ELSEVIER

Pattern Recognition Letters 17 (1996) 413–419

Pattern Recognition Letters

A computer vision approach to digit recognition on pulp bales

Jukka Heikkonen [a,*], Mika Mäntynen [b]

[a] European Commission, Joint Research Centre, Institute for Systems Engineering and Informatics, TP 361, I-21020 Ispra (VA), Italy
[b] Lappeenranta University of Technology, Department of Information Technology, P.O. Box 20, FIN-53850 Lappeenranta, Finland.

Received 20 December 1995

Abstract

This paper describes a computer vision approach for recognizing quality and producer information of pulp bales from digit series stamped on pulp bales. The digit recognition consists of three stages: segmentation of digit series, feature extraction, and classification. Segmentation of digit series is based on image thresholding and Randomized Hough Transform. Digit segmentation produces six digit windows. In feature extraction two band-pass derivative of Gaussian filters are used, and the resulting gradient field histograms are used after normalization in classification of digits. The digits in the test set can be classified 93% correct with a multiple layer perceptron network. Classification results with three other well know classifiers are also reported.

Keywords: Digit recognition; Randomized Hough transform; Neural pattern recognition

1. Introduction

Character recognition has been an important research field within computer vision. The research is motivated by strong demands in a number of fields for systems that can interpret optical data consisting of handwritten or machine printed characters (Ballard and Brown, 1982; Tappert et al., 1990). For instance in postal automation systems that can recognize handwritten or machine printed characters from letters and packages are apparently needed (Denker et al., 1989; Srihari, 1991). There are also many applications where the products should be identified via numerical codes or where the documents should be turned into an electrical form (Mori, 1992).

This paper describes a computer vision approach for recognizing quality and producer information of pulp bales moving on the conveyer. This information is needed for producing paper in a paper plant, and the information can be found from digit series of six digits stamped on the pulp sheet covering a pulp bale. The purpose of the machine vision system would be to replace humans in pulp bales recognition and to keep the factory information system up-to-date.

The quality of the digits stamped on the pulp bales are mostly moderate (Fig. 1). There are many sources for difficulties. For instance, image grey-level values of the digits and the pulp sheets may vary from one pulp bale to another typically depending on the moistness of the pulp sheets; the fonts of the digits vary slightly depending on pulp suppliers; the digits may be rotated with unknown degrees; the digits may be located anywhere on the visible site of a pulp bale; and on the pulp bales there are always other text and also

* Corresponding author. Email: jukka.heikkonen@jrc.it. On leave from Lappeenranta University of Technology, Department of Information Technology, P.O. Box 20, SF-53850 Lappeenranta, Finland.

0167-8655/96/$12.00 © 1996 Elsevier Science B.V. All rights reserved
SSDI 0167-8655(95)00136-0

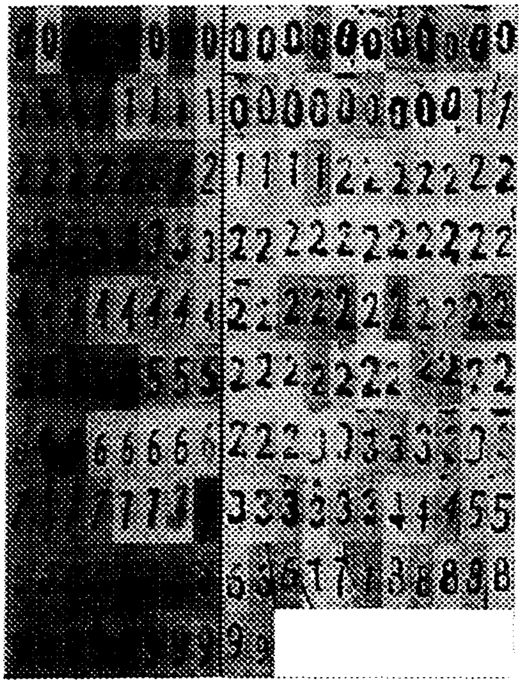

Fig. 1. Sample stamped digits to be recognized.

metal strapping wires disturbing the recognition task (see Fig. 2). Fortunately there is no need for size invariant digit recognition, because the sizes of the digits vary only slightly. However, the digits recognition system has to deal with the constraints listed above.

The digit recognition consists of three stages: digit segmentation, feature extraction, and classification. The segmentation of digits from pulp bales images and the feature extraction schema are described in detail in Sections 2 and 3 respectively. In Section 4 classification results obtained both with multiple layer perceptron network (Rumelhart, 1986) and three other well known classifiers are reported. Finally, Section 5 gives a conclusion.

2. Segmentation of digits

The first task is to locate the digit series in the image. Fig. 2 shows a pulp bale image. In this case the contrast between the digits (220351) and their surroundings is good as fortunately in most cases. However, the factory stamp clearly disturbs the segmentation task. The stamp is always in a random location and occasionally it can even overlap the digit series. In addition the metal straps may overlap the digit series causing false segment selections and errors in recognition. Fig. 3 illustrates that the quality difference between digits on dry on moist pulp sheets is remarkable. When a digit is stamped onto a moist pulp sheet, the ink spreads smoothly on the pulp sheet producing a messy digit. However, the segmentation method must locate digits on both types of pulp sheets.

Most segmentation algorithms are based on grey-level image thresholding and region growing (Zucker, 1976; Ohlander et al., 1978). In our segmentation approach the image is first thresholded with an initial threshold value and the areas (in pixels) of 8-connected segments in the resulting binary image are counted. The initial threshold value can be determined, for instance, from the grey-level histogram of the image (Luijendijk, 1991). Because an approximated area of a single digit can be estimated to be within a certain interval, the threshold value is next adjusted by increasing or decreasing it to maximize the number of segments whose areas fall within the estimated area interval of a single digit.

After thresholding each 8-connected region is considered as a candidate digit segment. Next for each candidate segment the topmost and the lowermost pixel in column direction (i.e. y-direction) is determined, and the (x,y)-coordinates of the pixels are stored into memory. This coordinate information is used to reduce the number of candidate segments to six segments via Randomized Hough Transform (RHT) (Xu, 1990; Kälviäinen, 1994).

The RHT first introduced a random sampling and a convergence mapping mechanics into the conventional Hough Transform method. For detecting curves which can be expressed by N-parameter equation $f(a_1,\dots,a_N,x,y)$ the basic idea of the RHT is the following. First the set P is formed from edge points of a binary image. Then N points (x_i,y_i) are picked randomly from the set P and the parameter space point $(a_1,\dots,a_N)$ is solved from the curve equation. The accumulator space is then updated as follows: if there exists a cell $G_i(a_{i1},\dots,a_{iN})$ in the accumulator such that $|a_{ij}-a_j| \leqslant a_{tj}$, $j=1,\dots,N$, then the score of G_i is incremented by one, otherwise a new cell $G(a_1,\dots,a_N)$ is created in the accumulator and its score is set to 1. Parameters a_{tj} are used to determine the matching criterion between the cells in the accumulator space and the calculated parameters. The

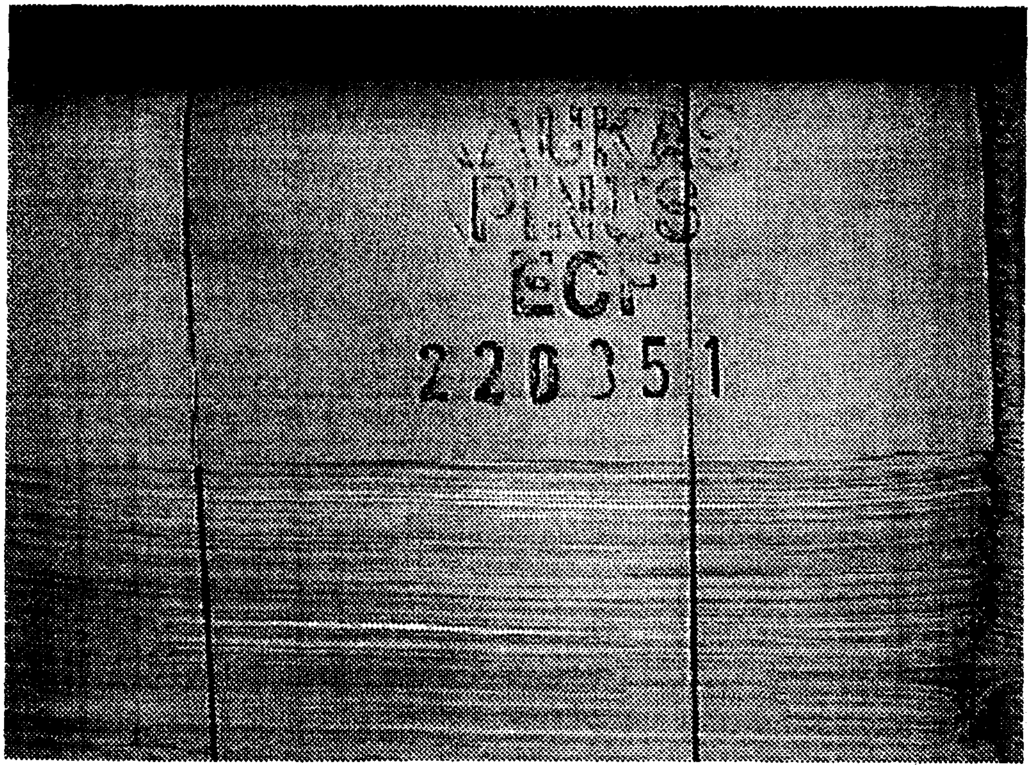

Fig. 2. Digit series and a factory stamp on a pulp bale

smaller the values of a_{tj} are the more accurate results are obtained but also the computation time increases. In RHT the above random sampling procedure is repeated until a predefined global maximum n_t in the accumulator space is detected. The cell which corresponds to this maximum gives the parameters of the detected curve.

The RHT is used to fit two lines through the topmost and the lowermost pixels of candidate segments (Fig. 4). For instance, the line passing through topmost digit segments points is determined by selecting three points randomly from the set of all topmost candidate segment points, and the line parameters are calculated in least square sense. A good parameterization of the line is $x \cos\theta + y \sin\theta = r$, which means that θ and r parameters should be determined. The parameters are accumulated, and the above process is repeated until a local maximum in the accumulator is found; this gives the θ and r parameters of the searched line. When θ is given in degrees and r in pixels, the cell matching parameters θ_t and r_t can be set to 3° and 3, respectively. In this application a suitable value for n_t is 5. When the cell corresponding to the maximum in the accumulator space is found, the final line parameter are determined from the average value of θ and r parameters that hit to the maximum cell. The line that passes through lowermost candidate segment points is determined in a similar manner. Finally, the line that is closer to its six closest candidate segment points is chosen to estimate the orientation of the digit series. When the orientation of the digit series is known, it is relative easy to determine the six digit windows because the estimated average height, width and area of a single digit are known a priori.

In Fig. 4 an example of digit localization is shown. Based on the orientation of a digit series and a single digit it is easy to determine the corresponding digit windows in the original gray level images. However, because of the camera lens causes small variations in the rotation angle of a single digit in a digit series, it is beneficial after localization of digit windows to determine the rotation angles of digits in each gray level digit window. The rotation angle, i.e. orientation of a single digit, ϕ, can be determined via second order

Fig. 3. A digit in a dry (left) and a moist pulp sheet (right).

moments (Teague, 1980)

$$\phi = \frac{1}{2} \tan^{-1} \left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}} \right), \tag{1}$$

where μ_{mn} are the second-order moments computed from the grey-level digit window. With the angle ϕ it is then possible to normalize each digit belonging to the same class to the same orientation, which should improve the digit recognition rates. Two examples of window normalization, i.e., straightening of digits, are shown in Fig. 5. It is important to notice that if there are disturbances, such as ink stains or straps overlapping the digit series, in the borders of the digit window, the value of ϕ does not necessarily give the correct rotation angle for the digit. However, it can be assumed that the absolute value of the rotation angle of the digit within the digit window cannot be greater than $10°$. By normalizing only those windows whose $|\phi|$ is less than $10°$ the effects of these disturbances can partially be avoided.

3. Feature extraction

The feature extraction schema selected must work with the constraints listed in the introduction. To select proper features for classification the following characteristics of the stamped digits should be noticed:

(1) the digits are not necessarily located in the middle of the digit window;
(2) there are often ink stains in the borders of the digit windows disturbing feature extraction; and
(3) some digits are broken into separate pieces and thus consist of several smaller regions.

A quite robust gradient field based image feature extraction schema for recognizing characters from gray level images was proposed by Holder et al. (1988). Gradient fields computed with Sobel masks have also been used for recognizing machine printed characters (Srihari, 1991). In this application gradient fields are also used as basic digit features for classification.

A gradient field for a digit is computed from a grey-level digit window using two band-pass derivative of Gaussian filters (Canny, 1986):

$$G_x(x,y) = G(y)\frac{\partial G(x)}{\partial x} \quad \text{and}$$

$$G_y(x,y) = G(x)\frac{\partial G(y)}{\partial y},$$

where

$$G(z) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-z^2}{2\sigma^2}\right). \tag{2}$$

These filters are sensitive to edges in the direction of their derivative component. The width of the filters, σ, is set to response to low frequencies in image windows to minimize distortion effects of dry and moist digits, small ink stains, and broken digits. Therefore we have set $\sigma = 12.5$. Because the filter outputs change only little in neighboring pixels, digit window is filtered only in 20×20 points, which also speedup the gradient field computations.

To remove the effects of variable illumination conditions, the gradient field of a digit window is normalized by dividing each gradient vector by the maximum magnitude gradient found in the window. The normalized gradient field is then thresholded so that gradient vectors whose magnitude is smaller than 0.3 are discarded (Fig. 6).

To reduce the number of features, gradient histograms for six overlapping regions (see Fig. 7) are finally computed on the basis of normalized and thresholded gradient vectors. The gradient directions are discretized into eight intervals:

$$[(k-1)\cdot 45°, k\cdot 45°], \quad k = 1,\dots,8,$$

and the number of gradients within each interval is counted for each region. The first four overlapping regions minimize the effects of local distortions (e.g. ink stains) and missing parts of digits. The 5th histogram is calculated from the middle section thus preventing distortions effects of the borders of the digit window. The 6th region contains the whole image window

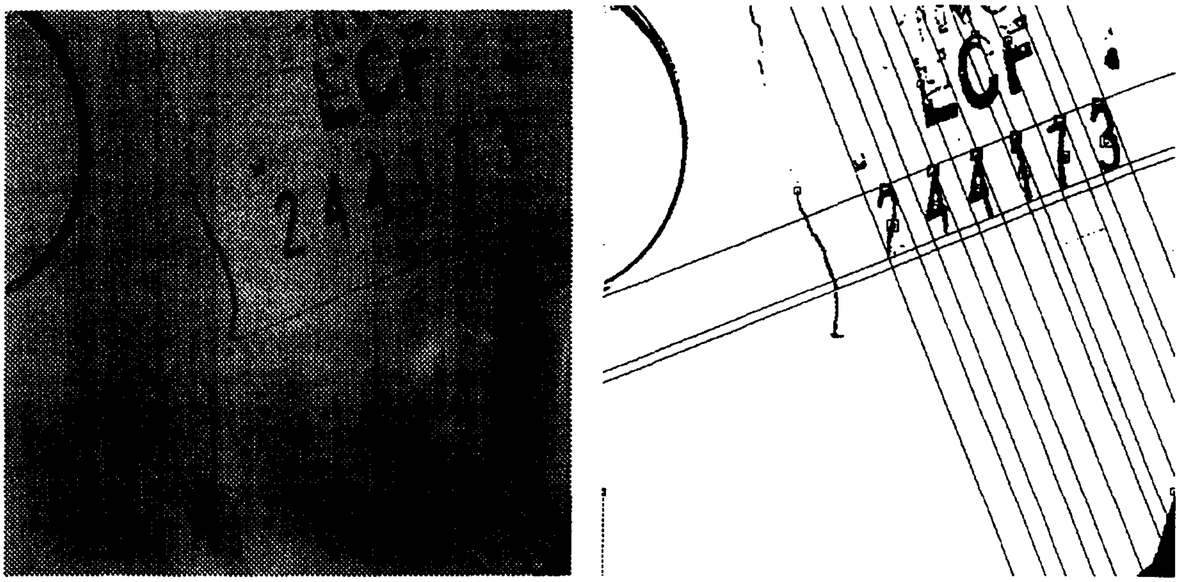

Fig. 4. An example of digit series segmentation.

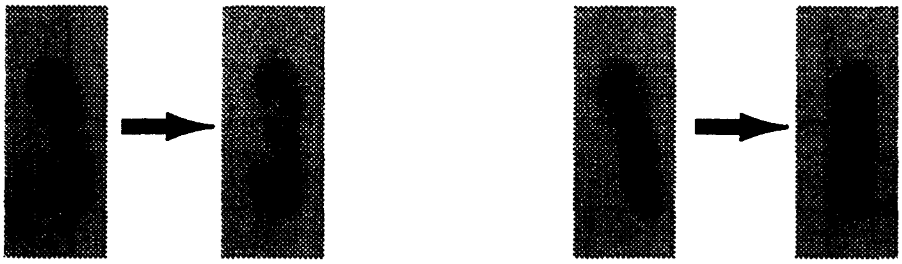

Fig. 5. Straightening of rotated digits within the digit windows.

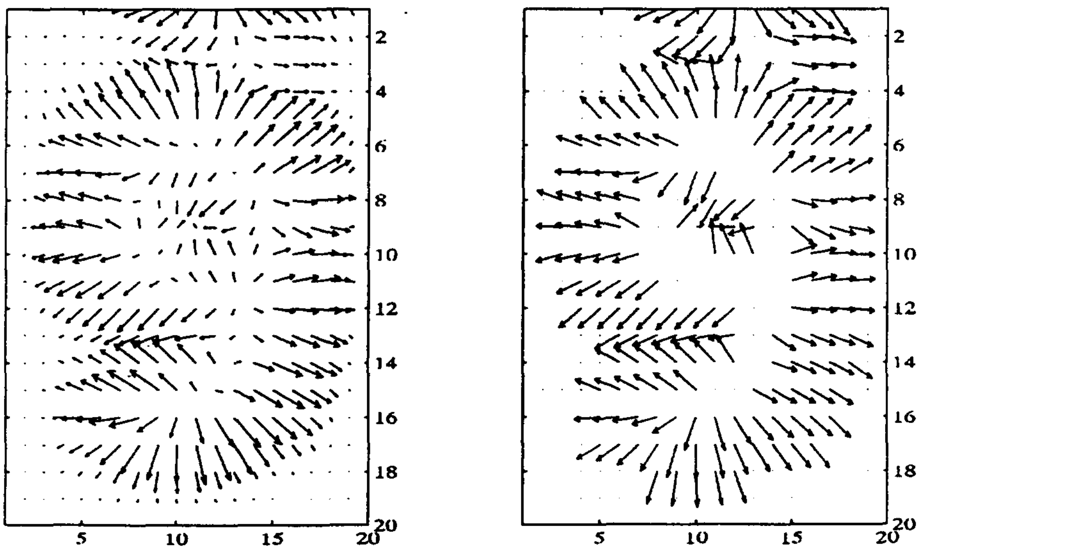

Fig. 6. Original (left), and normalized and thresholded (right) gradient fields.

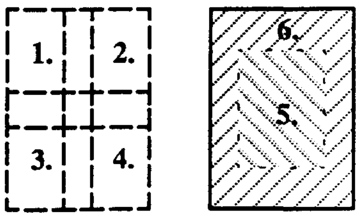

Fig. 7. Regions for gradient histogram calculations.

Table 1
Comparison of classifiers

Classifier	Results
K-NN ($K = 5$)	79%
LVQ	92%
TM	92%
MLP	93%

and provides translation invariant information. When each histogram entry is divided by the maximum entry value of its own histogram, 48 floating point feature values in the range of [0, 1] are obtained for the classification stage.

4. Classification and experimental results

Neural networks (NNs) have shown good performance in many pattern recognition tasks (see e.g. Lippman, 1989; Cun et al., 1989; Lampinen, 1992). Perhaps the most commonly used NN model for classification is multiple layer perceptron (MLP) network (Rumelhart, 1986). For this application a 3-layer perceptron network with 48 input, 16 hidden, and 10 output nodes was chosen as a classifier. The network was trained with backpropagation algorithm with adaptive learning rate and momentum (Vogl, 1988). Because there is no exact rule to determine the number of units in the network, the topology of the network was determined by testing several topologies and choosing the best performing one, i.e., the network which gave best classification results in the test set. The training and test material consists of pulp bale images of 512×512 pixels. In the learning set there were 240 digits and in the test set 120 digits. Table 1 shows the recognition results with the MLP network. The classification result can be regarded as a good one (even a human cannot classify all digits correct). Fig. 8 shows some digits in the test set that were unrecognized.

The recognition result of the MLP network was also compared to other well-know neural and non-neural classifiers. K-Nearest Neighbor (K-NN) classifier with $K = 5$ did not give good results, but plausible results were obtained with Learning Vector Quantization (LVQ) (Kohonen, 1992); the results are almost as good as with the MLP network (Table 1). The LVQ test was made with LVQ_PAK software package (Kohonen, 1992).

An alternative digit recognition scheme that seemed to work well is a simple template matching (TM) method (Ballard and Brown, 1982). In this method one template for each class is generated by calculating average pixel values of grey-level digit windows belonging to the same class. Test digit windows are then compared to templates using correlation variable r:

$$r(X,Y) = \frac{\sum_x \sum_y (\Delta I(x,y) * \Delta T(x,y))}{\sum_x \sum_y (\Delta I(x,y))^2 \sum_x \sum_y (\Delta T(x,y))^2}, \tag{3}$$

in which

$$\Delta I(x,y) = I(x,y) - \bar{I}(x,y),$$
$$\Delta T(x,y) = T(x-X, y-Y) - \bar{T}(x,y)$$

where I is a test digit window, (X,Y) is a displacement for calculating correlation, T is a template window (there are total of 10 templates), and $\bar{I}$ and $\bar{T}$ are the mean grey-level values of pixels in the test and template windows, respectively. The template with the highest correlation gives the class of the digit. One clear drawback of the template matching method is that its computation time is quite high, which makes it unsuitable for this application without any special hardware. However, the obtained classification result is as good as with LVQ method, and only a slightly worse than that of the MLP classifier.

5. Conclusions

This paper has described a computer vision system for digit series recognition on pulp bales. The digit recognition consists of three stages: segmentation of digit series, feature extraction, and classification. Segmentation of digit series is based on image thresholding and Randomized Hough Transform. Digit segmentation produces six digit windows. In feature extraction two band-pass derivative of Gaussian filters are

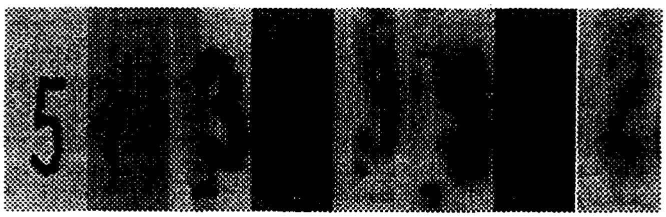

Fig. 8. Some digits that the MLP classifier did not recognize.

used, and the resulting gradient field histograms are used after normalization in classification of digits.

The classification result of the MLP network (digits are recognized 93% correct) can be recarded as satisfying one. The computation time for on-line digit series recognition (3–5 seconds with ordinary 486 micro computer) is also adequate for this application. The quality of stamped digits is such that recognition rate of 100% cannot be obtained. However, by improving the stamping process and by wrapping both dry and moist bales with try sheets, the digit recognition rate should raise from 93%. The recognition results should also improve by increasing the size of the learning set.

References

Ballard, D.H. and C.M. Brown (1982). *Computer Vision*. Prentice-Hall, Englewood Cliffs, NJ.

Canny, J. (1986). A computational approach to edge detection. *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (6), 679–698.

Cun, Y.L., L.D. Jackel, B. Boser, J.S. Denker, H.P. Graf, I. Guyon, D. Henderson, R.E. Howard and W. Hubbard (1989). Handwritten digit recognition: applications of neural network chips and automatic learning. *IEEE Commun. Mag.*, Nov. 1989, 41–46.

Denker, J.S., W.R. Gardner, H.P. Graf, D. Henderson, R.E. Howard, W. Hubbard, L.D. Jackel, H.S. Baird and I. Guyon (1989). Neural network recognizer for hand-written zip code digits. In: D. Touretzky, Ed., *Advances in Neural Information Processing Systems*. Morgan Kaufmann, San Mateo, CA, 323–331.

Holder, S. and J. Dengler (1988). Font- and sizeinvariant character recognition with grey level images features. *Proc. 9th Internat. Conf. on Pattern Recognition*, Rome, Italy, Vol. I, 252–254.

Kohonen, T., J. Kangas, J. Laaksonen and K. Torkkila (1992). LVQ_PAK: a program package for the correct application of Learning Vector Quantization algorithms. *Proc. Internat. Joint Conf. on Neural Networks*, Baltimore, MD, Vol. I, 725–730.

Kälviäinen. H. (1994). Randomized Hough Transform: new extensions. Doct. Tech. Thesis, Lappeenranta University of Technology, Department of Information Technology, Finland.

Lampinen, J. (1992). Neural pattern recognition: distortion tolerance by Self-Organizing Maps. Doct. Tech. Thesis, Lappeenranta University of Technology, Lappeenranta, Finland.

Lippman, R.P (1989). Pattern classification using neural networks. *IEEE Commun. Mag.*, Nov. 1989, 47–64.

Luijendijk, H. (1991). Automatic threshold selection using histogram based on the count of 4-connected regions. Pattern Recognition Lett. 12 (4), 219–228.

Mori, S., C.Y. Suen and K. Yamamoto (1992). Historical review of OCR research and development. *Proc. IEEE* 80 (7), 1029–1058.

Ohlander, R., K. Price and D.R. Reddy (1978). Picture segmentation using a recursive region splitting method. *Computer Graphics and Image Processing* 1, 169–182.

Rumelhart, D.E., R.J. Hinton and R.J. Williams (1986). Learning internal representations by error propagation. In: D.E. Rumelhart and J. McCelland, Eds., *Parallel Distributed Processing*. MIT Press, Cambridge MA, 318–362.

Srihari, S.N. (1991). Recognition of handwritten and machine printed text for postal address interpretation. *Pattern Recognition Lett.* 14 (4), 291–302.

Tappert, C.C., C.Y. Suen and T. Wakara (1990). The state of the art in on-line handwritten recognition. IEEE Trans. Pattern Anal. Mach. Intell. 12 (8), 787–808.

Teague, M.R. (1980). Image analysis via the general theory of moments. *J. Opt. Soc. Amer.* 70 (8), 920–930.

Vogl, T.P., J.K. Mangis, A.K. Rigler, W.T. Zink and D.L. Alkon (1988). Accelerating the convergence of the backpropagation method. *Biol. Cybernet.* 59, 257–263.

Xu, L., E. Oja and P. Kultanen (1990). A new curve detection method: Randomized Hough Transform (RHT). *Pattern Recognition Lett.* 11 (5), 331–338.

Zucker, S.W. (1976). Region growing: childhood and adolescence. *Computer Graphics and Image Processing* 6, 382–399.