MVA '96 IAPR Workshop on Machine Vision Applications, November. 12-14, 1996, Tokyo, Japan

Perturbation Models: A New Approach to Improving Handwriting Recognition

Thien M. Ha, Horst Bunke, and Guido Kaufmann
University of Berne
Institut für Informatik und Angewandte Mathematik
Neubrückstr. 10, CH-3012 Bern, Switzerland
Email: {haminh,bunke,kaufmann}@iam.unibe.ch

Abstract

This paper discusses the application of the perturbation method to handwriting recognition. First the standard pattern recognition operations of preprocessing, feature extraction, and classification are reviewed. Then we introduce the perturbation method as a new approach to overcoming the problems that result from the sequential architecture traditionally found in pattern recognition systems. Two case studies in handwriting recognition, namely, isolated numeral recognition and cursive handwriting recognition, are presented. Experimental results show that the perturbation method significantly improves the recognition rates of state-of-the-art systems.
Key Words: Handwriting recognition, isolated numeral recognition, cursive handwriting recognition, decision fusion, perturbation method, nearest neighbour rules, neural networks, hidden Markov models.

1 INTRODUCTION

Handwriting recognition is an important and necessary step in many document processing applications. For instance, we can think of the reading/processing of checks, mail addresses, tax forms, and census forms. Despite the fact that handwriting recognition is a subfield of pattern recognition and thus inherits its well-developed techniques, the problem remains extremely difficult due its variety in shape. There are many factors that contribute to this variety, the first of which is the writing style. Figure 1 shows a few examples of handwritten words and numbers in two styles, namely, printed and cursive. Printed style does not ensure that symbol patterns are disconnected; conversely, cursive style writing of a word may yield disconnected patterns [23]. Writing in cursive style creates patterns not included in the symbol set. Moreover, each writer has her own style, thus her own additional patterns, which may furthermore change with time, mood, stress, etc. Apart from style, [24] pointed out three main groups of factors that can account for the variety of handwriting, namely, the writer's personality, the circumstances at the writing time, and various technical aspects, such as paper, ink color, and writing instrument.

In this paper we discuss one method – called *perturbation method* – that has been experimentally shown to have the capability to cope well with some of the above problems caused by handwriting variability.The method has been applied to two tasks, namely, the recognition of isolated numerals and that of cursively handwritten words drawn from a small lexicon. In both cases, the perturbation method improves the recognition rates of state-of-the-art systems.

Section 2 reviews the standard pattern recognition paradigm for handwriting recognition. The perturbation method is then presented in Section 3. The next two sections are devoted to the applications of the perturbation method to the recognition of isolated numerals and cursive handwritten words, respectively. Section 6 concludes the paper.

2 PATTERN RECOGNITION PARADIGM

Handwriting recognition is a subfield of pattern recognition and thus inherits its techniques. A typical pattern recognition system operates in two phases, namely, training (learning) and recognition. In the training phase, the system learns from a large number of patterns for which the classes are known; in the recognition phase, the system is required to classify patterns for which the classes are unknown. The training typically consists of image preprocessing, feature extraction and feature storage. In the recognition phase, an unknown image is preprocessed, its features are extracted and compared to those learned in the training phase; see Fig. 2. The class that has the closest features will be

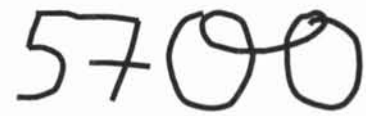
Style / Symbol Set	Printed	Cursive
Latin Alphabet		
Arabic Numerals		

Figure 1: Variety of handwriting.

selected as the recognition result. A very large number of methods exists for each of these operations. The choice of one method over another is eventually application-dependent. In the following we provide a brief overview of these operations in the context of handwriting recognition.

Handwriting recognition has traditionally been divided into two approaches, namely, statistical and structural. In the statistical approach, the pattern (character or word) is characterised by an ordered set of numerical values, whereas in the structural approach, the pattern is converted into a symbolic representation, such as a string, tree or graph. It is clear that the classifier and the type of features must be compatible, i.e., statistical, respectively structural, features require a statistical, respectively structural, classifier and vice versa.

The main goal of preprocessing is to eliminate undesirable effects. For instance, patterns in a real environment are usually distorted by various kinds of noise that should be filtered out. In character recognition, the actual size of the pattern is in many situations not relevant for the purpose of classification and therefore size normalisation is sometimes useful. Preprocessing can also be used to ease subsequent operations, such as image smoothing can contribute to the robustness of some feature extraction methods. The most common preprocessing methods are: various filters (smoothing, noise elimination) [15], binarisation [17], size normalisation [26], slant correction [21] and thinning [25].

Feature extraction provides a compact yet informative representation of a pattern. The success of a pattern recognition system is by large determined by the feature extraction [22]. Although there exist general statistical feature extraction methods, such as principal component analysis and discriminant analysis, experiments have shown that they are usually outperformed by extraction methods that take into account particular characteristics of the problem at hand. In the case of handwriting recognition, patterns are images the basic distinctive features of which are edges. Therefore, it is not surprising that feature extraction methods based on edges and contours prove powerful [7]. Moreover, structural features, such as arcs and holes, are high level representation of edges and contours. They are sometimes called topological since they are invariant with respect to topological transformations. Other feature extraction methods, such as projection, moments and Fourier descriptors, have also been used.

Classification consists in comparing the features provided by the feature extraction with those stored in the training phase. Statistical features can be compared by using various classification methods, such as, polynomial classifier, nearest neighbour, neural networks and hidden Markov models [26, 22, 8, 16, 20]. For structural features, a symbolic matching procedure (either exact or inexact) is needed [12, 2]. String, tree or graph matching is required as symbolic matching, depending on the representation of features.

Recently, a new paradigm appeared and intended to exploit the mutual advantages and drawbacks of several techniques to yield a better system [22]. It consists in combining the results from several independent systems each of which uses a different technique. Many different schemes to combine individual systems exist, but it seems that even the most primitive of them (using a voting scheme) can already give a better result than each of the individual techniques [18].

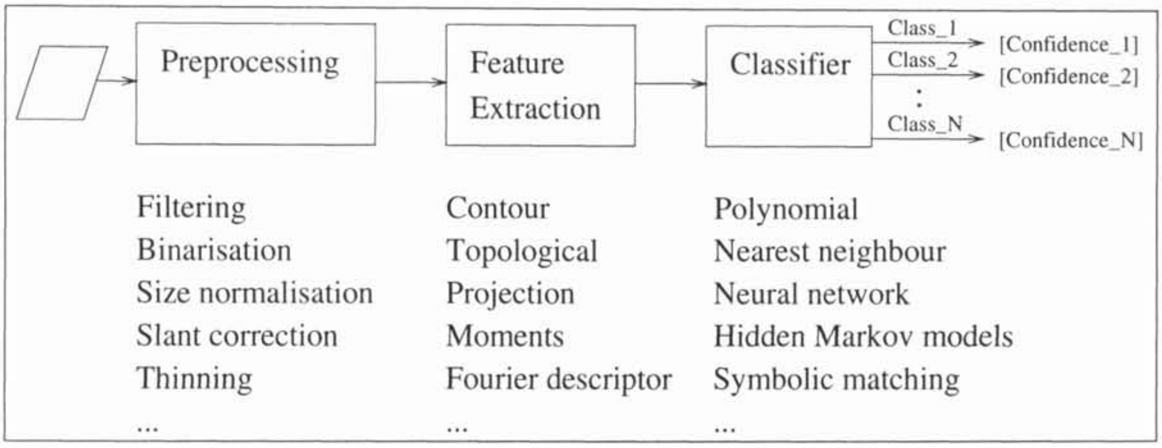


Figure 2: The standard pattern recognition paradigm.

3 PERTURBATION METHOD

The perturbation method results from the critical observation of the general pattern recognition paradigm shown in Fig. 2. The main flaw lies in its *serial* structure, i.e., an error in the preprocessing and/or feature extraction stage very likely leads to a wrong classification.

The perturbation method tackles the first step of the processing chain, namely, preprocessing whose main goal is to eliminate undesirable effects. Unfortunately, undesirable effects cannot always be defined in a clear and objective way. For instance, image filtering can be designed to fill-in gaps for broken characters but occasionally eliminates the hole of a loop thus destroying essential structural information. More generally, all transformations altering the standard form of an image are called perturbation models.

The perturbation method consists in applying a set of predefined inverse perturbation models to the input image (see Fig. 3). These inverse perturbations are independent of the input image and are expected to include the true perturbation that actually made the input image different from its standard pattern. We know that if an inverse perturbation actually corresponds to the true perturbation, the corresponding inversed image will be very close to the original standard pattern and could be easily recognised by some known method. Therefore, each inversed image is submitted separately to a conventional recognition system, the output score of which is then compared to the others. It is clear that among the scores, the one corresponding to the true perturbation can be expected best. Since each score is attached to a class, the recognition scheme is in fact a by-product of the reversing process [9].

4 ISOLATED NUMERAL RECOGNITION

In this section we describe the application of the previously mentioned methodologies to the problem of isolated handwritten numeral recognition. First, we consider the standard pattern recognition paradigm with its two approaches: structural and statistical. Experiments using the combination-based approach are also included. Then, our novel perturbation method is shown to have the capacity to improve the recognition rates of various implemented systems.

To compare statistical and structural approaches, we implemented four systems, the first two of which are structural based whereas the last two are statistical based, and tested them on the same database [10]. Figure 4 illustrates the four feature types, namely, quasi-topological, topological, projection-based, and contour-based. For each of the two structural features, an appropriate inexact matching algorithm is used as classifier. Both statistical systems make use of the distance-weighted k-nearest neighbor rule as classifier [6]. These four systems were compared by using 18468 samples (*br* directories) for training and 2213 samples (*goodbs* directories) for testing, both from the CEDAR database [13]. These data were collected from live mail in the U.S. and were thus totally unconstrained. The results are shown in Table 1 where the correct recognition rates are obtained at zero-rejection level (forced choice option). In general, it can be observed that statistical methods give much higher recognition rates than structural methods. This has also been observed by various other authors [24, 18]. Structural methods are appealing because they seem to match the way human beings read characters and perform quite well when the input data are of good quality, but usually fail in dealing with poor quality data (e.g., broken strokes, noisy data).

We also tested the combination-based approach and found out that the combination of the projection and the contour method using a weighted voting

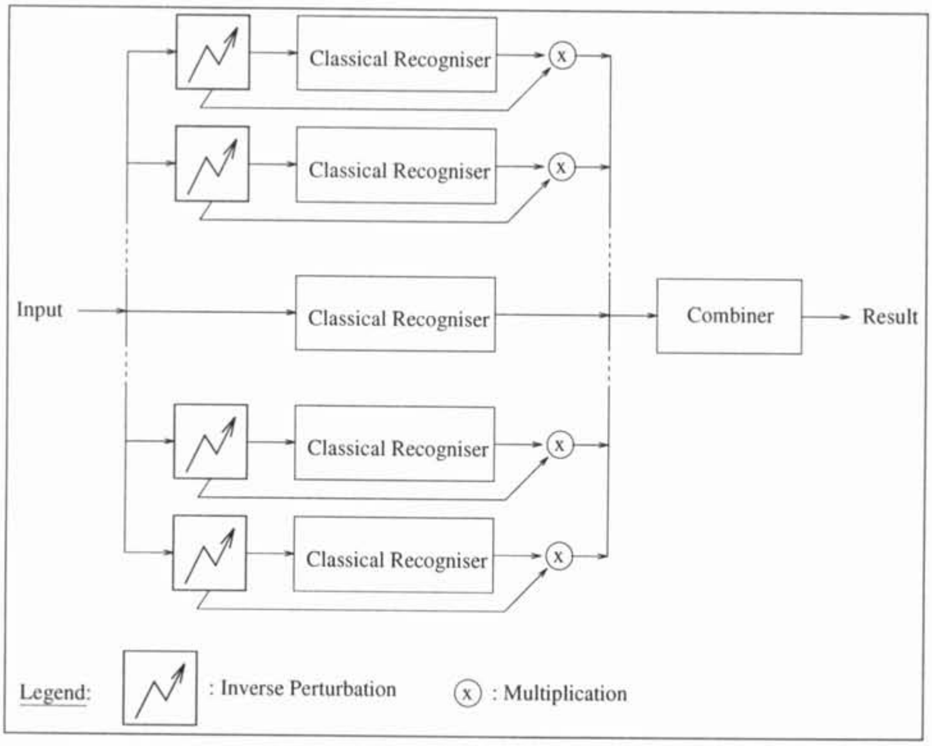


Figure 3: Perturbation-based recognition system.

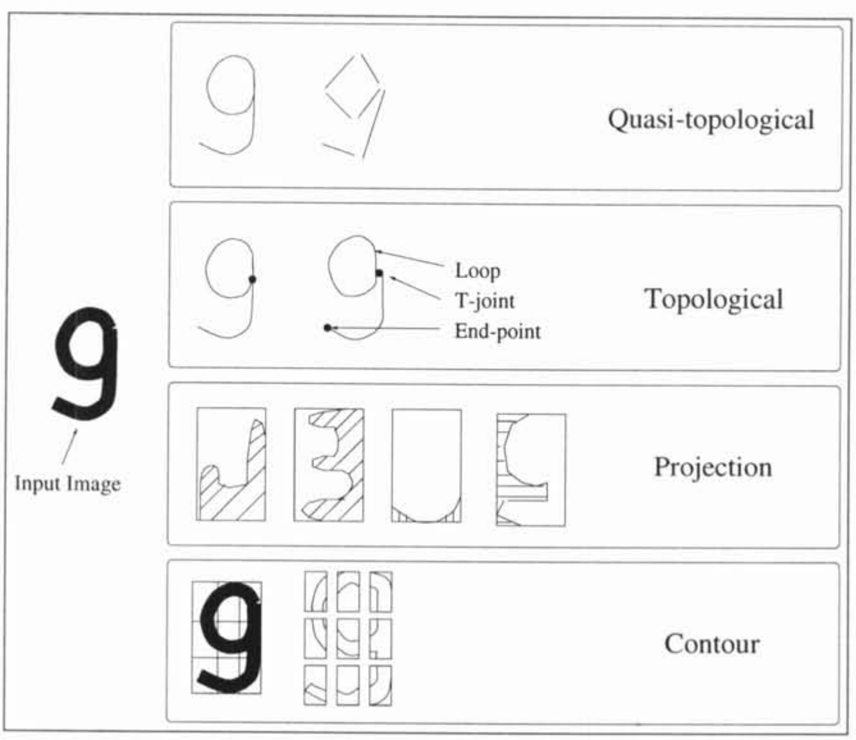


Figure 4: Feature Types.

Structural-based Systems	
Quasi-topological	91
Topological	95
Statistical-based Systems	
Projection	97.69
Contour	98.19
Combination-based System	
Projection & Contour	98.51

Table 1: Recognition Results.

scheme improved the recognition rate from 97.69% and 98.19% to 98.51% (see Table 1) [9, 10].

As mentioned in the previous section, there exist many statistical classifiers, such as polynomial and neural networks, that can be used for the same statistical features. Our experiments with neural networks show that their accuracy is comparable to nearest neighbour classifiers. The main differences lie in the training and recognition times. Neural networks are slower in training but faster in recognition than nearest neighbour classifiers. The figures would be different with optimised nearest neighbour classifiers, e.g., using editing and fast search [3]. These techniques would speed up the recognition time but would also increase the training time because they preprocess the training data. All in all, these techniques would make nearest neighbour classifiers similar to neural networks.

Applying the perturbation method requires the determination of a set of perturbation models. For isolated handwritten numerals, we have identified four geometric transformations, namely, rotation, slant, perspective view and shrink, as well as a stroke width transformation. Moreover, slant is decomposed into horizontal and vertical directions, whereas perspective view and shrink are each decomposed into horizontal, vertical, 1^{st} diagonal and 2^{nd} diagonal directions. The stroke width transformation is modelled by two morphological operators, namely, dilation and erosion. These result in a total of $T = 12$ perturbation models (Fig. 5).

The combiner in a perturbation-based recognition system (see Fig. 3) can take on different forms. Basically, it is a classifier that accepts as inputs the outputs of "classical" recognisers. We have experienced with various combination structures, such as weighted voting [9] and arithmetic averaging [11]. For other combination structures, see [22].

Finally, each perturbation model (e.g. rotation) can be applied with different parameter values (different rotation angles) each of which corresponds to one channel of Fig. 3. Usually, larger values are more heavily weighted down to penalise audacious hypotheses.

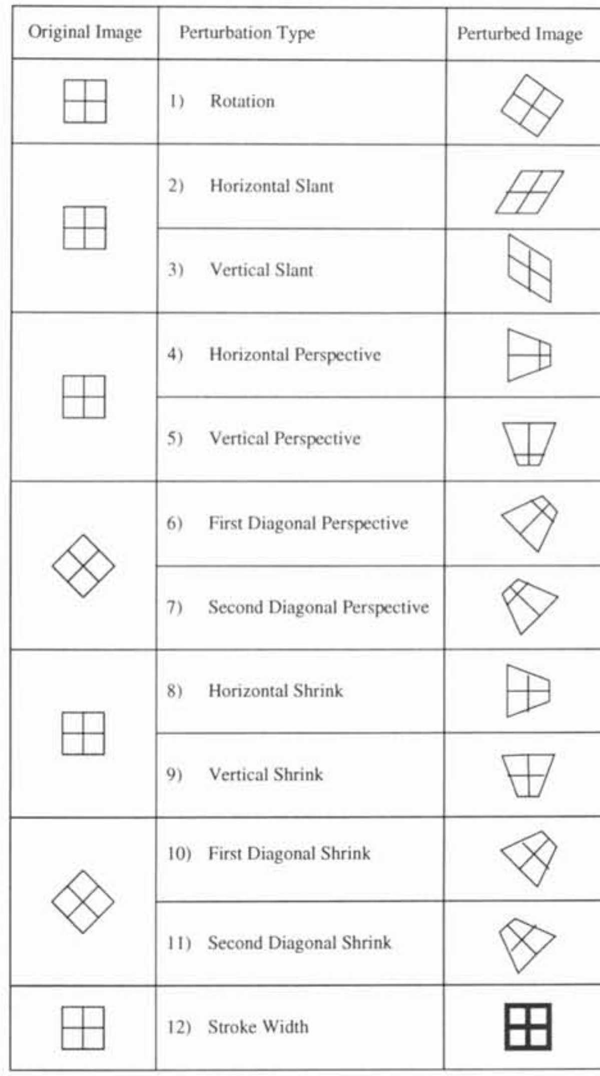


Figure 5: Perturbation models for isolated handwritten numerals.

We tested the perturbation method on two worldwide standard databases, namely, CEDAR and NIST [13, 27]. CEDAR was used in a pilot study whereas NIST, which is more than ten times larger, was used for large scale experiments. On CEDAR, the perturbation method boosted the recognition rate of the combination-based system from 98.51% to 99.10%, which is the highest rate ever published on this database [9]. On NIST-SD3, the perturbation method improved the recognition rate of the combination-based system from 99.45% to 99.54% (tested on more than 170000 numerals) [11]. Moreover, when the experimental conditions were set to those of the NIST conference (see Appendix A), i.e., training on NIST-SD3 and testing on NIST-SD7, the perturbation method improved the recognition rate of the combination-based system from 96.8% to

Rank	1	2	3	4	5	6	7	8	9	10
Recognition Rate	96.84%	96.65%	96.62%	96.57%	96.51%	96.33%	96.15%	96.12%	96.08%	95.90%

Table 2: Top-ten systems trained on NIST–SD3 and tested on NIST–SD7 at the NIST conference.

97.1%, outperforming all other systems; see Table 2. The performance on NIST–SD7 is lower than that on NIST–SD3 because these two databases were collected from two different populations of writers and therefore have very different statistical distributions [27].

5 CURSIVE HANDWRITING RECOGNITION

General cursive handwriting recognition is an extremely difficult problem even for human beings. It needs not only the ability to recognise characters and words but also knowledge about the syntax and even the semantics of the text. Modelling all these aspects is currently well beyond the state-of-the-art technology. Therefore we limit our discussions to the problem of recognising cursively handwritten words drawn from a small lexicon. More specifically, we consider the problem of classifying a word into one of the 26 German words that constitute the basic vocabulary, which allows – by concatenation – the construction of all German numeral amounts lower than one million. Although limited, the problem has interesting applications in the field of automatic bank check reading. (Notice that the problem is similar for other languages.) Since the problem fits in the standard pattern recognition framework, we first apply the standard method, i.e., preprocessing, feature extraction and classification, and then investigate the use of the perturbation method.

Preprocessing plays a much more important role for cursive handwriting than for isolated numerals because of a much greater variability. This is due to many reasons. Words are composed from an alphabet of 26 letters instead of the ten numerals. A letter may or may not have a descender/ascender part and can be written in lower- or upper-case. Two consecutive letters are not always connected in the same manner, depending on whether they are written in lower- or upper-case. The total number of classes of the lexicon (26) is also larger than that of numerals, thus increasing the risk of confusion. Words have different length, depending on the number of constituting letters. To reduce as much as possible these effects, our preprocessing comprised a series of normalising operations, namely, skew correction, slant correction, baseline detection, and size normalisation in the $x-$ and $y-$directions (Fig. 6).

In our system, features are represented by a dynamic sequence of vectors each of which contains the pixel values of the normalised image within a thin vertical strip. With the width of the strip being fixed, the sequence length (number of vectors) depends on the width of the normalised image.

Due to the dynamic nature of the above defined features, classification is performed via a dynamic comparison algorithm. In our work, each letter of the alphabet is represented by one hidden Markov model (HMM) and the HMM of a word is constructed by concatenation of individual letter HMMs [20]. The parameters of the letter HMMs are obtained in the training phase via the standard Baum-Welch algorithm. The classification of a word consists in comparing its feature sequence with all word HMMs of the lexicon, and choosing the closest one. The comparison is efficiently implemented using the Viterbi algorithm.

For our work we collected 13000 words from 500 different writers. Each writer had to write 26 words of our vocabulary once on a form. The 500 forms were divided into 5 sets with 100 forms each. For a given test set, the system was trained with the remaining 4 sets.

In principle, the perturbation method can be applied to any of the normalising operations. However, we have performed a series of experiments and it turned out that the size normalisation in the $y-$direction was the most important one. In the following, we provide some more details about the implementation of the perturbation method adapted to cursive handwriting.

The y-size of the word is determined by the middle area, that is bounded by the *lower* and the *upper baseline*. The lower baseline can usually be determined quite exactly because most writers try to follow a virtual baseline. The detection of the upper baseline, however, is much more difficult. To reduce this difficulty we used different approaches in the training and the recognition step. For the size estimation in the training step we took into account that our database consisted of 26 words of each writer. We assumed that all words of a writer had approximately the same height because they all had been written on the same form into given boxes. Instead of trying to determine the upper baseline for each word we estimated the middle area size by means of all instances of one writer. For this purpose we projected them horizontally and summed up the 26 histograms with the position of the lower baseline as a reference point. In the cumulative histogram we had the information of 144 characters with 24 descenders and 29 ascenders. The size of the middle area and the scale factor for the words of this writer could then be calculated easily.

For the recognition of a word from the test set of the database we were not allowed to use the information about other words from the same writer. So we needed a more sophisticated method for the

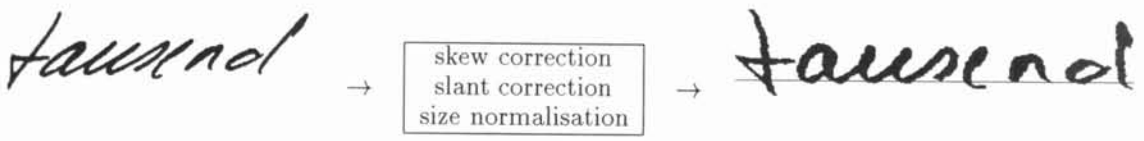


Figure 6: The applied preprocessing steps.

y-size estimation in this step. In contrast to other works we did not use a method based on a topological analysis of the contour line. Instead, we adapted at this point of our work the *perturbation approach*. For our system this meant that we had to find a y-scale factor range that should contain the real y-scale factor to normalise the word in y-direction. We introduced two factors $y_{minscale}$ and $y_{maxscale}$ that bound the perturbation range. For that purpose the method for the upper baseline estimation was expanded to get a *minimal* and a *maximal upper baseline*. The maximal upper baseline was positioned so high that the real upper baseline was certainly below it. Analogously, the minimal upper baseline was tried to placed below the exact position of the upper baseline (Fig. 7). With these two lines the new scale factors $y_{minscale}$ and $y_{maxscale}$ could be computed. We scaled our input image with different factors of this calculated scale range.

Following this approach we generated only a few inverse transformations when the writer was cooperative and the two upper baselines were close to each other. In more difficult cases we did a more conservative estimation by enlarging the range. The minimal and maximal upper baselines resulted from two combined methods (Fig. 8). The first one was based on a smearing approach. The second method looked for all local maxima in a word.

For the experiments the 500 forms were divided into 5 sets with 100 forms each. For a given test set, the system was trained with the remaining 4 sets. In all experiments we used the same trained system, whose words were normalised using all 26 instances of a writer for the size estimation. The differences of the tests only lay in the y-size estimations of the test words. In the first test a fixed normalisation was applied. Based on the two techniques described we computed one fixed estimation for the y-scale factor. The results are shown in Table 3. Using the proposed perturbation approach, the recognition rate was improved by nearly 6% in experiment 2 (see Table 4). In Table 5 finally the average number of perturbed patterns that were created in experiment 2 is presented. One can notice that in noisier data, where the recognition rate is below the average, more patterns have to be created. A more detailed description of this approach can be found in [14].

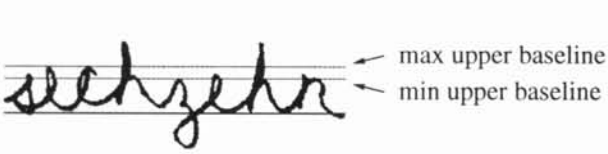


Figure 7: The estimated minimal and maximal upper baselines.

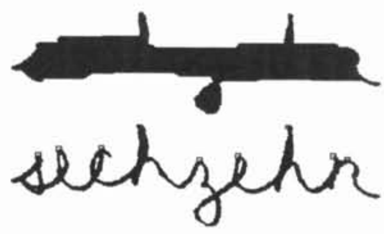


Figure 8: Smearing and local maxima extraction are the two basic techniques for the estimation of the upper baselines.

rank	set 0	set 1	set 2	set 3	set 4	avg.
1	78.8	83.7	86.0	83.1	85.3	**83.4**
2	88.1	90.5	92.6	90.9	92.5	90.9
3	90.7	93.4	94.2	93.7	94.4	93.3

Table 3: Recognition rate without perturbation approach.

rank	set 0	set 1	set 2	set 3	set 4	avg.
1	86.2	88.4	91.4	89.2	91.3	**89.3**
2	92.9	93.2	96.3	95.1	96.0	94.7
3	95.0	95.7	97.4	96.5	97.3	96.4

Table 4: Recognition rate with perturbation approach.

set 0	set 1	set 2	set 3	set 4	avg.
3.22	2.36	1.94	2.02	1.97	**2.30**

Table 5: Average number of test instances created by the perturbation module.

6 CONCLUSION

In this paper we have discussed the application of the perturbation method to handwriting recognition. First the standard pattern recognition operations of preprocessing, feature extraction, and classification were reviewed. Then we presented the perturbation method as a tool to tackle the flaw of the serial structure of the standard architecture. Two case studies in handwriting recognition, namely, isolated numeral recognition and cursive handwriting recognition, were then consecutively treated. Experimental studies showed that the perturbation method significantly improved the recognition rates of state-of-the-art systems. We strongly believe that the perturbation method is applicable to many other applications in pattern recognition as well.

Acknowledgments: This work is partly supported by the Union Bank of Switzerland Information Technology Laboratory (UBILAB) and the Swiss National Science Foundation.

A NIST Databases

Two databases, namely, SD3 and SD7, were provided by the American National Institute of Standards and Technology (NIST) in 1992 as parts of a conference to assess the state-of-the-art in isolated handwritten character recognition [27]. Twenty-nine groups from Europe and North America participated to compare the performance of their OCR systems. In total, 47 systems, both commercial and research, were presented. The databases contain isolated numerals (digits) as well as upper- and lower-case letters. Most systems used SD3 for training and SD7 for testing at the conference, although a few of them were trained using proprietary databases. The final report, including recognition results of all tested systems as well as their characteristics, is publicly accessible [27].

References

[1] H.S. Baird, "Document Image Defect Models," in H.S. Baird, H. Bunke, K. Yamamoto (Eds.): *Structured Document Image Analysis,* Springer Verlag, 1992, pp. 546-556.

[2] H. Bunke, "Syntactic and Structural Pattern Recognition," in *Handbook of Pattern Recognition and Computer Vision,* World Scientific, pp. 163-209, 1993.

[3] B.V. Dasarathy (Ed.), *Nearest Neighbor Pattern Classification Techniques,* IEEE Computer Society Press, 1991.

[4] B.V. Dasarathy (Ed.), *Decision Fusion,* IEEE Computer Society Press, 1994.

[5] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis,* John Wiley & Sons, 1973.

[6] S.A. Dudani, "The Distance-Weighted k-Nearest-Neighbor Rule," *IEEE Trans. on Systems, Man, and Cybernetics,* Vol. 6, No. 4, pp. 325-327, April 1976.

[7] J.T. Favata, G. Srikantan, and S.N. Srihari, "Handprinted Character/Digit Recognition Using a Multiple Feature/Resolution Philosophy," *Proc. of the Fourth Int. Workshop on Frontiers of Handwriting Recognition,* Taipei, Taiwan, Dec. 7-9, 1994, pp. 57-66.

[8] K. Fukunaga, *Introduction to Statistical Pattern Recognition,* 2^{nd} edition, Academic Press, 1990.

[9] Thien M. Ha and H. Bunke, "Handwritten Numeral Recognition by Perturbation Method", *Proc. of the Fourth Int. Workshop on Frontiers of Handwriting Recognition,* Taipei, Taiwan, Dec. 7-9, 1994, pp. 97-106.

[10] Thien M. Ha, D. Niggeler, H. Bunke, and J. Clarinval, "Giro Form Reading Machine," *Optical Engineering,* Vol. 34, No. 8, 1995, pp. 2277-2288.

[11] Thien M. Ha and H. Bunke, "Off-Line Handwritten Numeral Recognition by Perturbation Method". Submitted.

[12] J.S. Huang and K. Chuang, "Heuristic Approach to Handwritten Numeral Recognition," *Pattern Recognition,* Vol. 19, No. 1, pp. 15-19, 1986.

[13] J.J. Hull, "A Database for Handwritten Text Recognition Research," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 16, No. 5, pp. 550-554, May 1994.

[14] G. Kaufmann, H. Bunke, and Thien M. Ha, "Recognition of Cursively Handwritten Words Using a Combined Normalisation/Perturbation Approach," *The 5th International Workshop on Frontiers in Handwriting Recognition,* September 2-5, 1996, University of Essex, England, pp. 17-22.

[15] A.K. Jain, *Fundamentals of Digital Image Processing,* Prentice Hall, 1989.

[16] C. Lau (Ed.), "Neural Networks: Theoretical Foundations and Analysis," IEEE Press, 1992.

[17] Lee S.U. and Chung S.Y., "A Comparative Performance Study of Several Global Thresholding Techniques for Segmentation", Computer Vision, Graphics, and Image Processing 52, pp. 171-190, 1990.

[18] D.S. Lee, S.N. Srihari, "Handprinted Digit Recognition: A Comparison of Algorithms," *Pre-Proceedings of the 3^{rd} Int. Workshop on Frontiers in Handwriting Recognition,* Buffalo, New York, USA, May 25-27, pp. 153-162, 1993.

[19] R. Kasturi and L. O'Gorman (Eds.), *International Journal on Machine Vision and Applications,* Special Issue on Document Image Analysis Techniques, Vol. 5, No. 3, Summer 1992.

[20] L. R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, Vol. 77, No. 2, Feb. 1989, pp. 257-286.

[21] J. Schürmann et al., "Document Analysis – From Pixels to Contents," Proceedings of the IEEE, Vol. 80, No. 7, 1992, pp. 1101-1119.

[22] J. Schürmann, *Pattern Classification: A Unified View of Statistical and Neural Approaches,* John Wiley & Sons, 1996.

[23] J. C. Simon, "Off-line Cursive Word Recognition", Proceedings of the IEEE, Special Issue on Optical Character Recognition, Vol. 80, No. 7, July 1992, pp. 1150-1161.

[24] C.Y. Suen, C. Nadal, R. Legault, T.A. Mai, and L. Lam, "Computer Recognition of Unconstrained Handwritten Numerals," in *Proc. of the IEEE* ***80***, No. 7, 1992, pp. 1162-1180.

[25] C.Y. Suen and P.S. Wang (Eds.), Special Issue on "Thinning Methodologies for Pattern Recognition," International Journal of Pattern Recognition and Artificial Intelligence, Vol. 7, No. 5, Oct. 1993.

[26] J.R. Ullmann, *Pattern Recognition Techniques*, Butterworths, 1973.

[27] R.A. Wilkinson, J. Geist, S. Janet, P.J. Grother, C.J.C. Burges, R. Creecy, B. Hammond, J.J. Hull, N.W. Larsen, T.P. Vogl, and C.L. Wilson, *The First Census Optical Character Recognition Systems Conference*, The U.S. Bureau of Census and the National Institute of Standards and Technology, Technical Report #NISTIR 4912, Gaithersburg, MD, Aug. 1992.