Pergamon

Pattern Recognition, Vol. 28, No. 8, pp. 1237 1249, 1995
Elsevier Science Ltd
Pattern Recognition Society
Printed in Great Britain

0031-3203(95)00001-1

WAVELET DESCRIPTORS FOR MULTIRESOLUTION RECOGNITION OF HANDPRINTED CHARACTERS

PATRICK WUNSCH and ANDREW F. LAINE*
Computer and Information Sciences and Engineering Department, University of Florida, Gainesville, FL, U.S.A.

(*Received* 13 *August* 1993; *in revised form* 10 *January* 1995; *received for publication* 2 *February* 1995)

Abstract—We present a novel set of shape descriptors that represents a digitized pattern in a concise way and that is particularly well-suited for the recognition of handprinted characters. The descriptor set is derived from the wavelet transform of a pattern's contour. The approach is closely related to feature extraction methods by Fourier series expansion. The motivation to use an orthonormal wavelet basis rather than the Fourier basis is that wavelet coefficients provide localized frequency information, and that wavelets allow us to decompose a function into a multiresolution hierarchy of localized frequency bands. We describe a character recognition system that relies upon wavelet descriptors to simultaneously analyze character shape at multiple levels of resolution. The system was trained and tested on a large database of more than 6000 samples of handprinted alphanumeric characters. The results show that wavelet descriptors are an efficient representation that can provide for reliable recognition in problems with large input variability.

Shape representation   Wavelets   Multiresolution analysis   OCR   Neural networks

1. INTRODUCTION

Feature or descriptor extraction is a crucial processing step of shape recognition systems. In fact, what most distinguishes different recognition methodologies is the type of features selected for representation. In general, "good" features must satisfy the following requirements: first, *intraclass variance* must be small, which means that features derived from different samples of the same pattern class should be close (e.g. numerically close if numerical features are selected). Secondly, *interclass separation* should be large, i.e. features derived from samples of different classes should differ significantly.

In this paper we present a novel set of features that represents patterns in a concise way and that is particularly well-suited for recognition of unconnected handprinted characters. The recognition of handprint is an important subproblem of optical character recognition (OCR) with applications such as automatic ZIP-code reading or understanding annotations in technical drawings. The difficulty with handwriting recognition is large intraclass variance due to the shape variations caused by the distinct writing styles of individuals. Obviously, there is no tractable mathematical model that can describe such shape variations, and so it is impossible to find features that are invariant with respect to writing style. Nor can one formally prove that a particular feature set will exhibit small intraclass variance. Therefore, one can only aim at finding features that experimentally prove reasonably insensitive to shape distortions caused by individual writing style and that at the same time maintain the ability to separate samples of different classes.

Experience shows that shapes that "look alike" have similar low-frequency components in the Fourier domain. Low-frequency Fourier coefficients reflect the basic shape of a function, whereas the high frequency components represent the details. As far as handprinted character recognition is concerned, low-frequency components of a character turn out to be less sensitive to writing style variations. In fact, this has been the principal reason for the popularity of Fourier descriptors for handprinted character recognition.[1-4]

There is, however, a serious disadvantage inherent in Fourier descriptors. The frequency information obtained from a Fourier transform is global. Intuitively, a more localized frequency representation should be more effective for pattern recognition.

In the last few years, wavelet basis functions have become popular for accomplishing localized frequency analysis.[5,6] Unlike traditional short-time Fourier transforms, wavelet transforms have the capability of variable time-frequency localization. Furthermore, orthonormal wavelets with finite support provide a powerful mathematical tool for decomposing a function into a multiresolution hierarchy of different localized frequency channels. Such a decomposition allows us to simultaneously analyse a function at several levels of resolution and favors coarse-to-fine recognition strategies similar to those known to exist in the human visual system.[7]

* Author to whom correspondence should be addressed.

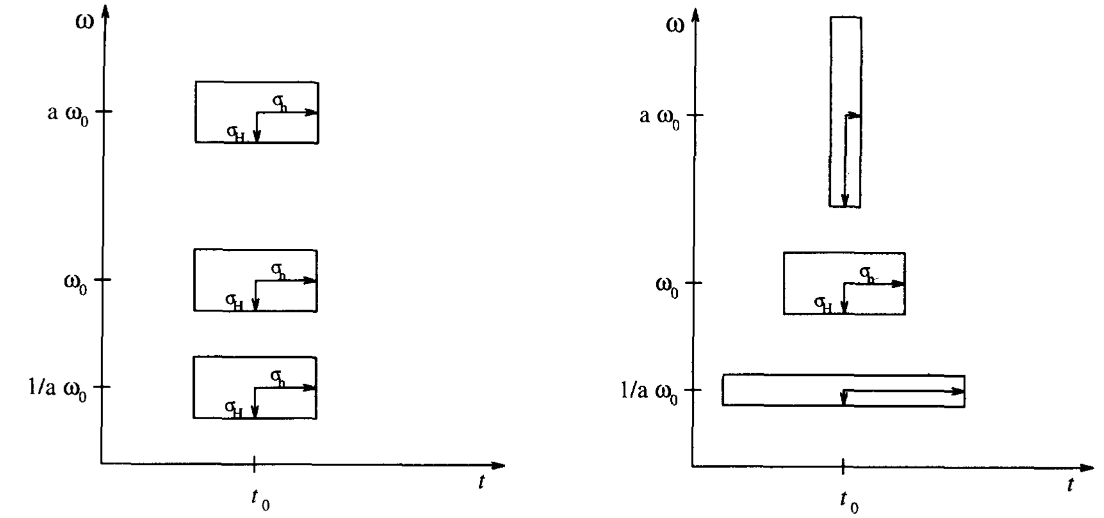

Fig. 1. Time-frequency resolutions for the STFT (left) and the wavelet transform (right).

This paper is organized as follows. Section 2 provides a brief overview of the major properties of a wavelet transform and shows how a set of multiresolution shape descriptors can be derived from the transform coefficients of a pattern contour. Next, we describe a character recognition system that uses a set of neural networks trained with wavelet descriptors to simultaneously analyse shape at multiple levels of resolution. In Section 4, we report the recognition results we obtained when training and testing the system with a large database of handprinted character samples. Finally, we summarize the significant results of this study and outline potential areas for future research.

2. WAVELET DESCRIPTORS FOR MULTIRESOLUTION SHAPE REPRESENTATION

In this section we show how to derive a set of multiresolution shape descriptors from wavelet coefficients of a pattern boundary. First we review the most important properties of a wavelet transform and describe how to derive a multiresolution hierarchy of localized frequency information by orthonormal wavelet decomposition. For a more rigorous treatment the reader is directed to papers by Mallat[8-10] and books by Chui[5] and Daubechies.[6] Since wavelet representations are not shift-invariant, we present appropriate normalization procedures in Section 2.2. We then show how an approximation of an original contour can be computed from a set of wavelet descriptors.

In order to carry out wavelet analysis, an initial functional representation of the input pattern is needed. Unlike most recognition problems, in the case of characters, the pattern contour contains all the information that is necessary for classification. Furthermore, contour representations require much less data than greylevel images.

A pattern contour can be represented as a closed parametric curve c in the complex plane $\mathbf{C}$, i.e.

$$c(t) = x(t) + jy(t), \quad 0 \le t \le T, \tag{1}$$

where j denotes the imaginary unit. Since a closed curve can be retraced infinitely often, c is a periodic function with period T. The wavelet transform of a curve c can be taken independently for each component

$$WTc(u) = WTx(u), \quad + j\,WTy(u). \tag{2}$$

2.1. *The wavelet transform*

2.1.1. *The continuous wavelet transform.* A shortcoming of the Fourier transform is that the frequency information it provides is global. It cannot be determined *where* the function exhibited a particular frequency characteristic. To overcome this limitation, short-time Fourier (STFT) transforms have been developed. The idea is to multiply a function to be analysed with a window function before computing its Fourier transform. A function* $h \in L^2(R)$ is called a window function if its "center" μ_h and "radius" σ_h exist:

$$\mu_h = \frac{1}{\|h\|_2^2} \int_{-\infty}^{\infty} t|h(t)|^2\,dt \tag{3}$$

$$\sigma_h = \frac{1}{\|h\|_2^2} \left\{ \int_{-\infty}^{\infty} (t - \mu_h)|h(t)|^2\,dt \right\}^{1/2}. \tag{4}$$

If $h(t)$ is a Gaussian, the corresponding STFT is called a *Gabor transform*. The difficulty with STFTs is that a window $h(t)$ of fixed size is accompanied by a fixed-size window $H(\omega)$ in the frequency domain.† However, what is actually needed is a long window to analyse slowly varying low-frequency components, and a narrow one to detect high-frequency bursts. A STFT imposes a fixed trade-off between time and frequency resolution within the time-frequency plane (Fig. 1).

* $L^2(R)$ denotes the space of measurable functions f that satisfy $\|f\|_2^2 = \int_{-\infty}^{\infty} |f(t)|^2\,dt < \infty$.

† A capital letter denotes the Fourier transform space, i.e. $h(t)$ and $H(\omega)$ are a FT-pair.

The wavelet transform, on the other hand, is based on dilations and translations of a prototype function $\psi \in L^2(R)$. These basis functions have short time resolution for high frequencies and longer time resolution for low-frequencies (Fig. 1). Such flexible time-frequency resolution is well suited for localized frequency analysis.

Specifically, a function $\psi \in L^2(R)$ is said to be a wavelet if it satisfies the following admissibility condition needed to obtain the inverse of the wavelet transform.

$$C_\psi = \int_0^\infty \frac{|\Psi(\omega)|^2}{\omega} d\omega < \infty. \tag{5}$$

Again, Ψ denotes the Fourier transformed of ψ. A wavelet basis is given by shifted and dilated versions of a basic wavelet ψ

$$\psi_{a,b} = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right). \tag{6}$$

The continuous wavelet transform (CWT) of a function $f \in L^2(R)$ is defined as

$$\mathscr{W}f(a,b) = \int_{-\infty}^{\infty} f(t)\psi_{a,b}\, dt = \langle f, \psi_{a,b} \rangle. \tag{7}$$

Hence the CWT decomposes a function f using a family of functions that are dilations and translations of some basic wavelet ψ. It can be shown that the transform described above provides frequency information of $F(\omega)$ within the frequency window‡ $[\frac{1}{a}(\mu_\Psi - \sigma_\Psi); \frac{1}{a}(\mu_\Psi + \sigma_\Psi)]$ which is localized in the time interval $[b + a(\mu_\psi - \sigma_\psi); b + a(\mu_\psi + \sigma_\psi)]$. Figure 1 (right) shows that the time-frequency window defined by the intervals above is narrow in time for high center frequencies μ_Ψ/a and wide for small center frequencies, while the area of the window remains constant $(4\sigma_\psi\sigma_\Psi)$.

For any ψ satisfying the admissibility condition an original function can be exactly reconstructed by the following inversion formula

$$f(t) = \frac{1}{C_\psi} \int_0^\infty \int_{-\infty}^{\infty} \mathscr{W}f(a,b)\psi_{a,b}(x)\, db \frac{da}{a^2}. \tag{8}$$

For practical purposes, the parameters, a and b are discretized. Next, we shall consider a compl.e representation built from a wavelet ψ such that the set $\mathscr{B} = \{\psi_{j,k}(x) = \sqrt{2^j}\psi(2^j x - k)\colon j, k \in \mathbf{Z}\}$ is an orthonormal basis of $L^2(R)$, i.e. the shift and scale parameters will be restricted to the dyadic values $a = \frac{1}{2^j}$ and $b = \frac{k}{2^j}$. These wavelets will decompose a function into a multiresolution hierarchy of localized frequency channels.

2.1.2. *Orthonormal wavelets and multiresolution decompositions.* The concept of multiresolution analysis is mathematically formalized as a nested sequence of subspaces $V_{2^j} \in L^2(\mathbf{R})$. A function $f \in L^2(\mathbf{R})$ is represented as a limit of successive approximations, each of which is computed by projecting f onto some V_{2^j}. The sequence of subspaces V_{2^j} must satisfy the following properties.

- Containment:

$V_{2^{j-1}} \subset V_{2^j}$ $\forall j \in \mathbf{Z}$ i.e. sequence $\{V_{2^j}\}$ is nested. This implies that a function approximated at resolution 2^j contains all the information of its lower resolution approximations.

- Completeness:

$\cap_{m\in\mathbf{Z}} V_m = 0$ $\cup_{m\in\mathbf{Z}} V_m = L^2(R)$. This property implies that if the resolution is increased to ∞, the approximation converges to the original function, whereas the approximated function converges to zero as the resolution approaches zero.

- Scaling property:

$f(x) \in V_{2^{j-1}} \Leftrightarrow f(2x) \in V_{2^j}$. This property states that the approximated functions are derived by scaling each other by the ratiò of their resolution levels.

Given such a sequence of subspaces, the approximation of a function f at resolution 2^j is defined as the projection of f onto V_{2^j}. Mallat[10] showed that for any sequence of subspaces satisfying the properties listed above, there exists a unique function $\phi \in L^2(R)$, called a *scaling function*, whose translations at scale 2^j form an orthonormal basis of V_{2^j}. In general ϕ is a lowpass filter and the multiresolution approximation of a function f is a sequence of smoothed versions of f. Let W_{2^j} be the orthogonal complement of V_{2^j} in $V_{2^{j+1}}$. Each scaling function has an associated wavlet function $\psi \in L^2(R)$ whose dilations and translations provide an orthonormal bases of W_{2^j}. Figure 2 shows an example of such a pair. Each $f_{2^N} \in V_{2^N}$ can then be decomposed by

$$f_{2^N} = g_{2^{N-1}} + g_{2^{N-2}} + \cdots + g_{2^{N-J}} + f_{2^{N-J}} \tag{9}$$

where $g_{2^j} \in W_{2^j}$ and $f_{2^j} \in V_{2^j}$. The dilations of ψ can be regarded as bandpass filters and the coefficient sequences of the g_{2^j} provide localized spectral information of f within the frequency bands $[2^j(\mu_\Psi - \sigma_\Psi); 2^j(\mu_\Psi + \sigma_\Psi)]$. The coefficients of the g_{2^j} can hence be interpreted as high-frequency details that distinguish the approximations of f at two subsequent levels of resolution. On the other hand, $f_{2^{N-J}}$, represents a coarser approximation of f_{2^N}.

Let $A_{2^j}f$ denote the operator that computes the approximation of f at resolution 2^j, i.e. that projects f onto V_{2^j}. Let D_{2^j} be the operator that computes the projection of f onto the subspaces W_{2^j}. If f is a discrete function, the decomposition described above can be achieved by successive convolutions of f with discrete filters, followed by resampling the approximated function by a coarser grid. Figure 3 visualizes this procedure for one level of analysis.

The resulting set of coefficients

$$WT_{2^J}\{f\} = (A_{2^{-J}}f, (D_{2^j}f)_{-J \le j \le -1}) \tag{10}$$

forms a pyramidal structure and is called a multi-

‡ μ_Ψ and σ_Ψ are, respectively, the "center" and "radius" of $\Psi(\omega)$ as defined in equations (3) and (4).

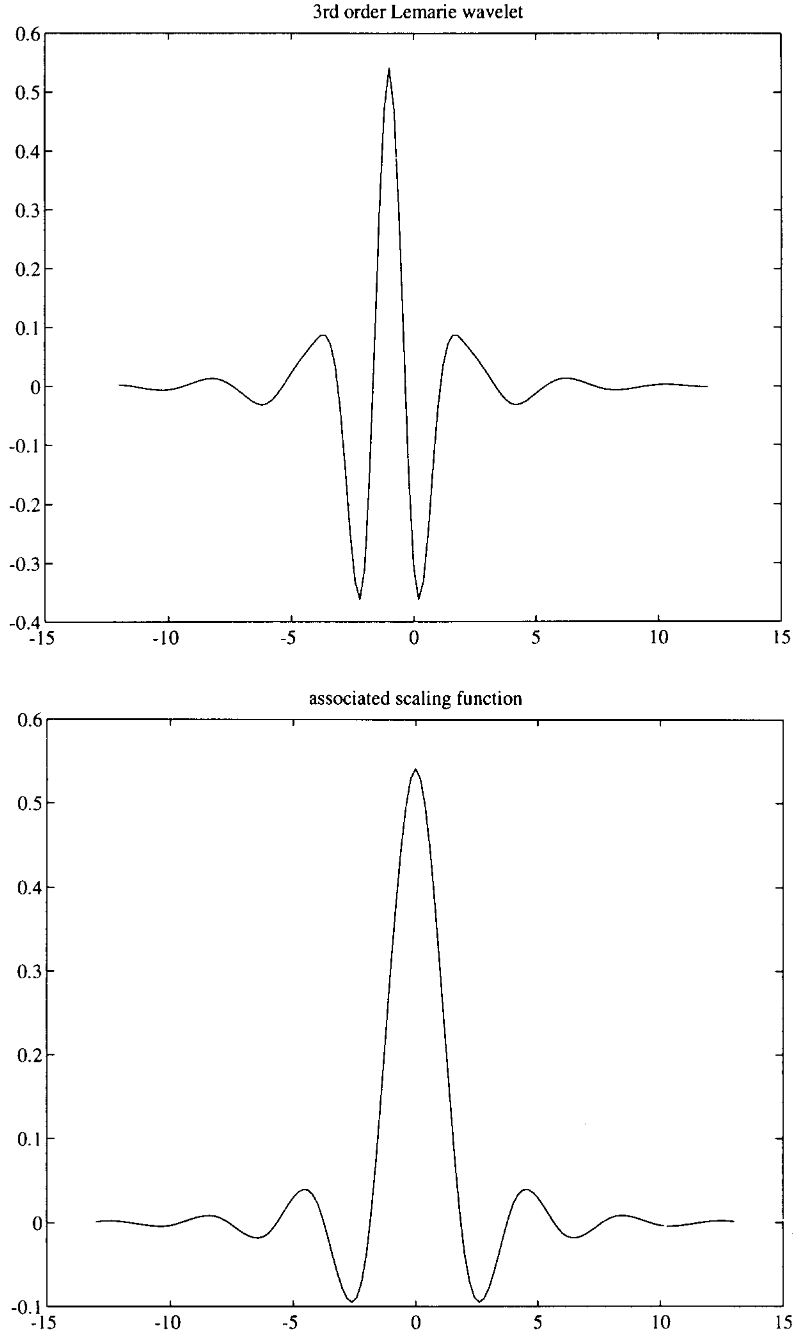

Fig. 2. Order-3 Lemarie wavelet (top) and its associated scaling function (bottom).

resolution representation of the discrete function* f. If the original function has N samples, then $A_{2^j}f$ and $D_{2^j}f$ shall have 2^jN samples. Thus the representation in (10) has the same total number of samples as the original function, i.e. the representation is non-redundant. Whenever we use the term "wavelet transform" in the sequel, we shall refer to the multiresolution pyramid defined in (10).

* Note that $A_{2^0}f = f$.

The original function can be *exactly* reconstructed from its wavelet representation by a procedure similar to the one outlined[8] in Fig. 3. The major difference is that approximations are upsampled rather than subsampled. If the reconstruction algorithm is applied to the low-frequency coefficients $A_{2^{-J}}f$ alone, we simply obtain a smoothed version of f.

Since the basic shape of an object is captured by the low-frequency components, our strategy for deriving a shape representation is to disregard the information within the high-frequency bands $(D_{2^j}f)_{-J\leq j\leq -1}$ of

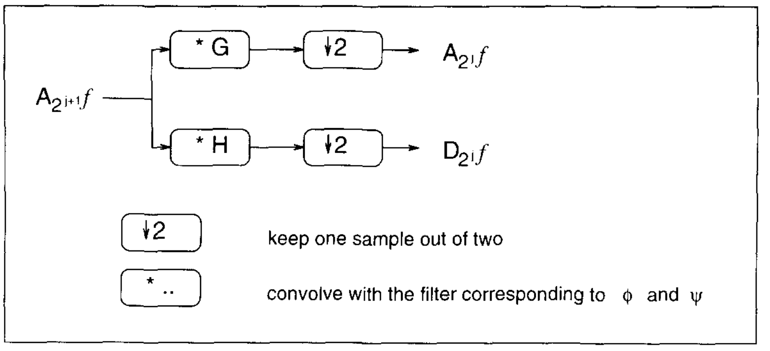

Fig. 3. Decomposition of the discrete approximation $A_{2^j}f$ into a coarser approximation $Af_{2^{j-1}}$ and detail coefficients $D_{2^{j-1}}$.

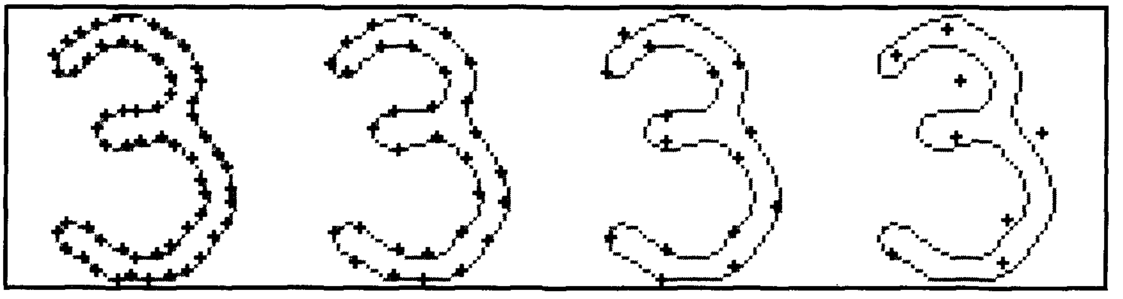

Fig. 4. Low-frequency wavelet coefficients $A_{2^J}c$ marked by crosses. $J = -2, \ldots, -5$.

the multiresolution representation and to use the low-frequency coefficients $V_{2^{-J}}f$ to accomplish recognition. In order to obtain features that represent an object at different resolutions, we choose the low-frequency components of distinct levels in the multiresolution hierarchy. Figure 4 shows the low-frequency wavelet coefficients (i.e. $A_{2^J}c$) of a character contour at the decomposition levels $J = -2, \ldots, -5$ [$c(t)$ denotes the parametric representation of the contour]. Note that at higher resolutions the sampling rate is high and the degree of smoothing low, whereas at low resolutions the degree of smoothing is high and the sampling rate low.

2.2. *Shift normalization*

Wavelet transforms have not found widespread use in the pattern recognition community because wavelet coefficients are not shift-invariant, i.e.

$$g(t) = f(t + \tau) \not\Rightarrow WTg(u) = WTf(u + \tau). \quad (11)$$

Figure 5 shows an example in the one-dimensional case. The two functions shown only differ by a horizontal shift, their shapes are identical. However, their low-frequency wavelet coefficients (marked, respectively, by boxes and crosses) do not correspond. The coefficient set of the translated function is *not* a translated set of the original coefficients. This is inherent in the ambiguity of the subsampling process, and is a well-known fact for traditional pyramidal multiresolution representations.[9]

In the case of a parametric curve, equation (11) implies that the resulting coefficients depend on the starting point chosen to trace a contour. Since there is no *a priori* way to determine a fixed starting point for a given shape, an appropriate normalization should be carried out *before* computing its wavelet transform. In the following paragraphs, we present two methods to normalize a contour representation of an object with respect to parameter shifts.

If two identical discrete curves are traced starting at different initial points, the resulting parametric functions differ by a cyclic parameter shift. We wish to find a complete representation that is invariant with respect to such shifts, i.e. given $f_n = g_{(n-n_0)\mathrm{mod}N}$, $\tilde{f}$ and $\tilde{g}$ have

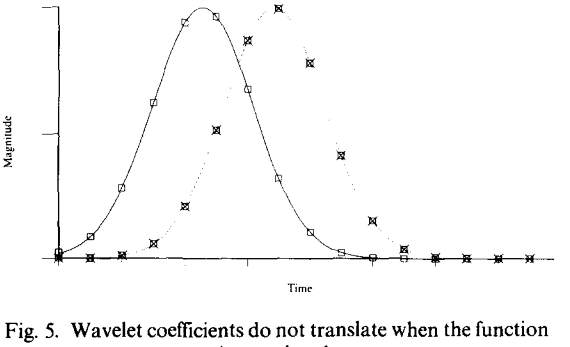

Fig. 5. Wavelet coefficients do not translate when the function is translated.

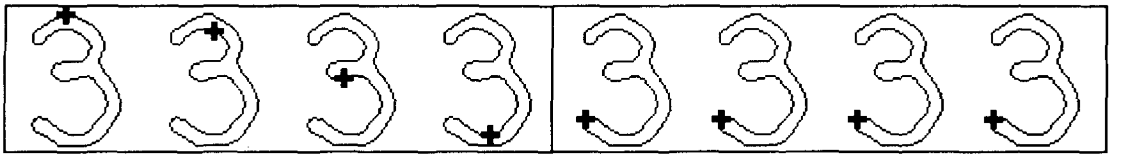

Fig. 6. Left: a sample character and its different starting points. Right: normalized starting points.

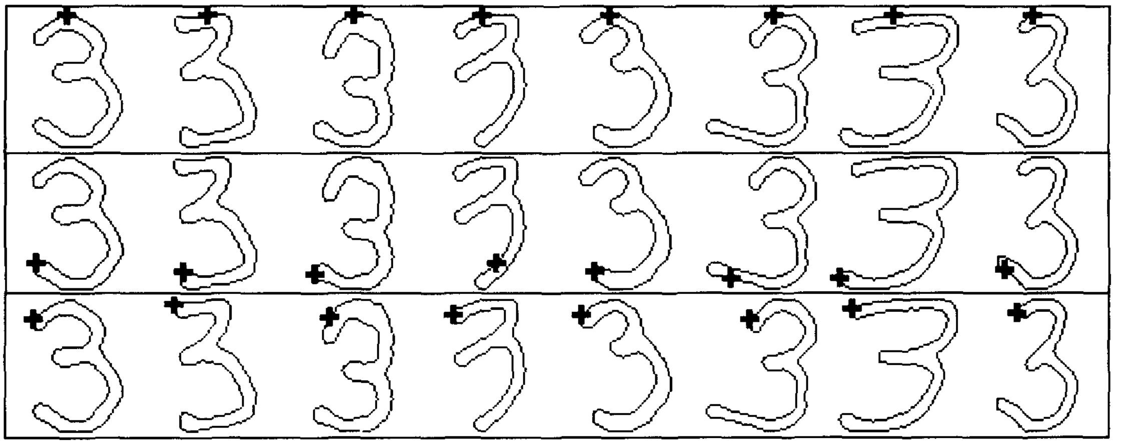

Fig. 7. Normalized starting points for different samples of the character '3'. Top row: starting points obtained by scanning from top to bottom. Middle row: starting points resulting from phase normalization. Bottom row: starting points determined with respect to the MBR.

to be derived such that

$$\tilde{f}_n = \tilde{g}_n \text{ and } \tilde{f}_n = f_{(n-v_f) \bmod N}, \quad \tilde{g}_n = g_{(n-v_g) \bmod N}. \tag{12}$$

The traditional approach to eliminate parameter shifts in a closed curve is to compute its Fourier transform. In the Fourier domain a spatial shift results in a frequency linear phase shift, i.e.*

$$g_n = f_{(n-n_0) \bmod N} \Rightarrow G_k = \exp(j2\pi k n_0/N) F_k, \quad n, k = 0, \ldots, N-1. \tag{13}$$

Obviously, the magnitudes of the Fourier coefficients $|G_k|$ and $|F_k|$ are independent of the parameter shift. However, these magnitudes are not a complete descriptor set and usually do not contain enough shape information for reliable recognition. The problem of complete frequency domain normalization for parameter shifts was first addressed by Crimmins[11] and later generalized by Arbter.[12,13] In our study, we used a special case of Arbter's Z-invariants.[12]

As stated in equation (13), the Fourier coefficients of two functions that only differ by a parameter shift, differ by a phase factor. From equation (12) we can obtain

$$G_1 = zF_1, \quad z = \exp(j2\pi n_0/N) \tag{14}$$

or, in polar coordinates,

$$G_1 = z|F_1| \exp(j\phi_1), \tag{15}$$

where ϕ_1 is the phase angle of F_1. Normalization is accomplished by multiplying each of the Fourier coefficients G_k with the phase of G_1^{-k}. This will eliminate the influence of any parameter shift. The normalized coefficients $\tilde{G}_k$ are then given by

$$\begin{aligned}\tilde{G}_k &= G_k \, arg(G_1)^{-k} \\ &= G_k z^{-k} \exp(-j\phi_1 k) \\ &= z^k F_k z^{-k} \exp(-j\phi_1 k) \\ &= \exp(-j\phi_1 k) F_k.\end{aligned} \tag{16}$$

The normalization of the Fourier coefficients of f is

$$\tilde{F}_k = F_k \, arg(F_1)^{-k} = \exp(-j\phi_1 k) F_k \tag{17}$$

which is exactly equal to $\tilde{G}_k$. Taking the inverse Fourier transform of the normalized coefficients, results in a shifted version of the original function. Equating the exponential factors in (13) and (17) one can obtain the spatial shift v_f introduced by the normalization equation (17),

$$v_f = \frac{\phi_1}{2\pi} N. \tag{18}$$

The technique described above uniquely determines a starting point for two identical periodic functions that differ only by a parameter shift (Fig. 6). Unfortunately, the situation is not as simple for handprinted characters, since we must determine a unique starting point for many samples of the same character. Each sample usually differs by a parameter shift and will have a slightly different shape due to the variations of individual handwriting style. Ideally one would need

* A capital letter denotes the Fourier transformed, i.e. f and F are an FT pair.

a way to map the starting points of different samples of the same character to corresponding contour points. However, there is no general way to find corresponding points on a character contour without *a priori* knowledge of the character. This is another form of the correspondence problem.[12]

Since the low-frequency Fourier coefficients of two similar shapes are close, the starting points determined by the normalization procedure described above should approximately correspond. Figure 7 illustrates this point. However, this normalization technique does not solve the correspondence problem *exactly*. For comparison, we considered a second starting point normalization strategy. Since orientation of each character is known, we can determine approximately corresponding points with respect to its minimum bounding rectangle (MBR). Specifically, we start the parameterization at the contour point that is closest to the top left corner of the MBR. The bottom row of Fig. 7 shows some examples of this normalization techniques. Extensive experiments have shown[14] that, in terms of classification accuracy, no significant difference between the two methods can be observed.

2.3. *Reconstruction from wavelet descriptors*

When identifying the resolution level of a wavelet descriptor set for a particular recognition problem, we must trade-off compactness of representation at the cost of loss of shape information.

A measure of the degree of information loss is the ability to reproduce an original pattern contour from its descriptor set.

Figure 8 shows reconstructions obtained from low-frequency wavelet coefficients across several levels of resolution. At higher resolution levels the approximation is nearly perfect, whereas the characters reconstructed from low resolution descriptors exhibit considerable distortion. However, for all resolution levels displayed in Fig. 8 the reconstructions preserve the basic shape of the original character. Note that some contours may not close since the number of discrete samples defining the contour may not be a power of two.

For comparison we also computed reconstructions for the same character contours shown in Fig. 8 obtained from truncated sets of Fourier coefficients. The results are shown in Fig. 9. The number of coefficients was the same as the reconstructions in Fig. 8 (from left to right: 72, 36, 18 and 9 complex numbers). The reconstructions obtained from Fourier coefficients are of comparable quality as the corresponding wavelet representation. However, as the recognition results in Section 4.2 shall demonstrate, wavelet coefficients are more reliable for recognition than Fourier descriptors. Although Fourier descriptors contain about same amount of shape information as low-frequency wavelet descriptors, Fourier descriptors exhibits larger intra-class variance and weaker interclass separation than corresponding wavelet descriptors.

3. A NEURAL NETWORK TOPOLOGY FOR MULTIRESOLUTION SHAPE RECOGNITION

In this section we describe the implementation of a character recognition scheme that relies upon wavelet descriptors to simultaneously analyse shape at multiple levels of resolution. Classification is accomplished by a set of neural networks, namely multilayer perceptrons (MLP). The recognition process consists of three stages: first, the input patterns are preclassified according to their topology. Next, the contour of an input shape is represented at multiple levels of resolution and each representation is classified independently. Finally the

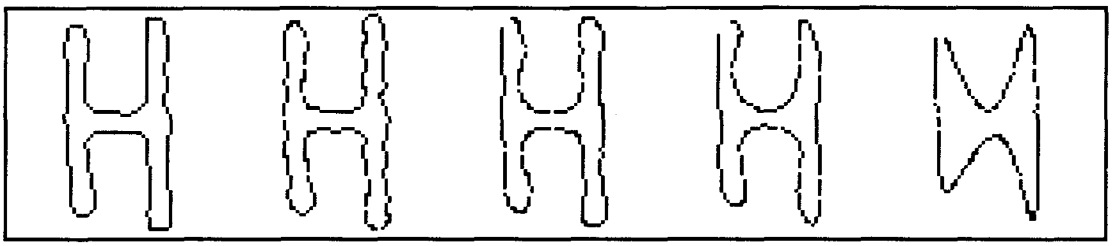

Fig. 8. Contours of a character approximated by incremental reconstruction across four levels of resolution. From left to right: original contour, level-2, 3, 4 and 5 reconstruction. (i.e., respectively, 72, 36, 18 and 9 complex descriptors).

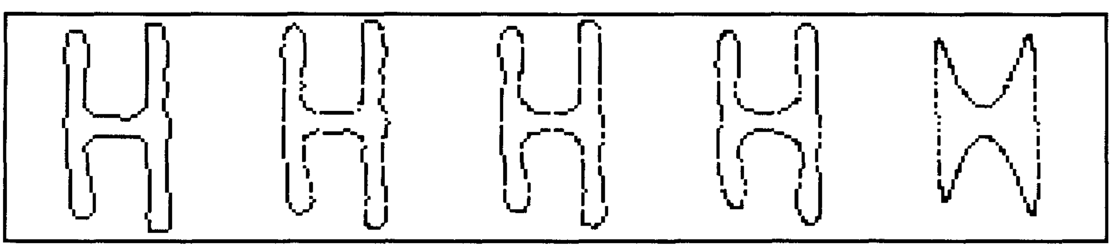

Fig. 9. Shapes reconstructed from the first k Fourier coefficients. From left to right: $k = 72, 36, 18, 9$.

classification results are combined to decide a final classification.

3.1. *Topological sorting*

Each character is *preclassified* according to its topology. By computing their Euler number, the alphanumeric characters of the Latin alphabet can be grouped into three distinct topology classes. Recall that the Euler number of a shape is defined as the number of connected components minus the number of holes. The characters of each topology class are analysed independently by distinct MLPs. Figure 10 shows the classifier structure. Note that the characters of the Latin alphabet are not distributed evenly among the three topology classes. Two thirds of the characters belong to class zero (no holes), whereas only nine and three characters belong to classes one and two, respectively.

There are two major advantages of such a two-stage scheme. First, it is obvious that the topological structure of a shape is one of the most important features for recognition and should therefore be considered by a classifier. Secondly, MLPs are trained by example and training algorithms usually require extensive computation. Moreover, training time does not increase linearly with the number of classes. Using this preclassification scheme, we needed only to train three "small to medium-sized" classifiers rather than a single complex one. This considerably reduced the overall training time required by our system classifiers.[14]

3.2. *Multiresolution recognition*

One of the motivations for using wavelet coefficients as contour descriptors is that the wavelet decomposition represents a function at different levels of resolution. There is substantial evidence that the human visual system uses similar multiscale representations.[7,9] By analogy our recognition algorithm takes into account contour representations of different resolutions. The strategy is the following: for each input pattern, wavelet descriptors at several levels of resolution are derived and classified independently. The classification results are then compared to decide a final match. Figure 11 illustrates the processing, where recognition is accomplished in three stages.

The results of the subclassifiers can be combined in many ways. In our implementation we used the following rule: If all classifiers reported the same result, the classification was accepted. If the classifiers did not agree, the pattern was rejected as "not classifiable". The advantage of this simple scheme is that the error probability is reduced considerably, because an error can only occur if *all* classifiers make the *same* confusion. Since the level of detail differs considerably at each resolution level, this case is not likely.

Besides reducing the error rate, an advantage of combining different recognition results, is that patterns which cannot be unambiguously identified are rejected rather than misclassified. This is usually not possible when an MLP classifier is used, because there is no straightforward way to train a network with an output node denoting a rejection. It is, however, extremely important that an OCR system has the capability to reject patterns rather than misclassifying them, because the cost associated with a rejection is usually much lower than that of a substitution error.[15] Imagine, for instance, an OCR system designed to read price labels in a supermarket. A rejection error means that the price label has to be read again, or in the worst case, that the price has to be entered manually. A substitution error, however, means that the customer will be charged a wrong price. Therefore, in practice, the failure to distinguish between rejections and substitutions is a severe shortcoming for any OCR system.

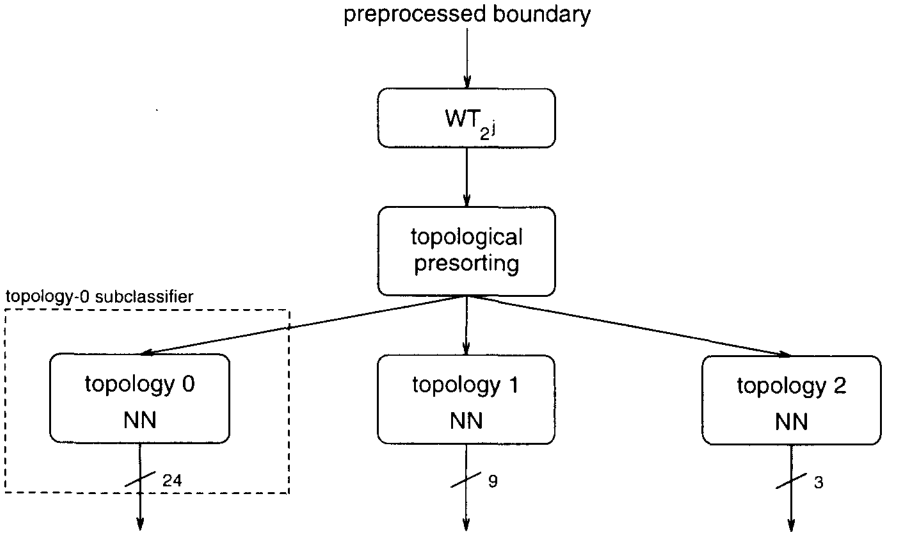

Fig. 10. Topology of a single resolution classifier.

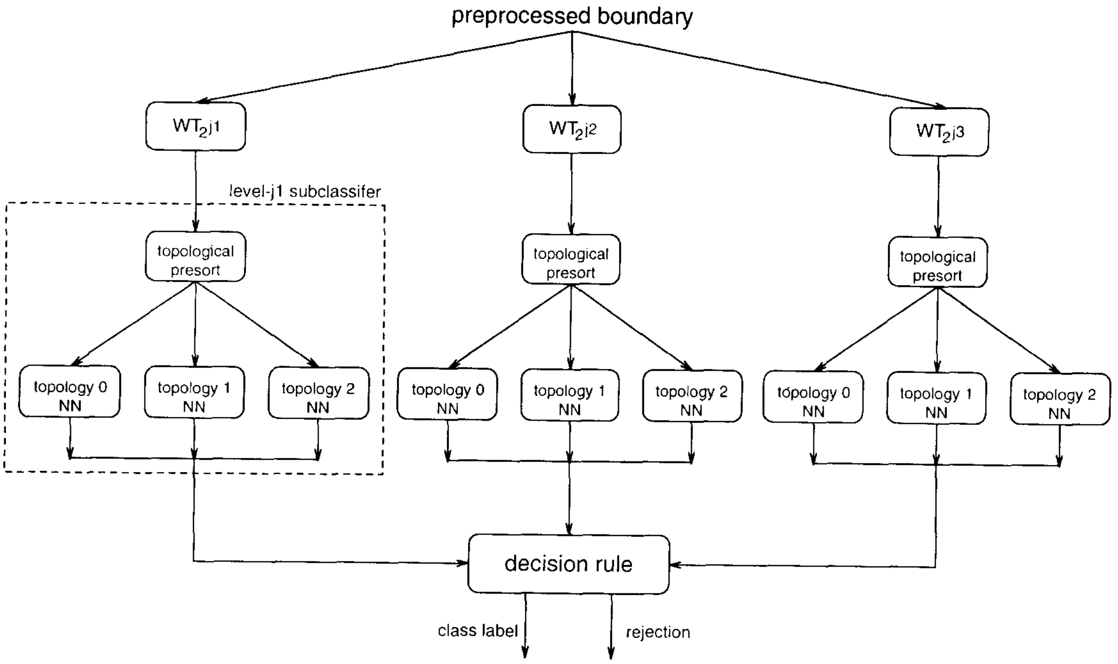

Fig. 11. A multiresolution recognition scheme that analyses a character at three levels of resolution.

3.3. *Determining the network structure*

It is well-known that multilayer perceptrons are capable of forming arbitrarily complex decision boundaries,[16] but performance heavily depends on the network architecture. A particular network architecture is defined by two parameters, namely the number of hidden layers and the number of nodes in each hidden layer. There are infinitely many network configurations that all perform the same classification task. Unfortunately, there is no mathematical rule or formula how to choose a network structure that is optimal for a given classification problem. Therefore, a "good" network structure must be determined experimentally. Given the large training times required in our study, an extensive search of an optimal network architecture was not feasible. Therefore, we imposed the following limits on the set of network structures considered in our investigation:

(1) *One* hidden layer and
(2) the number of hidden nodes close to $\log_2(N_O)$,

where N_O is the number of output nodes. The motivation for these constraints are discussed extensively in reference (14). In the next section we report recognition results obtained from the best performing network architectures.

4. EXPERIMENTAL RESULTS AND DISCUSSION

4.1. *The database*

In order to train and evaluate the performance of our recognition system, we used a large database of handprinted characters collected from previous related studies.[17–19] The database consisted of 6480 distinct samples of handprinted digits and upper-case letters. The samples were collected from fifteen different persons, each providing 12 samples of each pattern. The writers were asked to slash the character '0', to print '1' without serifs and to place horizontal bars on the letter 'I' (see Fig. 12). Such restrictions can also be found in standard character databases such as those by Munson or by Highleyman.[20] Unlike the Munson database, the samples of the character 'Z' were written without a horizontal stroke. The character '4' was required to be closed with a diagonal stroke. Similar definitions are common to most studies of OCR methodologies.[20]

Another constraint the writers had to meet, was to avoid gaps. This assumption is made in most recognition schemes based on contour features,[1,21,22] although it is rarely stated explicitly. Obviously, this constraint holds only in special applications. In Section 5.2 we outline how our recognition system can be modified to accommodate non-closed characters.

Fig. 12. Samples for '0', 'O', '1', 'I', '2', 'Z' and '4' as defined for our study.

The samples were collected on forms that allowed for simple segmentation. The forms were digitized with 300 dpi/8 bit resolution and normalized in size to 64 by 64 pixels by bilinear interpolation.[23,18]

4.2. *Wavelet descriptors vs Fourier descriptors*

First we conducted a preliminary study, the goal of which was to compare recognition results achieved with wavelet descriptors to those obtained with Fourier descriptors (FD). In this comparison we wanted to ensure that the classification was based on shape alone. Therefore we neither presorted the samples according to their topology, nor did we consider multiple levels of resolution. The set of Fourier descriptors used is described by Arber.[12] However, we omitted normalization for affine transformations, because the digits '6' and '9' cannot be distinguished in an affine-invariant space. For reference, we also collected the recognition results obtained with FD where the influence of an arbitrary starting point was eliminated by taking its magnitude. In order to avoid extensive training time, we limited the set of classes to the digit '0'–'9'. For each descriptor set, character shape was represented by the same number of features (36 real numbers). The number of training patterns was 1200. The classifiers were tested with 600 characters written by persons other than the training sample authors. Table 1 summarizes the error rates of the best performing classifiers.

These results suggest that wavelet descriptors are more efficient for recognition purposes than a pure frequency domain representation of the same dimensionality. Although a complete set of Fourier descriptor[12] was used, it yielded higher error rates. In order to rule out that this error rate was due to normal statistical variation, we tested whether the observed error rates differed at a 5% *significance level.*[24] The symbol '+' in the last column of Table 1 denotes this case.

If the character shape was represented by the magnitudes of the Fourier coefficients, the results were worse. This is intuitively clear, because despite being shift-invariant FD magnitudes lack completeness. Moreover, interclass separation seems to be poor for FD magnitudes, as suggested by a large number of training epochs.[14]

4.3. *Single resolution recognition*

Next we analysed the recognition results as function of decomposition level. We only considered descriptors of one resolution at a time. In this case the three-stage multiresolution recognition scheme shown in Fig. 11 reduces to the two-stage scheme depicted in Fig. 10. In this study we used two distinct test sets. First, we tested the classifier with samples taken from the same individuals that provided the training samples. Here, the application in mind is a system that only needs to recognize a few distinct handwriting styles and that must be retrained, when used by a different individual. The second test set was designed to simulate a more general application where the system is only trained once and is then expected to recognize samples of unknown handwriting. It is intuitively clear, that the second application is more difficult, and indeed, the recognition results differed considerably. In both test situations the number of training and test samples were identical. Table 2 summarizes the results for the best performing configurations.

Testing the classifier with familiar handwriting yielded error rates of less than 1% for each decomposition level. In order to determine the quality of the observed

Table 1. Wavelet descriptors vs Fourier descriptors

Features	L	H	Errors	% Error	% Correct	5% Sig.
Level-4 WD	36	6	0	0	100	
FD	36	5	8	1.33	98.67	+
FD (mag)	36	5	13	2.17	97.83	+

Recognition results for the digits '0'–'9', where represented by wavelet and Fourier descriptors, respectively. The second column shows the number of features (number of input nodes). Column three contains the number of hidden nodes that yielded best recognition.

Table 2. Recognition rates as a function of decomposition level

Level	L	NN_0	NN_1	NN_2	Errors	% Error	% Correct	5% Sig.	95% Conf.
3	72	5	5	2	13	0.60	99.40	+	[0.3, 0.9]
4	36	4	5	1	6	0.28	99.72		[0.1, 0.5]
5	18	5	2	2	10	0.46	99.54	–	[0.2, 0.8]
3	72	5	2	2	38	1.75	98.24	+	[1.2, 2.3]
4	36	4	5	1	27	1.25	98.75		[0.8, 1.7]
5	18	2	5	2	35	1.62	98.38	–	[1.1, 2.2]

Top: results for test samples of familiar handwriting. Bottom: results for test samples of unfamiliar handwriting. The second column shows the number of features (number of input nodes). Columns three through five show the number of hidden nodes for each subnet that yielded the best recognition results.

Table 3. Multiresolution recognition results

Levels	L	Subst.	Rej.	% Subst.	% Rej.	% Correct	95% Conf.
3/4	108	3	13	0.14	0.60	99.26	[0.0, 0.3]
4/5	54	0	17	0	0.79	99.21	—
3/5	90	1	17	0.05	0.79	99.16	[0.0, 0.1]
3/4/5	126	0	29	0	1.34	98.66	—
3/4	108	13	39	0.60	1.81	97.59	[0.3, 0.9]
4/5	54	9	45	0.42	2.08	97.50	[0.1, 0.7]
3/5	90	7	60	0.32	2.78	96.90	[0.1, 0.6]
3/4/5	126	6	83	0.28	3.84	95.88	[0.1, 0.5]

Top: results for test samples of known handwriting. Bottom: results for test samples of unknown handwriting. The second column shows the number of features (number of input nodes). Columns three through five show the number of hidden nodes for each subnet that yielded the best recognition results.

error estimate, we computed 95% *confidence intervals*.[24] Due to the large number of test samples (2160 distinct test samples were used), the confidence intervals observed were rather narrow. Table 2 shows that in 95% of all cases we will observe error rates of less than 1% for familiar handwriting. Unfortunately it is hard to compare these results to others reported in the literature, since in most papers the assumptions and restrictions on the test data are only incompletely stated or very different from ours. However, we can directly compare these results to those of a previous study conducted with the same database.[18] The features used in this early study were wavelet coefficients derived from a character's grey level image. The number of substitution errors* was 35 which corresponded to an error rate of 1.43%.† Note that the feature vector dimension was 192. Using wavelet descriptors, we obtained a significantly smaller error rate (0.28%) while only using 36 features. Thus, despite a more than five-fold data reduction in representation, the error rate was reduced by more than a factor of five.

Table 2 also shows the recognition rates observed when tested with samples of unfamiliar handwriting. These error rates were about three times greater than those for known handwriting. Still, the number of errors were about the same as those reported in[18] for recognition of *familiar* handwriting. As far as decomposition level is concerned, it turned out that level-4 WD were the best compromise between compactness of representation and preservation of shape information. The recognition results obtained with level-3 WD are only slightly worse, but the difference is statistically significant. Surprisingly, level-5 WD yielded better recognition results than those of decomposition level three, although the character shape was represented by only 18 features (real numbers). In fact, we observed that the level-5 error rates did not differ significantly from those obtained by level-4 descriptors.

* Digits and letters were classified by different perceptrons.

† A recent study[19] showed that using hexagonal wavelets these error rates can be reduced by about 25%.

4.4. *Multiresolution recognition*

The results discussed in the previous section suggest that WD are a concise shape representation that is very effective for recognition. Next, we exploited the possibly most powerful aspect of wavelet descriptors: hierarchical multiresolution representation. Table 3 summarizes the results when testing the multiresolution classification scheme with both familiar and unfamiliar handwriting.

A significant conclusion that can be drawn from Table 3 is that it is clearly advantageous to simultaneously analyse multiple resolutions when recognizing handprinted characters. Note that this was not obvious *a priori*. If the same misclassifications had occurred at all resolutions, it would have been senseless to consider more than one decomposition level, since most errors would have remained undetected. Our results demonstrate that this is not the case. The gain in reliability is achieved at the expense of a somewhat lower rate of correct classifications. This is, of course, due to the rejections that occur if the distinct classifiers have contradictory outputs. But, as mentioned earlier, the distinction between rejections and substitution errors is important for most OCR applications.

The data in the top part of Table 3 can be directly compared with the recognition rates reported by Laine *et al.*[18] Note that for all cases shown in Table 3, the number of features remained smaller than those used by the classifier described in reference (18). Thus, representing a character at multiple resolutions allowed us to maintain a rate of correct classifications of over 99% while avoiding substitution errors. More importantly, we could also reduce the error rate below 1% for samples of unfamiliar handwriting. In this case the rejection rates were, of course, somewhat higher.

5. CONCLUSIONS AND FUTURE RESEARCH

5.1. *Summary*

In this paper we introduced a novel set of features that are well-suited for representing digitized handprinted characters. Features are derived by computing the wavelet transform of a character contour. Since only the low-frequency bands of the transform coefficients are included, the features are insensitive to the shape variations caused by the writing styles of different persons. Representations by wavelet transform coefficients depend on the parameterization starting point of a function. Appropriate normalizations yield a shift-invariant multiresolution representation for characters of known orientation.

The most powerful aspect of a wavelet representation is, however, the decomposition of a function into a multiresolution hierarchy. We described the implementation of a character recognition system that uses WD to simultaneously analyse character shape at multiple resolutions. Recognition is accomplished by several neural nets, namely multilayer perceptrons, that are independently trained with WD obtained from distinct resolution levels. The system was trained and tested with a large database of handprinted characters. Significant results of the study are listed below:

- wavelet descriptors are a compact representation for digitized characters that contain sufficient shape information to allow for reliable recognition. Even when characters were described with only 1/32th (level-5 WD) of their original data, recognition rates did not significantly degrade;
- wavelet descriptors are insensitive to individual writing style variations. Confronted with unfamiliar handwriting, the recognition system continued to exhibit low error rates;
- although closely related to Fourier descriptors, wavelet descriptors are a significantly better shape representation for handprinted characters than a complete Fourier descriptor set of the same dimensionality;
- multiresolution recognition is a powerful methodology to increase recognition reliability. In contrast to most single scale techniques, it enabled the system to reject patterns which could not be unambiguously classified, and thus considerably reduced the rate of substitution errors observed;
- although the multiresolution recognition scheme reduced the error rate at the expense of rejections, the rate of correct classifications remained satisfactory.

5.2. *Future work*

Since we derived shape descriptors from the character contour, we had to impose the constraint that character contours be closed. For practical applications, this restriction should be overcome. This issue is discussed in greater detail in reference (14). It is relatively straightforward to lift the constraint for pen-based input, since the data may be naturally split into a set of distinct strokes. The major modifications of our system would be to skeletonize the input by a thinning algorithm and presort the samples according to their number of strokes rather than their Euler number. However, the solution of this problem is more difficult if only a greylevel image is available, and may require knowledge-based preprocessing.

Nevertheless, wavelet descriptors are a compact shape representation that are well-suited for applications with large input variability. It would be interesting to apply WD to recognition problems other than OCR. Since the wavelet transform can also be executed in the Fourier domain, it is straightforward to apply the normalization suggested by Arbter[12] to derive *affine-invariant wavelet descriptors*. Thus wavelet descriptors can easily be generalized to serve as a shape representation for a variety of additional recognition problems.

Acknowledgements—This work was sponsored in part by the National Science Foundation, Grant No. IRI-9111375, and the Engineering Foundation, Grant No. RI-A-91-09.. The authors are most thankful to Prof. Bjorn Jawerth, Dept. of Mathematics, University of South Carolina, for his helpful suggestions and insight.

REFERENCES

1. C. T. Zahn and R. Z. Roskies, Fourier descriptors for plane closed curves, *IEEE Trans. Comput.* **21**, 269–281 (March 1972).
2. G. H. Granlund, Fourier preprocessing for handprint character recognition, *IEEE Trans. Comput.* **21**, 195–201 (February 1972).
3. E. Persoon and K.-S. Fu, Shape discrimination using Fourier descriptors, *IEEE Trans. Systems, Man Cybernet.* **7**, 170–179 (March 1977).
4. M. Shridhar and A. Badreldin, High accuracy character recognition algorithm using Fourier and topological descriptors, *Pattern Recognition* **17**, 515–524 (1984).
5. Charles K. Chui, *An Introduction to Wavelets.* Academic Press, San Diego, California (1992).
6. Ingrid Daubechies, *Ten Lectures on Wavelets.* CBMS-NSF regional conference series in applied mathematics, **61** (1992).
7. David Marr, *Vision.* W. H. Freeman, San Francisco (1982).
8. S. Mallat, A theory for multiresolution signal decomposition: the wavelet representation, *IEEE Trans. Pattern Analysis Mach. Intell.* **11**, 674–693 (July 1989).
9. S. Mallat, Multifrequency channel decomposition of images and wavelet models, *IEEE Trans. Acoustics, Speech Signal Process.* **37**, 2091–2110 (December 1989).
10. S. Mallat, Multiresolution approximation and wavelet orthonormal bases of $L^2(R)$, *Trans. Am. Math. Soc.* **315**, 69–87 (September 1989).
11. T. R. Crimmins, A complete set of Fourier descriptors, *IEEE Trans. Systems, Man Cybern.* **12**, 236–58 (1982).
12. K. Arbter, Affine-invariant Fourier descriptors, in J. C. Simon, ed., *From Pixel to Features*, pp. 153–164. Elsevier Science Publishers B. V., Amsterdam (1989).
13. K. Arbter, W. E. Snyder, H. Burkhardt and G. Hirzinger, Application of affine-invariant Fourier descriptors to recognition of 3-D objects, *IEEE Trans. Pattern Analysis Mach. Intell.* **12**, 640–647 (July 1990).
14. P. Wunsch, Wavelet descriptors for multiresolution representation and recognition of handprinted characters. Master's thesis, University of Florida (August 1993).
15. Theo Pavlidis, Recognition of printed text under realistic conditions, *Pattern Recognition Lett.* **14**, 317–326 (April 1993).
16. R. P. Lippman, An introduction to computing with neural nets, *IEEE ASSP Mag.* **4**, 4–22 (1987).
17. Andrew F. Laine, William E. Ball and Arun Kumar, A multi-scale approach for recognizing complex annotations in engineering drawings, in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, Lahaina, Hawaii (June 3–6 1991).

18. A. F. Laine, S. Schuler and V. Girish, Wavelet representations for recognizing complex annotations, *Mach. Vision Applic.* **6**, 110–123 (1993) (Special Issue on Document Image Analysis and Techniques.)
19. A. F. Laine and S. Schuler, Hexagonal wavelet representations for recognizing complex annotations, in *Proc. IEEE Conf. Comput. Vision Pattern Recognition*, Seattle, Washington (June 20–24 1994).
20. C. Y. Suen, M. Berthold and S. Mori, Automatic recognition of handprinted characters—the state of the art, *Proc. IEEE*, **68**, 469–485 (April 1980).
21. T. Taxt, J. Olafsdottir and M. Daehlen, Recognition of handwritten symbols, *Pattern Recognition* **23**, 1155–1166 (November 1990).
22. G. Savarayudu and I. Sethi, Walsh descriptors for polygonal curves, *Pattern Recognition* **16**, 327–336 (1983).
23. H. S. Baird, S. Kahan and T. Pavlidis, Components of an omni-font page reader, in *IEEE Comput. Soc. Conf. Pattern Recognition*, pp. 344–347 (1986).
24. I. F. Blake, *An Introduction to Applied Probability*. Wiley and Sons, New York (1979).

About the Author—PATRICK WUNSCH received his intermediate diploma ("Vordiplom") from *Technische Universität München*, Germany, in 1991. After being awarded a Fulbright grant in 1992, he transferred to the Research Center for Computer Vision and Visualization at the University of Florida, where he received the M.S. degree in Summer 1993. He then joined DLR, the German Aerospace Research Establishment, where he is presently pursuing a Ph.D. degree at the Institute for Robotics and System Dynamics. His research interests include computer vision, pattern recognition and neural networks.

About the Author—ANDREW F. LAINE received his B.Sc. degree from Cornell University, in 1983 and 1989 received the M.Sc. and D.Sc. degrees in computer science from Washington University in St Louis. He is currently an Associate Professor of Computer and Information Sciences at the University of Florida, Gainesville. His research interests include methods of multiresolution analysis applied to problems in computer vision, pattern recognition and medical imaging. Dr Laine has chaired the SPIE conference on "Wavelet Applications in Signal and Image Processing", in San Diego, CA, 1993–1995. He also served as Guest Editor for the *Journal of Mathematical Imaging*, Special Issue on wavelet theory, 1993.