HMM Based Handwritten Text Recognition Using Biometrical Data Acquisition Pen

Ondřej Rohlík, Pavel Mautner and Václav Matoušek*
Dept. of Computer Science and Engineering
University of West Bohemia in Pilsen
www.kiv.zcu.cz

Jürgen Kempf
Dept. of Microelectronic Systems
University of Applied Sciences in Regensburg
www.fh-regensburg.de

Abstract

Development of new text and graphical input devices is considered to be important part of human-computer interaction by many researchers worldwide. The paper presents our experience with the text recognition methods that we have developed for a new designed electronic pen that produces signals corresponding to the movement of the pen on paper. Signals are described by a set of primitives and hidden Markov models are used for word recognition. Results of tests are discussed as well as other possible application areas of our electronic pen.

1 Introduction

There are many commercial systems designed for text recognition worldwide. They can be divided into off-line and on-line. Off-line (static) recognition deals with the recognition of text that is written on a paper-like medium—thus, the input of off-line recognizers is pictorial representations of text. In on-line (dynamic) recognition the data of a written text are recorded during writing. Hence, the additional timing information, i.e. on-line information, of the writing can be utilized to recognize the written text [1].

Off-line and on-line recognizers require different hardware. While off-line recognizers only need a scanner to digitize written text, on-line recognizers require a transducer that records the writing as it is written. In order to use on-line recognition methods, words must be written using special input device—this restricts the number of practical applications of on-line recognition techniques. The technology with which tablet digitizers are currently constructed is either electromagnetic/electrostatic or pressure-sensitive. Other approach to acquire data suitable for on-line recognition assumes that the text is written by special instrumented pen that enables tracing its location and/or movement.

The obvious disadvantage of all current devices that are used for data acquisition is the limited mobility [4] of a system composed of two or more parts (e.g. a pen with a tablet or a pen with an infrared transmitter and several receivers).

Hundreds of papers reporting the progress in handwritten text recognition have been published through-out more than thirty years of active research in the domain. In both off-line and on-line systems, the ultimate objective is to convert analog handwritten sentences or phrases (from off-line or on-line sources) into a digital form (ASCII) [3].

The handwritten text recognition problem is usually divided into isolated character recognition and word recognition. Isolated character recognition is considered an easier problem by many researchers as it does not require word segmentation in order to extract characters to be recognized. On the other hand, there are only a few applications that utilize isolated character recognition methods—Graffiti-like gesture recognizers used in handheld computers and hand-printed words recognizers.

Handwritten word recognition is a much more active research area. A straightforward approach to the recognition of handwritten words is to segment the words into single characters and then to classify these segments. This works well with high quality input only but not with cursive script, which is common in real applications. Sometimes it is even impossible to segment the word into discrete characters because of the inaccurate writing, especially on word endings, where sloppy writing often leads to the omission of characters. It is, therefore, desirable to avoid explicit segmentation. In speech recognition Hidden Markov Models (HMM) are used to solve the segmentation problem. The utilization of HMMs in handwriting recognition has also been reported [5].
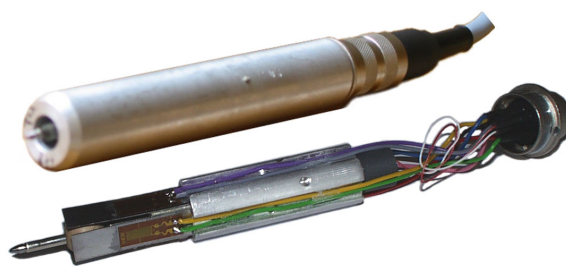

Figure 1: Biometrical pen—with and without cover

To avoid disadvantages of current on-line data acquisition systems mentioned above we have constructed an unique pen that integrates all the electronic devices needed for data acquisition inside the pen. The paper reports our experience with the handwritten text recognition methods developed to process the signals produced by the pen.
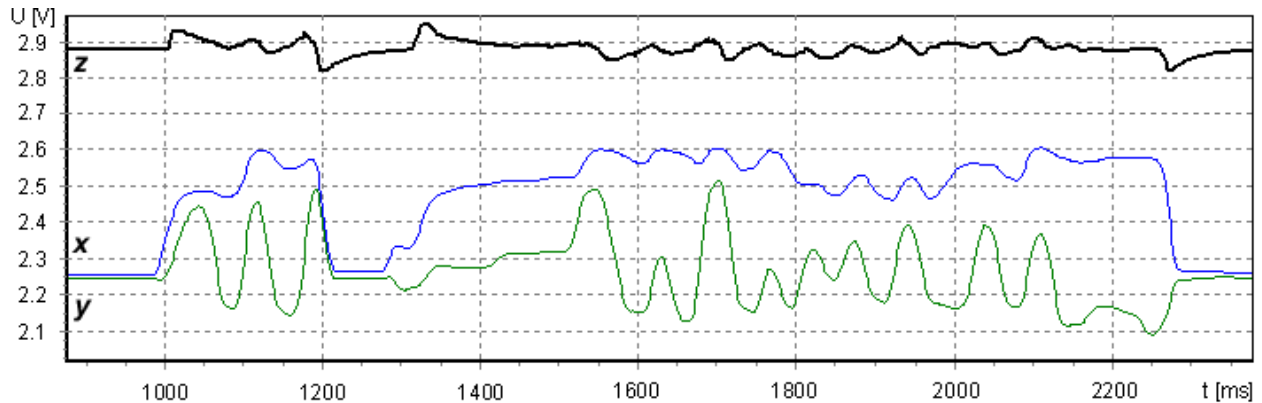
Figure 2: Signals produced by the pen

Both the pen and the signals it produces are described in Section 2 of this paper. Due to the unique design of the pen the signals we obtain are entirely different from the data acquired using a scanner, tablet or coordinate system. That is why we could not use already published methods and had to propose new methods for the text recognition. These methods are described in detail in Section 3. Results of the word recognition experiments and suggestions for future work are discussed in Section 4 and Section 5, respectively.

2   Pen Description

As mentioned above, data acquisition is performed by a special electronic pen which was built at the University of Applied Sciences in Regensburg during the spring of 2002 (Fig. 1). The pen consists of two pairs of mechanical sensors that measure the horizontal and vertical movements of the ballpoint nib and a pressure sensor that is placed in the top part of the pen. The pen produces a total of three signals (Fig. 2). The upper signal corresponds to the pressure sensor and the other two correspond to the horizontal and vertical movements of the pen. The data (Fig. 2) were acquired while writing the word "Dobrou" (Fig. 3), which means "good".

Figure 3: Word "Dobrou"

Four strain gauge sensors that measure the horizontal and vertical movements of the pen are located near the pen nib and are placed orthogonal to each other. The signal produced by the horizontal pair of sensors is called x and the one produced by the vertical sensors y. Each pair of sensors is connected to a Wheatstone bridge. Therefore there is only one output signal corresponding to the horizontal movement of the pen (x) and one corresponding to the vertical movement (y). We cannot provide more detailed description of the pen because of the patent pending status.

We carried out a number of experiments trying to reconstruct the trajectory of the pen by computing the velocity and position of the pen nib but the results were

not good (this only works if the text or drawing is very large). This is why we could not use existing text recognition algorithms and had to develop our own.

It may seem that there is no point in using a pen that produces signals that can hardly be used for pattern recognition. Our pen, however, has two significant advantages. It is possible to write with this pen even without having access to a computer (e.g. on a train or during a lecture) and without installing any other hardware (tablet, IR receivers, etc.). In addition, while writing with the pen both the electronic and paper versions of the document are available at once.

Let us look at the signals (Fig. 2) in greater detail. Note the behavior of the pressure signal—the test person wrote the word (Fig. 3) in two parts—first the letter "D" and then the rest of the word "obrou". The beginning of the writing as well as the end can easily be identified using the first-order difference of the pressure signal.

Note also that the x signal (corresponding to the horizontal movement) does not fall below the quiescent value. There are two reasons for this. Firstly, the test person held the pen slightly slanted to the right, therefore one of the sensors measuring the nib movement was still under certain pressure. Secondly, the word was written from left to right, hence the left sensor was stimulated much more than the right one.

3   Method

We started our work trying to use the dynamic time warping techniques (DTW) for word recognition, but the results were not good enough. Moreover, DTW can hardly be used for a large vocabulary, which makes this method unsuitable for possible commercial utilization. That was the reason why we decided to use the hidden Markov models (HMM), which yield good results in speech recognition [7].

An essential part of our work was the preprocessing and choice of primitives (called observations in HMM terminology) necessary for the description of the signal. Experiments have shown that the best results can be obtained by concentrating on the x and y signals only. We defined four types of primitives (Tab. 1). Our final choice may seem to be too simple, but surprisingly many experiments proved it to be the best choice for our purpose.

Signal processing starts with filtration, continues with the detection of the beginning and the end of the

Primitive (Observation)	Signal Trend	
	x	y
0	↗	↗
1	↗	↘
2	↘	↗
3	↘	↘

Table 1: Primitives

word (using the pressure signal); finally the pair of x and y signals is transformed into a sequence of primitives (observations). Depending on the number of changes in the primitive sequence the proper size of HMM is set (size of arrays A, B and π). We use left-to-right model as in speech recognition (Fig. 4). For HMM training we use the Baum-Welch algorithm with modification according to [2]. The Backward algorithm is used for recognition.

After the training phase we obtain the optimal HMM coefficients for each word (λ). The next step is the decomposition of word model λ (arrays A, B and π respectively) into several models of letters. This process is semiautomatic—an expert has to locate the boundaries between the letters in the signal of each training word. Arrays A, B and π are decomposed as shown in Fig. 5. Locations of the boundaries between the letter in the arrays A, B and π are based on the number of changes in the sequence of primitives (observations). Using this technique over the whole training set we obtain several models for each letter. There are three types of letter models—letter at the beginning of the word, letter in the middle and letter at the end of the word. If there are two models of one letter that has the same size of arrays A, B and π, then it is useful to combine (average) them. The fewer letter models the faster the recognition will be.

The recognition process presumes that models for all the words in the vocabulary will be composed using the
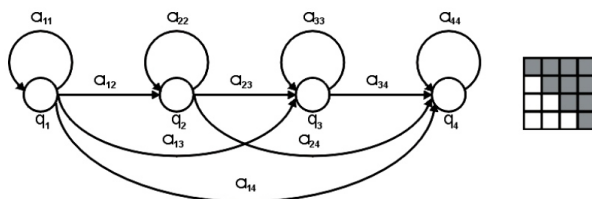

Figure 4: Left-to-right HMM

letter models as basic structural elements. The letters are composed in a way opposite to the way in which they were decomposed during the training phase. Then the most suitable word from the vocabulary is chosen using the Backward algorithm.

4 Experimental Results

We tested the method described in Section 3 on data sets of various sizes (Tab. 2). The recognition rate was between 82 and 90 percent depending on the number of different words used for training, number of training phases and size of the vocabulary. In order to achieve experimental results quickly we decided to use words consisting of 15 different letters only (a, d, e, f, h, i, m
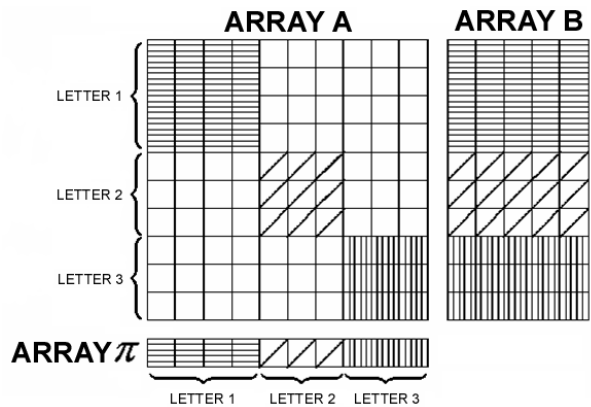

Figure 5: Decomposition of word HMM arrays into letter HMM arrays

n, o, r, s, t, u, v, y). This limitation helped us to reduce the complexity of tagging the training set of words.

Vocabulary size	1649	2198	5129
Recognition rate (%)	88	90	82
Recognition time (min)	17–26	27–49	360

Table 2: Recognition results

5 Conclusion and Future Work

Our paper summarizes our experience with the new pen developed by our team. The original intention of our pen has been the signature verification, where we yield 1.3% equal error rate (EER) on skilled forgeries [6]. We tried to find out whether the prototype of our electronic pen is also suitable for handwritten text recognition.

Results achieved so far show that HMM based algorithms can be used for text recognition but they are too time-consuming and recognition rate is about 5–10 percent below the accuracy of state-of-the-art handwritten text recognition systems. Therefore, it is necessary to speed up the algorithm to achieve on-line recognition. Unfortunately, techniques commonly used in speech recognition cannot be used to speed up the handwritten text recognition. For this purpose we will have to find our own algorithms.

We will also improve the recognition accuracy. In the new version of the pen (just being built) we use different sensors for x and y signals acquisition. The first experiments show that x and y signals are much better (than used to be), therefore we expect higher recognition accuracy of the new system. For further improvement of the system we plan to incorporate language models into the classifier.

To prove that the method is suitable for commercial applications it is necessary to build a very large vocabulary (more than 10,000 words) i.e. get more training data (27^3 triplets of letters) and tag them.

It is assumed that the pen will also find other applications in e.g. tele-robotics and robot sensing (haptics).

References

[1] S. Jaeger. *Recovering Dynamic Information from Static, Handwritten Word Images*. Ph.D. thesis, Ulm, 1998.

[2] T. Kanungo. Hidden Markov Models *available at http://www.cfar.umd.edu/∼kanungo*, 1998.

[3] R. Plamondon, S. N. Srihari. On-Line and Off-Line Handwriting Recognition: A Comprehensive Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(1), 2000.

[4] J. Preece. *Human-Computer Interaction*. Addison-Wesley, Essex, 1994.

[5] S. Rieck and M. Schüßler. Recognizing Handwritten Address Words with Hidden Markov Models. *Knstliche Intelligenz*, Volume 2, 1999.

[6] O. Rohlik et al. A new approach to signature verification: Digital data acquisition pen. *Neural Network World*, 5(2), 2001.

[7] L. R. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition*. PTR Prentice-Hall, New Jersey, 1993.