Influence of Zoning on Whole Word Recognition

Nasser Sherkat
Intelligent Recognition Systems, School of Computing and Informatics
The Nottingham Trent University, Burton Street, Nottingham, NG1 4BU, UK
Nasser.Sherkat@ntu.ac.uk

Abstract

This paper presents the investigations carried out in improving the performance of a holistic cursive word recogniser using approximating splines. The holistic cursive word recogniser and the spline based zone line extraction mechanism are described. A series of tests are carried out using 3600 word images from 18 writes and a 200 word lexicon to compare the performance of the histogram based zone extraction with that of the spline based approach and the results are presented. It is found that the spline based zone extraction significantly improves the recognition results. The future directions of the research for further improvement of recognition performance are presented.

1. Introduction

When the original document images are of poor quality or handwriting style is such that segmentation would be particularly prone to errors holistic word recognition provides a number of advantages over a segmentation based approach. One such recogniser has been developed and presented previously [7, 8]. The holistic recognizer is based on three landmark features namely, holes, vertical bars and cups (HVBC). The recognition algorithms are driven by a lexicon, which can be automatically constructed from an ASCII word list. The effective extraction of these features heavily relies on the ability of correctly identifying the various zones of a given word. The previous reporting of the HVBC recogniser relied on a histogram based zoning algorithm. Although reasonable results could be obtained from that method, it was acknowledged that the histogram method lacked sensitivity as it assumed a straight line baseline. Work in parallel has been carried out to investigate the means of developing a more sensitive and realistic handwriting baseline extraction technique. As a result a spline based zone extraction mechanism has been developed. This mechanism does not assume a straight base line and has been shown to provide a very effective means of detecting the different zones of handwriting. Hitherto the HVBC and the zoning extraction method had not been combined. This paper presents the results, for the first time, of experiments carried out to assess the effect of spline based zoning on the recognition results. The paper is structured as follows. An overview of the two main parts namely the HVBC recogniser and the zoning mechanism are presented. The main features of each part are highlighted before proceeding to experiments and the results.

2. The holistic Recognizer

The system has been designed to recognize unconstrained, lower-case Roman script from any writer. The recognizer accepts off-line (static) word image data as input. The input image is in facsimile format with a resolution of 200x100 dpi. It is assumed that all word images are bi-level. In order to operate, the recognizer must first be supplied with a template database containing a list of words together with the corresponding sets of vertical bars, loops and cups that the words are expected to contain. Vertical bars, loops and cups were selected for recognition purposes. This selection is based on previous research in recognition of on-line cursive handwriting [7]. These features, taken together, are discriminative in the sense that if all of the required bars, loops and cups can be extracted from a word image, then that word image can usually be correctly recognized. This is also likely to be the case in systems with large template databases, where the possibility of confusion between similar words is high. Also, it would be expected that the features are relatively independent, i.e. using all three features together should provide more discriminatory information than using any single feature or pair of features. The extraction of these features is also relatively straightforward and computationally inexpensive. In addition, the features have formed part of several effective recognition systems developed by other researchers. For instance, Hull et al. [6] describes a segmentation based recognizer that uses horizontal strokes, concavities and loops as discriminatory features, while Gorsky [1] describes a holistic recognizer that employs line segments, ascenders, descenders and loops.

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE

2.1 Vertical bar extraction

Vertical bars are identified by examining the number of black pixels per unit area in each of the three zones of a word image. The number of black pixels per unit area is referred to as the pixel density. The horizontal positions within each zone at which the pixel density reaches a maximum value are taken as likely candidates for the position of vertical bar sections. When these bar sections have been identified in all three zones, bar sections in different zones that are close to each other are joined up to form full length bars. Bar sections are extracted from all three zones, rather than from the word image as a whole, in order to minimise the effect of word slant. The steps for extraction of vertical bars are presented in [8].

2.2 Loop extraction

A number of problems can occur when trying to extract basic loops from a word image. First, non-basic loops are very common, and may be difficult to distinguish from basic loops. Non-basic loops can be caused inadvertently by, for example, t-bars intersecting with main letter strokes, or can be a deliberate mannerism of a particular writer. For example, some writers habitually write certain letters, such as \`f', \`j, \`k', \`q' or \`z', with looped ascenders or descenders, while it is also common for vertical strokes, such as those that occur in the letters \`l' and \`i', to be written with a loop. Secondly, loops in handwritten text may have small gaps in them. In this paper, loops that are entirely surrounded by black pixels is referred to as closed loops, while loops that are almost entirely surrounded by black pixels, i.e. loops with small gaps, is referred to as open loops. Open loops can make recognition problematic. For example, it may be difficult to distinguish between a loop and a cup. Also, letters written near to each other can produce unexpected open loops. For further details of how open and closed loops are dealt with please refer to [8].

2.3 Cup Extraction

Extracting basic cups from a word image is a difficult problem, as non-basic cups are common and may be difficult to distinguish from basic cups. The strategy used for dealing with this problem is the same as the one used for loop extraction, i.e. initially all cups are found and then a number of heuristics are applied in an attempt to eliminate the non-basic ones. Experimentation has shown that this approach is effective, except in the case of cups in ligatures. These are non-basic, as ligatures do not appear in database templates. Also, they can be difficult to distinguish from basic cups, and hence it is not straightforward to eliminate them using the heuristics mentioned above. For further details please refer to [8].

3. Word Template Database

In this system, the database template for a word is constructed by concatenating templates representing the word's constituent letters. The template for a letter is derived by applying the feature extraction procedures, described above, to a representative set of letter images taken from the development set. This allows the most common features that occur in each letter to be determined. Some letters, most notably \`f', \`z' and \`s', can assume several common forms, or allographs, in handwritten text. Currently, various ad-hoc methods, which are described further below, are used to construct and represent the templates for these letters (see figure 1).

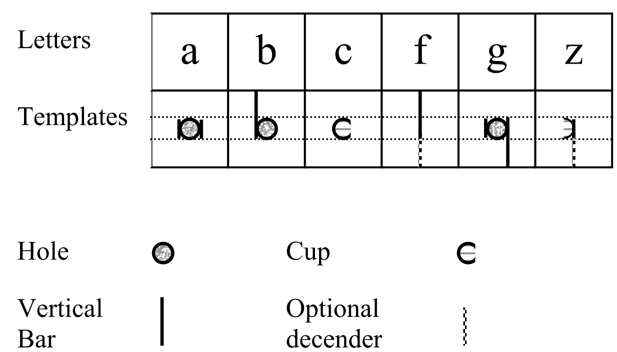

Fig. 1: Sample letter templates

If a feature is included in a letter template, then its properties, such as size and position, are chosen so that they are an average of the properties of all instances of the letter in the reference data set. For instance, the relative width of a letter is calculated by averaging the relative widths of all instances of the letter in the data set. All features are assumed to start and end vertically precisely at zoning lines. For instance, all descenders representing vertical bars are assumed to extend from the bottom line to the upper base-line. Also, the vertical bars in the letters \`f' and \`z' are marked as having optional descenders. This means that these bars can be considered as mid-zone bars or as descenders. It was necessary to adopt this representation because handwritten \`f's and \`z's commonly occur as allographs with and without descenders. The template for the letter \`s' was difficult to choose, as \`s's in handwritten words have several commonly occurring allographs. A representation consisting of a single loop with mid-zone bars on either side was eventually chosen, as this seemed to occur the most often in practice.

1520-5263/05 $20.00 © 2005 IEEE

4. Feature Set Matching

In order to recognize a word image, the list of features extracted from the image must be compared with all of the feature sets stored in the word template database. For each database entry, a score is required that numerically assesses the similarity between the entry and the set of features extracted from the word image. Similarity scores are calculated so that the higher the score, the greater the degree of similarity. In our system, the normalised string edit distance [2], a variant of the string edit distance [5], is used to calculate the required numerical measure of similarity.

5. Zoning Methodology

Line extraction and zone estimation has been subject of much direct or indirect (as part of a larger system) research. A comprehensive review of the state of the art with respect to this field has been presented in [9]. The script on a page is segmented at local horizontal extrema into small parts, the 'ink-boxes' representing the connectivity information. Each ink-box is attributed by its bounding box, the amount of ink covered by the box as well as a representing point and a tolerance measure. The representing point is chosen with respect to the desired zone-line, e.g. the bottom of the box during estimation of the baseline. Contextual information provided by overall statistics and previously obtained guidelines is used to adjust both representative point and tolerance. In order to eliminate undue oscillation of the approximating cubic spline function [9] (denoted by g(x); spline for short), the spline's control points are obtained by resampling the representative points and their associated tolerances.

The resampling ratio and the tolerances determine the overall smoothness of the resulting spline. The reach of the triangular weighting function used for weighted averaging during resampling models the distance of influence of individual ink-boxes. Boxes representing the relatively rare descenders, for example, define the height of the lower zone over a wider distance than the many boxes representing middle zone segments. The estimate of the bottom line should consequently employ a larger distance of influence. As the bottom line is also expected to be smoother than the half-line, the resampling ratio should be smaller. Resampling also compensates for an overall upwards or downwards trend of the script by projecting the ordinate of the representative points along the assumed axis of the script. Initially, the linear regression of the ink-boxes is employed as an estimate of the axis. The centre-line (i.e. the zone-line representing the centre of the middle zone) is then estimated and used as a more accurate description of the axis (see section 5.1.2 for details).

5.1.1 Line-Segmentation

From the list of all ink-boxes, the line-segmentation step attempts to coarsely identify the majority of boxes forming the topmost line of text. Based on these topmost boxes, the zone-lines are estimated. As the line of text is then encompassed by the top and bottom zone-line, ink-boxes belonging to the topmost text-line can be more accurately identified and removed from the set of ink-boxes. The method continues to identify, approximate and remove the next-highest lines of text until no more boxes remain.

The simplest way to segment pages into lines of text is horizontal projection of the ink into a (vertical) histogram (Fig. 2), in which gaps indicate line-breaks. This method, however, depends on horizontal text. Even if skewed text is identified and corrected beforehand, the lines must be clearly spaced, as the histogram might otherwise not contain the required gaps. Additional spurious gaps might appear if diacritical marks are written far above the rest of the text-line. If the text is enclosed in boxes, the histogram might not contain any gaps at all, rendering segmentation impossible.

The 'Vertical Descent' segmentation has therefore been developed, which assumes that lines are arranged relatively smoothly and near-horizontally, and that a text-line 'hides' most of subsequent lines if one descends vertically. Only the few segments 'seen' from above are considered for the topmost text-line (Fig. 2). This set can contain spurious boxes ('seen' through the gaps in the topmost text-line or 'around' the line) and usually

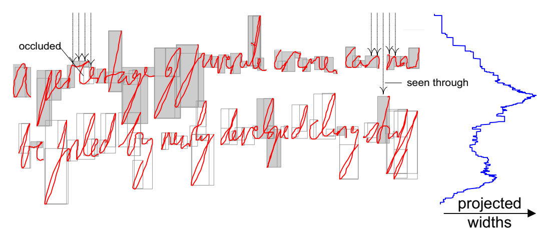

Fig. 2: 'Vertical Descend' marks ink-boxes of the top line.

does not include all boxes that actually belong to the line of text (the 'occluded' boxes). The approximating spline function based on the initially selected boxes, however, is precise enough to serve as a preliminary indication of the text's location. Spurious boxes far below the topmost line can thus be identified and ignored and the guideline updated accordingly. The set is then extended by boxes near to the new guideline, thus including originally occluded boxes. The guideline is then made to follow the boxes more closely by increasing the resampling rate, which allows identification of discontinuities. Discontinuities usually occur if the set

1520-5263/05 $20.00 © 2005 IEEE

describes two different lines of text with large gaps in the higher line. The guideline might then alternate between the higher line and inside the gaps the lower line. Detected discontinuities are thus used to restrict the set to ink-boxes of the higher line and the guideline is updated accordingly. The set is again amended by boxes surrounding the guideline in order to contain the majority of boxes that form the line of text.

5.1.2 Zone-line Estimation

Individual zone-lines are estimated based on the set of boxes. First the centre-line is established and used as the text axis during resampling of the remaining splines. The centre line is expected to be best described by the centroid of boxes forming the middle zone. Centroids of Ascenders and descenders would unduly influence the spline and should be discarded. They could be identified by contextual information such as the average height of all boxes in the line of text or their position in relation to the text. The average height is easily obtainable from the set of boxes and is used to modify the tolerances assigned to each representative point. Increasing the tolerance allows modelling the certainty to which a box is not an ascender (or descender) and thus the degree to which it is "ignored". Avoiding a crisp decision also limits the effect of incorrectly identified ascender/descender-boxes, which might contain useful information about the centre-line.

Contextual size-information, however, is unable to distinguish between middle-zone boxes and diacritical marks or 'open-ended loops' attached to an ascender/descender (Fig. 3A). This often leads to a distorted first guideline g1, which can nevertheless provide the contextual position-information needed to identify and eliminate the affected boxes for calculation of g2. In some cases, the elimination of outliers described above is affected by errors propagated from a heavily distorted first guideline (Fig. 3B). The second guideline is therefore used to restrict the estimation of the final centre-line to middle-zone boxes only. Various strategies could be adopted for the extraction of the remaining zone-lines. The half-line for example, could be extracted first. Ink-boxes that extend above the half-line could then be identified as ascenders whose upper borders define the top-line in turn. An error in the half-line, however, might affect the top-line as well. In order to minimise such dependencies and error-propagation all zone-lines and their intermediate guidelines are calculated independent of other zone-lines, referring only to the centre-line during resampling. Their estimation follows similar reasoning and utilises the same contextual information as demonstrated for the centre line. Two to three intermediate guidelines are used per zone-line (see [9]) for details).

The obtained zone-lines exhibit the following properties. As the splines are approximating the representative points, the zone-line does usually not reach all extrema found in the script if they are of inconsistent height (e.g. the tops of 'f' and 'l' in). This compromise can not only deal with contradicting ascenders/descenders but also restricts the impact of potentially misclassified boxes. If no boxes can be found within the distance of influence, the outer zone-lines

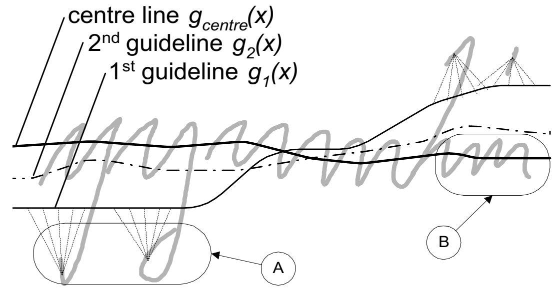

Fig. 3: Estimation of the Centre-line via guidelines

approach the respective inner lines. While the outer lines could be kept at a minimum distance from the middle zone, this would require an empirically obtained average height of the outer zones. The heuristics, however, might not correspond to the script at hand. Using empirical heights would also fail to indicate the lack of precise information, which could be used to adjust the method classifying points.

6. Results and Discussion

The tests originally carried out using the histogram zoning [8] were repeated with the zoning replace by the spline based zoning. Again this time the base word set was used to construct the 200 entry database. A summary of the results is presented in Table 1.

Writer	Slant Degrees	Top (%)	Top (%) Sp	Top 5 (%)	Top 5 (%) Sp
1	5 right	58.29	67.37	74.87	78.23
2	10 right	78.00	82.09	90.00	93.45
3	20 right	40.00	68.67	63.00	76.04
4	0	54.65	72.05	85.35	89.00
5	0	66.00	83.02	84.50	87.98
6	10 left	67.35	74.78	87.24	91.23
7	10 right	42.13	58.24	72.59	75.34
8	5 left	85.43	88.87	93.07	98.87
9	0	70.35	77.87	90.95	95.23
10	5 right	51.76	69.34	78.39	85.67
Overall	-	62.41	74.23	82.08	87.10

Table 1 Summary results for 200 words(Sp=Spline)

When used with a template database containing 200 entries, as expected the influence of spline zoning is marked. The overall top rank and top 5 ranks have increased by 11.99% and 5.02% respectively. As depicted in the table recognition results are less significantly influenced by the slant of writing in the case of spline based zoning. In order to maintain the same parameters no correction was made for slanted writing during the detection of vertical bars. However, it is apparent that the spline based zoning is less susceptible to slants. This is deemed to be due to the fact that the histogram zoning would have difficulty in correctly zoning a slanted vertical bar due to the reduced effective height in its projection. Furthermore, the spline zoning benefits from the 'sentence view' when determining zoning for a word rather than the localised and limited view available from a histogram. It must be stressed that due to the significant variation of unconstrained handwriting baseline a sentence view based on histograms is not practical.

Further template databases containing 400, 600, 800 and 1000 word entries were created. These databases were constructed by selecting four additional disjoint sets of 200 words from a 15,000 word dictionary. The additional sets were randomly selected. The test results are presented in table 2.

Data base size	Top (%)	Top (%) Sp	In top 5 (%)	In top 5 (%) Sp	In top 50 (%)	In top 50 (%) Sp
200	62.41	74.23	82.08	87.10	95.57	97.90
400	57.17	65.67	77.15	81.56	93.46	94.56
600	53.90	58.86	74.08	78.86	91.70	93.87
800	51.38	55.67	68.42	72.49	90.59	91.72
1000	49.47	52.61	64.71	68.56	89.28	90.04

Table 2 The effect of database size on HVBC with and without spline zoning

In both cases the recognition rates deteriorate markedly with increase in the database size. This is not far from expectation since an increase in the number of words means that there is a higher likelihood that words with similar shapes are present. This is an inherent feature of this holistic recogniser. However as can be seen, in general, the use of spline zoning yields a better rate of recognition though this influence fades as the lexicon size increases (see fig 4). So far the flow if information has been top-down in the sense that zoning information has guided the classification of zoning for the various extracted features. Work is currently ongoing to investigate the possibility of refining the zoning through the findings of the feature extraction and recognition phases.

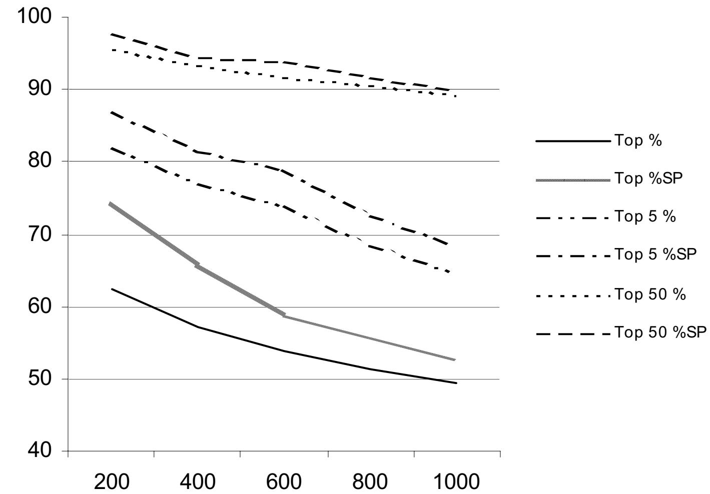

Fig. 4: Comparison of histogram and spline zoning as the lexicon grows.

7. References

[1] N. D. Gorsky (1994) Off-line recognition of bad quality handwritten words using prototypes. In S. Impedevo (Ed.), Fundamentals in Handwriting Recognition, NATO ASI Series F, 124, (pp. 199-217). Berlin: Springer-Verlag.

[2] A. Marzal and E. Vidal (1993) Computation of normalised edit distance and applications. IEEE Trans. PAMI, Vol. 15, no. 9, September 1993.

[3] G. Seni, V. Kripasundar and R.K. Srihari (1995) Generalizing edit distance for handwritten text recognition. Proceedings of SPIE - The International Society for Optical Engineering, 1995, Vol 2422, pp. 54-65, 1995. ISBN 0 8194 1769 6

[4] A. W. Senior (1994) Off-line cursive handwriting recognition using recurrent neural networks. PhD thesis, Trinity Hall, Cambridge University, Cambridge, UK, 1994.

[5] R. A. Wagner and M.J. Fischer (1974) The string-to-string correction problem. Journal ACM, Vol. 21, No. 1, Januray 1974, pp. 168-173.

[6] J. J. Hull, T.K. Ho, J. Favata, V. Govindaraju and S. Srihari (1992) Combination of segmentation- based and wholistic handwritten word recognition algorithms. S. Impedovo and J.C Simon (Eds.), From Pixels to Features III: Frontiers in Handwriting Recognition (pp. 261-272). Holland: Elsevier Science.

[7] R.K. Powalka, N. Sherkat, R.J. Whitrow, Word Shape Analysis for a Hybrid Recognition System, International Journal of Pattern Recognition, Vol.30, No.3, pp.421-445, 1997

[8] N. Sherkat, R. J. Whitrow, Whole Word Recognition in Facsimile Images, IEEE proceedings of Fifth International Conference on Document Analysis and Recognition ICDAR'99, pp 547-550, India, September 1999, ISBN 0-7695-0318

[9] A. Hennig,, N. Sherkat, Exploiting Zoning based on Approximating Splines in Cursive Script Recognition Journal of Pattern Recognition, Vol. 35, pp 445-454, 2002

Proceedings of the 2005 Eight International Conference on Document Analysis and Recognition (ICDAR'05)
1520-5263/05 $20.00 © 2005 IEEE