A Multiresolution Approach to On-line Handwriting Segmentation and Feature Extraction

Claudio De Stefano°, Marco Garruto* and Angelo Marcelli*
°DAEIIMI – Universita’ di Cassino, ITALY
*DIIE – Universita’ di Salerno, ITALY
{destefano@unicas.it;marcounisa@libero.it;amarcelli@unisa.it}

Abstract

We present an approach to on-line handwriting segmentation into elementary strokes. The segmentation is achieved by exploiting curvature information extracted from the electronic ink at different level of resolution. Such information is then combined into a saliency map, through which the segmentation points are eventually found. Information from the saliency map is also used to select the optimal resolution to be used for describing the curvature of each stroke. Experiments conducted by encoding stroke curvature information into strings have shown that similar strings are associated with similar parts of the original ink, independently of the actual word they belong to.

1. Introduction

Many studies on handwriting generation have shown that complex movements like handwriting can be seen as a composition of elementary movements, or strokes, each corresponding to an elementary shape [1-3]. According to this approach, thus, handwriting generation is the result of a complex motor program that generates the appropriate sequence of strokes needed to draw the sequence of elementary shapes forming the handwriting. Thus, handwriting recognition may be seen as a bottom-up process that extracts the strokes from the ink, encodes their features and eventually performs the classification by comparing the description of the specimen with those of a set of allographs.

The main problem to deal with while pursuing such an approach is that of variability, which emerges from four different factors: posture, neuro-biomechanical noise, style and sequencing. Posture refers to changes in size, position, orientation and slant of the handwriting that mainly depends and the postural condition of the writer. Neuro-biomechanical factors greatly affect the quality of handwriting by modifying both the motor control program and the production of individual strokes. As a matter of fact, fluency in handwriting emerges from the time superimposition of strokes due to anticipatory effects, so that the actual trajectory for drawing a stroke depends on both the previous and the successive ones [3]. Style refers to the various models that are associated with a single character by different writers. Finally, sequencing refers to the variation in the order of individual strokes during handwriting. To deal with such variability, several approaches exploiting either temporal or shape information, have been proposed to locate and represent the individual strokes [4]. All of them, though, use some thresholding technique to decide whether or not a change in the observed signal (speed or curvature, mainly) corresponds to a segmentation point. Thus, setting the value for the threshold represents the major challenge for those methods, and seems to be the main reason of the erratic behaviour they sometimes exhibit.

We propose to address the segmentation problem by mimicking the model proposed in [5,6] to describe the early stages of visual processing. According to such a model, the visual information registered at any given time in a scene is processed at decreasing level of resolution and the results propagated along a pyramidal structure. The number of levels along which a given processing propagates its effects represent a measure of the relevance, or saliency, of the processed visual information. The saliency map is then obtained by associating to each visual element in the scene its saliency, so as the highest values in the map “point to” the most relevant elements in the scene. Many experiments have shown that this model fits more than satisfactorily when compared with human eye-movements triggered only by preattentive stimuli, while top-down, model-driven processes are in place during attentive vision tasks [7]. Following this approach, therefore, the segmentation problem can be reformulated as an early, preattentive scene analysis problem. The scene the system is looking at is the electronic ink, and the information we extract from the ink, of which we want to estimate the saliency at different level of resolution, is the
curvature. Thus, the segmentation points should correspond to the highest values of the saliency map. By analogy with the experimental results on preattentive visual tasks, the obtained segmentation should be much more invariant with respect to non-significant shape variations or changes in the writing speed.

This note presents a segmentation algorithm we have developed by following this paradigm. The method is based upon a frequency analysis of the discrete-time sequences \( x(n) \) and \( y(n) \), representing the \((x,y)\) coordinates of the points collected by the input device, respectively. The analysis is performed on a multi-scale representation of the original curve, obtained by filtering the Discrete Fourier Transform (DFT) of the sequences \( x(n) \) and \( y(n) \). After the application of the filter, new sequences \( x'(n) \) and \( y'(n) \) containing a smaller number of points are obtained by applying the Inverse Discrete Fourier Transform (IDFT).

In practice, at each scale, we obtain a filtered version of the original curve containing a smaller number of points: this multi-scale representation is then used to build a saliency map, that is a suitable data structure able to highlight the points of the original curve in which significant curvature variations are recorded at different scale. These points represent the desired segmentation points. The saliency of the segmentation point is also used to select among the different representations of the ink, the one to be adopted for describing the shape of the obtained strokes in terms of their curvature. A simple encoding of such feature is then proposed to show the invariance of the obtained decomposition.

The remaining of the paper is organized as follows: Section 2 describes the segmentation method, while Section 3 illustrates the selection of the representation and the feature extraction. Experimental results and some concluding remarks are eventually left to Section 4.

2. Stroke segmentation

The input provided by the tablet is a sequence of points in the \((x,y)\) plane corresponding to the uniform time sampling of the ink produced by the writer. These points are not uniformly distributed along the line but spatially concentrated in correspondence of minima of writing speed. For this reason, in the first step of our method, we preprocess the original set of points by adding new points in such a way to obtain a 8-connected line in the \((x,y)\) plane. In this way, we get rid of the information relative to the dynamics of the handwriting generation process, that, as already noted, significantly varies among writers. Thus, the remaining of the process relies upon static information carried by the shape of the ink, under the assumption that it exhibits a much smaller amount of variations in comparison with the dynamic one. In the sequel, we will denote by \( N \) the number of the points provided by the tablet, by \( M \) the number of points obtained at the end of the first step, and by \( x(n) \) and \( y(n) \) \((n=1..M)\) the coordinates on the \((x,y)\) plane of the points belonging to the line \( A \) obtained after the first step. Note that \( M >> N \), where the magnitude of the inequality depends on the resolution of the tablet and on the writing speed.

The aim of the second step of our method is that of performing a spatial frequency analysis of \( A \) in order to obtain a multi-scale representation of the original curve. To this purpose the DFTs \( X(K) \) and \( Y(K) \) \((K=1..M)\) of the sequences \( x(n) \) and \( y(n) \) are computed. At each scale, the desired representation of the original curve is obtained by applying the IDFTs to the first \( T \) elements of the sequences \( X(K) \) and \( Y(K) \): the smaller the value of \( T \), the coarser the approximations of \( A \), while the opposite is true for values of \( T \) close to \( M \). We have chosen values of \( T \) in the range \((3,N)\): this choice seems reasonable because \( N \) is the original number of points provided by the tablet and 3 is the minimum number of points to describe a curve. At the end of this step we obtain different representations \( A_i \) \((i=1..N-2)\) of the original curve containing a number of points ranging from 3 to \( N \).

These representations are used in the third step of our method to build a saliency map, that in our case assumes the form of an histogram whose peaks identify the regions of the original curve in which curvature changes occur at different scales: these are the regions in which most likely a decomposition point should be placed. In this step, the points belonging to each representation \( A_i \), are processed in the following way:

a) an arclength representation is computed: this representation is a function \( \alpha(\lambda) \) where \( \lambda \) is the curvilinear abscissa of a point, and \( \alpha \) is the angle formed by the two segments connecting that point with the previous and the successive one, respectively. Note that the range of values for the curvilinear abscissa has been normalized with respect to the length of the curve, so that \( \lambda \) ranges between 0 and 1, and partitioned in \( N \) intervals of the same size. In this way, independently on the actual length of each curve \( A_i \), the arclength representations have the same scale on the \( x \)-axis.

b) the representation is parsed and the local maxima are recorded.

Finally, an histogram reporting the average values \( <\alpha(\lambda)> \) of the local maxima are computed. The histogram is obtained by averaging the values \( \alpha(\lambda) \) of the local maxima of the curves \( A_i \) \((i=1..N-2)\). These values measure the saliency of each region. It is worth noticing that the size of such regions depends on the value of \( N \), i.e. from the tablet resolution: the greater the value of \( N \), the smaller the region, but obviously the higher the computational cost of
the method. The desired segmentation points are then located in the regions corresponding to values of the histogram bigger than the its average value.

Figure 1a-d illustrates the results of the segmentation. The algorithm for selecting the most suitable representation for the curve and the following feature extraction are described in the following Section.

3. Stroke description

As mentioned before, we want to describe the shape of each stroke by means of its curvature. The segmentation algorithm, on the other hand, does not provide segmentation points, but rather segmentation intervals, i.e. regions of the ink in which the segmentation points should lie. Since we have many representations of the ink that contribute to those intervals, we have to select one of them for describing the shape of the strokes. Intuitively, we would like to select the representation corresponding to the smallest resolution at which the less salient segmentation points are still detectable: selecting a representation corresponding to a higher resolution would be too coarse, and therefore would possibly hide some relevant features, while one corresponding to a lower resolution would be too fine, and therefore too sensitive to non-salient changes in the shape.

The algorithm for selecting the most suitable representation exploits once again the saliency map, but in a different context. In particular, for each representation \( \alpha_i \), it computes the distance between the vector \( \langle a(\lambda) \rangle \) and \( \langle a(\lambda) \rangle \), i.e. the difference between the curvature observed at that scale and the saliency map. According to the model, such difference should be very high in correspondence of the lowest resolutions, get smaller as far as the resolution approaches the “right” one and then increase again as the resolution becomes too high. Therefore, to select the most suitable representation we find the best fit of the distances with a parabola and select the scale corresponding to the vertex of the parabola. Fig. 1e shows the histogram of the distances, the best fitting parabola, and the selected resolution, at scale 83. Once this resolution has been selected, the shape of the stroke is described by computing the change of the curvature along the curve at that resolution. Those changes are then encoded by quantizing the angle into 16 different directions, as shown in fig. 1f. The results show that similar part of the ink, such as the two downwarding strokes of the character “u” are encoded into similar strings, namely MMNO and MMNP. Similarly, the ascenders of the character “b” (KMMMMMN) and “l” (KL MMMMO).

4. Experiments and concluding remarks

We have run two set of experiments to validate the results of the proposed method. The first one has been carried out by using a set of almost 1,000 words produced by the same writer provided by the Handwriting Recognition group at IBM T.J. Watson Research Center, the second one by using 2,000 words produced by 10 different writers, collected in our lab by using a Wacom PL 100V tablet with a cordless stylus and a sampling rate of 100 hz. Each writer was required to drawn 10 times a set of 20 words without any specific instruction or model to adhere.

Before describing our results, it is worth noticing that performance evaluation of a segmentation method is extremely difficult because the "true" segmentation is generally not available. As a consequence, we have set up an experimental procedure, which is articulated in two phases: segmentation and performance evaluation. In the segmentation phase, each word in the data set decomposed by using our method. During the performance evaluation phase, the segmentation points were compared with those manually entered by a human expert. It was assumed that a segmentation point was correctly located if it was close to where the human expert believed there was a junction between strokes. Thus, the performance of the segmentation can be expressed as the percentage of words correctly decomposed over the total number of words. The experiments on the IBM database have shown that the suggested method produces correct decomposition in 99.53% of the cases. As with regards to the robustness of the feature extraction, once the segmentation was validated, the strings encoding the features extracted from each stroke were recorded and compared. Two strings were considered as the same if there were at most two different symbols for each stroke. This choice for the string matching follows from the observation that the anticipatory effects mainly influence the beginning and the end of the strokes and therefore the first and the last symbols of each string. The experiments on our database have shown that only on 56 words the method provided different string, and that this in 53 words this is due to different sequence of strokes used by the writer to write the same word. In conclusion, the experiments conducted till now, despite the simplicity of the feature extraction and the string matching algorithm, confirm that the proposed method provides very stable and consistent...
Figure 1. The proposed method. a) The original input. b) A graphical rendering of the multiscale representation: at each scale, the darkest/brightest points correspond to most salient curvature changes. c) The saliency map: the bold line represents the value used for thresholding the histogram and selecting the segmentation points. d) The obtained segmentation. e) The histogram of the differences between the curvature at each scale and the saliency map and the best fitting parabola. The selected resolution correspond to the vertex of such a parabola. f) The string encoding the change of curvature along the curve at the selected resolution.

segmentation results, and therefore seems a promising way to extract from handwriting basic, writer-specific writing units.

Further development will address the problem of using such basic writing units to perform the recognition of any words that can be produced by means of such writing units, thus encompassing the disadvantages of both holistic and analytical methods for cursive handwriting recognition. As with respect to holistic methods, it should be possible to recognize any word composed of means of such writing units, not only the ones belonging to a given, size limited dictionary. Similarly, analytical methods may benefit as well from the segmentation and description of the individual strokes provided by our method, in that it should not be necessary to segment the word into characters, since, as suggested by psychological experiments in human character recognition, the basic writing units constituting the graphic alphabet of a writer may correspond to groups of letters that are produced as single, specific elementary movements.

References


