Text Localisation and Handwriting Recognition

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

July 11, 1996

Abstract

This report discusses several general aspects of text localisation and handwriting recognition. In particular, we consider their applications to extraction and recognition of numerals written on Giro forms. For text localisation, we investigate the problem of extracting text printed or written inside boxes on forms. We review a number of representative methods for solving this problem, describe the implementation of one of them, and present some experimental results obtained on real data. For handwriting recognition, we first present the general methodology, including a new approach called perturbation method. Then we explain how the general methodology is applied to three subproblems in handwriting recognition, namely, the recognition of isolated numerals, that of numeral strings, and the recognition of cursively handwritten words drawn from a small lexicon. Experimental results show that our systems are either equivalent to or better than state-of-the-art systems.

CR Categories and Subject Descriptors: I.5.0 [Pattern Recognition]: General; I.5.1 [Pattern Recognition]: Models; I.5.2 [Pattern Recognition]: Design Methodology;
Key Words: Text localisation, handwriting recognition, isolated numeral recognition, numeral string recognition, cursive handwriting recognition, Giro form, decision fusion, perturbation method, nearest neighbour rules, neural networks, hidden Markov models.

Contents

1 INTRODUCTION

Text localisation and handwritten character recognition are two of the most important topics that cover a wide spectrum of problems in document processing. Indeed, the state-of-the-art in the field of document processing, as witnessed by [32, 33, 45, 3, 30, 31], divides most complex tasks into three subtasks, namely, layout analysis, character recognition and contextual processing. Layout analysis consists in localising text or graphics parts of interest on a document. Character recognition reads text parts and provides us with strings of characters. In many applications, text parts consist of names and/or addresses and can thus be verified against a directory of valid names and addresses by a contextual processing step. In other applications, strings of characters obtained from text parts can be parsed and checked to verify whether they comply with the grammar of the language in use.

This report deals with two issues, namely, text localisation and handwriting recognition. It has been recognised that text localisation is a problem which is strongly dependent on the particular document to be processed [55] and usually requires the introduction of *a priori* knowledge about the document to be actually useful in practice. Therefore, we focus in our study on one particular problem, viz. that of text written or printed within a box. Although this may seem very restrictive, the problem has many real practical applications, such as the processing of checks and various forms, for example, tax forms. The second issue addressed in this report is handwriting recognition. Unlike machine printed character recognition for which many commercial products already exist, handwriting recognition is still in its infancy. However, both machine printed and handwritten character recognition have an advantage over layout analysis in the sense that they are better defined problems and thus have a well established theoretical framework, viz. that of pattern recognition [11, 56, 17].

Handwriting is characterised by its variety in shape. There are many factors that contribute to this variety, the first of which is the writing style in use. Fig. 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 [51]. 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, [53] 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. Due to this complexity, we divide the problem of handwriting recognition into three subproblems, namely, the recognition of isolated numerals, that of numeral strings, and the recognition of cursively handwritten words drawn from a small lexicon.

The next section addresses text localisation and the following sections address the three aforementioned subproblems of handwriting recognition.

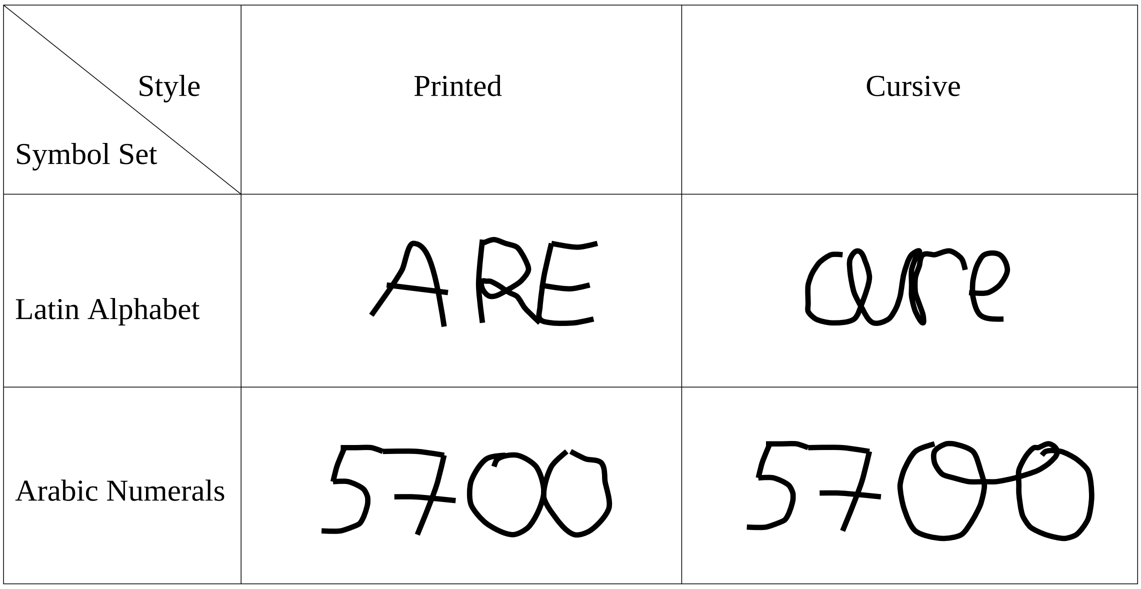

Figure 1: Variety of handwriting.

2 TEXT LOCALISATION

We address the problem of extracting text written or printed inside a box. As mentioned in the last section, this problem is highly dependent on the particular kind of document to be processed. Therefore, we took one kind of document, viz., the Giro forms used in Switzerland as experimental data in our study. First, the structural properties of the Giro forms are given, particularly the region that contains the boxes circumscribing the amount of money (Section 2.1). We will also present a number of deviations of real Giro forms from the standard, which illustrate the difficulties of the problem. Next, a brief overview of potential techniques for solving this problem is given (Section 2.2). Finally, we present the adopted technique, its implementation and experimental results.

2.1 Structural Properties

In this section we describe the region of a Giro form that contains two boxes circumscribing the amount of money. (For a more general description of the whole Giro form, please see [19].) This region is composed of two boxes: the left and the right one. The left one circumscribes the 'francs' part and the right one the 'cents' part of the amount. See Fig. 2 for a few examples. The two boxes are always horizontally aligned. Their size is defined by the Postes, Téléphones et Télégraphes (PTT) of Switzerland, but no tolerances are given nor is the line thickness.

Even for this relatively simple problem, we can already see a number of difficulties. For instance, in Fig. 2 (b), the two boxes touch each other. The amount of money also touches the bounding boxes (Fig. 2 (b)). Moreover, the box sizes and

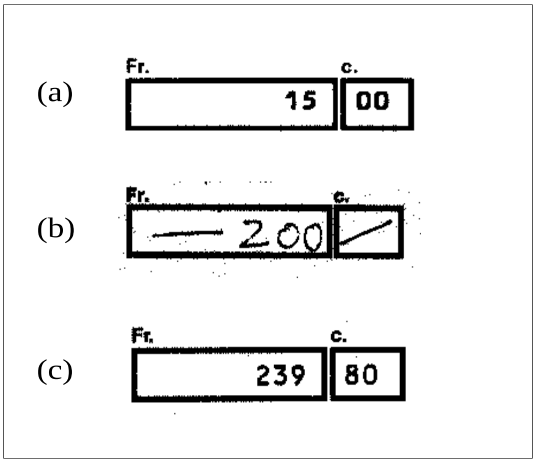

Figure 2: Various instances of boxes circumscribing the amount of money on Giro forms

line thickness may vary in a quite large range of values (see Fig. 2).

2.2 Overview of Techniques

There exist many techniques to solve this kind of problem, most of them are from the field of technical drawings analysis [15]. For instance, the connected component analysis (CCA) may be used to separate black areas that do not touch each other. If the digits do not touch their bounding box and the two bounding boxes do not touch each other, the CCA is able to assign distinct labels to these areas and the text localisation problem is solved by simple comparison of relative positions of black areas. Unfortunately, if some touches are present, this technique fails.

The second technique consists of three steps. The first step thins the image, i.e., reduces the lines to one pixel width. Next, the thinned image is approximated by a set of line segments. Last, these segments are matched against the a priori knowledge about the region, such that long horizontal and vertical segments are assigned to bounding boxes and short segments to digits. This technique is more robust than the CCA. However, thinning may introduce artifacts that have to be eliminated by using some heuristics and special treatments are required to deal with the case of touching bounding boxes.

The third technique also consists of three steps. First, both inner and outer contours of the image are detected (contour tracking). The second and third steps are similar to those of the second technique, i.e., polygonal approximation and matching. This technique has the same properties as the second one.

The fourth technique consists of three steps. First, black pixels of the image are projected on the horizontal and vertical axes, resulting in two profiles. High values in a profile reflect the presence of a line in the perpendicular direction. Therefore, the second step analyses the two profiles to detect the bounding boxes and consequently the digits which lie inside them (last step). This technique is robust against noise because it is based on projection (integration effect). The main drawback is that it is sensitive to the orientation of the image and thus deskewing is required if the original image is tilted. However, such a deskewing is in any case necessary for most document processing operations and very powerful techniques already exist for this purpose [1].

These four techniques are, to our knowledge, representative of a large number of published methods. Their performance can be expected good if there is no touching between bounding boxes and digits. In the presence of these problems, we can a priori exclude the CCA. The next section describes in more detail the fourth technique, which is the simplest among the last three techniques, together with experimental results obtained from real data.

2.3 Projection-Based Technique

The projection-based technique consists of three steps as follows.

1) Project the black pixels in the image on the vertical and horizontal axes. This provides us with two profiles which will be analysed separately. High values in a profile reflect the presence of a line in the perpendicular direction.

2) For each profile, we use a priori information about the boxes (size and tolerance) to determine the circumscribing lines. This is achieved by computing the derivative of the profile the high value of which reveals the potential presence of an edge in the perpendicular direction. The positions at which the derivative of the profile take high values are checked against the a priori size and tolerance information to determine the final result. Thus, circumscribing lines of the two boxes are localised.

3) The amount in francs and cents is finally obtained by extracting all pixels lying inside the two bounding boxes.

The algorithm has been developed and tested using 48 real Giro forms at 200 and 300 dpi resolutions. After some parameter adaptations, the algorithm works correctly on all these cases. One problem that has not been mentioned so far in this report is text that stretches out of its bounding box. This problem can not be solved by the currently implemented version and requires more sophisticated methods, see [18] for an example.

3 HANDWRITING RECOGNITION

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. 3. The

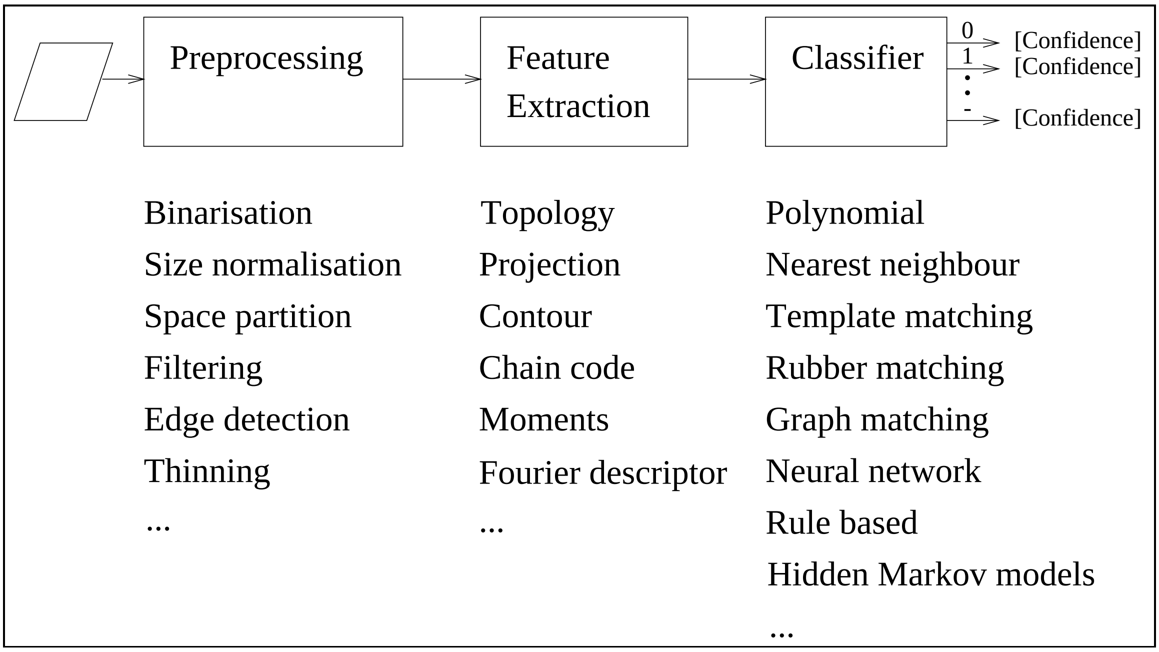

Figure 3: The standard pattern recognition paradigm.

class that has the closest features will be 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.

Handwriting recognition has traditionally been divided into two approaches.

1. *Statistical* approach: the pattern (character or word) is characterised by an ordered set of numerical values.

2. *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 feature type must be compatible, i.e., statistical, respectively structural, features require a statistical, respectively structural, classifier.

Recently, a new paradigm appeared and intended to exploit the mutual advantages and drawbacks of several techniques to yield a better system. It consists in combining the results from several independent systems each of which uses a different technique; see Fig. 4. 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 compared to the individual constitutive techniques [40].

In the following section, we review the three standard operations of a pattern recognition system, namely, preprocessing, feature extraction and classification. The new paradigm of combination can be considered as an extension of the standard pattern recognition paradigm where the combiner is simply a new classifier. The next

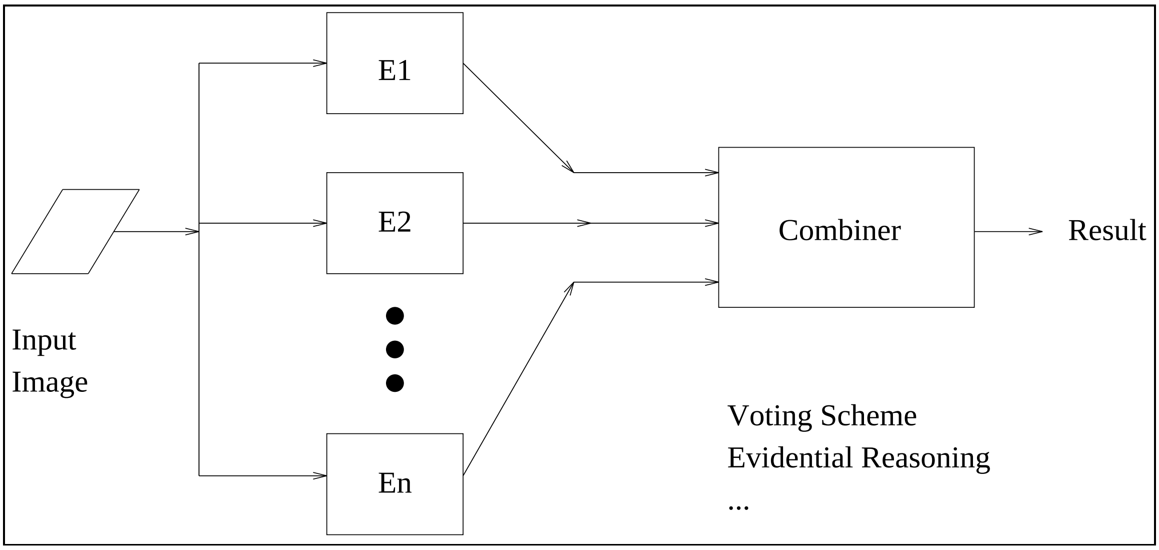

Figure 4: Decision Fusion.

section presents a novel approach, called perturbation approach, which constitutes our main contribution to the design methodology of pattern recognition systems. Finally, we briefly discuss the concept of reliability.

3.1 Pattern Recognition Components

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) [34], binarisation [39], size normalisation [56], slant correction [49] and thinning [54].

Feature extraction provides a compact yet informative representation of the pattern. The success of a pattern recognition system is by large determined by the feature extraction [50]. 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 [13]. Moreover, structural features, such as arcs and holes, are high level representation of edges and contours. By extension, it can be expected that they would yield even better results.

Classification consists in comparing the features provided by the feature extraction with those stored in the training phase. For structural features, a symbolic matching procedure (either exact or inexact) is needed [28, 4]. String, tree or graph matching is required as symbolic matching, depending on the representation of features. Statistical features on the other hand can be compared by using various classification methods, such as, polynomial classifier, nearest neighbour, neural networks and hidden Markov models [56, 17, 38, 47].

3.2 Perturbation Method

Perturbation method results from the critical observation of the general pattern recognition paradigm. The main flaw lies in its *serial* structure, i.e., an error in any part of its structure would lead to definitely wrong results.

Perturbation method tackles the first component of the 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 eliminate the hole of a loop thus destroying essential structural information. More generally, all transformations that alter the standard form of an image are called perturbation models.

Perturbation method consists in applying a set of predefined inverse perturbation models to the input image (see Fig. 5). 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.

3.3 Reliability

In many practical applications, such as automatic bank check reading, it maybe more costly to make a wrong decision than to withhold making a decision (reject). With a reject option, the optimality of a decision rule is defined by the tradeoff between error rate and reject rate, i.e., the decision rule should be designed such that it minimises the error rate for a given reject rate, or vice versa. The optimum decision rule with a reject option for a Bayes classifier consists simply in rejecting the pattern if its highest *posterior* probability is below some predefined threshold $(1-t), t \in [0, 1-\frac{1}{N}]$, where N is the total number of classes [6, 7].

Although the theory is well established, practical problems arise because of the

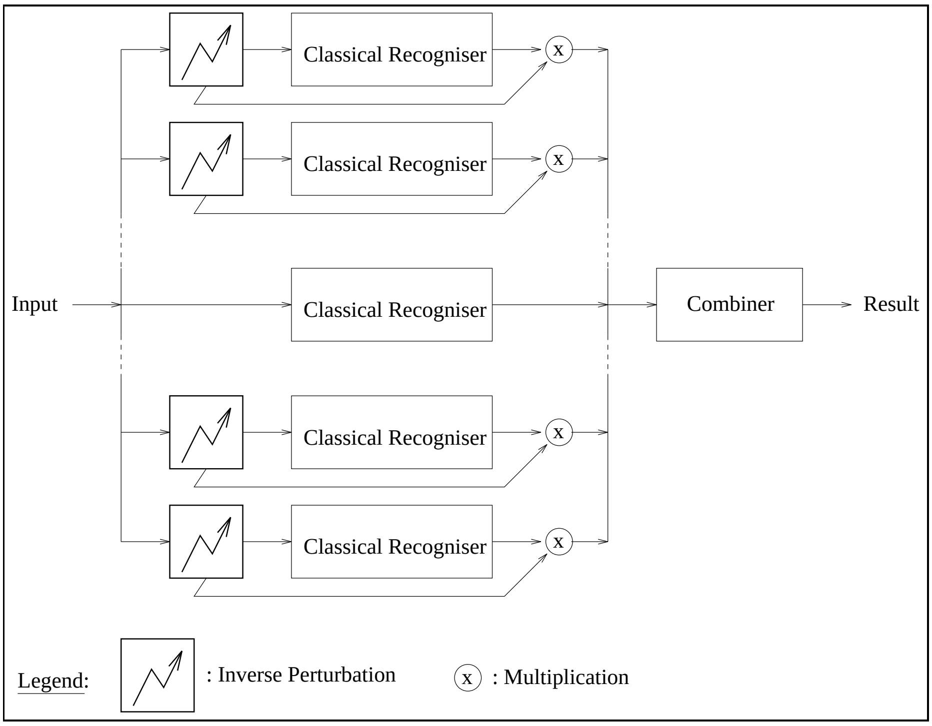

Figure 5: Perturbation-based recognition system.

difficulty in computing the *posterior* probabilities. Two approaches have been pursued. In the first approach, the rule is adapted to the specific classifier, which is usually different from the optimum Bayes classifier. The most prominent example is the adaptation to the nearest neighbour rule leading to the $(k, l)-$Nearest Neighbour Rule [27]. This rule searches for the k nearest neighbours to the test pattern, and chooses the class that is most heavily represented if more than l votes are cast to this class and reject otherwise (qualifying majority). In the second approach, the *posterior* probabilities are first estimated via some statistical estimators, e.g. polynomial-based estimator and neural networks [50, 48]. Then, these estimates are improved by an additional stage, called *confidence mapping*, by using the statistics on the outputs of the first stage estimators [50].

4 ISOLATED NUMERAL RECOGNITION

In this section we discuss 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,

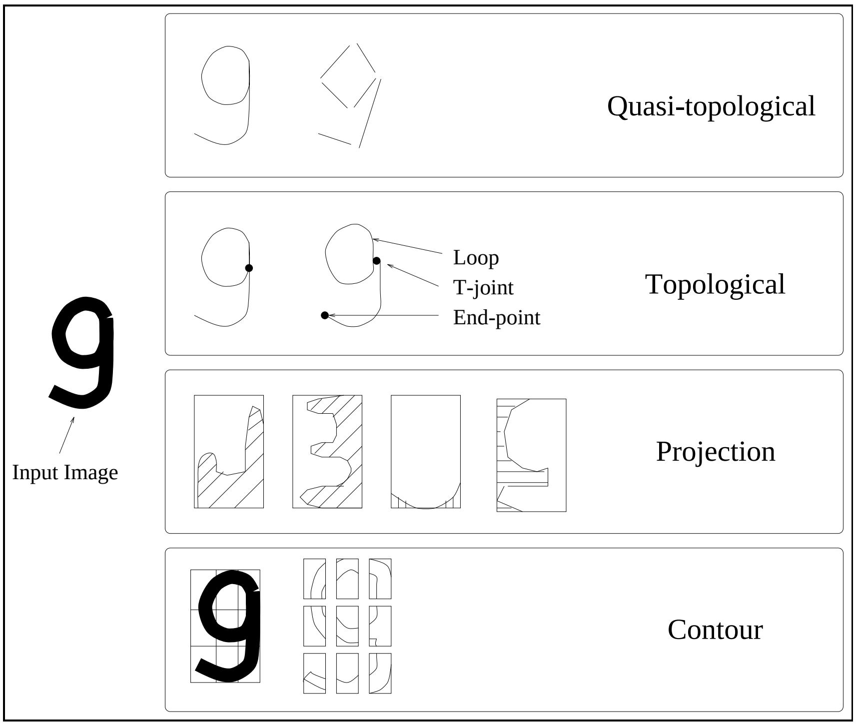

Figure 6: Feature Types.

the first two of which are structural based whereas the last two are statistical based, and tested them on the same database [22]. Fig. 6 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 [12]. 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 [29]. 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 [53, 40]. 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 scheme improved the recognition rate from 97.69% and 98.19% to 98.51% (see Table 1) [22].

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

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.

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 [9]. 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. 7).

We tested the perturbation method on two worldwide standard databases, namely, CEDAR and NIST [29, 57]. 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 reached on this database [20]. On NIST, the perturbation method improved the recognition rate of the combination-based system from 99.45% to 99.54% [24, 26].

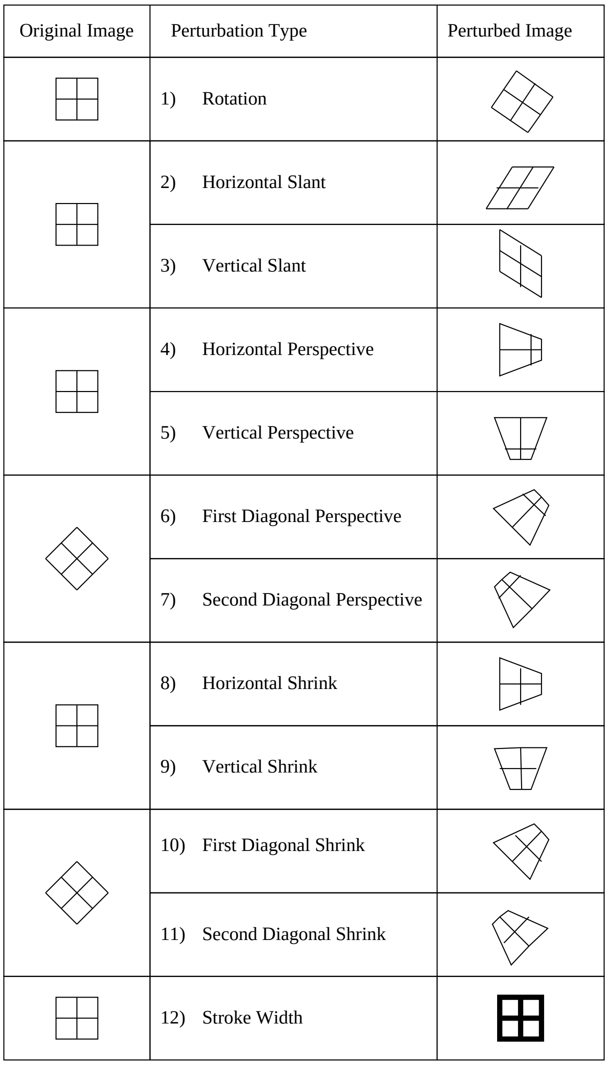

Figure 7: Perturbation models for isolated handwritten numerals.

5 NUMERAL STRING RECOGNITION

The problem of recognising numerals strings differs from that of isolated numerals mainly because it cannot be cast into the standard framework where the input pattern is classified into one of a finite number of classes. Indeed, a numeral string can represent any integer number from zero to infinity, or at least a very large value. Therefore the problem is no longer within the standard framework, and some extensions are necessary. Two approaches have been proposed to deal with this problem, namely, discrete and continuous. See [5] for a recent survey.

In the discrete approach, which includes both segmentation-based and segmentation-free methods, the numeral isolation (separation) takes place at a number of points where the image exhibits some special characteristics. For instance, the analysis of the vertical projection of black pixels provides a simple (but not always correct

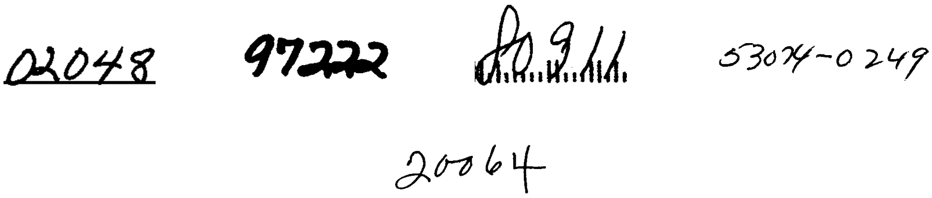

Figure 8: Numeral String Samples from CEDAR Database

or sufficient) way to segment the input string into numerals or groups of numerals. Other special characteristics are line end-points, crossings, and T-joints [46]. For segmentation-based methods, these special points are carefully chosen so that they provide correct cuttings of the input image without wrongly splitting any numeral across its body. This task is difficult and has been shown to be unreliable [16]. The segmentation-free methods remedy these problems [46], [16]. Each split is validated by submitting its left- and right-part to an isolated recognition system. The resulting confidences are then used as a measurement of the likelihood of the split being correct. Thus, all potential splitting points are examined and the system chooses the subset to be the final result that maximizes a global likelihood. Therefore the choice is more rational than in the segmentation-based methods where heuristics prevail, at the price of a higher computational complexity, however. Another important drawback of the segmentation-free methods is that the segmentation error rate increases very fast with the number of numerals composing the whole string [46, 36]. See also [8, 14, 37, 42, 52] for other versions of segmentation-free methods.

In the continuous approach, a sliding window scans the input image from left to right. Each position of the window defines a sub-image which is extracted and analyzed. Since it is generally impossible to have a window width that fits all individual numerals in a string, some mechanism must be provided to handle this variability. For instance, the window width is chosen to be roughly two times the average width of the individual numerals so as to be almost certain that it covers any numeral in a string at least once through the scanning [43]. Another method consists in using a recurrent neural network (a general architecture that includes time-delay neural networks) to encode information between neighboring vertical strips (thin vertical sliding windows) [41]. Other similar mechanisms have also been used [35, 44].

At present time it is not clear which approach, discrete or continuous, is better. Interestingly enough, most discrete methods report their results using the CEDAR database [29] whereas the continuous methods tend to use the NIST database [57] (see Figs. 8 and 9). The former contains *totally unconstrained* data collected from live mail whereas the latter contains *slightly constrained* data (the writers were requested to write the numeral strings in preprinted boxes). There seems to be a strong correlation between data and methods.

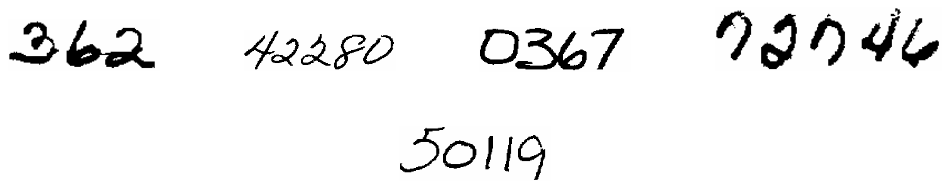

Figure 9: Numeral String Samples from NIST SD3 Database

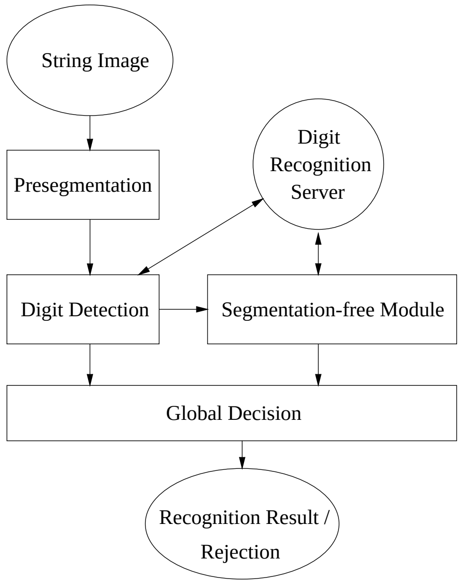

Figure 10: System Architecture.

The system presented in this report is based on the discrete approach, but we will present our results on both the CEDAR and NIST databases, thus providing a first cross-validation. The system architecture appropriately combines segmentation-based and segmentation-free methods. This combination takes advantages of one method to compensate for the drawbacks of the other, and vice versa. The basic idea consists in alleviating the importance of heuristics in the segmentation-based methods by means of a segmentation-free method while avoiding the main drawbacks of the latter, namely, its fast-dropping segmentation rate and its excessive computational burden for long numeral strings. More precisely, our system is composed of a cascade of two methods, the first – called presegmentation in the following – is segmentation-based, whereas the second is segmentation-free. In contrast with the usual segmentation-based methods where the numeral string is segmented into individual numerals, our presegmentation *only* attempts to segment it into groups of numerals. Each group should represent an arbitrary integer number of numerals, i.e. one, two, three, ..., numerals. All groups of numerals (with unknown lengths) are then separately recognized and the results are eventually merged together yielding the final interpretation; see Fig. 10. To avoid the two aforementioned drawbacks of segmentation-free methods, each output of presegmentation (a partial image) is first submitted to a 'Digit Detection' module and only if the partial image is rejected do we go to the 'Segmentation-free' module. For further details, see [21].

We have implemented the above system and carried out extensive experiments on data from both the CEDAR and NIST database. Recognition rates of 83.6% (CEDAR) and 92.7% (NIST) at the string level were achieved. These results compare favorably to other published methods. Since our system – based on a discrete approach – was developed for totally unconstrained data (CEDAR), it could also handle slightly constrained data (NIST).

6 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 the 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 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 framework, we first apply the standard method, i.e., preprocessing, feature extraction and classification, and then investigate the use of perturbation method.

Preprocessing plays a much more important role for cursive handwriting than for

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

Table 2: Recognition rate without perturbation approach.

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

Table 3: Recognition rate with perturbation approach.

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 comprises a series of normalising operations, namely, skew correction, slant correction, baseline detection, and size normalisation in the $x-$ and $y-$directions.

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. 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 [47]. 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. The results are shown in Table 2.

In principle, the perturbation method can be applied to any of the normalising operations. For instance, to apply the perturbation method to slant correction, we generate three normalised images the first of which is obtained using the usual estimate of the slant angle α whereas the remaining two images are obtained using $\alpha \pm \delta$. These three images are then fed to the same HMM-based classifier and their results are combined. The method is similar for any other normalising operations. For further details, see [23, 25]. We have performed a series of experiments and found out that the size normalisation in the $y-$direction is the most important one. Table 3 reports the recognition rates obtained by applying the perturbation method on size normalisation in the $y-$direction. A remarkable improvement of 6% in the recognition rate is observed.

7 CONCLUSION

In this report we have discussed text localisation and handwriting recognition. For text localisation, we considered the problem of extracting text printed or written inside boxes on Giro forms. We reviewed a number of representative methods for solving this problem, implemented one, and tested it on real data. For handwriting recognition, we first presented the standard general methodology, including our novel perturbation method and a brief discussion on reliability. Three subproblems of handwriting recognition, namely, the recognition of isolated numerals, that of numeral strings, 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.

References

[1] H.S. Baird, "Anatomy of a Versatile Page Reader", in *Proc. of the IEEE **80**,* No. 7, 1992 pp. 1059-1065.

[2] 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.

[3] H.S. Baird, H. Bunke and K. Yamamoto (Eds.), *Structured Document Image Analysis*, Springer-Verlag, 1992.

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

[5] R.G. Casey and E. Lecolinet, "Strategies in Character Segmentation: A Survey," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 1028-1033.

[6] C.K. Chow, "An Optimum Character Recognition System Using Decision Functions," *Institute of Radio Engineers (IRE) Transactions on Electronic Computers* ***6***, pp. 247-254, December 1957.

[7] C.K. Chow, "On Optimum Recognition Error and Reject Tradeoff", *IEEE Transactions on Information Theory,* Vol. IT-16, No. 1, pp. 41-46, January 1970.

[8] G. Congedo, G. Dimauro, S. Impedovo, and G. Pirlo, "Segmentation of Numeric Strings," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 1038-1041.

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

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

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

[12] 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.

[13] 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.

[14] A. Filatov, A. Gitis, and I. Kil, "Graph-based Handwritten Digit String Recognition," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 845-848.

[15] A.J. Filipski and R. Flandrena, "Automated Conversion of Engineering Drawings to CAD Form," in *Proc. of the IEEE* ***80***, No. 7, 1992, pp. 1195-1209.

[16] H. Fujisawa, Y. Nakano, and K. Kurino, "Segmentation Methods for Character Recognition: From Segmentation to Document Structure Analysis", *Proceedings of the IEEE,* Special Issue on OCR, Vol. 80, No. 7, pp. 1079-1092, July 1992.

[17] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second edition, Academic Press, 1990.

[18] V. Govindaraju and S.N. Srihari, "Separating handwritten text from interfering strokes," in *From Pixels to Features III: Frontiers in Handwriting Recognition*, S. Impedovo and J. C. Simon (eds.), Elsevier Science Publishers B.V., Amsterdam, The Netherlands, pp. 17-28, 1992.

[19] Thien M. Ha and H. Bunke, "Model-Based Analysis and Understanding of Check Forms," *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 8, No. 5, 1994, pp. 1053-1080.

[20] 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.

[21] Thien M. Ha, D. Niggeler, and H. Bunke, "A System for Segmenting and Recognising Totally Unconstrained Handwritten Numeral Strings," *Proceedings of the third International Conference on Document Analysis and Recognition,* Aug. 14-16, 1995, Montréal, Canada, pp. 1003-1009.

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

[23] 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. To appear.

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

[25] G. Kaufmann, H. Bunke, and Thien M. Ha, "A Perturbation-Based System for Recognising Handwritten German Words," *DAGM-Symposium, Mustererkennung 1996,* Sept. 11-13, 1996, Universität Heidelberg, Germany. To appear.

[26] H. Bunke and Thien M. Ha, "On Perturbation Method in Handwriting Recognition," Machine Vision and Applications' 96, November 12-14, 1996, Tokyo, Japan. To appear.

[27] M.E. Hellman, "The Nearest Neighbor Classification Rule with a Reject Option," *IEEE Transactions on Systems, Science, and Cybernetics,* Vol. SSC-6, No. 3, pp. 179-185, July 1970.

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

[29] 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.

[30] *Proceedings of the Second International Conference on Document Analysis and Recognition,* Tsukuba, Japan, Oct. 20-22, 1993.

[31] *Proceedings of the third International Conference on Document Analysis and Recognition,* Montréal, Canada, Aug. 14-16, 1995.

[32] T. Pavlidis and S. Mori (Eds.), *Proceedings of the IEEE,* Special Issue on Optical Character Recognition, July 1992.

[33] L. O'Gorman and R. Kasturi (Eds.), *IEEE Computer,* Special Issue on Document Image Analysis Systems, July 1992.

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

[35] J. Keeler and D.E. Rumelhart, "A Self-Organizing Integrated Segmentation and Recognition Neural Network," in *Advances in Neural Information Processing Systems,* Vol. 4, J.E. Moody, S.J. Hanson, and R.P. Lippmann (Eds.), San Mateo, CA : Morgan Kaufmann, 1992, pp. 496-503.

[36] F. Kimura and M. Shridhar, "Segmentation-Recognition Algorithm for Zip Code Field Recognition", *Machine Vision and Applications,* Vol. 5, No. 3, Springer Verlag, pp. 199-210, Summer 1992.

[37] F. Kimura, Y. Miyake, and M. Shridhar, "Handwritten ZIP Code Recognition Using Lexicon Free Word Recognition Algorithm," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 906-910.

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

[39] 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.

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

[41] S.W. Lee and E.J. Lee, "Integrated Segmentation and Recognition of Connected Handwritten Characters with Recurrent Neural Network," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 413-416.

[42] E. Lethelier, M. Leroux, and M. Gilloux, "An Automatic Reading System for Handwritten Numeral Amounts on French Checks," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 92-97.

[43] G.L. Martin, M. Rashid, and J.A. Pittman, "Integrated Segmentation and Recognition Through Exhaustive Scans or Learned Saccadic Jumps," *International Journal of Pattern Recognition and Artificial Intelligence,* Vol. 7, No. 4, 1993, pp. 831-847.

[44] O. Matan, J.C. Burges, Y. Le Cun, and J.S. Denker, "Multi-Digit Recognition Using a Space Displacement Neural Network," in *Advances in Neural Information Processing Systems,* Vol. 4, J.E. Moody, S.J. Hanson, and R.P. Lippmann (Eds.), San Mateo, CA : Morgan Kaufmann, 1992, pp. 488-495.

[45] 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.

[46] H. Nishida and S. Mori, "A Model-Based Split-and-Merge Method for Recognition and Segmentation of Character Strings", in *Advances in Structural and Syntactic Pattern Recognition*, H. Bunke (Ed.), World Scientific, pp. 300-309, 1992.

[47] 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.

[48] M.D. Richard and R.P. Lippmann, "Neural Network Classifier Estimate Bayesian *a posteriori* Probabilities," *Neural Computation **3**,* 1991, pp. 461-483.

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

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

[51] 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.

[52] N.W. Strathy and C.Y. Suen, "A New System for Reading Handwritten ZIP Codes," Proceedings of the third International Conference on Document Analysis and Recognition, Aug. 14-16, 1995, Montréal, Canada, pp. 74-77.

[53] 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.

[54] 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.

[55] S. Tsujimoto and H. Asada, "Major Components of a Complete Text Reading Systems", in *Proc. of the IEEE **80**,* No. 7, 1992, pp. 1133-1149.

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

[57] 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.