PERGAMON

Pattern Recognition 33 (2000) 1967–1973

PATTERN RECOGNITION
THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Holistic recognition of handwritten character pairs

Xian Wang, Venu Govindaraju*, Sargur Srihari

Center of Excellence for Document Analysis and Recognition, State University of New York at Buffalo, Buffalo, NY 14260, USA

Received 20 July 1998; received in revised form 27 April 1999; accepted 21 September 1999

Abstract

Researchers have thus far focused on the recognition of alpha and numeric characters in isolation as well as in context. In this paper we introduce a new genre of problems where the input pattern is taken to be a pair of characters. This adds to the complexity of the classification task. The 10 class digit recognition problem is now transformed into a 100 class problem where the classes are $\{00, \ldots, 99\}$. Similarly, the alpha character recognition problem is transformed to a 26×26 class problem, where the classes are $\{AA, \ldots, ZZ\}$. If lower-case characters are also considered the number of classes increases further. The justification for adding to the complexity of the classification task is described in this paper. There are many applications where the pairs of characters occur naturally as an indivisible unit. Therefore, an approach which recognizes pairs of characters, whether or not they are separable, can lead to superior results. In fact, the holistic method described in this paper outperforms the traditional approaches that are based on segmentation. The correct recognition rate on a set of US state abbreviations and digit pairs, touching in various ways, is above 86%. © 2000 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Handwriting recognition; Holistic; Character recognition; Segmentation; Digit recognition; GSC; Feature vectors

1. Introduction

Researchers have thus far focused on the recognition of alpha and numeric characters in isolation as well as in context. In this paper we introduce a new genre of problems where the input pattern is taken to be a pair of characters. This adds to the complexity of the classification task. The 10 class digit recognition problem is now transformed into a 100 class problem where the classes are $\{00, \ldots, 99\}$. Similarly, the alpha character recognition problem is transformed to a 26×26 class problem, where the classes are $\{AA, \ldots, \dot{Z}Z\}$. If lower-case characters are also considered the number of classes increases further. The justification to adding to the complexity of the classification task is described in this paper. There are many applications where the pairs of characters occur naturally as an indivisible unit. Therefore, an approach which recognizes pairs of characters, whether or not they are separable, can lead to superior results. In fact, the holistic method described in this paper outperforms the traditional approaches that begin with segmentation.

There are many applications that require the recognition of unconstrained handwritten words. A word can be either purely numeric as in the case of a ZIP Code, or purely alphabetic as in the case of US state abbreviations (Fig. 1) or mixed as in the number of an apartment (e.g., 1A). In general, a character string recognizer has many applications. The applications include, but are not limited to, reading bank checks, reading tax forms and interpretation of postal addresses.

The task becomes particularly challenging when adjacent characters in a character string are touching. Unlike purely alphabetic strings where joining of the characters is natural and takes place by means of ligatures, the joining of numerals in a numeric word and the uppercase characters in an abbreviation are accidental. The various ways in which two digits can touch are categorized in Fig. 2. Some of the categories lend themselves to natural segmentation, whereas for some the holistic approach is the only option available.

* Corresponding author. Tel.: + 1-716-645-6164ext.103; fax: + 1-716-645-6176.

E-mail address: govind@cedar.buffalo.edu (V. Govindaraju).

0031-3203/00/$20.00 © 2000 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.
PII: S0031-3203(99)00204-6

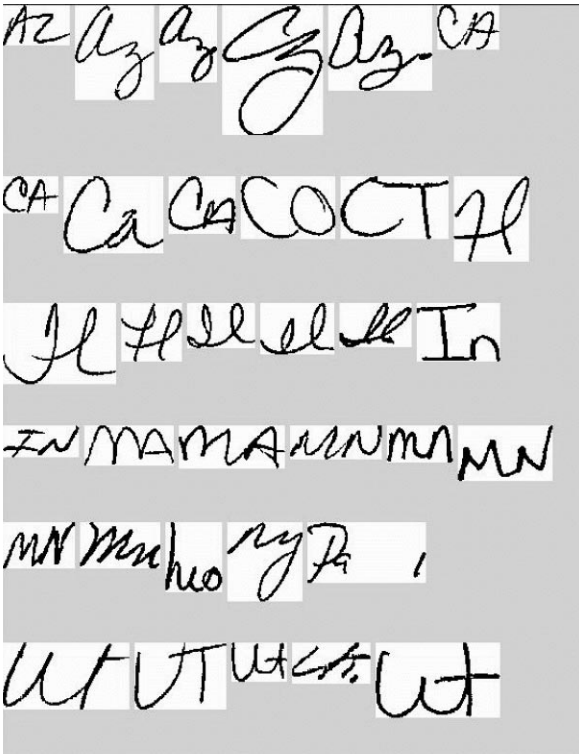


Fig. 1. Pairs of alpha characters that appear in state abbreviations in the US mail.

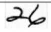
Category	Examples		Frequency
Single Point Touching			55%
Ligature Touching			18%
Multiple Points Touching			10%
Overlap			2%
Ligature Overlap			4%
Noise			9%
Broken			2%

Fig. 2. Categories of TDPs based on the manner in which the two digits touch and their frequency of occurrence in a set of 2778 samples.

Researchers have addressed the issue of touching characters in various ways, all of which rely upon segmenting the touching characters into individual characters [1–6]. While the segmentation approach does prove effective for many applications, it often leads to additional problems. Improper segmentation tends to leave some of the newly separated characters with artifacts, such as a character might end up losing a piece of its stroke to the adjacent character, or might end up with an additional ligature. Either way, subsequent recognition of the character is rendered inaccurate. Furthermore, finding the precise splitting path that separates the touching characters is non-trivial (Fig. 3).

The domain of our investigation will draw illustrations and testing samples from the USPS mail stream where numeric strings and state abbreviations are abundant. Numeric strings are present in ZIP codes, street numbers, apartment and suite numbers, and box numbers. 25,686 touching digit pairs collected from over 200,000 addresses reveals that about 15.5% of the numeric strings have touching digits in them. 13.1% are touchings digit pairs and about 2.4% have three or more touching digits. A distribution of the classes of touching digits pairs in the training set is shown in Fig. 4.

The predominance of touching digit pairs as opposed to triples and quadruples justifies our emphasis on the digit pair problem.

Henceforth, we will refer to a pattern of character pairs as CP and digit pairs as DP. The characters that

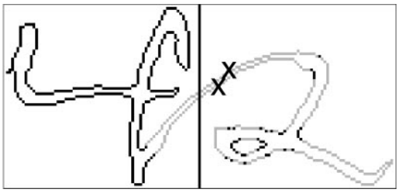


Fig. 3. There is no simple splitting line that will segment the two touching digits without leaving artifacts on the separated digits.

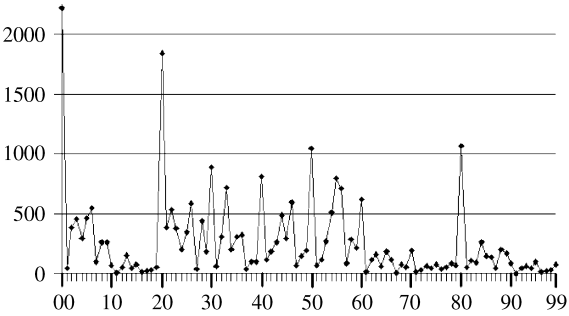


Fig. 4. Distribution of 25,686 TDPs in training set collected from more than 200,000 US mailpieces. It is to be noted that pairs with ‘0’ are very frequent.
The 25 most frequently occurring digit pairs are:
00, 02, 03, 05, 06, 20, 21, 22, 23, 25, 26, 28,
30, 32, 33, 35, 36, 40, 44, 46, 50, 54, 55, 56,
60, 80

constitute the pair could be separated or touching. One of the advantages of the holistic approach described in this paper is that no distinction needs to be made among touching and non-touching pairs. A segmentation based method must necessarily make such a determination.

While the holistic method applies equally to both touching and non-touching characters, its advantage over traditional segmentation-based methods is more pronounced among the touching character pairs. Such pairs will be referred to as TCP and TDP for touching character pairs and touching digit pairs, respectively, and will be the focus of this paper. We will assume that the pattern in question has two touching characters. We will use the holistic paradigm of recognition which has been successfully used in word recognition, but has not been applied to the TCP/TDP problem.

Input to the TDP problem is a pair of digits extracted from numeric strings in an address and the expected output is its classification as one of possible 100 classes. Input to the TCP problem is a pair of characters that represent the US state abbreviation as it appears on an address. The expected output is the identity of the state name. A small subset of the 676 (26 × 26) possible abbreviations are valid. The 50 states and 12 special territories give a set of 62 possible classes. The method described in this paper outperforms the traditional segmentation-based approaches in both cases.

2. Background of segmentation-based methods

A survey of segmentation strategies is provided by Casey and Lecolinet [7]. Fig. 5 shows some examples with the output of a segmentation-based method described in Ref. [3]. Common approaches are often heuristic. While they have the advantage of efficiency, their accuracy is limited. Vertical histograms have been widely used for segmentation but are error-prone.

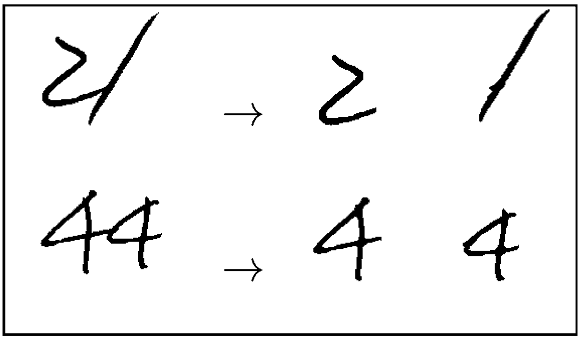


Fig. 5. Examples of touching digits and the desired results from a segmentation-based method.

Another method uses the upper and lower profiles in conjunction with a set of heuristics to determine the segmentation points [4]. A segmentation path is constructed upward from the highest point on the lower profile of a numeral or downward from the lowest point in the upper profile. The approach is further refined in Ref. [5]. A graph based method has been described in Ref. [6]. This method analyzes the contours of the connected numerals. Vertically oriented edges derived from adjacent strokes form the vertices of the graph, which are potential points of segmentation. There are other methods where a recognizer is used to aid in the segmentation process [8]. Potential segmentation points are validated by submitting the resulting segments to a digit recognizer. The recognizer provides a means of quantifying the "goodness" of a segmentation point, otherwise unavailable to methods using heuristics alone. However, frequent calls to a recognizer during the course of segmentation makes the process inefficient.

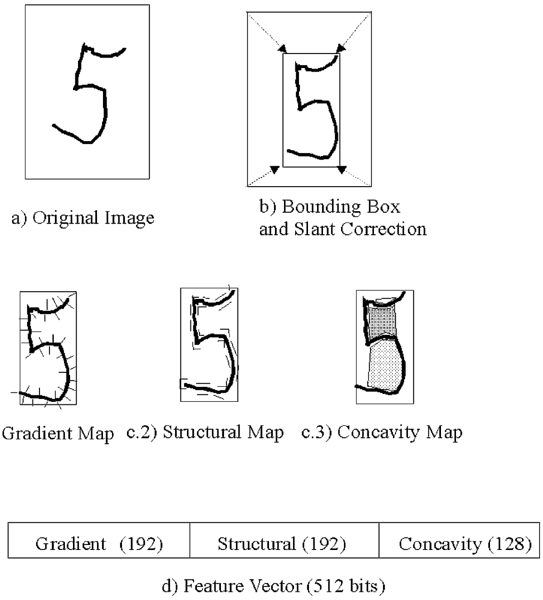
We will compare the holistic method introduced in this paper with two segmentation-based methods. The method described by Fenrich [9] and the method described by Shi and Govindaraju [3].

Fenrich describes a recognition aided iterative method is used. The number of digits in a numeric string is initially estimated from the aspect ratio of the digit string and successive estimates are obtained by a linear regression model. Digits that are recognized with high confidence are removed after each iteration. The effective contribution of the removed digits to the density of the digit string is recorded. This information is fed to a least-squares linear model. By setting the density to zero (all digits are removed), the least squares equation can estimate the number of digits in the string. Connected digit components are split into required number of digits. The segmenter has a correct segmentation rate of 93.33% when the input is specified as a 5 or 9 digit ZIP Code, and is 83.03% correct when the number of digits has to be estimated.

Shi and Govindaraju describe a method of splitting digits that follows the contours of the strokes. Significant right turning points together with their opposite contour points are located to divide the contour into contour pieces. Contour pieces are then classified as belonging to one or the other digit. A vertical line bisecting the image is used as the guide. A contour piece is classified as part of the left digit if the *center of mass* of the piece is on the left of the line, otherwise, as part of the right digit. In most cases, the two digits have similar widths and the touching strokes lie close to the center line. Images of segmented digits may be recovered from the segments by drawing line segments from each contour point to its opposite contour point, or the modified opposite contour point if the line segment to the opposite point is too long.

3. Methodology

Since the character recognizer chosen does not in anyway limit the holistic paradigm presented in this paper, we will describe the methodology using the GSC character recognizer [1]. The Gradient, Structural, Concavity, (GSC) features are symbolic multi-resolutional features (Fig. 6). **G**radient of the image contour captures the local shape of a character. The Gradient features are extended to **S**tructural features by encoding the relationships between strokes. **C**oncavity features capture the global shape of characters. Features at the three levels, **G, S, C** are combined in a k-nearest-neighbor classification method to produce a multi-resolution digit recognizer. **G** features are the finest and the **C** features are the coarsest.


a) Original Image

b) Bounding Box and Slant Correction

c.1) Gradient Map c.2) Structural Map c.3) Concavity Map

d) Feature Vector (512 bits)

Fig. 6. Gradient, structural, and concavity features are mapped on a binary feature vector of 512 bits.

Several modifications are made to the GSC isolated character recognizer to support the holistic recognition of TDPs and TCPs.

1. *Grid sizes*: The original GSC recognizer designed for isolated characters uses 12 binary values for Gradient and Structural features and eight binary values for Concavity features. Using a 4×4 grid of 16 cells, this leads to $(12 + 12 + 8) \times 16 = 512$ bit feature vector.

   The original GSC recognizer would compute the score of a class by the following formula: $Score = (n_{11} + n_{00}/S)/512$, where n_{11} represents the bit positions where the binary values are '1' in both the test sample and a prototype of the class, and n_{00} represents the bit positions the binary values are '0' in both the test sample and prototype of the class. S is typically chosen in the interval $[1, \dots, 5]$ to bias the score in favor of the matches.

   We use a 4×6 grid instead of the 4×4 grid used for isolated character recognition. The reasoning behind this modification is obvious. Since the pattern of concern is two characters side by side, the width of the pattern is going to be larger than the height. We do experiment with other sizes of grid and present the results.

   Given that the GSC recognizer is modified to accommodate variable size grids, the size of the feature vector is not always 512. The formula is empirically determined as $Score = (n_{11} + n_{00} \times 4/11)/Vector\ Length$. We have experimented with various grid sizes for reasons described in the previous section. Results of the experiments are presented in Table 1. The 4×6 grid gives the best results. It is interesting to note that the 4×8 grid, which is essentially puting two 4×4 grids side by side is not the best choice.
2. *Weighted scoring*: If the width of the pattern is divided into three zones, it is our conjecture that the most useful features are present in the left and right zones, while the central zone merely contains the ligature that joins the two characters (Fig. 5). Further, given the various ways in which the joining of two

Table 1
Recognition result of GSC features in different grid size based on 2488 learning samples and 226 testing samples

Grid size	G vector	S vector	C vector	Total vector	Correct	Weighted
4 × 4	192	192	128	512	79.2%	
4 × 6	288	288	192	768	82.3%	85%
6 × 9	648	648	432	1728	78.8%	
4 × 8	384	384	256	1024	79.2%	81.4%
4 × 7	336	336	224	896	79.6%	
6 × 6	432	432	288	1152	77.0%	

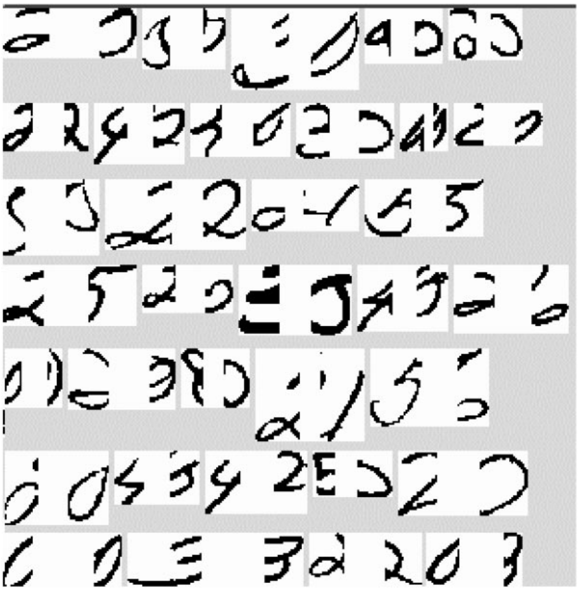


Fig. 7. TDP image samples with the central zone suppressed. The recognition rate achieved by people on the set is the same as that achieved by our program.

characters can occur (Fig. 2), it is best to disregard the central portion altogether. To this purpose, we weight the contribution of different zones of the pattern to minimize the effects of the central zone.

We can suppress the third and fourth columns of the grid based on the observation that it carries little useful information. The cognitive reasoning for suppressing the information in the central zone is supported by Fig. 7. Ten human subjects when asked to classify the TDPs in the figure had about the same recognition rate as our program, 85% (Table 1).

Variable weights to the G, S, and C features was also experimented. Recognition rates are shown in Table 2. Best results are achieved by giving the maximum weight to the Concavity features. Given the complex shape of the TDP pattern, this result does seem reasonable. The coarsest features are the most important. Furthermore, the computation of score de-emphasizes the third and fourth columns and emphasizes the first, second, fifth and sixth columns.

The formula for scoring is empirically modified as

$$Score = \frac{n_{11} + 3/11 \times n_{00} + w_g \times gn_{11} + w_s \times sn_{11} + w_c \times cn_{11}}{Total\ Features},$$

where gn_{11}, sn_{11}, cn_{11} refer to the bits matching among the G, S, and the C vectors, respectively, while considering only the first, second, fifth and sixth columns. w_g, w_s, and w_c are the weights.

Table 2
Recognition rates with weighted values for the G, S, and C features based on 2488 learning samples and 226 testing samples

G-weight	S-weight	C-weight	Correct	Error
w_g	w_s	w_c	(%)	(%)
0	0	0	82.3	17.7
0.5	0	1.0	84.1	15.9
0.5	0.1	1.0	84.5	15.5
0.5	0.2	1.2	85.0	15.0

4. Experimental details

Touching digit pairs were manually selected from postal images. Touching digit strings with more than two digits, digits with underlines and highly noisy strings were eliminated. The frequency of occurrence of the various digit pairs is weighted by the number of records in the ZIP Code in the USPS database. The database is organized by ZIP Codes. Each ZIP Code has records corresponding to delivery addresses that exist in the ZIP Code. There are about 43,000 ZIP Codes in the United States. Of these, some ZIP Codes are more common in the mailstream than others as they have a larger number of delivery addresses. Surely, the volume of mail destined for Manhattan, NY is far greater than the volume destined for Boise, ID. ZIP Code.

Note the frequency of digit pairs which have a ‘0’ (Fig. 8). This also conforms to our experience that people tend to write 0’s so that it touches its adjacent numeral with a ligature.

Table 3 shows the results for test samples that have more than 100 training samples. In our experiments we use the scheme of randomly assigning every 10th sample of a class to the testing set and the remaining 9 are put into the training set. We observe that when the number of training samples are more than 100, the recognition accuracy is very high for the test patterns of the class. Further, on such classes, the recognition accuracy is far superior to that of a segmentation-based approach.

In a set of 23,579 state abbreviations, about 20% had touching characters. The recognition rate on touching samples was 86.9% and on the non-touching samples it was 88.8%.

A comparison of the holistic method described in this paper with segmentation-based methods described in Refs. [3,9] are shown in Table 4. The holistic method outperforms the segmentation-based methods in all categories except for the case when there is a peculiar overlapping of strokes. It is especially superior in the category of touching digit pairs that are touching at multiple points.

Table 5 shows the results of comparison between the segmentation method proposed in Ref. [3] and the

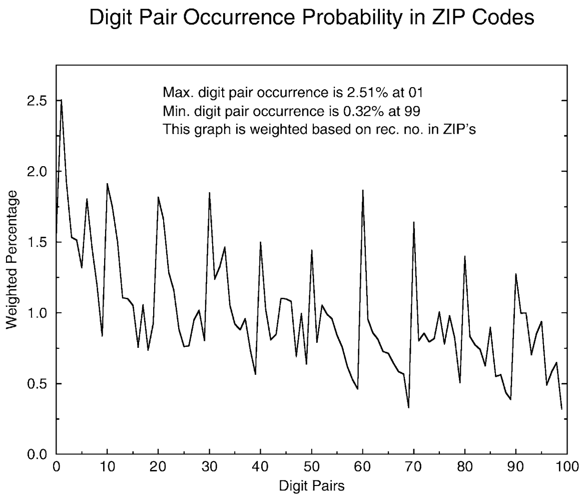


Fig. 8. Digit pair frequencies indicate that '0' is the most common digit in a pair of digits.

Table 3
Recognition rates for classes with sufficient number of training samples

Class	Training samples	Test images	Correct rate
00	418	43	42 (97.7%)
20	452	48	46 (96.8%)
21	102	11	10 (90.9%)
22	161	17	17 (100%)
26	184	18	18 (100%)
30	188	19	18 (94.7%)
33	105	10	10 (100%)
40	147	15	13 (86.7%)
46	148	15	14 (93.3%)
50	276	29	26 (89.7%)
55	127	13	12 (92.9%)
56	104	10	9 (90.0%)
80	407	43	42 (97.7%)
Total	2,819	291	277 (95.2%)

holistic method described in this paper using image sets taken from the CEDAR CD-ROM.

As an interesting related experiment, the methodology described was applied to digit pairs which were not touching. 2117 images (pairs of digits that are not touching) were extracted at random from USPS mailpiece images for the purpose. 703 images were used for the purpose of testing. The remaining were added to the training samples. The division between testing and training was done by ensuring that there are "enough" training samples for each "test" image. Recognition rate of 84.2% was achieved. This is comparable to the recognition rate achieved for digit pairs that are touching.

5. Summary

We have demonstrated that a holistic method of recognizing touching digit pairs and touching character pairs (such as in state abbreviations) is a viable approach. It gives a recognition rate of 86.8% on a set of touching digit string images and a recognition rate of 86.9% on a set of touching characters (state abbreviations). The recognition rate on digit pairs that are not touching is 84.2%.

The choice of features used, and the weighting of the various cells in the grid so as to give less importance to the central zones of the patterns, perhaps has support in how humans recognize touching digit pairs.

Future work will involve collecting sufficient number of training samples for all classes. A combination strategy that uses the results of both the segmentation and holistic methods is being investigated.

Table 5
Our holistic method outperforms a segmentation based method [3] when tested on the CEDAR CD-ROM test images

Image set	Number of images	Method in Ref. [3]	Holistic
BHA	628	80.4%	83.5%

Table 4
Recognition rate comparison across various categories of TDPs

Category	Images	Method in Ref. [9]	Method in Ref. [3]	Holistic
All	2778	81.6%	83.5%	85.1%
Single point touching	1537	83.9%	85.4%	86.4%
Ligature touching	484	82.2%	85.5%	86.0%
Multiple point touching	272	75.0%	73.5%	85.3%
Overlaps	175	87.4%	92.0%	86.3%
Noisy	256	68.8%	72.7%	73.0%
Broken	54	88.9%	87.0%	90.7%

References

[1] John T. Favata, Geetha Srikantan, A multiple feature/resolution approach to handprinted digit and character recognition, Int. J. Imaging Systems Technol. 7 (1996) 304–311.

[2] Oivind Due Trier, Anil K. Jain, Torfinn Taxt, Feature extraction methods for character recognition — a survey, Pattern Recognition 29 (4) (1996) 641–663.

[3] Z. Shi, V. Govindaraju, Segmentation and recognition of connected handwritten numeral strings, Pattern Recognition 30 (9) (1997) 1501–1504.

[4] M. Shridhar, A. Badreldin, Recognition of isolated and simply connected handwritten numerals, Pattern Recognition 19 (1986) 1–12.

[5] F. Kimura, M. Shridhar, Segmentation-recognition algorithm for handwritten numeral strings, Mach. Vision Appl. (5) (1992) 199–210.

[6] J.M. Westall, M.S. Narasimha, Vertex directed segmentation of handwritten numerals, Pattern Recognition 26 (10) (1993) 1473–1486.

[7] R.G. Casey, E. Lecolinet, Strategies in character segmentation: a survey, Proceedings of the Interrational Conference on Document Analysis and Recognition, Montréal, Canada, 1995, pp. 1028–1033.

[8] S. Seshadri, D. Sivakumar, A technique for segmenting handwritten digits, Pre-Proceedings of the Interrational Workshop on Frontiers in Handwriting Recognition III, Buffalo, New York, USA, May 25–27, 1993, pp. 443–448.

[9] R. Fenrich, in: S. Impedovo, J.C. Simon (Eds.), Segmentation of Automatically Located Handwritten Numeric Strings, From Pixels to Features III: Frontiers in Handwriting Recognition, Elsevier, Amsterdam, 1992, pp. 47–59.

About the Author—XIAN WANG is a Research Scientist in Center of Excellence for Document Analysis and Recognition (CEDAR), State University of New York at Buffalo. He received his Ph.D in Signal and Information processing from Department of Electronic Engineering, Tsinghua University in 1994. Prior to joining CEDAR, he was an Invited Research Officer at Institute for Posts and Telecommunications policy (IPTP), Ministry of posts and Telecommunications of Japan. His research interests include Document Image Analysis and Recognition, USPS and Japanese Mail Address Recognition and Interpretation.

About the Author—VENU GOVINDARAJU received his Ph.D in Computer Science from the State University of New York at Buffalo in 1992. He has coauthored more than 90 technical papers in various International journals and conferences and has one US patent, he is currently the associate director of CEDAR and concurrently holds the research associate professorship in the department of Computer Science and Engineering, State University of New York at Buffalo. He is the associate editor of the Journal of Pattern Recognition and the area chair of the IEEE SMC technical committee for pattern recognition. Dr. Govindaraju has been a co-principal investigator on several federally sponsored and industry sponsored projects. He is presently leading multiple projects on postal applications. He is a senior member of the IEEE.

About the Author—SARGUR SRIHARI received his BE in Electrical Communication Engineering from the Indian Institute of Science, Bangalore, in 1970, and a Ph.D in Computer and Information Science from the Ohio State University in 1976. He is presently a University Distinguished Professor of SUNY. He is the founding director of CEDAR, Center of Excellence for Document Analysis and Recognition, located at the State University of New York at Buffalo. The work of CEDAR has led to systems deployed for the United States Postal Service, Australia Post, and the Internal Revenue Service. Dr. Srihari is a fellow of the IEEE and editor-in-chief of the International Journal of Document Analysis and Recognition.