MORPHOLOGICAL SYSTEMS FOR CHARACTER IMAGE PROCESSING AND RECOGNITION

Ping-Fai Yang and Petros Maragos
Division of Applied Sciences, Harvard University, Cambridge, MA 02138, USA

ABSTRACT

In this paper we applied min/max signal operations, common in morphological image analysis, to both feature extraction and classification of character images. We propose a system that computes an improved version of the *morphological shape-size histogram*, which reduces sensitivity to stroke thickness, size and rotation. For pattern classification we introduce the class of *min-max classifier*, which generalizes Boolean DNF functions for real-valued inputs. A *Least Mean Square* (LMS) algorithm was used for practical training of min-max classifiers. Experimental results show that min-max classifiers were able to achieve error rates that are comparable to neural networks trained using back propagation. The main advantage of the min-max/LMS algorithm is its simplicity and faster speed of convergence.

1. INTRODUCTION

Two separate stages of operations are usually distinguished in character recognition systems. In the first one the raw input image is preprocessed and features are extracted. They are designed to capture important information in the input character. Good feature vectors can greatly enhance the operation of the second stage, the classifier. In this paper we describe the application of min/max signal operations, common in morphological image analysis, to both of these stages.

For feature extraction we provide improved versions of the *morphological shape-size histogram* [1, 2] (also called "pattern spectrum in [2]). The shape-size histogram is an important tool for shape representation in morphological image analysis. It is derived from the morphological size distributions [3, 1] (also called "granulometries" in [3]), which were introduced by Matheron as an axiomatic definition of "size". The improvements we made reduced the sensitivity of the shape-size histogram to stroke thickness, size, and orientation, thus making it a better feature for character images.

For character classification we employed *min-max classifiers* which we introduced in [4]. These functions generalize Boolean functions by replacing AND/OR with MIN/MAX operations, thereby allowing real valued inputs within the range $[0,1]^d$. Another motivation for working with the min-max classifiers is their close relation to a large class of nonlinear signal/image operators known as morphological filters, which are defined via min-max operations on their inputs. As discussed in [1] and [5], these min-max morphological operators can be applied to a broad variety of feature extraction and shape analysis/detection tasks in images or arbitrary geometrical objects.

This research was supported by a National Science Foundation Presidential Young Investigator Award under Grant MIPS-86-58150 with matching funds from Xerox, and in part by the ARO Grant DAALO3-86-K-0171 to the Brown-Harvard-MIT Center for Intelligent Control Systems.

2. FEATURE EXTRACTION VIA SHAPE-SIZE HISTOGRAMS

The shape-size histogram is obtained by smoothing the image I via opening and closing ($\bullet$) filters by structuring elements

$$nB = \underbrace{B \oplus \ldots \oplus B}_{n} \text{ for } n \geq 1; \qquad 0B = \{(0,0)\}$$

that are multiscale dilations ($\oplus$) of a unit-size convex shape B. The size parameter is n. The difference in areas of the image I smoothed at successive sizes gives the shape-size histogram:

$$SH_I(n,B) = A[I \bullet (n+1)B] - A[I \bullet nB], \text{ for } n \geq 0 \quad (1)$$

We employed only the closing part of the shape-size histogram since it gives information about holes and cavities in the thin character images.

The direct application of this shape descriptor as a feature may not, however, be desirable. For instance, Fig. 1 illustrates some problems that may occur. The shape-size histograms calculated using a 2×2 square structuring element for images are very sensitive to stroke thickness, size, and orientation. To reduce the effects of these on the feature, we employed a system for extracting enhanced shape-size histograms which consists of three steps:

1. thin the character;
2. find the area functions $A[I \bullet nB]$ of the thinned images;
3. calculate the *normalized shape-size histogram* from the area function.

2.1. Thinning

Before thinning, the character is dilated with a 3×3 pixel square to fill up small gaps. We employed the *homotopic*

0-7803-0946-4/93 $3.00

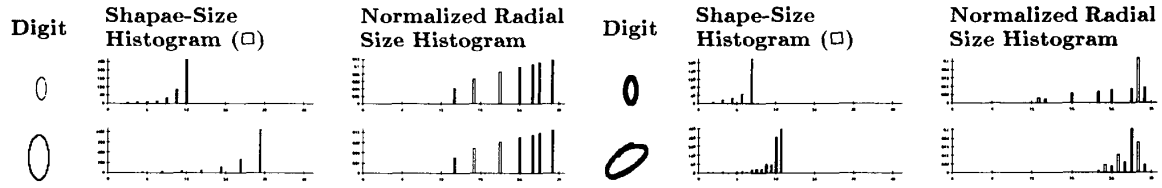


Figure 1: Sample zeroes together with their pattern spectra. The second and fifth columns are size histograms calculated using a 2×2 square. The third and sixth columns show better results using the normalized radial size histogram.

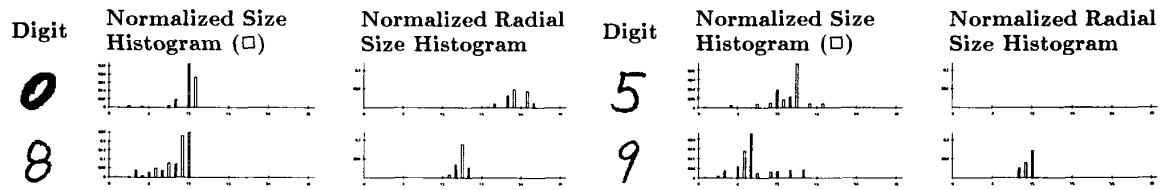


Figure 2: Examples of features employed in the experiments. The second and fifth columns contain normalized size histograms calculated using a 2×2 square while the third and sixth are normalized radial size histograms.

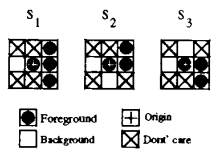


Figure 3: Structuring elements used in the thinning operations.

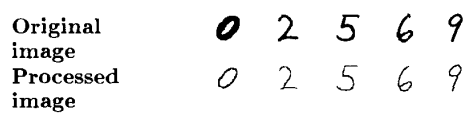


Figure 4: Examples of images before and after homotopic thinning.

thinning algorithm described in [6]. The basic operation is the *thinning* operator

$$I \mapsto I \odot \{S_i\} = I \setminus \bigcup_i (I \otimes S_i).$$

The set mapping $I \otimes S_i = (I \ominus A_i) \cap (I^c \ominus B_i)$ is called the *hit-miss transform* [1]. For the structuring elements $S_i = (A_i, B_i)$ we employed (see Fig. 3), the hit-miss transform detects the boundary of the image. Therefore, the set difference ($\setminus$) removes parts of the image ($\bigcup_i (I \otimes S_i)$) while preserving four-connectivity of the foreground I. In each step of the algorithm, the image is thinned used $\{S_1, S_2, S_3\}$ and three rotated versions

$$I \mapsto (((I \odot \{S_i\}) \odot \{S_i\}_{rot90^\circ}) \odot \{S_i\}_{rot180^\circ}) \odot \{S_i\}_{rot270^\circ}$$

This is iterated under convergence (i.e. no change in the image). Figure 4 shows some sample digits before and after homotopic thinning.

2.2. Normalization of Area Function

After the thinning operation, the area functions are calculated for $n = 0, \ldots, N_b$. The maximum size N_b is determined by the relative size of the image with respect to the structuring element B. Suppose the dimension of the bounding box of the image I is $i_w \times i_h$, and that for the structuring element $b_w \times b_h$; then $N_b = \max(i_w - b_w, i_h - b_h)$. Intuitively, the parameter N_b is the structuring element size n that is required to produce the convex hull of the largest image that fits within a box of size $i_w \times i_h$ (i.e. the box itself). Since we have features of varying lengths, it is convenient in practice to rescale them to one uniform size. To facilitate this we first interpolate $A[I \bullet nB]$, which is defined for integer scales n, to a piecewise constant function defined over the continuous size parameters $0 \leq s \leq N_b$. The domain of the interpolated area function $A(s)$ is then rescaled to $[0, M]$ and resampled at integer values $m = 0, \ldots, M$. The net effect of these operations is expressed in the formula:

$$A'(m) = A\left(I \bullet \left\lfloor \frac{mN_b}{M} \right\rfloor B\right), \text{ for } m = 0, \ldots, M$$

The normalized shape-size histogram $SH'_I(m, B)$ is calculated from $A'(m)$ as in Eq. (1) with the A replaced by A'. Finally, the amplitude is normalized by dividing $SH'_I(m, B)$ with the area of the maximally closed image. Examples of this normalized size histogram is shown in Fig. 2.

Further improvement to reduce sensitivity to rotation is provided by the *radial size histogram* [2] (see Fig. 1, 2). It replaces the closing by a 2-D element B with an intersection of closings by the four directed vectors $\{\rightarrow, \nearrow, \uparrow, \nwarrow\}$. Figure 1 shows the remarkable improvement of this feature over the plain shape-size histogram. Experience has also shown that the radial size histogram is useful for detecting

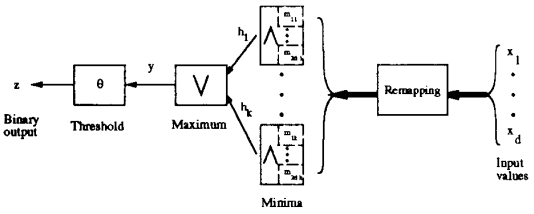


Figure 5: The connections between different modules within a min-max classifier.

holes in the digit. Examples of radial size histogram for different digits are shown in Fig. 2.

3. MIN-MAX CLASSIFIERS

3.1. Definition

The signal flow diagram of a min-max classifier is shown in Fig. 5. Its first layer consists of k minima functions

$$h_j = \min_{i \in I_j} \ell_i, \quad j = 1, \ldots, k$$

calculated over a set of *literals* (ℓ_i) equaling either the input x_i or its complement $1 - x_i$. It is more convenient to introduce the complemented variables by remapping the input $\vec{x} \in [0,1]^d$ to a $2d$-dimensional one that includes both uncomplemented and complemented variables. Therefore, h_j's are uncomplemented minima of these $2d$ remapped variables. The subsets I_j are controlled via *mask* variables m_{ij} in the following fashion

- x_i is included in I_j if $m_{ij} \geq 0$,
- x_i is excluded from I_j if $m_{ij} < 0$.

In the second stage, the maximum function outputs the largest of all the h_j's. It is then thresholded at θ to obtain a binary output. Therefore, a min-max function is fully specified by the $2dk$ mask variables and the threshold θ.

3.2. Training

For training the min-max classifier we employed a *Least Mean Square* (LMS) algorithm [7]. The basic update equation is

$$\vec{p}(t+1) = \vec{p}(t) - 2\mu(z(t) - d(t))\frac{\partial z(t)}{\partial \vec{p}} \tag{2}$$

where $\vec{p}$ are the parameters that define the system (for the min-max classifier $\vec{p} = (m_{ij}, \theta)$). The speed of convergence is controlled by the convergence parameter μ.

Specifying Eq. 2 to the min-max classifier, we have to calculate two types of partial derivatives, $\frac{\partial z}{\partial \theta}$ and $\frac{\partial z}{\partial m_{ij}}$. The first one can be calculated using the definition of the thresholding function, whereupon one obtains

$$\frac{\partial z}{\partial \theta} = \begin{cases} -\frac{1}{2\beta} & \text{if } |y - \theta| \leq \beta\ , \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

We have approximated the derivative of the threshold, a delta function ($\delta(y - \theta)$), with a finite impulse of width β. The derivative with respect to the mask variables are calculated using the chain rule

$$\frac{\partial z}{\partial m_{ij}} = \frac{\partial z}{\partial y}\frac{\partial y}{\partial h_j}\frac{\partial h_j}{\partial m_{ij}}$$

The expression on the right hand side involves derivatives of rank order operations. These are calculated by using *implicit* definitions for the maximum and minimum functions. Salembier [8] also used a similar approach in his work on adaptive morphological filtering. For instance, the inputs and output of the maximum function satisfy

$$G(y, h_1, \ldots, h_k) = \sum_{j=1}^{k} \{U_3(y - h_j) - 1\} + \frac{G_e}{2} = 0. \tag{4}$$

In what follows we require two types of step functions

$$U_2(x) = \begin{cases} 1 & \text{if } x \geq 0, \\ 0 & \text{if } x < 0. \end{cases} \qquad U_3(x) = \begin{cases} 1 & \text{if } x > 0, \\ \frac{1}{2} & \text{if } x = 0, \\ 0 & \text{if } x < 0. \end{cases}$$

The variable G_e is the number of inputs h_j that are equal to the output y. Using the fact $\frac{dG}{dh_j} = 0$, and again approximating delta functions by finite impulses, we obtain

$$\frac{\partial y}{\partial h_j} \approx \begin{cases} \frac{1}{N_{\max}} & \text{if } 0 \leq y - h_j \leq \beta \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

where $N_{\max}$ is the number of h_j's such that $y - h_j \leq \beta$. The result for the minimum function is similar. Its implicit representation is

$$L(h_j, x_1, \ldots, x_d, m_{1j}, \ldots, m_{2dj}) = \sum_{i=1}^{2d} \{U_2(m_{ij})[U_3(x_i - h_j) - 1]\} + \frac{L_e}{2} = 0. \tag{6}$$

The quantity L_e is the number of inputs in I_j that are equal to the minimum. After some simple calculations we find the approximate derivative:

$$\frac{\partial h_j}{\partial m_{ij}} \approx \begin{cases} -\frac{1}{2N_{\min}} & \text{if } |m_{ij}| \leq \beta \text{ and } = x_i \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where $N_{\min}$ is the number of inputs such that $m_{ij} \geq 0$ and $|x_i - h_j| \leq \beta$.

4. EXPERIMENTAL RESULTS

Three different sets of data were used in our experiments. In increasing order of difficulty: the first set consisted of 0's and 1's, the second set 0's and 6's, and the third 6's and 8's. For each digit we extracted 20 components of normalized radial size histogram and 20 components of Fourier descriptors [9] forming feature vectors of 40 components. There are 600 of each type of digits in the training data set and 200 of each in the test set. Results for the min-max classifiers are shown in Table 1. To gauge the performance of the min-max classifiers, we used neural networks trained using the back propagation algorithm [10]. The networks have one input layer, one hidden layer and one output layer. The number

Min-Max

Digits	% error (train)	% error (test)
0,1	0.083	0.25
0,6	1.783	1.35
6,8	15.35	15.05

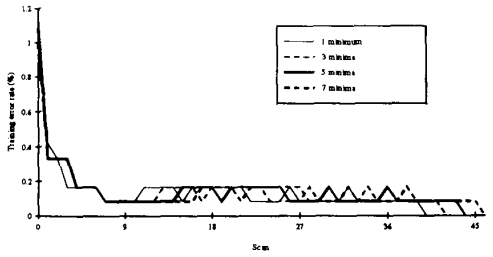
Neural Networks

Digits	% error (train)	% error (test)
0,1	0.083	0
0,6	1.517	1.5
6,8	18.4	18.15

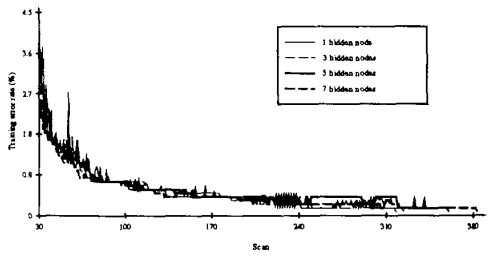
Table 1: The top table shows typical error rates obtained in our experiments using min-max classifier with LMS algorithm. For comparison the lower table shows error rates using neural networks on the same data sets. The features are normalized radial histogram with Foruier descriptors.

of hidden nodes is kept the same as the number of minima in the min-max classifier. Typical error figures for the neural networks are also shown in Table 1. Note that the min-max classifiers were able to achieve error rates that are comparable to or better than those generated using neural networks. The first advantage of the min-max/LMS algorithm is its faster speed of convergence compared to the neural network/back propagation approach. For example, Fig. 6 shows typical convergence plots of LMS and back propagation. The former takes about 10 times less steps to obtain a minimum error rate. Another advantage of the min-max classifier is its simplicity, since it involves only min-max operations. Neural networks, however, require calculation of the computational costly logistic function.

The experimental results on handwritten recognition provides evidence on the practical utility of morphological systems for feature extraction and the min-max classifier. Our immediate goal is to develop a full fledged character recognition system by generalizing the min-max classifier to multiple classes.



(a)



(b)

Figure 6: Convergence plots for min-max/LMS and neural net/back propagation for the 0-1 classification problem. (a) Typical convergence plots for the LMS algorithm. The minimum error rate (0%) is attained within 40 scans. (b) Convergence plots for the Back Propagation on the same data set. After 40 scans the typical error rate was 2%. The minimum error rate (0.083%) was achieved around 400 scans.

REFERENCES

[1] J. Serra, *Image Analysis and Mathematical Morphology.* New York: Academic Press, 1982.

[2] P. Maragos, "Pattern spectrum and multiscale shape representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, pp. 701–716, 1989.

[3] G. Matheron, *Random Sets and Integral Geometry.* New York: Wiley, 1974.

[4] P.-F. Yang and P. Maragos, "Learnability of min-max pattern classifiers," in *SPIE Vol. 1606 Visual Communications and Image Processing '91: Image Processing*, pp. 294–308, 1991.

[5] P. Maragos and R. W. Schafer, "Morphological systems for multidimensional signal processing," *Proceedings of the IEEE*, vol. 78, no. 4, pp. 690–710, 1990.

[6] D. S. Bloomberg, "Connectivity-preserving morphological image transformations," in *SPIE Vol. 1606 Visual Communications and Image Processing '91: Image Processing*, 1991.

[7] B. Widrow and S. D. Stearns, *Adaptive Signal Processing.* Englewood Cliffs, NJ: Prentice-Hall, 1985.

[8] P. Salembier, "Structuring element adaptation for morphological filters," *Journal of Visual Communication and Image Representations*, vol. 3, no. 2, pp. 115–136, June 1992.

[9] C. T. Zahn and R. Z. Roskies, "Fourier descriptors for plane closed curves," *IEEE Computer*, vol. C-21, pp. 269–281, Mar. 1972.

[10] D. E. Rumelhart, J. L. McClelland, and the PDP Research Group, *Parallel distributed processing: explorations in the microstructure of cognition.* Cambridge, MA: MIT Press, 1986.