The New Text and Graphical Input Device: Compact Biometrical Data Acquisition Pen

Ondrej Rohlik, Pavel Mautner, Vaclav Matousek[*] & Juergen Kempf[†]

[*]University of West Bohemia, Faculty of Applied Sciences, Univerzitni 22, 30614 Pilsen, Czech Republic

rohlik|mautner|matousek@kiv.zcu.cz

[†]University of Applied Sciences Regensburg, Prüfeninger Strarüße 58, 93049 Regensburg, Germany

juergen.kempf@micro.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. Our paper presents our experience with the text recognition methods that we have developed for a new designed electronic pen that produces three 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 and other application areas of the pen are proposed.

Keywords: HCI, handwritten text recognition, OCR, HMM, biometrics

1 Introduction

There are many commercial systems designed for text recognition worldwide. Most of them are based on optical character recognition (OCR) methods. These systems use as an input device a scanner or a pen with a tablet or a pen with an infrared transmitter and several receivers. The obvious disadvantage of these devices is the limited mobility of a system composed of two or more parts (Preece, 1994). Our approach is brand-new. We have constructed a special pen that integrates all the electronic devices needed for data acquisition inside. 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 methods known from the OCR systems and had to propose new methods for 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.

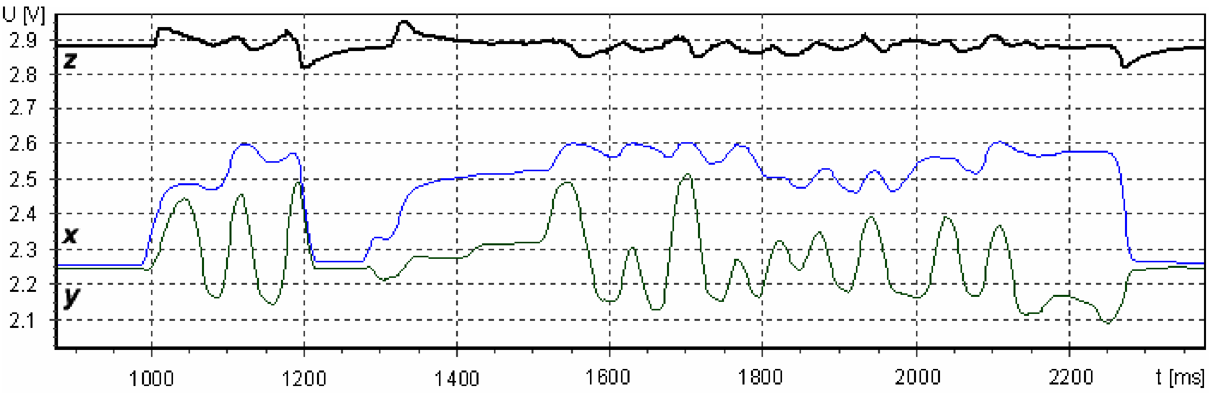

Figure 2: Signals produced by the pen.

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”.

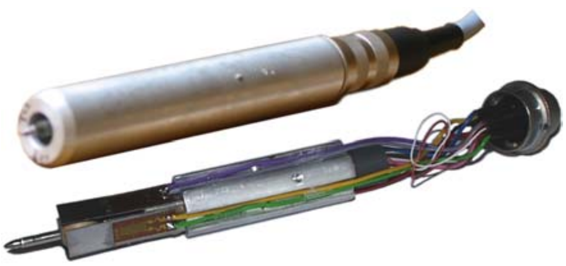

Figure 1: Biometrical pen – with and without cover.

Four 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 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 poor. This only works if the text or drawing is very large. Nevertheless, the results are not very good. That is why we could not use existing text recognition algorithms (used in OCR systems) 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.

Figure 3: Word “Dobrou”.

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.

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.

Primitive (Observation)	Signal Trend	
	x	y
0	↑	↑
1	↑	↓
2	↓	↑
3	↓	↓

Table 1: Primitives.

Signal processing starts with filtration, continues with the detection of the beginning and the end of the 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 (Rabiner & Juang, 1993). 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.

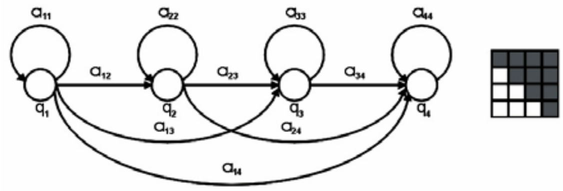

Figure 4: Left-to-right HMM.

The recognition process presumes that models for all the words in the vocabulary will be composed using the 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 (Fig. 6). Then the most suitable word from the vocabulary is chosen using the Backward algorithm.

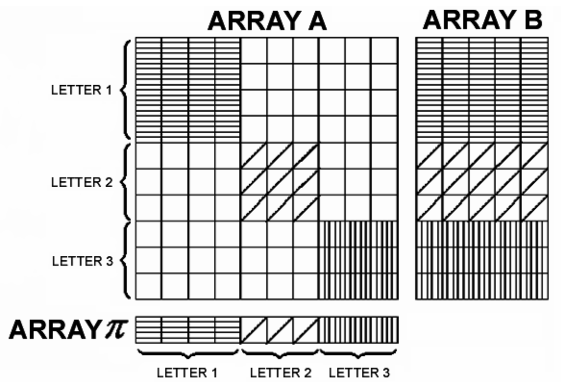

Figure 5: Decomposition of word HMM arrays into letter HMM arrays.

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 n, o, r, s, t, u, v, y). This

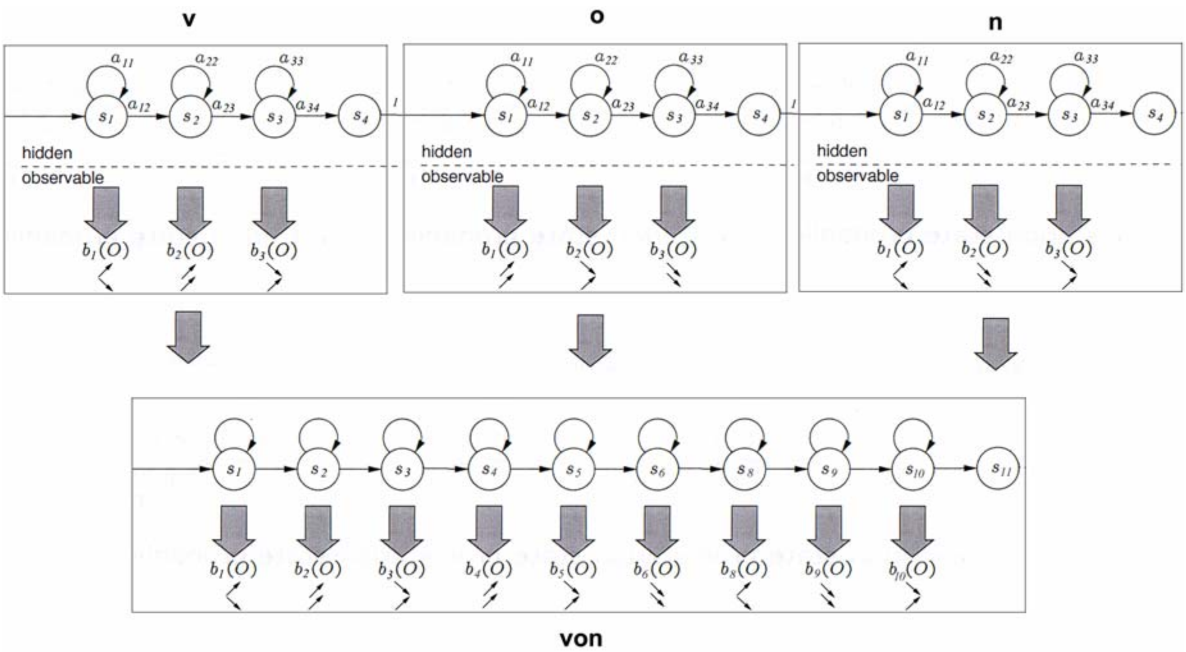

Figure 6: Composition of word HMM from letter HMMs

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: Experimental results.

5 Conclusion and Future Work

Our paper summarizes our experience with the new pen developed by our team. We proved that the prototype of our electronic pen is 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. 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. To prove that the method is suitable for commercial applications it is necessary to build a very large vocabulary (more than 10,000 words), get more training data (27^3 triplets of letters) and tag them. In order to improve the recognition rate we plan to incorporate language models into the classifier. Special attention will be paid to signature analysis and signature verification, too.

References

Preece, J. (1994), *Human-Computer Interaction,* Addison-Wesley, Essex.

Rabiner L. R. & Juang B. H. (1993), *Fundamentals of Speech Recognition,* PTR Prentice-Hall, New Jersey, 1993.

Rohlik, O., Mautner, P., Matousek, V., Kempf, J., & Weinzierl, K. (2001), A New Approach to Signature Verification: Digital Data Acquisition Pen, *Neural Network World* 5, 493-501.