Pattern Recognition. Vol. 19, No. 1, pp. 1-12, 1986.
Printed in Great Britain.

0031-3203/86 \$3.00 + .00
Pergamon Press Ltd.

RECOGNITION OF ISOLATED AND SIMPLY CONNECTED HANDWRITTEN NUMERALS

M. SHRIDHAR and A. BADRELDIN
Department of Electrical Engineering, University of Windsor, Windsor, Ontario, Canada N9B 3P4

(*Received 28 November 1984; received for publication 22 May 1985*)

Abstract—In this paper the authors describe the results of their investigation into the development of a recognition algorithm for identifying numerals that may be isolated or connected, broken or continuous. Using a structural classification scheme, the recognition algorithm is derived as a tree classifier. In an extensive test experiment, an accuracy of 99% was realized with isolated numerals. When connected numerals were also included a recognition accuracy of 93% was obtained.

Handwritten numerals   Broken numerals   Connected numeral strings   Topological features
Segmentation algorithm   Tree classifier

1. INTRODUCTION

The recognition of handwritten numerals continues to be a challenging problem, since any practical recognition system requires: (a) that the algorithms be writer-independent; (b) the algorithms be robust and yield high accuracies ($\simeq 99\%$); (c) the algorithms successfully segment connected numerals into their individual units and recognize them.

A survey of existing literature reveals a significant amount of research to deal with the first two requirements outlined above.[1] A number of features both in the topological or spatial domain, as well as in the frequency domain, have been proposed[2-11] for uniquely characterizing the numerals.

Also, a variety of classifiers, such as distance, tree, structural and syntactic classifiers, have been proposed for developing a decision strategy that used the features of the numerals to yield a recognition.[12-15]

However, most of these algorithms assume that the handwritten numerals are continuous, isolated and completely described by their boundaries. Such assumptions are highly restrictive, since most of the handwritten numerals encountered in the real world are often broken, and further, two adjacent numerals are often connected.

In this paper the authors describe the results of their investigation into the development of a recognition algorithm that can deal with isolated broken numerals as well as connected strings of numerals.

In the first phase of this study the authors describe a set of topological features to uniquely characterize isolated numerals. It is assumed that the numerals may be continuous or broken. The second phase of this deals with the segmentation of connecting strings of numerals into their individual units. The segmentation technique utilizes a hierarchical approach.

2.1. FEATURE EXTRACTION

All of the features associated with the numerals are derived from the left and right profiles of their external contours. These profiles may be determined from the stored boundaries of the digitized image or directly from the thresholded image through scanning operations. Figure 1 displays a typical set of profiles $\{LP(k), RP(k); k = 1, 2, \ldots, NH\}$ for the numeral 5, where $LP(k)$ and $RP(k)$ are the left and right profiles of the numeral, and NH is the height of the character. From these profiles a set of features are defined for a unique characterization of the various numerals. These are as follows.

(1) *Character widths.* The character width at any location k is defined as

$$W(k) = RP(k) - LP(k).$$

(2) *Ratio.* The ratio is defined as

$$R = NH/W_{\mathrm{MAX}}$$

where NH is the height of the character and W_{MAX} is the maximum width and is given by

$$W_{\mathrm{MAX}} = \max_k (W(k)).$$

(3) *Location of extrema.* The following locations are defined on the left and right profiles:

$$L_{\mathrm{MAX}} = \text{location of } \max_{k \in R_1} \{LP(k)\};$$

$$R_{\mathrm{MAX}} = \text{location of } \max_{k \in R_1} \{RP(k)\};$$

$$L_{\mathrm{MIN}} = \text{location of } \min_{k \in R_1} \{LP(k)\};$$

$$R_{\mathrm{MIN}} = \text{location of } \min_{k \in R_1} \{RP(k)\};$$

where R_1 is the range of k in which the extrema are sought.

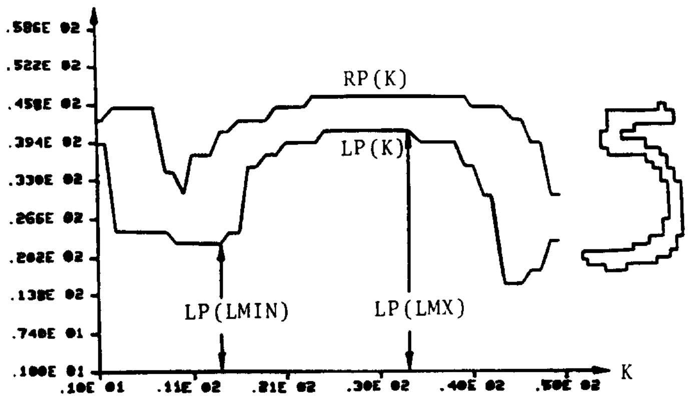

Fig. 1. Left and right profiles of the numeral 5.

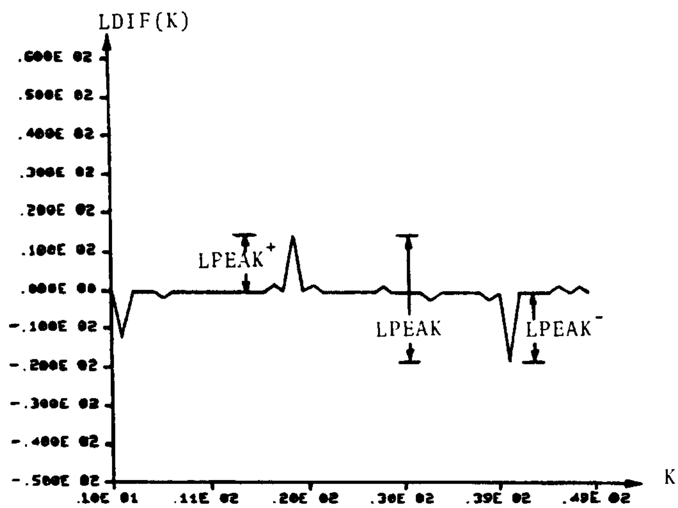

Fig. 2a. First difference of the left profile of the numeral 5.

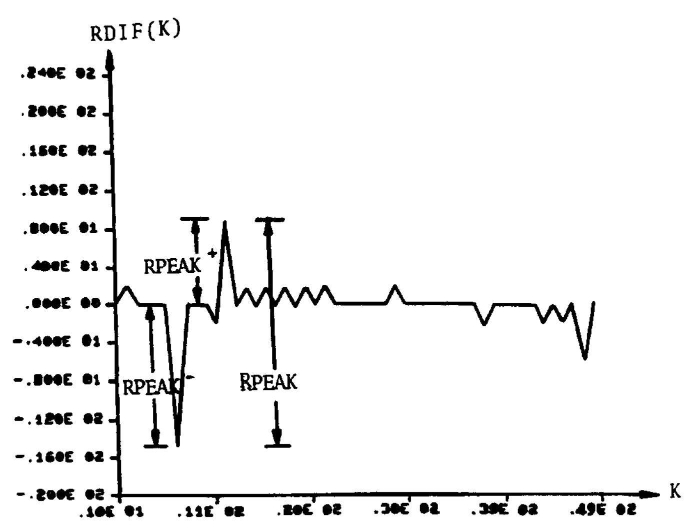

Fig. 2b. First difference of the right profile of the numeral 5.

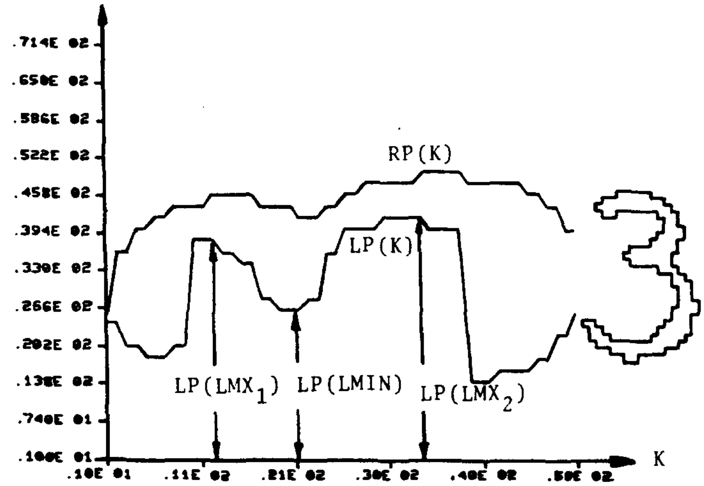

Fig. 3. Left and right profiles of the numeral 3.

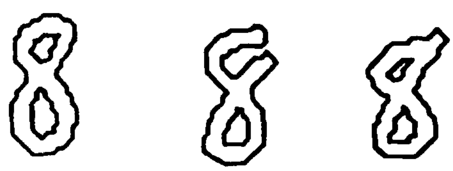

Fig. 4. Specimens of the numeral 8.

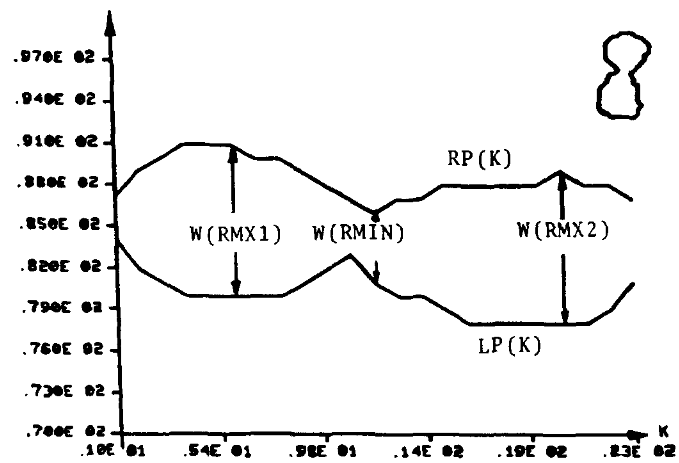

Fig. 5. Left and right profiles of the numeral 8.

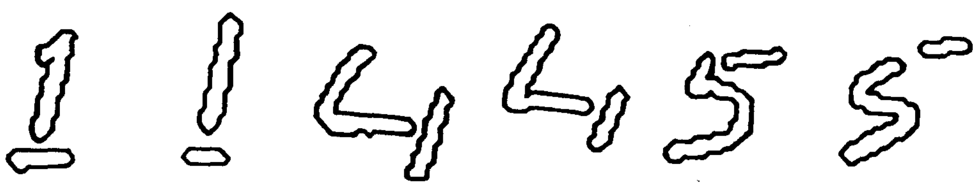

Fig. 6. Specimens of broken numerals.

(4) *Discontinuities in character profiles.* These are quantified by using the first difference values of the left and right profiles. Thus

$$L_{PEAK}^{+} = \max_{k \in R_2} \{L_{DIF}(k)\}$$

$$L_{PEAK}^{-} = \min_{k \in R_2} \{L_{DIF}(k)\}$$

$$R_{PEAK}^{+} = \max_{k \in R_2} \{R_{DIF}(k)\}$$

$$R_{PEAK}^{-} = \min_{k \in R_2} \{R_{DIF}(k)\}$$

where R_2 is the range of k in which the discontinuities are sought, and

$$L_{DIF}(k) = LP(k) - LP(k-1)$$

$$R_{DIF}(k) = RP(k) - RP(k-1)$$

are the first differences of the left and right profiles, respectively.

A single measure of discontinuity is defined as

$$L_{PEAK} = |L_{PEAK}|^{-} + |L_{PEAK}|^{+}$$

and

$$R_{PEAK} = |R_{PEAK}|^{-} + |R_{PEAK}|^{+}$$

Figures 2a and 2b display L_{PEAK} and R_{PEAK} for the numeral 5 shown in Fig. 1.

2.2. DETERMINATION OF SUB-CLASSES

The sub-classes of each numeral must be properly identified so as to handle a wide variety of writing styles exhibited by the population.

In this approach, a clustering analysis is carried out by utilizing the numeral specimens in the training set. Those numerals that show significant similarity in their feature properties are grouped together. Features

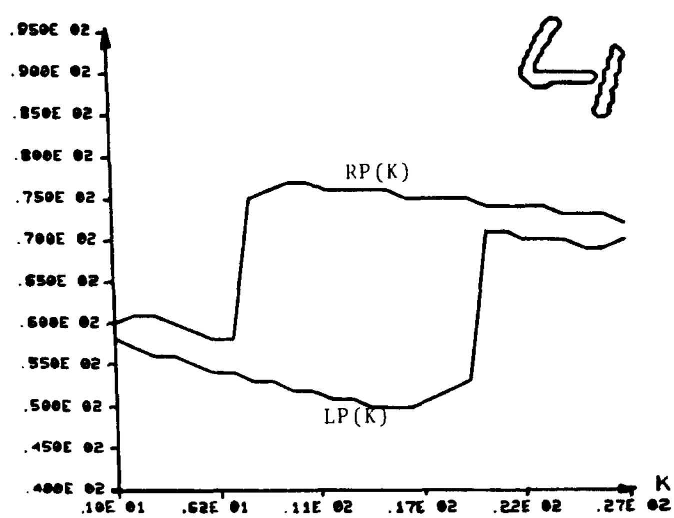

Fig. 7a. Left and right profiles of the broken numeral 4.

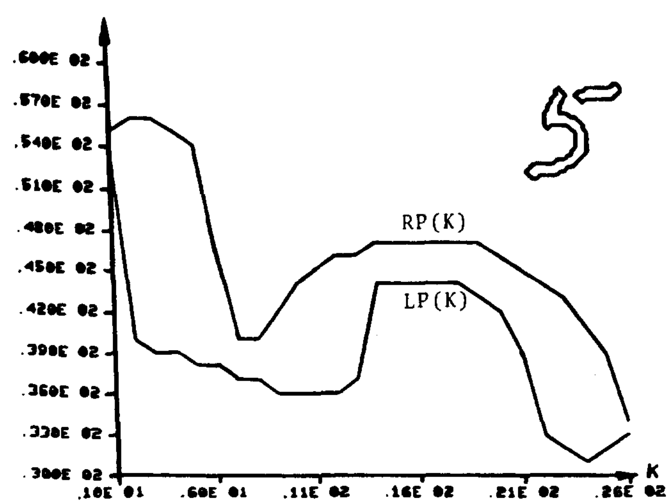

Fig. 7b. Left and right profiles of the broken numeral 5 (with vertical gap).

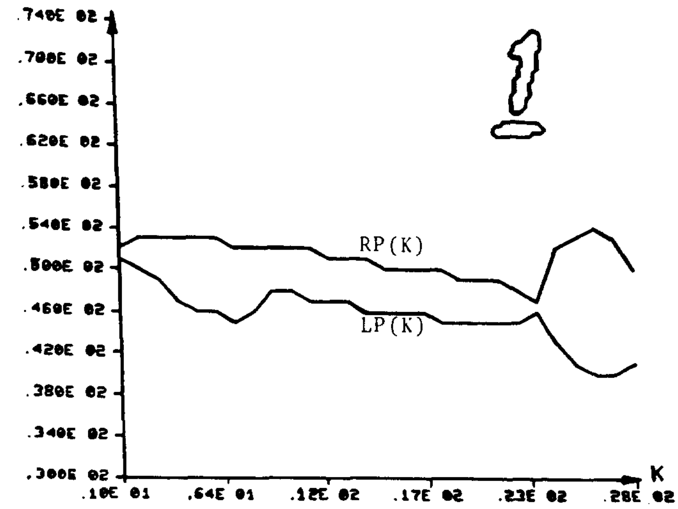

Fig. 7c. Left and right profiles of the broken numeral 1.

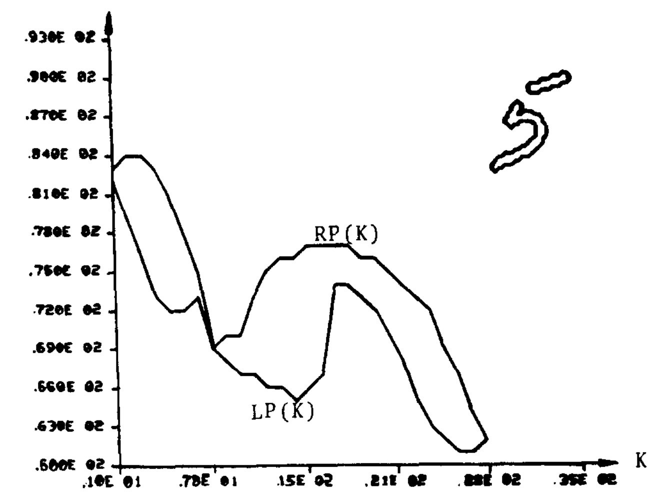

Fig. 7d. Left and right profiles of the broken numeral 5 (with horizontal gap).

that show significant variability are utilized to derive the different sub-classes for each numeral. The procedure will be illustrated by considering different examples. Figures 2a and 2b display the first differences of the left and right profiles for the numeral 5, shown in Fig. 1. An observation of Figs 1, 2a and 2b reveals the following: (1) a discontinuity in the right profile near the top; (2) a discontinuity in the left profile near the bottom; (3) the left profile attains a minimum before it attains its maximum.

The above observations may be quantified as (see Figs 1, 2a and 2b):

$$L_{\mathrm{PEAK}} > 10,\ 25 < R_2 < 47;$$

$$R_{\mathrm{PEAK}} > 10;\ 2 < R_2 < 25;$$

$$L_{\mathrm{MAX}} > L_{\mathrm{MIN}};\ 10 < R_1 < 40.$$

In a similar manner, numeral 3, whose profiles are shown in Fig. 3, may be seen to possess:

$L_{\mathrm{PEAK}} > 10$ (discontinuity in the left profile); $R_{\mathrm{PEAK}} < 10$ (right profile is relatively smooth); $L_{\mathrm{MAX1}} < L_{\mathrm{MIN}} < L_{\mathrm{MAX2}}$ (there are two maxima and one minimum in between). The range over which the observation is valid is $1 < R_1 < 50$.

Another illustration will consider the numeral 8. Figure 4 displays a typical set of specimens. The numeral shown in Fig. 4a may be described by the following feature properties:

$L_{\mathrm{PEAK}} < T_1$ (left profile is relatively smooth); $R_{\mathrm{PEAK}} < T_2$ (right profile is relatively smooth); $W(R_{\mathrm{MAX1}}) < W(R_{\mathrm{MIN}}) < W(R_{\mathrm{MAX2}})$ (the width of the numeral is minimum in the middle region, as shown in Fig. 5); where T_1 and T_2 are specified threshold values. The range R_1 for the above features is (1,23), where 23 is the height of the numeral.

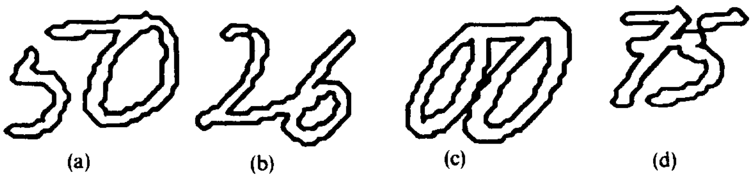

Fig. 8. Connected strings of numerals.

While a majority of the specimens will have the above features, there are significant exceptions. The numeral specimen in Fig. 4b, for example, has a break in the right profile (a relatively common occurrence in the numeral 8) and hence R_{PEAK} for this type of 8 will exceed T_2. Thus a second sub-class for this numeral must be defined. In a similar manner, the numeral specimen shown in Fig. 4c will have a different left profile and hence a third sub-class must be recognized. Therefore, the numeral 8 is characterized by three sub-classes.

This procedure was used for each numeral and appropriate sub-classes were then determined. Table A1 in Appendix 1 lists all the features that numerals (normalized to a height of 50 units) possess, while Table A2 in Appendix 2 lists the description of each numeral and its sub-classes in terms of these features. For the sake of brevity, only a partial listing is provided in Table A2. However, these sub-classes account for more than 90% of the numeral specimens commonly encountered.

2.3. FEATURES FOR BROKEN NUMERALS

An analysis of the numeral specimens collected from the participating individuals revealed that the most commonly occurring broken numerals were 1, 4 and 5. A typical collection is shown in Fig. 6. It is assumed that the breaks in the numerals are due to the writing habits of the individuals and not to the use of faulty writing implements.

A study of the broken numerals, such as the ones shown in Fig. 6, clearly reveals the existence of a vertical or a horizontal gap separating the two segments of the broken numeral. The left and right profiles will not be affected by the presence of a vertical gap between the two segments. This is clearly the case for the broken specimens of the numeral 4 and some of the numeral 5. Figures 7a and 7b display the left and right profiles for the broken numerals 4 and 5 with vertical gaps.

When horizontal gaps are encountered, as in 1 and 5, left and right profiles will be undefined at the locations corresponding to the gap between the two segments. By simply skipping these locations and storing the profile information only at locations where they are defined, the features can be properly evaluated. Figures 7c and 7d display the left and right profiles for the broken numerals 1 and 5 of the second type.

2.4. SEGMENTATION OF CONNECTED NUMERAL STRINGS

In this section, a procedure for segmenting connected numeral strings will be presented. The following assumptions are made in the derivation of the procedure: (1) the individual numerals in the string are

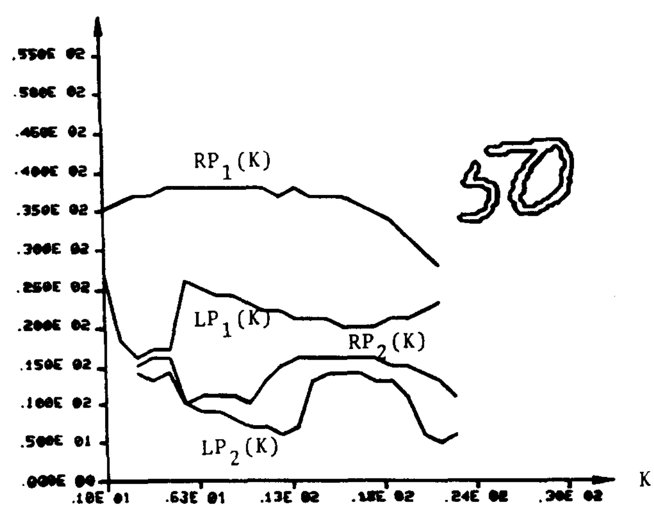

Fig. 9. Profiles of the connected string shown in Fig. 8a.

simply connected; (2) the individual numerals in the string are well separated and have no overlaps; (3) the number of numerals (connected or not) in the field is known *a priori*.

Figure 8 displays connected strings of two numerals that fall within the above assumptions and which can be segmented by the proposed procedure.

2.4.1. *Segmentation algorithm*

The algorithm for segmenting connected numeral strings will be presented by considering strings of two numerals. The extension to the general case is straightforward. The algorithm is a hierarchical approach based on the level of difficulty in segmentation.

The numeral string (connected or not) is imaged, framed and finally thresholded to yield a binary image. A border following algorithm[16] is then applied to the thresholded image to yield a new image with the boundaries (enclosing the character) taking on the highest (whitest) gray level. The algorithm consists of the following steps.

(1) First, the assumption that the two numerals are not connected is tested by initiating a vertical scan (starting in the middle of the frame) and counting the number of times the border element is encountered. If no border element is encountered, then the two numerals are isolated and their profiles are evaluated as described in the earlier sections.

(2) If only two border elements are encountered, then the two numerals are connected. The frame is then divided into two fields and the profiles in the left and right fields are evaluated as before. It is noted that additional sub-classes may have to be defined for each numeral, in view of the shape distortion caused by the connecting tail.

(3) If more than two border elements are encountered then the following steps are implemented.

(a) Initiate a vertical scan at a point to the left of the centre and count the border elements. If the number of border elements is still greater than 2, start the vertical scan at a point to the right of the centre. This procedure is repeated by moving further to the left or right and initiating a new scan until only two or zero border elements are encountered. If the procedure fails after a prescribed number of vertical scans, then step (b) is implemented. Figure 9 displays the left and right profiles of the two numerals shown in Fig. 8a using the above technique.

(b) The procedure outlined in 3(a) could fail if the connected numeral strings are slanted, as in Figs 8(c, d). It is possible to argue that the segmentation of slanted numeral strings can be realized by scanning at an angle. However, non-vertical scans (except for scans at 0° and ±45°) are not straightforward, owing to the spatial quantization inherent in the image array. As an illustration, scans at an angle of about 11° from the vertical would require 3 pixels vertical and 1 pixel diagonal scans. Considering that the numerals have an average height of only 30–35 pixels, significant errors in segmentation will arise with this scanning scheme.

In order to overcome this problem, a new procedure called the 'Hit and Deflect Strategy' (HDS) is proposed. In this procedure (under the assumptions of Section 2.4) a vertical scan is initiated at a point corresponding to the maximum peak in the bottom profile of the lower half of the image. The bottom profile and the initial scanning point are shown in Figs 10a and 10b for the numeral strings of Figs 8c and 8d, respectively.

During this initial vertical scan a border element is encountered. At this point the scanning point is moved (deflected) to the left or right or up as follows.

(i) If the pixel to the left of the border element is '0' and the pixel to the right of the border element is 'background', then the scanning point is moved to the right.

(ii) If the pixel to the left of the border element is 'background' while the pixel on the right is '0'

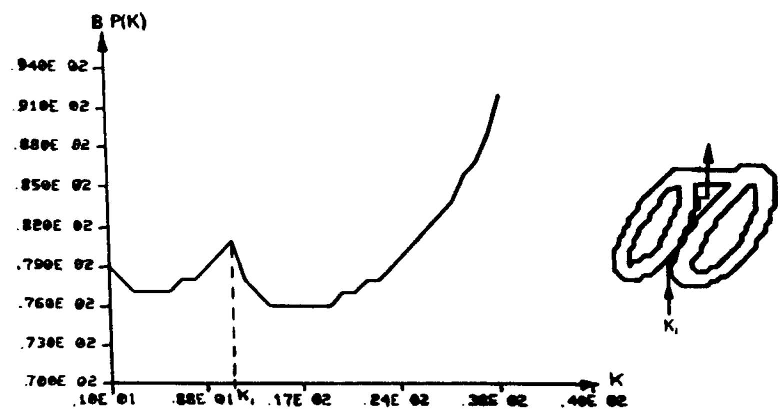

Fig. 10a. Bottom profile of the connected string shown in Fig. 8c.

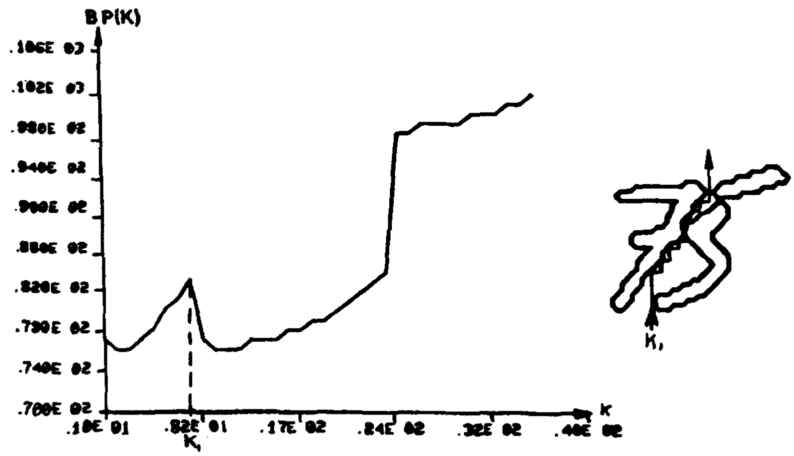

Fig. 10b. Bottom profile of the connected string shown in Fig. 8d.

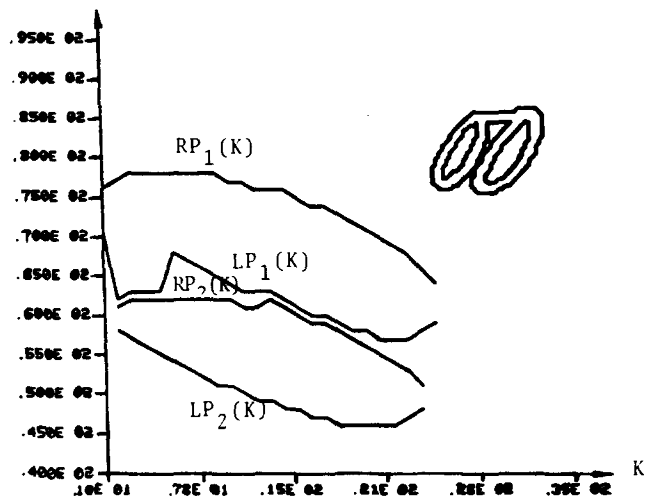

Fig. 10c. Profiles of the connected string shown in Fig. 8c.

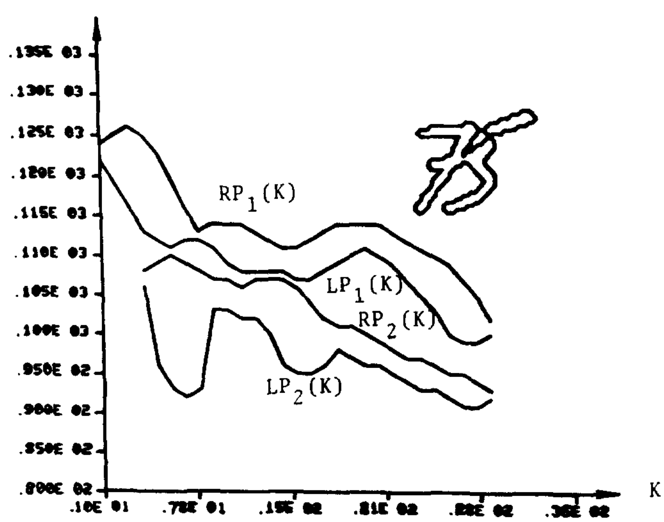

Fig. 10d. Profiles of the connected string shown in Fig. 8d.

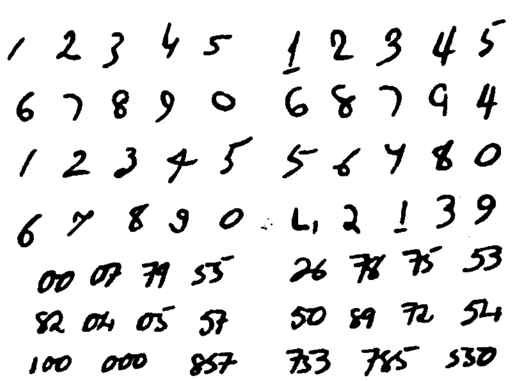

Fig. 11. Specimens of the handwritten data.

then the scanning point is moved to the left.

(iii) If the pixels on the left and the right of the border element are 'zeros' or 'border elements', the scanning point is moved up.

The above procedure HDS is repeated until the top frame line is encountered. The procedure is shown in Figs 10a and 10b, and the left and right profiles of the numeral strings are displayed in Figs 10c and 10d, respectively.

(iv) If the peak of the bottom profile is less than a threshold value T ($T = NH/4$, where NH is the height of the numeral), the same procedure is repeated by initiating the scan at a point corresponding to the deepest valley in the top profile of the upper half of the image.

3. RECOGNITION ALGORITHM

The recognition algorithm utilizes a tree structure to identify the unknown numeral specimens. The algorithm is initiated by determining the ratio of height to maximum width. If the ratio exceeds 2.5, then the test numeral is identified as numeral 1 and the algorithm terminates. Otherwise, the feature L_{PEAK} is evaluated and compared against a threshold to determine the appropriate sub-groups to which the unknown numeral belongs. Other features are then evaluated and, depending on their values, the algorithm proceeds down the tree until an identity is established.

4. TEST RESULTS

Five hundred test specimens were obtained from 20 individuals. The specimens included isolated numerals, broken numerals and connected strings of two numerals. Figure 11 displays typical specimens of the handwritten data used. The recognition algorithm was applied to the test specimens and an accuracy of 93% was realized. It must be pointed out that additional sub-classes had to be defined to take into account the extensions associated with the connected strings. However, when only isolated numerals were considered the recognition accuracy increased to nearly 99%.

5. CONCLUSIONS

(1) A new set of topological features are shown to provide unique and reliable description of handwritten numerals.

(2) The algorithm is shown to perform with high accuracy even when the numerals are broken.

(3) The segmentation algorithm is shown to yield proper segmentation of a class of connected numeral strings.

(4) Accuracies as high as 98% are shown to be feasible with the proposed algorithm.

SUMMARY

The recognition of handwritten numerals continues to be a challenging problem, since any practical recognition system requires: (a) that the algorithms be writer-independent; (b) the algorithms to be robust and yield high accuracies; (c) the algorithms to successfully segment connected numerals into their individual units and recognize them.

In this paper the authors describe the results of their investigation into the development of a recognition algorithm that can deal with isolated broken numerals as well as connected strings of numerals.

In the first phase of this study the authors describe a set of topological features to uniquely characterize isolated numerals. It is assumed that the numerals may be continuous or broken. The topological features are derived from the left and right profiles of the thresholded image of the numerals. Using a structural scheme, features are defined and these are combined to yield a logical description of the different numerals and their sub-classes. The recognition algorithm utilizes a tree classifier to determine the identity of the test numeral. Tests with 500 numeral specimens yielded an overall accuracy of 99%.

The second phase of this deals with the segmentation of connecting strings of numerals into their individual units. The segmentation technique utilizes a hierarchical approach starting with the hypothesis that the numerals in the field are fully isolated. If the hypothesis fails, then a test to determine whether the numerals are non-touching and simply connected is carried out. Progressively the tests are extended to include the possibility that the numerals may be connected and touching. Throughout the phase, it is assumed that the numerals in the string are non-overlapping. The recognition algorithm was again applied, and a recognition accuracy of 93% was achieved.

Acknowledgement—This work was supported by NSERC (Natural Sciences and Engineering Research Council) of Canada.

REFERENCES

1. C. Y. Suen, M. Berthod and S. Mori, Automatic recognition of handprinted characters—The state of the art, *Proc. IEEE* **68**, 469–487 (1980).
2. M. Shridhar and A. Badreldin, Handwritten numeral recognition technique by tree classification method, *J. Image Vision Comput.* **2**, 143–149 (1984).
3. M. Shridhar and A. Badreldin, A high accuracy syntactic recognition algorithm for handwritten numerals, *IEEE Trans. Syst. Man Cybernet.* **SMC-15**, 152–158 (1985).
4. T. Pavlidis and F. Ali, Computer recognition of handwritten numerals by polygonal approximations, *IEEE Trans. Syst. Man Cybernet.* **SMC-6**, 610–614 (1975).
5. H. Freeman, Boundary encoding and processing, *Picture Processing and Psychopictorics*, Lipkin and Rosenfeld, eds, pp. 241–306. Academic Press, New York (1970).
6. M. Shridhar and A. Badreldin, High accuracy character recognition algorithm using Fourier and topological descriptors, *Pattern Recognition* **17**, 515–524 (1984).
7. E. Persoon and K. S. Fu, Shape discrimination using Fourier descriptors, *IEEE Trans. Syst. Man Cybernet.* **SMC-7**, 170–179 (1977).
8. C. T. Zahn and R. Z. Roskies, Fourier descriptors for plane closed curves, *IEEE Trans. Comput.* **C-21**, 269–281 (1972).
9. G. H. Granlund, Fourier preprocessing for handprint character recognition, *IEEE Trans. Comput.* **C-21**, 195–201 (1972).
10. M. K. Hu, Visual pattern recognition by moment invariants, *IRE Trans. Inf. Theory* **IT-8**, 179–187 (1962).
11. M. T. Y. Lai and C. Y. Suen, Automatic recognition of characters by Fourier descriptors and boundary line encodings, *Pattern Recognition* **14**, 383–393 (1981).
12. Q. R. Wang and C. Y. Suen, Analysis and design of a decision tree based on entropy reduction and its application to large character set recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 406–417 (1984).
13. T. Young and T. Calvert, *Classification, Estimation and Pattern Recognition*. Elsevier, Amsterdam (1974).
14. K. S. Fu, *Syntactic Pattern Recognition and Applications*. Prentice-Hall, Englewood Cliffs, NJ (1982).
15. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles*, 3rd Edn. Addison-Wesley, Reading, MA (1979).
16. A. Chottera and M. Shridhar, Feature extraction of manufactured parts in the presence of spurious reflections, *Can. Elect. Engng J.* **7**(4), 29–34 (1982).

About the Author—M. SHRIDHAR received his M.S. from the Polytechnic Institute of New York in 1967 and his Ph.D. from the University of Aston in Birmingham (U.K.) in 1969.

He is currently a Professor of Electrical Engineering at the University of Windsor, Windsor, Ontario, Canada.

Dr Shridhar's research interests are in the area of signal processing with applications to robotics with vision, speaker, speech and pattern recognition.

About the Author—AMIRA BADRELDIN received her M.A.Sc. in System Design from the University of Waterloo, Ontario, Canada, and her Ph.D. in Electrical Engineering from the University of Windsor in 1985.

Her research interests are in the area of image processing, pattern recognition and expert systems.

APPENDIX 1

Table A1. Features used in the recognition of the numerals 0, 1, ..., 9

No.	Label	Features corresponding to label
1	a_1	True if $L_{PEAK} < 10$; $2 \le R_2 \le 50$
2	a_2	True if $L_{PEAK} < 5$; $2 \le R_2 \le 10$
3	a_3	True if $L_{PEAK} > 5$; $2 \le R_2 \le 15$
4	a_4	True if $L_{PEAK} > 10$; $2 \le R_2 \le 15$
5	a_5	True if $L_{PEAK} > 10$; $2 \le R_2 \le 20$
6	a_6	True if $L_{PEAK} > 5$; $2 \le R_2 \le 25$
7	a_7	True if $L_{PEAK} > 5$; $5 \le R_2 \le 15$
8	a_8	True if $L_{PEAK} > 5$; $5 \le R_2 \le 35$
9	a_9	True if $L_{PEAK} > 10$; $5 \le R_2 \le 40$
10	a_{10}	True if $L_{PEAK} > 10$; $10 \le R_2 \le 30$
11	a_{11}	True if $L_{PEAK} > 10$; $15 \le R_2 \le 40$
12	a_{12}	True if $L_{PEAK} < 5$; $25 \le R_2 \le 50$
13	a_{13}	True if $L_{PEAK} > 10$; $30 \le R_2 \le 50$
14	a_{14}	True if $L_{PEAK} < 5$; $30 \le R_2 \le 50$
15	a_{15}	True if $L_{PEAK} < 5$; $35 \le R_2 \le 50$
16	a_{16}	True if $L_{PEAK} > 10$; $35 \le R_2 \le 50$
17	a_{17}	True if $L_{PEAK} > 5$; $40 \le R_2 \le 50$
18	b_1	True if $R_{PEAK} > 10$; $2 \le R_2 \le 50$
19	b_2	True if $R_{PEAK} > 10$; $2 \le R_2 \le 15$
20	b_3	True if $R_{PEAK} < 10$; $2 \le R_2 \le 30$
21	b_4	True if $R_{PEAK} < 5$; $2 \le R_2 \le 45$
22	b_5	True if $R_{PEAK} < 10$; $25 \le R_2 \le 45$
23	b_6	True if $R_{PEAK} > 10$; $25 \le R_2 \le 50$
24	b_7	True if $R_{PEAK} < 5$; $25 \le R_2 \le 50$
25	b_8	True if $R_{PEAK} > 10$; $30 \le R_2 \le 50$
26	b_9	True if $R_{PEAK} > 5$; $35 \le R_2 \le 50$
27	b_{10}	True if $R_{PEAK} > 10$; $35 \le R_2 \le 50$
28	b_{11}	True if $R_{PEAK} > 5$; $40 \le R_2 \le 50$
29	c_1	True if R_{MIN} $(1 \le R_1 \le 30)$ is less than R_{MAX2} $(R_{MIN} \le R_1 \le 30)$ and greater than R_{MAX1} $(1 \le R_1 \le R_{MIN})$
30	c_2	True if R_{MIN} $(10 \le R_1 \le 40)$ is less than R_{MAX2} $(R_{MIN} \le R_1 \le 40)$ and greater than R_{MAX1} $(1 \le R_1 \le R_{MIN})$
31	c_3	True if R_{MIN} $(10 \le R_1 \le 45)$ is less than R_{MAX2} $(R_{MIN} \le R_1 \le 45)$ and greater than R_{MAX1} $(1 \le R_1 \le R_{MIN})$
32	d_1	True if $RP(R_{MIN}) = RP(R_{MAX})$ where R_{MIN} is in the range $5 \le R_1 \le 25$, and R_{MAX} in the range $1 \le R_1 \le R_{MIN}$
33	d_2	True if $RP(R_{MIN}) = RP(R_{MAX})$ where R_{MIN} is in the range $5 \le R_1 \le 25$, and R_{MAX} is in the range $R_{MIN} \le R_1 \le 40$
34	e_1	True if $L_{MAX} < L_{MIN}$; where L_{MAX} is in the range $1 \le R_2 \le 30$, and L_{MIN} is in the range $1 \le R_2 \le L_{MAX}$
35	e_2	True if $L_{MAX} < L_{MIN}$; where L_{MAX} is in the range $10 \le R_2 \le 30$, and L_{MIN} is in the range $10 \le R_2 \le L_{MAX}$
36	e_3	True if $L_{MAX} < L_{MIN}$; where L_{MAX} is in the range $10 \le R_1 \le 30$, and L_{MIN} is in the range $10 \le R_1 \le L_{MAX}$
37	e_4	True if $L_{MAX} < L_{MIN}$; where L_{MAX} and L_{MIN} are in the range $15 \le R_1 \le 45$
38	e_5	True if $L_{MAX} < L_{MIN}$; where L_{MAX} and L_{MIN} are in the range $20 \le R_1 \le 50$
39	e_6	True if $L_{MAX} < L_{MIN}$; where L_{MAX} and L_{MIN} are in the range $40 \le R_1 \le 50$
40	f_1	True if $R_{MIN} < R_{MAX}$; where R_{MIN} is in the range $1 \le R_1 \le 30$, and R_{MAX} is in the range $1 \le R_1 \le R_{MIN}$
41	f_2	True if $R_{MIN} < R_{MAX}$; where R_{MIN} and R_{MAX} are in the range $20 \le R_1 \le 35$
42	f_3	True if $R_{MIN} < R_{MAX}$; where R_{MIN} and R_{MAX} are in the range $35 \le R_1 \le 50$
43	g_1	True if $W(20) \ge W(40)$
44	g_2	True if $W(25) \ge W(10)$
45	g_3	True if $W(25) \ge W(40)$
46	g_4	True if $W(25) \ge W(45)$
47	h	True if Ratio > 2.5
48	i	True if $W_{MIN} < W_{MAX1}$ and $W_{MIN} < W_{MAX2}$; where $W_{MIN} = \min_{10 < J < 40} \{W(J)\} = \{W(L_{MIN})\}$; $W_{MAX1} = \max_{1 < J < L_{MIN}} \{W(J)\}$; $W_{MAX2} = \max_{L_{MIN} < J < 50} \{W(J)\}$

APPENDIX 2

Table A2. Description of each numeral and its sub-classes.

Numeral	Description of the numeral in terms of the features given in Table A1
0	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{b}_7 \vee \bar{a}_{12}) \wedge a_8 \wedge (d_1 \vee d_2)\ \bar{a}_4 \wedge \bar{a}_{16} \wedge \bar{a}_{11} \wedge g_4$
0	$a_1 \wedge \bar{b}_1 \wedge \bar{h} \wedge (g_2 \wedge g_3) \wedge e_3 \wedge f_1$
6	$a_1 \wedge b_1 \wedge \bar{a}_3 \wedge (\bar{a}_{13} \wedge \bar{b}_8) \wedge e_3 \wedge f_1$
1	$a_1 \wedge \bar{b}_1 \wedge h$
1 1	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{b}_7 \vee \bar{a}_{12}) \wedge a_8 \wedge (d_1 \vee d_2) \wedge \bar{a}_4 \wedge \bar{a}_{16} \wedge \bar{a}_{11} \wedge \bar{g}_4$
1 1	$a_1 \wedge \bar{b}_1 \wedge \bar{h} \wedge (\bar{g}_2 \vee \bar{g}_3) \wedge a_2 \wedge (\bar{a}_{17} \vee b_{11}) \wedge \bar{a}_8$
2	$\bar{a}_1 \wedge b_1 \wedge b_3 \wedge a_5 \wedge (b_9 \vee \bar{a}_{15}) \wedge f_3$
2	$a_1 \wedge b_1 \wedge \bar{b}_3 \wedge (a_{13} \vee b_8) \wedge a_{14} \wedge b_{10} \wedge \bar{e}_2$
2 2	$\bar{a}_1 \wedge b_1 \wedge b_3 \wedge \bar{a}_5 \wedge \bar{a}_{10} \wedge (a_{13} \vee b_8) \wedge \bar{e}_2$
2	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{b}_7 \vee \bar{a}_{12}) \wedge a_8 \wedge (\bar{d}_1 \wedge \bar{d}_2) \wedge a_{13} \wedge f_3$
3	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{b}_7 \vee \bar{a}_{12}) \wedge a_8 \wedge (d_1 \vee d_2) \wedge \bar{a}_4 \wedge a_{16} \wedge c_2 \wedge \bar{f}_3$
3 3	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{b}_7 \vee a_{12}) \wedge a_8 \wedge (d_1 \vee d_2) \wedge a_4 \wedge a_{13} \wedge \bar{f}_3$
4 4	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{a}_{12} \vee \bar{b}_7) \wedge \bar{a}_8$
4 4 4	$\bar{a}_1 \wedge b_1 \wedge \bar{b}_3 \wedge (b_9 \vee \bar{a}_{15}) \wedge \bar{a}_7 \wedge \bar{b}_7 \wedge \bar{a}_5 \wedge \bar{b}_5$
5	$\bar{a}_1 \wedge b_1 \wedge \bar{b}_3 \wedge (b_9 \vee \bar{a}_{15}) \wedge \bar{a}_7 \wedge b_7$
5 5	$\bar{a}_1 \wedge b_1 \wedge \bar{b}_3 \wedge (b_9 \vee \bar{a}_{15}) \wedge a_7 \wedge a_{13} \wedge e_1 \wedge e_6$
5 5	$\bar{a}_1 \wedge \bar{b}_1 \wedge \bar{b}_4 \wedge e_4 \wedge b_9 \wedge \bar{f}_3 \wedge \bar{b}_6 \wedge c_1$
6 6	$a_1 \wedge b_1 \wedge \bar{a}_3 \wedge (a_{13} \vee b_8) \wedge a_{14} \wedge \bar{b}_{10}$
6	$\bar{a}_1 \wedge b_1 \wedge \bar{b}_3 \wedge (\bar{a}_{15} \vee b_9) \wedge \bar{a}_7 \wedge \bar{b}_7 \wedge \bar{a}_5 \wedge b_5 \wedge \bar{b}_2$
6	$a_1 \wedge \bar{b}_1 \wedge \bar{h} \wedge (\bar{g}_2 \vee \bar{g}_3) \wedge f_2 \wedge e_3 \wedge \bar{f}_1 \wedge \bar{b}_4 \wedge \bar{a}_9 \wedge \bar{g}_1$
7 7	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (a_{12} \wedge b_7) \wedge a_7$
7 7	$\bar{a}_1 \wedge b_1 \wedge \bar{b}_3 \wedge (a_{15} \wedge \bar{b}_9) \wedge a_6 \wedge \bar{c}_2$
8 8	$\bar{a}_1 \wedge b_1 \wedge \bar{b}_3 \wedge (a_{15} \wedge \bar{b}_9) \wedge a_6 \wedge c_2 \wedge e_5$
8	$a_1 \wedge \bar{b}_1 \wedge \bar{h} \wedge (\bar{g}_2 \vee \bar{g}_3) \wedge f_2 \wedge \bar{e}_3$
8	$a_1 \wedge b_1 \wedge \bar{a}_3 \wedge (\bar{a}_{13} \wedge \bar{b}_8) \wedge \bar{e}_3$
9 9 9	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (a_{12} \wedge b_7) \wedge \bar{a}_7$
9 9	$\bar{a}_1 \wedge \bar{b}_1 \wedge b_4 \wedge (\bar{a}_{12} \vee \bar{b}_7) \wedge a_8 \wedge (d_1 \vee d_2) \wedge \bar{a}_4 \wedge a_{16} \wedge \bar{c}_2$

Where '$\wedge$' is the logical 'AND',

'$\vee$' is the logical 'OR',

and '-' is the logical 'NOT'.