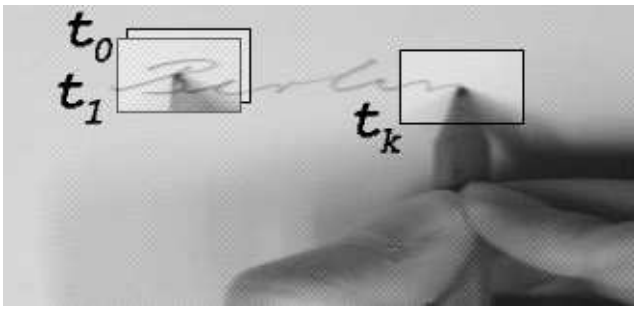

Figure 1. Pen tracking: The search window for the template matching is shown for time steps t_0, t_1 and the current frame t_k during the writing of the word *"Berlin"*.

As, however, this technique is extremely error prone a final image showing the current word completely needs to be acquired. This off-line information is then used to improve the pen trajectory extraction process. Using an existing recognition system results on a small writer dependent task are presented.

3. Video-based Handwriting Recognition

The proposed system for video-based on-line handwriting recognition consists of four major building blocks. Firstly, the trajectory of the pen is automatically extracted from the image sequences recorded during the observation of the writing process. We apply an approach for pen tracking which is slightly different the one used in the original study [13]. Secondly, the raw samples of the pen position are pre-processed in order to reduce the influences introduced by the method used for pen tracking. Thirdly, robust features are calculated for stroke-like segments of the pen trajectory. Finally, sequences of feature vectors are fed into a statistical recognition module for handwriting based on Hidden Markov Models (HMMs).

3.1. Pen Tracking

The image sequences required for video-based pen tracking are captured using a standard video camera working in interlaced PAL mode. We obtain 50 frames per second with a resolution of 768 $\times$ 288 pixels. The camera is positioned approximately 50 cm above the writing surface at an angle of 65 degrees. The writing space visible in the video images covers an area of approximately 18 $\times$ 13 cm so that the writing of multiple word instances or even short portions of text can be observed.

The method applied for video-based pen-tracking is based on the system developed by Munich & Perona [13],

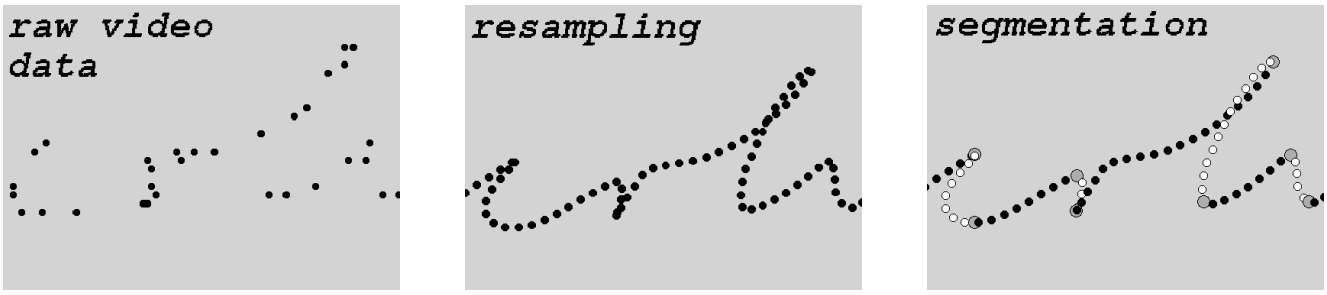

Figure 2. Pre-processing: From left to right the raw samples obtained from the pen tracker, the resampled data, and the segmented trajectory are shown.

except for changes in the Kalman filter realization and the pen-up/down classification used. Additionally, a completely automatic method for initial pen-finding is employed.

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 is reduced by predicting the expected position of the pen tip in the succeeding video image using a Kalman filter. For improved efficiency the kinematic motion model is split into a horizontal and a vertical component which are treated independently [8]. The real position is then estimated by performing a template matching in a small region around the predicted position as shown in Fig. 1. The size of this search window is determined from the prediction error estimated by the Kalman filter model and typically measures 120 by 60 pixels.

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 pen-up phase. However, for the recognition process the distinction between the two writing modes is extremely important. Therefore, we developed an approximate method for deciding between pen-up and pen-down strokes. Assuming that the ink trace will usually be dark against the light paper background a binarization of the image is performed in a region 30 by 30 pixels around the estimated pen position. The threshold for binarization is automatically determined from the intensity histogram of the region, assuming a bi-modal distribution of light and dark grey levels [15]. If intensity values falling below this threshold can be found along the line connecting adjacent pen positions, the corresponding trajectory segment will be classified as pen-down and as pen-up otherwise. In order to prevent occlusions by the pen tip or the writer's hand we perform the pen-up/down classification with a time-delay of approximately 1 second, so that the current position of writing is sufficiently far away from the part of the image that has to be analyzed.

3.2. Pre-processing

Before feature vectors are calculated some preprocessing steps are applied to the trajectory data, in order to make them to a certain degree independent of the device or method used for pen-tracking. Therefore, not only data obtained from the pen-tracking process described in the previous section can be used in the development of a video-based handwriting recognition system but also also traditional trajectory data obtained from a tablet digitizer.

First, the geometric distortion of the estimated pen positions caused by the viewing angle of the camera is corrected. Secondly, the trajectory is smoothed by replacing each sample point with the mean value of its neighbors. Then it is spatially resampled in order to compensate for device specific sampling frequencies and variations in the velocity of handwriting. For the placement of the intermediate sample points the orientation $\vec{v}(t)$ of a resampled segment t is computed from the current direction $\vec{z}_t$ to the following original sample point and the orientation of the preceeding resampled segment $t-1$ by applying a momentum term.

$$\vec{v}(t) = \alpha \frac{\vec{z}_t}{||\vec{z}_t||} + (1-\alpha)\vec{v}(t-1)$$

This technique is applied in order to further increase the smoothness of trajectories recorded with low sampling frequencies like 50 Hz.

The third step consists of partitioning the trajectory into stroke-like segments. Assuming that a high curvature or a low velocity signalize start or end points of handwritten strokes, the segmentation is based on the local curvature and the vertical velocity of the pen movement [16].

The effects of these pre-processing methods on video-based pen trajectories are illustrated in Fig. 2.

3.3. Feature Extraction

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. As suggested by [2] an overlap of 50% of the segments was used for calculating 17-dimensional feature vectors, which consist of 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 also approximate time-derivatives of these features are calculated. Inspired by [2] we additionally use 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 $t-2$.

3.4. Statistical Modeling and Recognition

Configuration, training, and decoding of Hidden-Markov-Models (HMMs) was carried out in the framework of the methods and tools provided by the ESMERALDA development environment [3]. As general setup of the HMMs we used semi-continuous systems 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 re-estimation. 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. For decoding of the enlarged recognition model standard Viterbi beam-search is used. The adaptation of model parameters to the specific characteristics of the test condition is achieved using Maximum Likelihood Linear Regression (MLLR) [10] which adapts the mixture means with affine transformations. Due to the low complexity of our current models and for efficiency reasons we apply the simplified approach described in [4].

4. Results & Discussion

As only limited conclusions can be drawn from performance measured on writer dependent experiments we decided to collect data for a writer independent evaluation. 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 collect on-line handwriting data for training purposes using a digitizing tablet (Wacom UltraPad A3). A total of 13 writers were asked to write the names of 201 large German cities well separated and in random order on A4-size paper. For collecting a test-set of video-based data this procedure was repeated with four different writers using visual pen tracking.

The evaluation results obtained for different training and testing conditions are summarized in table 1. First, we carried out a cross-validation experiment only on data collected via the digitizing tablet. Using the data of 7 writers for training and of 6 different writers for testing an error rate of 6.0% could be achieved. For the evaluation of our approach on video-based trajectory data the training set was enlarged to comprise the data of 11 writers. On the independent test set of video-based material from 4 different writers an error rate of 24.5% was reached. 3.4% of these errors were introduced by the video-based pen tracking.

In order to improve the recognition quality on video-data we carried out some additional experiments using adaptation. In the first one the data of the test writers were split into halves. Writer dependent supervised adaptation was then carried out on the first set and recognition on the second one. The overall error rate could be reduced by 10%. As supervised estimation techniques are always limited in their applicability in a second experiment we performed unsupervised adaptation directly during the recognition process. As this method will in general result in less improvement of recognition quality the resulting error rate of 23.6% is quite satisfactory.

Though the lexicon used in our experiments is relatively small the task is still challenging in several respects: it is writer independent, handwriting is unconstrained, and there exists a mismatch between training and testing conditions. The inherently reduced quality of the video-based trajectory data will probably account for most of the increased error rate compared to the cross-validation experiments on tablet data only. Furthermore, as the training set is still fairly small, we can not yet benefit from advanced modeling techniques such as context dependent units or allograph models.

In order to improve the recognition accuracy on video data we will therefore aim at collecting more training material. Our first experiments using adaptation also demonstrated that it will be beneficial to further investigate techniques dealing with mismatched training and testing conditions which are known from speech recognition research.

5. Conclusion

We proposed a complete on-line handwriting recognition system based on visual input. Pen trajectories are automatically extracted from image sequences recorded during the observation of the writing process. The pre-processing and

Training	Adaptation	Test	Error Rate
7 Wacom	*none*	6 Wacom	6.0%
11 Wacom	*none*	4 Video	24.5%
11 Wacom	*un*supervised, writer-specific, incremental	4 Video	23.6%
11 Wacom	*none*	50% of 4 Video	24.3%
11 Wacom	supervised, writer-specific, 50% of 4 Video	50% of 4 Video	22.0%

Table 1. Evaluation: Error rates achieved for different training and testing conditions (number of writers and data sets used) and adaptation methods applied.

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. We presented evaluation results on a writer independent unconstrained handwriting recognition task that demonstrate the success of the techniques applied. First experiments indicate that further improvements can be expected from the use of adaptation techniques.

Acknowledgments

We 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.

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] 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.

[3] G. A. Fink. Developing HMM-based recognizers with ESMERALDA. In V. Matoušek, P. Mautner, J. Ocelíková, and P. Sojka, editors, *Lecture Notes in Artificial Intelligence*, volume 1692, pages 229–234, Berlin Heidelberg, 1999. Springer.

[4] A. Fischer and V. Stahl. Database and online adaptation for improved speech recognition in car environments. In *Proc. Int. Conf. on Acoustics, Speech, and Signal Processing*, Phoenix, Arizona, 1999.

[5] J. Hu, M. K. Brown, and W. 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. In *Proc. 7th Int. Workshop on Frontiers in Handwriting Recognition*, pages 249–260, Amsterdam, September 2000.

[7] F. Jelinek. *Statistical Methods for Speech Recognition*. MIT Press, Cambridge, MA, 1997.

[8] M. 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] A. Kosmala, J. Rottland, and G. Rigoll. Large vocabulary on-line handwriting recognition with context dependent Hidden Markov Models. In *Mustererkennung 97, 19. DAGM-Symposium Braunschweig*, Informatik aktuell, pages 254–261. Springer, Berlin, 1997.

[10] C. J. Leggetter and P. C. Woodland. Maximum Likelihood Linear Regression for speaker adaptation of continuous density Hidden Markov Models. *Computer Speech & Language*, 9:171–185, 1995.

[11] J. Makhoul, T. Starner, R. Schwartz, and G. Chou. On-line cursive handwriting recognition using Hidden Markov Models and statistical grammars. In *Human Language Technology*, pages 432–436. Morgan Kaufmann, 1994.

[12] M. E. Munich. *Visual Input for Pen-based Computers*. PhD thesis, California Institute of Technology, Pasadena, California, Jan. 2000.

[13] M. E. Munich and P. Perona. Visual input for pen-based computers. In *Proc. Int. Conf. on Pattern Recognition*, volume 3, pages 33–37, Vienna, Austria, 1996.

[14] K. A. Nathan, J. R. Bellegarda, D. Nahamoo, and E. 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.

[15] N. Otsu. A threshold selection method from gray-level histograms. *IEEE Trans. on Systems, Man, and Cybernetics*, 9:62–66, 1979.

[16] R. Plamondon and F. J. Maarse. An evaluation of motor models of handwriting. *IEEE Trans. on Systems, Man, and Cybernetics*, 19(5):1060–1072, 1989.

[17] 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.

[18] C. C. Tappert, C. Y. Suen, and T. Wakahara. The state of the art in on-line handwriting recognition. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 12(8):787–808, 1990.