Text Localisation and Handwriting Recognition: Application to Numeral Extraction and Recognition *

Thien M. HA and Horst BUNKE
University of Berne
Institut für Informatik und Angewandte Mathematik
Länggassstr. 51, CH-3012 Bern, Switzerland
Email: haminh@iam.unibe.ch

October 27, 1994

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 standard general methodology, then apply it to handwritten numeral recognition, and finally propose a new approach, called perturbation-based, that allowed us to boost the recognition rate up to 99.1% on a world-wide standard database.

CR Categories and Subject Descriptors: I.4.3 [Image Processing]: Enhancement; I.4.4 [Image Processing]: Restoration; I.4.5 [Image Processing]: Reconstruction; I.4.7 [Image Processing]: Feature Measurement; I.4.10 [Image Processing]: Image Representation; I.5.0 [Pattern Recognition]: General; I.5.1 [Pattern Recognition]: Models; I.5.2 [Pattern Recognition]: Design Methodology; I.5.4 [Pattern Recognition]: Applications; I.5.m [Pattern Recognition]: Miscellaneous.
Additional Key Words: Text localisation, handwriting recognition, isolated numeral (digit), GIRO form, geometric correction, inverse process, invariants, projections, contour, size and shape, statistical, structural, classifier design, feature evaluation, perturbation method.

*This work was partly supported by the UBILAB (Union Bank of Switzerland Information Technology Laboratory) and the Swiss National Funds.

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 [IEEE 92, IEEE Comp. 92, MVA 92, BaBuYa 92, ICDAR 93], 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 consists of names and/or addresses and can thus be verified against a directory of valid names and addresees 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: text localisation and some preliminary studies on handwriting recognition. It has been recognised that text localisation is a problem which is strongly dependent on the particular document to be processed [TsAs 92] and usually requires the introduction of a priori knowledge about the document to be actually useful in practice. Therefore, we will 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 [DuHa 73, Ul 73, Fu 90]. The next section addresses text localisation and Section 3 deals with handwriting recognition.

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 described (Section 2.2). Finally, we present the adopted technique, its implementation and experimental results.

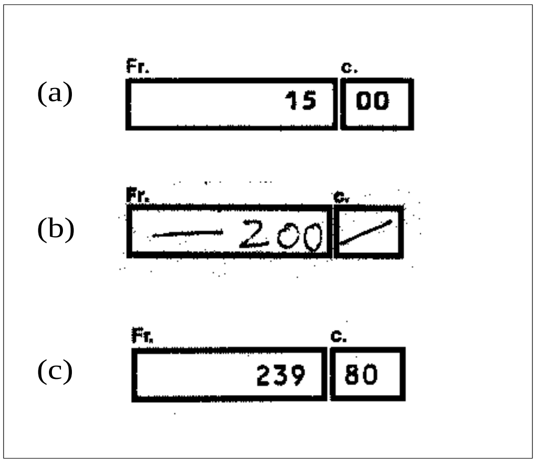

Figure 1: Various instances of boxes circumscribing the amount of money on Giro forms

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 [HaBu 94b].) 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. 1 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. 1 (b), the two boxes may touch each other. The amount of money can also touch the bounding boxes (Fig. 1 (b)). Moreover, the box sizes and line thickness may vary in a quite large range of values (see Fig. 1).

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 [FiFl 92]. 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 artefacts 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 a 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 [Ba 92a].

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, circumsribing 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 developped 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

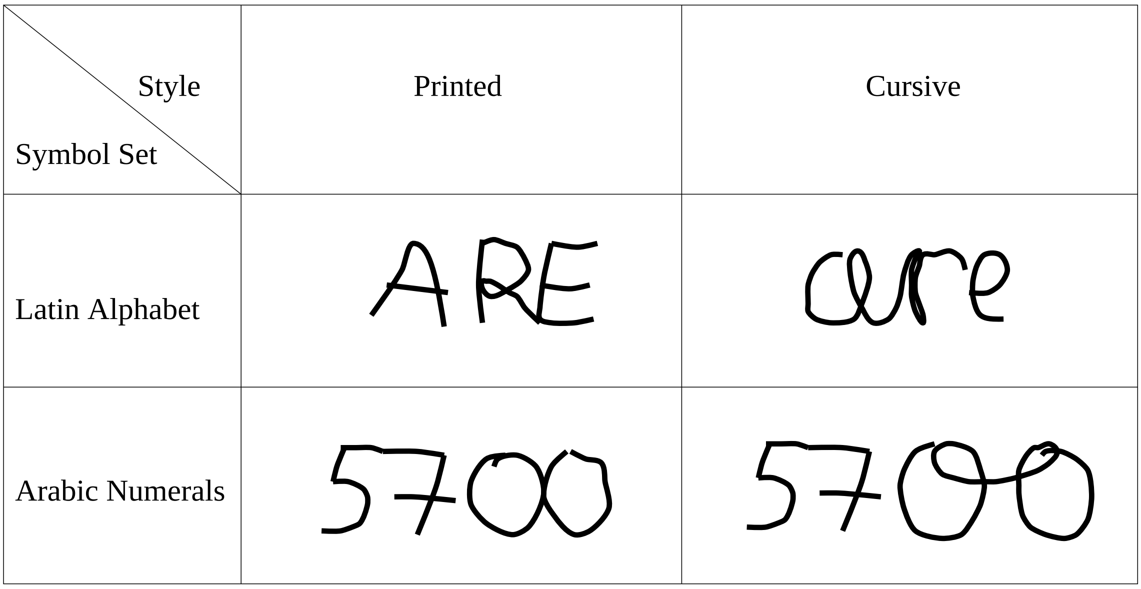

Figure 2: Variety of handwriting.

[GoSr 92] for an example.

3 HANDWRITING RECOGNITION

Handwriting is characterised by its variety in shapes. There are many factors that contribute to this variety, the first of which is the writing style in use. Fig. 2 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 [Si 92]. 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, [Suen et al. 92] pointed out three main groups of factors that can account for the variety of handwriting, namely, the writer's personnality, the circumstances at the writing time, and various technical aspects, such as paper, ink color and writing instrument. In the sections that follow, we will first present the standard general methodology in handwriting recognition and then focus on the particular but important case of isolated Arabic numeral recognition. In Section 3.3, we introduce a totally new approach – called perturbation-based approach – that endeavours to explicitly model writing habits and styles. This approach allowed us to boost the recognition rate of our isolated numeral recognition system to more than 99% on the CEDAR database (a worldwide standard database [Hu 94]).

3.1 Standard General Methodology

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. The 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 traditionnally 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 string, tree or graph.

It is clear that the classifier and the feature type must be compatible, i.e., statistical, respestively 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. 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 [Lee 93].

In the following sections, we review the three *standard* operations of a pattern recognitoin 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.

3.1.1 Preprocessing

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) [Ja 89], binarisation [LeCh 90], size normalisation [Ul 73], slant correction [Schürmann et al. 92] and thinning [SuWa 93].

3.1.2 Feature Extraction

Feature extraction provides a compact yet informative representation of the pattern. The success of a pattern recognition system is by large determined by the features extraction. Unfortunately, there exist no general feature extraction methods and the design of features remains a highly heuristical activity. Even in the limited scope of handwriting recognition, a very large number of feature types has been proposed [MoSuYa 92]. In the following, we will discuss two statistical and two structural feature types.

Fig. 3 illustrates the four selected feature types, the first two of which are structural and the others are statistical. The first feature type is obtained by first decomposing the line image into x-y-monotone segments. The x-y-monotony of a segment is characterised by the fact that its x and y coordinates are non-increasing or non-decreasing when we follow the segment from one end to the other. These segments are then used to build up a high level representation of the line image, called description by quasi-topological features [NiMo 92, Ni 94]. The second feature type – called topological – consists of *loops*, *end points* and *T-joints*, each augmented by some metrics information such as position and size [HuCh 86, Zi 93]. The third feature type – called projection-based – consists of a collection of directional projections of black pixels (Radon transform) and directional profiles of the outer contour. These projections and profiles are considered as one-dimensional signals and subsampled yielding the feature vector of the input image [Lee 93, HaBu 94a]. The last feature type – called contour-based – is obtained by first detecting the contour points and then analysing them separately in nine (3x3) overlapping regions [HaBu 94a, Gr 94]. For each region, a directional histogram is computed. All channels of the nine histograms are concatenated into a single feature vector.

3.1.3 Classification

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 [Bu 93]. 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 [Ul 73, Fu 90, La 92, Ra 89].

3.2 Isolated Numeral Recognition

This section presents the application of the standard general methodology to isolated Arabic numeral recognition. We discuss four independent systems based on the four feature extraction methods previously mentioned, as well as a combination method based on a weighted voting scheme. Experimental results obtained using the CEDAR database will be reported.

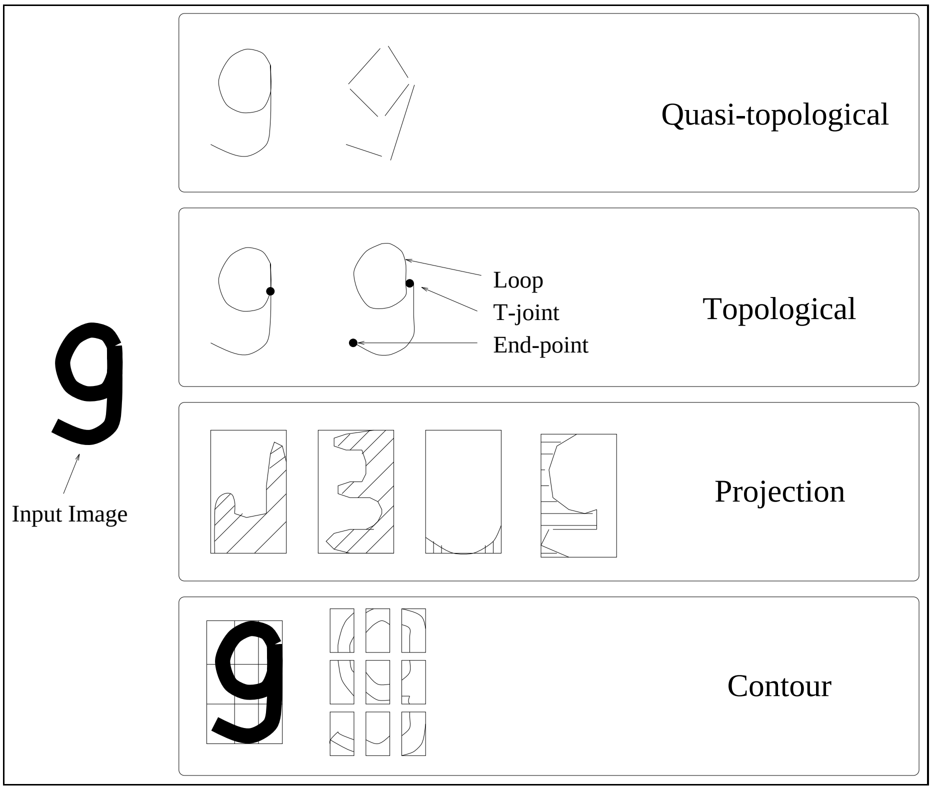

Figure 3: Feature Types.

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.

In general, the result of text localisation may provide us with images such as those in Fig. 2, i.e., numerals may be touched. Therefore, prior to the recognition of isolated numerals, the image must somehow be segmented into individual numerals. This is, in itself, a very difficult problem the discussion of which is out of scope of this report; see [Fe 92, KiSh 92, Ni 94]. In the following, we will only consider the images that represent isolated numerals.

In the first system, the input image is preprocessed by filtering to eliminate small holes and speckle noise, and the description by quasi-topological features is extracted from the filtered image. This description is then used, in the recognition phase, for comparison with a number of prototype descriptions in a knowledge base, obtained in the training phase [NiMo 92, Ni 94]. The second system makes use of topological features, such as loops, end points and T-joints and classifies the input pattern by an inexact graph matching procedure [HuCh 86, Zi 93]. The third system extracts projection-based features in the form of a feature vector and the classification method is based on the distance-weighted K–Nearest Neighbour rule [HaBu 94a, Du 76]. The last system is similar to the the third one, except for the features, which are contour-based [HaBu 94a, Gr 94].

These four systems were tested by using 18468 samples (*br* directories) for training and 2213 samples (*goodbs* directories) for testing, both from the CEDAR database [Hu 94]. 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-rejetion 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 [Suen et al. 92, Lee 93]. Structural methods are appealing because they seem to match the way human beings read charaters 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 recogniton rate from 97.69% and 98.19% to 98.51% (see Table 1).

3.3 Perturbation-based Approach

In this section we present a new approach to handwritten numeral recognition. Our approach is based on the modelling of human writing habits and thus tackles the difficult cases in totally unconstrained data. The writing habits are modelled via a set of geometric transformations (e.g., rotation and slant) called perturbation models. These models aim at representing the deviations of written numerals from what we consider 'standard patterns'. Assuming that these models cover a large spectrum of writing habits, their use in the design of a recognition system requires a kind of *reversing process*, capable of correctly bringing a written numeral back to one of its standard forms.

Many approaches have been proposed to solve the reversing problem. For instance, [Schürmann et al. 92] normalise the input image according to a set of reversing parameters computed from the image itself. In [YaYaSa 90] the reversing parameters are determined so as to equalise a density function. These normalisation procedures have all contributed to improving the recognition rate significantly. However, they include a number of drawbacks that will be detailed in Section 3.3.1. Translation- and rotation-invariant transforms have also been proposed as an indirect way to cope with the reversing problem [Ja 89]; they are generally limited to simple (geometric affine) transformations and are not complete. Elastic matching is another promising indirect way to cope with this problem. However, it suffers from two shortcomings, namely, that prototypes have to be handcrafted and that the definition of cost functions is not obvious [Revow 93, Wa 93].

In our approach, the reversing process is model-based. That is, in contrast with the standard pattern recognition paradigm, where normalisation is performed a priori, we first apply a fixed, predefined set of hypothetical inverse perturbations to the image. Then all perturbed versions are submitted separately to a same type of conventional numeral recogniser. These two steps provide us with a set of results which are subsequently combined. The idea behind this scheme is that at least one of the hypothetical inverse perturbations will correctly bring the input image back to one of its standard forms. Such a perturbation should manifest its evidence by outperforming the others in terms of output score. Thus the reversing process is not exclusively based on image data but utilises model information, too. In this report we limit ourselves to describing the task of recognising the input numeral without identifying the best reverse perturbation model(s) explicitly.

Section 3.3.1 presents the basic observations that led us to the approach proposed in this report. The general architecture of our system is described in Section 3.3.2. Details of the new recognition method are provided in Section 3.3.3. Section 3.3.4 presents our experiments using the CEDAR database. We discuss our results in

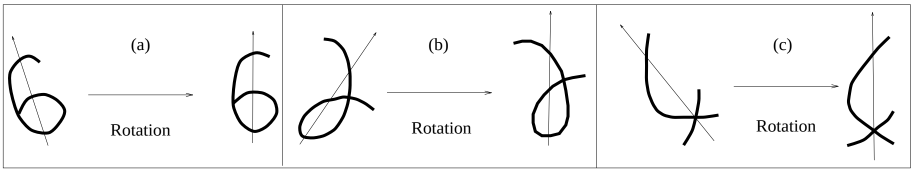

Figure 4: Observations.

Section 3.3.5 and conclude the new approach in Section 3.3.6.

3.3.1 Observations

The key idea of the perturbation approach lies in the process of reversing an input image back to one of its standard forms. In the present section, we will show, through examples, that

> **Observation 1**
> *the reversing process can not always be correctly achieved by using the input data alone.*

To support our assertion, we will give some examples where a reversing process using input data alone fails. By 'input data alone' we mean all techniques based on the pixel distribution disregarding the nature of the alphabet being considered. For instance, line regression is a powerful method to determine the rotation angle of a character [Schürmann et al. 92]; it is based on pixel distribution and is independent of the alphabet (be it Arabic numerals, Roman characters, Greek characters, etc.). For instance, in Fig. 4a, the detection of the rotation angle by line regression yields a correct result for the numeral '6'. However, in Figs. 4b and 4c, line regression fails to detect the correct rotation angles of the numerals '2' and '4'. Notice that in the case of Fig. 4b, the result would be correct if we were considering the Greek alphabet (symbol γ). These simple observations show that the input data alone may fail to provide the correct reversing parameters. In order to overcome this problem, the reversing process must take into account the nature of the alphabet in use.

Although we have taken the rotation operation as an example, it should be clear that the argumentation is applicable to other geometric transformations, e.g., slant and perspective views. In general we observe that

> **Observation 2**
> *most perturbations are due to writing habits and styles.*

Without referring in detail to other disciplines, such as graphology and calligraphy, it can readily be seen that a number of simple geometric transformations (such as rotation, slant, perspective view and shrink) can account for a wide variety of handwriting habits and styles (see Fig. 5).

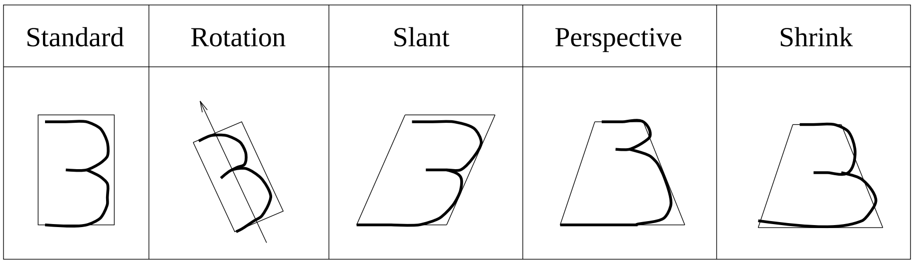

Figure 5: Examples of styles.

3.3.2 Perturbation-based Recognition

In this section we describe our recognition scheme, which is based on Observation 1. In order to solve the problem that the reversing process cannot always be correctly achieved using the input data alone, we propose the reversing process to be model-based, i.e., taking into account the nature of the alphabet being considered. Since we are primarily interested in designing a recognition system and not concerned with explicitly recovering the standard form, the reversing process will be implicit.

The basic idea consists in applying a set of predefined inverse perturbations to the input image (see Fig. 6). 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.

In reality, a written pattern may result from a mixture of many perturbation types (mixed style). So the determination of the true perturbation may prove difficult. However, if the goal is to design a recognition system, then a weighted-voting scheme is sufficient to determine the most probable class of the input image. Notice in Fig. 6 that each output score is downweighted by a factor which is inversely proportional to the perturbation degree. This is meant for avoiding that an input image is transformed into a completely different pattern without being penalised.

3.3.3 Parametrisation

The previous section proposed a new recognition scheme in generic form, i.e., without specifying the exact parameters. By using Observation 2 on writing habits and styles, we propose a parametrisation based on four geometric transformations, namely, rotation, slant, perspective view and shrink. Moreover, slant is decomposed into horizontal and vertical directions, whereas perspective view and shrink are each

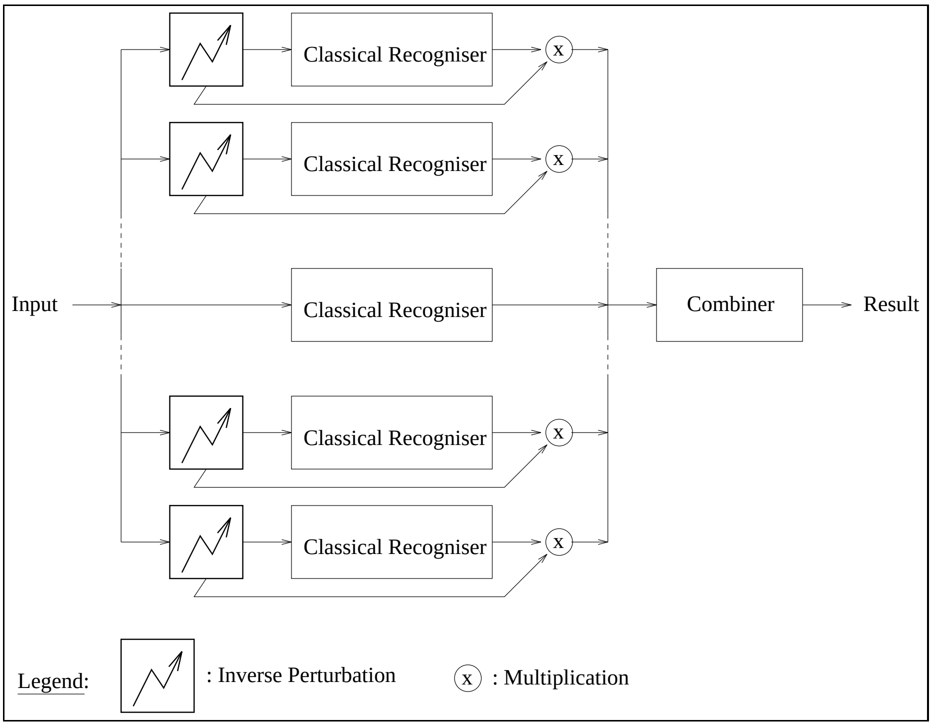

Figure 6: Perturbation-based recognition system.

decomposed into horizontal, vertical, 1^{st} diagonal and 2^{nd} diagonal directions. This results in a total of $T = 11$ perturbation types (Fig. 7).

Let $I(x, y)$ be an image and $I'(x', y')$ its perturbed version defined by the perturbation type $t(t = 1, .., T)$ and degree $\delta \in [-1, 1]$, i.e., $I(x, y) = I'(x', y')$ where $(x, y) = f(x', y'; t; \delta; M_t)$. For each perturbation type t, the transformation f depends on an empirically determined scalar M_t, which represents the maximum extent of that perturbation. For instance, the rotation angle of the first perturbation type $(t = 1)$ is defined by $\theta = \delta \cdot M_1$, i.e., M_1 is the rotation angle for $\delta = 1$. Note that if the perturbation degree $\delta = 0$, then all perturbation types become the identity transformation, i.e., $(x, y) = (x', y')$.

In order to catch the true perturbation parameter δ, we can try to cover the parameter space (t, δ) by sampling δ (t is already discrete) with a small sampling interval Δ, i.e., $\delta = k \cdot \Delta, k = \pm 1, \pm 2 \, .. \, \pm K; K \cdot \Delta \leq 1$. However, this approach is very time consuming since for each additional value of δ , $T = 11$ channels have to be added to the system. Eventually, we decided to use a relatively large sampling interval $(\Delta = 0.2)$ and to limit the range of values for k to $k = \pm 1, \pm 2$, resulting in an actual range of $[-0.4, 0.4]$ for δ. This means that we do not attempt to detect the true perturbation parameters but simply make the input image closer to one of its standard forms, instead. The penalisation factor p is related to the perturbation degree by:

$$p = (p_0)^{-|k|} \; ; p_0 > 1 \tag{1}$$

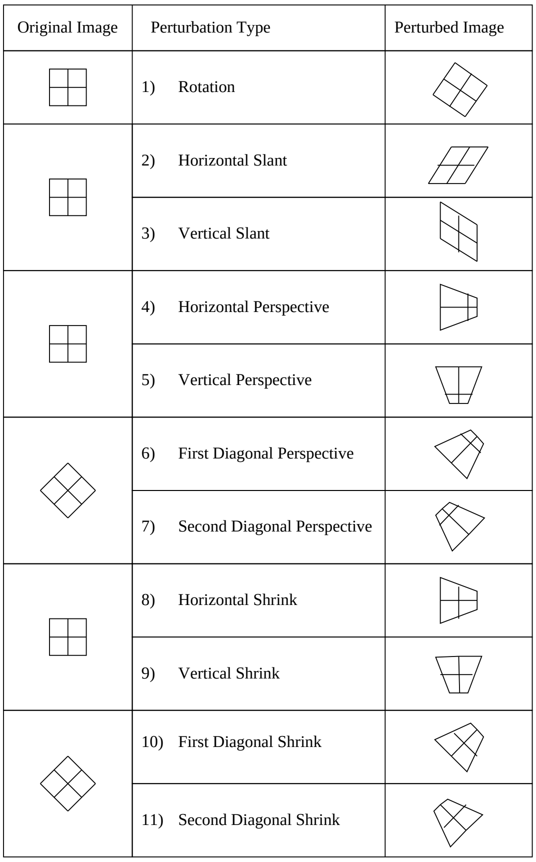

Figure 7: Parametrisation.

In short, the perturbation-based recognition is characterised by the following parameters:

$$\begin{array}{rcl} M_t & : & maximum\ extent\ of\ perturbation\ type\ t; \\ & & t = 1, .., T; T = 11. \\ \Delta & : & sampling\ interval\ for\ \delta; \Delta = 0.2. \\ K & : & maximum\ perturbation\ degree; K = 2. \\ p_0 & : & penalisation\ coefficient; see (1). \end{array}$$

Taking into account the middle channel of Fig. 6 ($\delta = 0$), the total number of channels is $2K \cdot T + 1$.

3.3.4 Experiments

We tested our new algorithm on the same data as in Section 3.2. As "Classical Recogniser" (see Fig. 6) we considered three potential systems, namely, the two statistical and the combination-based ones, all mentioned in Section 3.2. They will be called respectively, C1, C2 and C3 in the following.

Classical Systems	
C1 (Projection)	97.69
C2 (Contour)	98.19
C3 (Projection & Contour)	98.51
Perturbation-based Systems	
P1 (Projection)	98.59
P2 (Projection & Contour)	99.09
Published Systems	
S1 [Lee 93]	98.87
S2 [Lee 93]	98.46
S3 [Lee 93]	98.33
S4 [Lee 93]	97.92
S5 [Knerr 93]	97.6
S6 [Revow 93]	97.5

Table 2: Recognition rate on *goodbs* data of the CEDAR database.

In order to test the proposed perturbation-based method, we built two systems, which will be called P1 and P2, respectively. C1 – which is the weakest method among C1, C2, and C3 – was adopted as "classical recogniser" in P1, while C3 – the strongest method among C1, C2, and C3 - was used in P2. In both P1 and P2, a distance-weighted k-NN classifier was used as "combiner" according to Fig. 6. To reduce the computation time, we use a three-stage scheme for both systems. In the first stage, the system operates without perturbation using a very high rejection threshold (90%). The rejected cases go to the second stage where perturbation is limited to the first degree ($k = \pm 1$), and only if the score is still lower than the rejection threshold do we go to the third stage where second degree perturbation is used ($k = \pm 2$) in a cumulative way with the 1^{st} and the 0^{th} perturbation degrees. After the 3^{rd} stage, the best class is chosen for forced choice option. Otherwise, an additional threshold is used to reject the input numeral. Recognition rates for both systems P1 and P2 at zero rejection level are reported in Table 2. The rejection-versus-error curves of the best classical system C3 and the best perturbation based system P2 are plotted in Fig. 8.

3.3.5 Discussions and Future Research

The results in Table 2 show that the perturbation approach significantly improves the recognition rate for both classical recognisers C1 and C3. Further improvements are still possible by refinement of some system parameters.

To the best of our knowledge, the second perturbation-based system P2 outperforms all published results on the same data. However since the size of the test set (2213) is rather small, these statistics are not reliable and further tests on larger

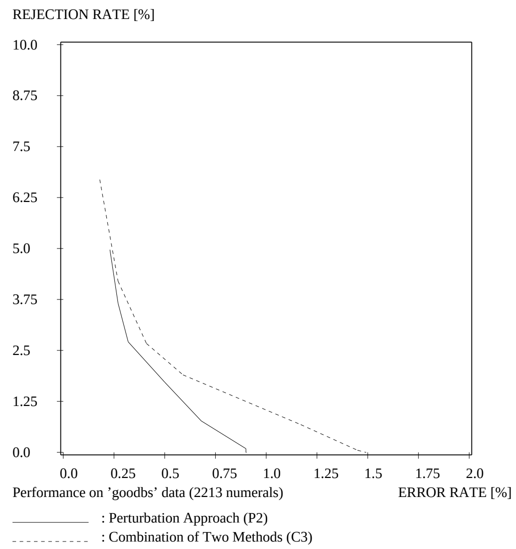

Figure 8: Rejection-versus-error curves.

databases must be performed for better comparison.

The main drawback of the perturbation approach is its computational burden. For example, on a Sun Sparcstation 10, method P2 needs 6 seconds for one numeral on average as opposed to 1.5 seconds needed by C3. However, the structure of the system is highly parallel and therefore is perfectly suited for hardware implementation.

In a broad sense, the perturbation approach can be considered as a kind of 'generalised correlation'. The latter comprises a very large amount of variants, ranging from simple cross-correlation functions (perturbation is limited to translation) to more sophisticated methods, such as the one reported in [YaYaYa 93]. In this work, the authors apply perturbation at the local stroke level according to some rules defined for Japanese characters.

We believe that the perturbation approach opens a new research perspective. For example, virtually any recognition method can be embedded in a perturbation-based system. Moreover, other perturbation models can be designed to improve the modelling of writing habits and styles. Last, the perturbation models can be used to generate more samples from a small training set thus artificially enriching the training data [Ba 92b, Ki 93].

3.3.6 Conclusion

We have presented a new approach to handwriting recognition based on writing habits and styles. The new approach constitutes a shift from the standard pattern recognition paradigm where normalisation is the first step prior to feature extraction and classification. The new approach replaces normalisation by a set of perturbators modelling human writing habits and styles. As a result, the subsequent operations are repeated for each habit and style yielding a set of results that eventually are submitted to a vote.

Computer simulation on totally unconstrained handwritten numerals from the CEDAR database showed that the new approach significantly improves the recognition rates of conventional recognition methods. In our experiments, we obtained 99.1% correct recognition rate, which is the highest rate ever reached on this database.

4 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, then applied it to handwritten numeral recognition, and finally proposed a new approach, called perturbation-based, that allowed us to boost the recognition rate up to 99.1% on a worldwide standard database.

References

[Ba 92a] H.S. Baird, "Anatomy of a Versatile Page Reader", in [IEEE 92], pp. 1059-1065.

[Ba 92b] 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.

[BaBuYa 92] H.S. Baird, H. Bunke and K. Yamamoto (Eds.), *Structured Document Image Analysis*, Springer-Verlag, 1992.

[Bu 93] H. Bunke, "Syntactic and Structural Pattern Recognition," in *Handbook of Pattern Recognition and Computer Vision,* World Scientific, pp. 163-209, 1993.

[DuHa 73] R.O. Duda and P.E. Hart, *Pattern Classification and Scene Analysis*, John Wiley & Sons, 1973.

[Du 76] 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.

[Fe 92] R. Fenrich, "Segmentation of Automatically Located Handwritten Numeric Strings", 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. 47-59, 1992.

[FiFl 92] A.J. Filipski and R. Flandrena, "Automated Conversion of Engineeirng Drawings to CAD Form," in [IEEE 92], pp. 1195-1209.

[Fu 90] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, second edition, Academic Press, 1990.

[GoSr 92] 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.

[Gr 94] A. Grossmann, "Handwritten Numeral Recognition by a Contour Method," Semester Project Report, IAM-PR-500, Institut für Informatik und angewandte Mathematik, University of Berne, Switzerland, 1994.

[HaBu 94a] Thien M. Ha and H. Bunke, "Handwritten Numeral Recognition by Perturbation Method", accepted for publication in the *Proceedings of the Fourth International Workshop on Frontiers of Handwriting Recognition,* Taipei, Taiwan, Dec. 7-9, 1994.

[HaBu 94b] Thien M. Ha and H. Bunke, "Model-Based Analysis and Understanding of Check Forms," accepted for publication in the *International Journal of Pattern Recognition and Artificial Intelligence,* 1994.

[HuCh 86] J.S. Huang and K. Chuang, "Heuristic Approach to Handwritten Numeral Recognition," *Pattern Recognition,* Vol. 19, No. 1, pp. 15-19, 1986.

[Hu 94] 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.

[ICDAR 93] *Proceedings of the Second International Conference on Document Analysis and Recognition,* Tsukuba, Japan, Oct. 20-22, 1993.

[IEEE 92] T. Pavlidis and S. Mori (Eds.), *Proceedings of the IEEE,* Special Issue on Optical Character Recognition, July 1992.

[IEEE Comp. 92] L. O'Gorman and R. Kasturi (Eds.), *IEEE Computer,* Special Issue on Document Image Analysis Systems, July 1992.

[Ja 89] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, 1989.

[Ki 93] N. Kita, "Making a Personal Recognition Dictionary from Characters Automatically Generated by Using Handwriting Model," Proceedings of the 2nd Int. Conf. on Document Analysis and Recognition, Tsukuba, Japan, Oct. 20-22, 1993, pp. 76-81.

[KiSh 92] 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.

[Knerr 93] S. Knerr and S. Sperduti, "Rejection Driven Hierarchy of Neural Network Classifiers," Proceedings of 1993 International Symposium on Nonlinear Theory and its Applications, Vol. 3, Hawaii, 1993.

[La 92] C. Lau (Ed.), "Neural Networks: Theoretical Foundations and Analysis," IEEE Press, 1992.

[LeCh 90] 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.

[Lee 93] 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.

[MoSuYa 92] S. Mori, C.Y. Suen, and Y. Yamamoto, "Historical Review of OCR Research and Development", in [IEEE 92], pp. 1029-1058.

[MVA 92] 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.

[Ni 94] D. Niggeler, "Handwritten Numeral String Recognition", Master Thesis, Institut für Informatik und angewandte Mathematik, University of Berne, Switzerland, 1994.

[NiMo 92] 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.

[Ra 89] 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.

[Revow 93] M. Revow, C.K.I. Williams and G.E. Hinton, "Using Mixtures of Deformable Models to Capture Variations in Hand Printed Digits," *Pre-Proceedings of the third Int. Workshop on Frontiers in Handwriting Recognition,* Buffalo, New York, USA, May 25-27, pp. 142-152, 1993.

[Schürmann et al. 92] J. Schürmann et al., "Document Analysis - From Pixels to Contents," Proceedings of the IEEE, Vol. 80, No. 7, 1992, pp. 1101-1119.

[Si 92] 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.

[Suen et al. 92] C.Y. Suen, C. Nadal, R. Legault, T.A. Mai, and L. Lam, "Computer Recognition of Unconstrained Handwritten Numerals," in [IEEE 92], pp. 1162-1180.

[SuWa 93] 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.

[TsAs 92] S. Tsujimoto and H. Asada, "Major Components of a Complete Text Reading Systems", in [IEEE 92], pp. 1133-1149.

[Ul 73] J.R. Ullmann, *Pattern Recognition Techniques*, Butterworths, 1973.

[Wa 93] T. Wakahara, "Handwritten Numeral Recognition Using LAT with Structural Information," Pre-Proceedings of the third Int. Workshop on Frontiers in Handwriting Recognition, Buffalo, New York, USA, May 25-27, 1993, pp. 164-174.

[YaYaSa 90] H. Yamada, K. Yamamoto and T. Saito, "A Nonlinear Normalization for Handprinted Kanji Character Recognition - Line Density Equalization," Pattern Recognition, Vol. 23, No. 9, 1990, pp. 1023-1029.

[YaYaYa 93] M. Yasuda, K. Yamamoto and H. Yamada, "Effect of Perturbated Correlation Method for Optical Character Recognition," Proceedings of the 2nd Int. Conf. on Document Analysis and Recognition, Tsukuba, Japan, Oct. 20-22, 1993, pp. 830-833.

[Zi 93] M. Zimmermann, "Handwritten Character Recognition", Semester Project Report, IAM-PR-92476, Institut für Informatik und angewandte Mathematik, University of Berne, Switzerland, 1993.