RECOVERING DYNAMIC INFORMATION FROM STATIC HANDWRITING

G. BOCIGNONE, A. CHIANESI, L. P. CORDELLA† and A. MARCELLI
Dipartimento di Informatica e Sistemistica, University of Naples "Federico II", Via Claudio, 21, 80125 Naples, Italy

Abstract — It is generally agreed that the advantage of on-line character recognition methods with respect to off-line ones mostly relies on the availability of dynamic information. This mainly concerns the order in which the strokes forming characters have been drawn. In this paper we present and discuss a method which attempts, in the off-line case, to recover part of the lost script dynamics. The method makes it possible to reconstruct one of the most likely trajectories followed by the writer while drawing characters. It is based on a suitable implementation of good continuity criteria which take into account direction, length and width of the strokes making up characters. The algorithm either subdivides or unfolds the digital ribbon forming isolated characters as well as groups of connected characters, by solving the ambiguities which arise at every joint. Experimental results are reported and discussed.

OCR Handprint Handwriting Off-line recognition Dynamic information Thinning Skeleton shape fidelity

1. INTRODUCTION

In many image processing applications (OCR, storage, analysis and transmission of line drawings, text and map processing, etc.), the digital representation of the objects of interest in the image can be considered as being made of ribbons evolving and intersecting in a two-dimensional (2D) space.

In the case of handwritten characters, this assumption implies thinking of the writing process as a complex human movement of trajectory formation, which sequentially structures a ribbon shape as a function of time. In this framework, space oriented models have been proposed to interpret handwriting as a mechanism capable of generating smooth trajectories by composing basic curve elements.\(^1\,\,^3\)

In general, a distinction is made in the literature between on-line and off-line script, depending on the type of information available for the purpose of machine recognition.\(^3\,\,^9\) For instance, dynamic information about the order in which the line pieces making up the characters (strokes) are drawn by the writer is available and exploited in the on-line case, whereas only the static result of the writing process is available in the off-line case. Therefore, it is agreed that the advantage of on-line over off-line systems mostly depends on the possibility to capture temporal or dynamic information. This information or at least part of it would be very useful also in the static case.

In this paper we present and discuss a method which attempts, in the off-line case, to recover part of the lost script dynamics, exploiting only information available at the visual level. For instance, given a character like the simple one in Fig. 1, the problem we want to solve is that of recovering the order in which the strokes making up the character were drawn. In the given example, the most likely order is the one indicated by the sequence of arrows.

The complexity of the task is evident if we consider that the figures we are dealing with are represented by self-intersecting and joining 2D digital ribbons. At each joint a complex decision about trajectory direction is necessary in order to unfold or divide the ribbon.

The basic assumption underlying our method is that subsequently drawn ribbon strokes share more similar features than strokes which, although joining, are not temporally subsequent.

The method we propose has been developed in the framework of a handwritten character recognition system with the purpose of recovering information useful to obtain more effective decompositions and descriptions of characters in terms of meaningful parts. It has been shown to be effective both in the case of isolated handprinted characters and in the case of connected sequences of characters (handwritten words).

Problems arising especially in this second case will be discussed in the following sections.

In the analog plane, a character, as well as a handwritten word, can be seen as a set of complex curves (in particular a single one) and can then be described for instance in terms of a set of functions \(\alpha_i(l)\), where \(\alpha_i\) is the direction of the tangent to the \(i\)th curve at every point, and \(l\) is a curvilinear coordinate along the curve.

The fact that movements made by humans for drawing characters can be described as piecewise continuous trajectories,\(^1\,\,^2\,\,^5\) implies continuity cond-
2. PRELIMINARY RIBBON REPRESENTATION

According to the considerations illustrated above and to previous work on elongated figure description and ribbon modelling,\(^7,8\) we chose the skeleton as the most convenient initial representation of a ribbon. In order to preserve information about ribbon thickness, the skeletonization algorithms computing the Medial Axis Transform (MAT) of a figure\(^9\) must be used. In particular, we have experimented with algorithms that convert an 8-connected figure into an 8-connected digital curve having unit width throughout and outlining a generalized axis of symmetry of the figure (e.g. references \((10, 11))\). This kind of transformation is information-preserving, and hence reversible, since it associates to every skeleton pixel a label specifying the distance of the pixel from the background.

The MAT of characters is a digital curve that is not necessarily simple, i.e., it can be made of parts that join or cross (Fig. 2). In the following, MAT pixels will be termed end points (EP), normal points (NP) or branch points (BP), depending on whether the number of MAT pixels in their \(3 \times 3\) neighborhood is one, two or greater than two.

Unfortunately, thinning algorithms, including medial axis transformation, although appealing techniques for representing ribbon-like figures in a compact way, exhibit a number of unwanted features that require their output to be further processed before it can be used.

In our case, a first problem is that of computing the direction of the digital lines merging at a line joint, so as to evaluate the smoothness of the transition from one line piece to the other. This cannot be reliably done on the digital skeleton looking at a small window around the junction pixels. A solution is to approximate the MAT with a piecewise continuous curve. Note that some sort of approximation is in any case generally necessary in order to use the skeleton of a figure for description and recognition purposes.

A second, more complex problem is that thinning techniques typically give rise to lines whose shape does not always faithfully reflect that of the original figure, even in the case of ribbon-like figures. These shape distortions depend on the limited locality of the criteria according to which skeleton pixels are selected, while they are independent of the metric used in the digital plane for thinning. Since they are mainly located in correspondence to line joints, they can dramatically affect the possibility of reliably applying good continuity criteria. It is absolutely necessary to avoid or correct such distortions in order to obtain significant results, as it can be easily seen from the examples given in the figures of this paper.

As for the first problem, we adopted a polygonal approximation. The procedure reduces the digital line to a polyline made of \(k\) generally connected segments \(E_i, \ i = 1, \ldots, k\), bounded by \(k\) vertices \(V_i\). A great many algorithms have been proposed in the literature for approximating a digital curve with a polygonal
line. The main results of the polygonal approximation relevant to our purposes are:

- the reduction of the noise introduced by both digitization and thinning processes;
- the reduction of the amount of data to be processed;
- the possibility of describing the MAT in terms of angles and sides.

However, in order not to destroy possibly important information about the shape of the curves representing characters, the approximation has to guarantee that significant curve inflections are preserved. To this end, we consider only non-collinear points as candidate vertices\(^{(12)}\) and adopt a low value for the approximation tolerance. This choice does not prevent us from obtaining a significant reduction of the number of skeleton points, but suggests the choice of a merging technique to perform the approximation, since it is less computationally expensive with respect to splitting techniques when the number of sides of the polygon is not particularly small\(^{(12,13)}\).

The approximation of the MAT is performed by following all its branches according to a breadth first strategy, while locating the vertices of the approximating polygonal. Tracing is started from the most left EP of every connected set of pixels or, in the absence of end points, from the top left point. If the most left EP has an \(x\)-distance from the top left NP greater than a threshold depending on the type of script, the latter point is taken as the starting point; this may be the case with some connected sequences of letters. These choices have been suggested from the simple observation that writing using the latin alphabet normally develops from left to right and that characters are most frequently drawn starting from their left side. This behavior is quite general, independent of the type of character (handprinted or cursive written) and of aspects such as culture and nationality. It is clear however that no choice is absolutely unquestionable. Whatever the choice, a number of counterexamples could be found. We made a choice that is in accordance with a simple heuristics and, to our experience, leads to maximize the positive results.

On the other hand, our approach is aimed not to recover the “true” order and direction in which all the strokes making up a character were drawn, but to properly connect or separate contiguous line pieces at junctions. We believe in fact, that recovering the right information about which are the line pieces presumably included between a pen-down and a pen-up, even independently of the direction in which they were traced by the writer, is an already interesting result which can be profitably used in the framework of a handwritten character recognition system\(^{(14,15)}\).

It may be possible to locate with higher reliability the starting point of the line tracing, by making more
complex hypotheses on the handwriting process and assuming suitable information from the gray level image, as very recently proposed in reference (16). The output data structure of the approximation procedure is a directed graph whose nodes are the vertices \( V_i \) of the polygonal and the arcs represent its sides.

The following information is associated to each node:

1. number of the vertex in the tracing order;
2. \( X, Y \) coordinates of the vertex;
3. 8-distance of the vertex from the background (label);
4. number and list of both father nodes and son nodes;
5. number of components the node belongs to.

In addition, the algorithm gives two arrays whose generic elements, length\([i, j]\) and width\([i, j]\), respectively represent the length and the average width of the side between two subsequent vertices \( V_i \) and \( V_j \). In the following, we will call PMAT the polygonally approximated MAT.

Since the decision about dividing or unfolding the ribbon at each branch point is taken by using good continuity criteria, it is important that the PMAT is not affected by significant shape distortions that, by modifying the geometry of the polygonal, would cause a wrong decision to be taken. As already said, thinning techniques typically give rise to distorted representations of ribbon shape. Most typical distortions are:

1. Spurious whiskers generated in the presence of insignificant protrusions, noise or convexities of the ribbon edge.
2. Spurious BPs generated because of the splitting of an ideal single branch point at ribbon crossings. Consequently, a spurious branch not corresponding to an actual ribbon stroke but to a virtual one is generated.
3. Spurious inflections at the junction of strokes.
4. Conjoint effects of types (2) and (3).

Figures 3(a) and (b) show the MAT and the PMAT of a character, respectively. It can be noted that shape distortions are mostly generated in correspondence of branch points, where the continuity criteria have to be checked. It is therefore fundamental to avoid or eliminate distortions for any method based on good continuity criteria to be effective. We have shown in reference (17) how the information held by the labeled MAT is sufficient for locating and correcting such distortions. The algorithms proposed (which will not be discussed here), exploit information about labels of the vertices and relative direction and length of PMAT sides. In this way it is possible to recognize and eliminate spurious whiskers, to merge split BPs, and to correct anomalous inflections of the MAT. Figure 3(c) shows the PMAT of a character after having corrected the shape distortions of types (1)–(3) introduced by thinning.

3. DIVIDING AND UNFOLDING THE RIBBON

Once the corrected PMAT of the ribbon (CPMAT) has been obtained, the segments joining at each BP are considered, in order to group or split them to form sets of connected simple curves. The goal is to reconstruct the stroke sequence of the ribbon according to the
The other two features we adopt to evaluate the degree of similarity between segments joining at a BP are segment length and width. By using these features, for every pair of segments joining at a BP, a good continuity score \( G = A(\alpha) + L(\sigma) + W(\delta) \) is computed, where:

- \( \alpha \) is the angle between the segments;
- \( \sigma \) is the ratio between the length of the shortest and the longest segment of the pair;
- \( \delta \) is the absolute value of the difference between the average widths of the two segments.

The meaning of \( G \) is that the likelihood that two segments belong to the same stroke is considered a linear combination of functions. We included \( L(\sigma) \) among them, because we found experimental evidence that this is one of the criteria used by man, especially in the case of handprintings. Depending on writing characteristics, however, this function may be of minor importance.

The functions \( A, L \) and \( W \) have been empirically evaluated, as illustrated in Section 4, and each of them

![Fig. 4. A handwritten numeral. The results of the various phases of the process are shown. The last picture of the set shows the polygonal subdivided as to represent the characters and the order assigned to the vertices. The first digit of the label assigned to the vertices identifies a component as found by the procedure.](image-url)
Fig. 5. An example of results obtained with handwritten words.
Fig. 6. Some handprinted characters and the final polygonals representing them.

provides as output a value ranging from 0 to 1 which represents the normalized probability that, according to each of the three similarity criteria adopted (i.e. direction, length and width similarity), independently considered, a pair of segments is merged.

The dividing and unfolding procedure is continued for every BP of the CPMAT of a character by taking into account the number k of segments joining at the BP and the value of the score $G$ assigned to each pair of such segments. The two most common cases are $k = 3$ and 4.

For $k = 3$, the pair of segments with the highest score $G$ is kept together and the third segment is cut at the BP. This is compatible with knowledge about the handwriting process and namely with the hypothesis that during writing the number of pen-up and pen-down is minimized.

For $k = 4$, the pair of segments with the highest score $G$ is kept together and either unfolded with respect to the remaining pair or divided from it, according to the case. The latter is also considered united, without further calculations being made for the moment. The rationale behind this behavior is that at the intersection among four segments, two types of decision can be taken: either two segments lying on opposite sides of the junction or two segments on the same side are kept together. In the former case, the most simple hypotheses on the handwriting process lead to the conclusion that we are in the presence of a crossing. In the latter case, exemplified by a $K$-junction, the less compromising decision about the pair of remaining segments, on the basis of the presently available information, is that of keeping them together for the moment and deciding in the following, on the basis of more global information, whether they are actually part of a single stroke or must be split.

Any other solution does not seem in accordance with the possible hypotheses on the writing process. Some examples of characters and short words processed according to our method are shown in Figs 4, 6.

4. EXPERIMENTAL RESULTS

The experimental work has been carried out by using a data set of 10,000 handwritten characters produced by 20 writers. Each writer was requested to fill in a set of forms with predefined square boxes of size $0.25 \times 0.25$ in. The writers were only requested to write characters inside the boxes, without imposing either any character model or constraints on the writing instrument. Upper and lower case letters and numerals were allowed. The character set was uniformly distributed over the various classes and included both handprinted and cursive letters. Each writer was also requested to write a few cursive words, but the results obtained with this material were only used for qualitative evaluations.

Each form was digitized using a flat-bed scanner with a resolution of 300 dpi. The data set was divided into training and test sets, comprising $1/3$ and $2/3$ of the characters, respectively.
In order to quantitatively evaluate the trend of the functions $A$, $L$ and $W$ used to implement good continuity criteria, the training set was considered and six parameters were evaluated for every pair of segments joining at a BP: $x$, $\sigma$ and $\delta$ defined in the previous section, and three boolean variables $v_1(x)$, $v_2(\sigma)$, $v_3(\delta)$ specifying whether, according to the opinion of a human observer, the pair of segments should be split or merged into the same stroke, taking into account $x$, $\sigma$ and $\delta$ independently. The value false was assigned to $v_i$ if the segments were not to be merged together.

The histograms of $A(x)$, $L(\sigma)$ and $W(\delta)$ were then computed. In particular, $x$ was uniformly sampled with a $5^\circ$ resolution in the range $90^\circ$ to $180^\circ$, $\sigma$ spanned in the range 0 to 1 and $\delta$ assumed discrete values from 1 to 5, the latter being the maximum value observed in the whole data set. The function $A$ was defined as

$$A(x) = N_1(x)/[N_1(x) + N_2(x)]$$

where $N_1(x)$ is the number of times $v_1(x)$ was true and $N_2(x)$ the number of times $v_1(x)$ was false.

By using a linear regression, a curve was fitted to these points, obtaining the trend of $A$ shown in Fig. 7. As can be seen, for $x < 110^\circ$ the probability of merging two segments is equal to 0, whilst for $x > 160^\circ$ the probability is 1.

The functions $L$ and $W$ were defined analogously to $A$. In the case of the considered training set, the histogram of $L$ showed a linear trend for $0.2 < \sigma < 0.9$, while $W$ showed a linear trend for $1 < \delta < 5$.

By running the programs on the test set, about 97% of correct results were obtained. The errors were essentially found in correspondence either to shape distortions of the PMAT not properly corrected or to situations where $G$ assumed very similar values for the different pairs of segments joining at the BP.

The computational cost of the whole process, excluding medial axis transformation, ranges from 10 to 20% of the time needed for computing the MAT, depending on the character considered. Some of the partial times are, in fact, more or less independent of the character, but others, like time for correcting distortions, depend on its shape. In any case the average time computed over a set of characters distributed among the various classes with the typical frequency occurring in a text page, was nearer to the lower bound.

5. DISCUSSION

The method illustrated in the previous sections allows the reconstruction of one of the possible trajectories followed by the writer, and thus to recover part of the lost dynamic information, starting from static handwriting. It is mainly based on a shape preserving representation of the lines making up handwritten characters and on the implementation of a set of suitable good continuity rules.

It has to be noted however, that at the connection between subsequent characters in a word, cases may occur where good continuity criteria alone do not always lead to the expected conclusions about the way of coupling strokes at a junction. This happens especially when some line piece is retraced. Better results could only be obtained by incorporating in the system some more knowledge about the handwriting process. The convenience of doing this depends on the characteristics of the recognition system which follows.

As the experiments illustrated in the paper have shown, in the large majority of cases, our method makes it possible to properly connect, at junctions, pieces of curve which have been drawn successively to each other, but does not always guarantee that the
recovered drawing direction is the one actually followed by the writer: it could be the opposite one. This is not a problem if information about drawing direction is only partially exploited by the recognition system, as in the case of the system we are developing. On the other hand, the drawing direction is in many cases writer dependent and, to ascertain it in questionable cases, it would be necessary to search for special clues on the gray level image of the script, as recently suggested in reference (16). Some of these techniques could be integrated in our method, if so desired.

With respect to other approaches using good continuity criteria applied to the skeleton of characters (e.g. reference (18)), the distinctive features of our method are:

1. To outline that the shape distortions introduced by thinning techniques at the intersection among strokes can heavily compromise the possibility of successfully applying good continuity criteria and thus to carry out a technique for correcting such distortions before further processing.

2. To evaluate the smoothness of the digital lines by deriving information from a very wide neighborhood of the skeleton branch points and not simply from their nearest neighbors. This is obtained by polygonally approximating the skeleton and by suitably using information about the direction of the sides of the polygon (see Section 3).

3. To use, among good continuity criteria, not only the information about stroke direction, but also some more information useful to evaluate the homogeneity between contiguous stroke pieces, such as information about their thickness and relative length.

4. To have experimentally evaluated the form of the functions \( A, L \) and \( W \) used to implement good continuity criteria, so as to take advantage of the knowledge of the human behavior.

The proposed method not only allows more reliable decompositions and descriptions of characters but, because of its generality, it can also be useful for processing and interpreting other important ribbon like shapes such as maps and engineering drawings.

6. SUMMARY

In this paper a method for recovering part of the dynamic information in the case of off-line character recognition has been presented. The method makes it possible to reconstruct one of the possible trajectories followed when drawing characters. In fact, from a static point of view, characters and more generally handwritten script appear as a set of self-intersecting and joining 2D ribbons. Therefore, a crucial point in order to accomplish the task, is that of dividing or unfolding the ribbon at every joint.

First of all, the ribbon is substituted by a reliable representation, which we have called CPMAT. On the basis of suitable good continuity criteria, any connected sequence of segments making up the CPMAT is unfolded or divided into components, according to the case. In the former case, every crossing is changed into an "overlap" and the CPMAT is transformed into a simple curve. In the latter case, at a junction where a pen-up (pen-down) reasonably occurred during writing, the CPMAT is "separated" into components. Part of the lost dynamic information is then restored, allowing a more effective description of characters for recognition purposes. Experimental results are reported and discussed.

REFERENCES


About the Author—Giuseppe Boccignone was born on 30 July 1959. He received the laurea degree in physics from the University of Turin, Italy, in 1985. In 1986 he joined Olivetti Corporate Research, Ivrea, Italy. From 1990 to 1992, he was chief researcher of the Computer Vision Lab at C.R.I.A.I., Naples, Italy. He is currently at Research Labs of Bull HN Information Systems, Milano, Italy, leading projects on biomedical images. Since 1987 his main research interests have been in the field of medical image processing, optical character recognition, texture analysis and shape description.

About the Author—Angelo Chianese was born in Naples, Italy, in 1954. He graduated in electronic engineering in 1980 at University of Naples and is presently Associate Professor of Computer Science at the same University. He has worked on computer networks, parallel architectures, computer vision, optical character recognition and more recently, concurrent object oriented programming.

About the Author—Luigi P. Cordella was born in Milano, Italy, on 26 November 1938. He is Professor of Computer Science at the Faculty of Engineering of the University of Naples, Italy. He has been active in the field of image processing and pattern recognition since 1972. His present research interests are in the field of optical character recognition, shape analysis, document and map processing, parallel computer architectures for vision.

About the Author—Angelo Marcelli was born on 20 March 1957. He received the laurea degree in electronic engineering in 1983 and the Ph.D. in electronics and computer engineering in 1987, both from University of Naples. Currently he is Associate Researcher at the Faculty of Engineering of the University of Naples, Italy. His research interests are in optical character recognition, texture analysis, shape description and expert systems.