March, 1995

LIDS-P 2295

Research Supported By:

ARPA grant F49620-93-1-0604
ARO grant DAAL03-92-G-0115
AFOSR grant F49620-92-J-0002

Multiscale Object Recognition and Feature Extraction Using Wavelet Networks (U)*

Jaggi, S.
Karl, W.C.
Krim, H.
Willsky, A.S.

March 1995

LIDS-P-2295

Multiscale Object Recognition and Feature Extraction Using Wavelet Networks (U) *

Seema Jaggi
W. Clem Karl Hamid Krim
Alan S. Willsky
Laboratory for Information and Decision Systems
EECS, Massachusetts Institute of Technology
Cambridge, Massachusetts, 02139
Telephone: (617) 253-3816
Telefax: (617) 258-8553
Email: jaggi@mit.edu

ABSTRACT

In this work, we present a novel method of object recognition and feature generation based on multiscale object descriptions obtained using wavelet networks in combination with morphological filtering. First, morphological filtering techniques are used to obtain structural information about the object. Then, wavelet networks are used to extract or capture geometric information about an object at a series of scales. A wavelet network is of the form of a 1-1/2 layer neural network with the sigmoid functions replaced by wavelet functions. Like neural networks, wavelet networks are universal approximators. In contrast to neural networks, the initialization of a wavelet network follows directly from a commonly known transformation, namely, the discrete dyadic wavelet decomposition. In contrast to a dyadic wavelet decomposition, the wavelet parameters are then allowed to vary to fit the data. Although developed in the context of function approximation, wavelet networks naturally fit in this object recognition framework because of the geometric nature of the network parameters (i.e. translations, rotations, and dilations). Wavelet networks are the basis for a hierarchical object recognition scheme where the wavelet network representation of the object at each scale is a feature vector which may be used to classify the object. At coarse scales, the feature vector is used to narrow the field of possible objects and to yield pose information. This information may also be used to generate candidate matches between the data and more detailed object models. The wavelet network representation at finer scales is then used to identify the object from this reduced space of possible objects. In keeping with our proposed integrated approach to ATD/R, we demonstrate how wavelet networks may be applied to anomaly suppression in laser range images by fitting a multiresolution wavelet basis to the data in conjunction with the expectation-maximization (EM) algorithm. Here, the wavelet network is used to refine the chosen wavelet basis. We demonstrate each of these applications on both simulated and model board data.

*This work was supported by the Advanced Research Projects Agency through Air Force Grant F49620-93-1-0604 and the Army Research Office though Grant DAAL03-92-G-0115 and the Air Force Office of Scientific Research through Grant F49620-92-J-0002.

1 INTRODUCTION

In this work, we present some preliminary steps towards a novel method of feature generation and object recognition. Our goal is to extract features that have the unique property of being both geometric and multiscale in nature. In particular, the features considered in this work will be based on multiscale object descriptions obtained using wavelet networks in combination with morphological filters. Our features incorporate geometric information since human perception and human object recognition is based on geometric information. Further, our features include multiscale information to yield hierarchical object representations. Such features are amenable to multiscale object recognition schemes where coarse scale features are used to refine fine scale searches. This combination of morphological and wavelet analysis yields a representation which is hierarchical as well as intuitively shape-based.

Morphological processing [1–7] is commonly used to probe the underlying shape or texture of a set (e.g. a binary image) or a function (e.g. a gray-scale image). The output of morphologically processing an image with a structuring element of any given shape and size is highly dependent on the specified structuring element. It follows that, in order to gain insight into the overall structure of the image, the image must be processed with a family of structuring elements. As (geometric) fine scale features are removed from the image, the image may be subsampled [8,9]. This process yields a morphological pyramidal image decomposition [6] which is similar in flavor to the Laplacian pyramid representation developed by Burt and Adelson [10]. However, in contrast to multiresolution techniques based on linear filtering, multiresolution morphological processing preserves the shape and edge information of the underlying object. For these reasons it appears that morphological processing will be a powerful tool in developing our geometrically based features. For this work, morphological processing will be used to keep detail at any given geometric scale from corrupting features at other scales. Thus, we use morphological processing to obtain a geometric multiscale representation of binary images. That is, the object is decomposed into its canonical parts via morphological filtering.

Wavelet networks, recently introduced by Zhang and Benveniste [11,12], are of the form of a 1-1/2 layer neural network with the sigmoid functions replaced by wavelet functions. Like neural networks, wavelet networks are universal approximators. In contrast to neural networks, the initialization of a wavelet network follows directly from a commonly known transformation, namely, the discrete dyadic wavelet decomposition. In contrast to a dyadic wavelet decomposition, the wavelet parameters are allowed to vary to fit the data. Although developed in the

context of function approximation, wavelet networks naturally fit in this object recognition framework because of the geometric nature of the network parameters (i.e. translations, rotations, and dilations). For this work, the canonical parts of an object (as determined through morphological processing) are represented in terms of the wavelet network decomposition.

The organization of this paper is as follows. We begin by discussing wavelet networks as applied to function approximation in Section 2. Next, in Section 3 we discuss the use of morphological filtering to obtain a multiresolution representation of the image. Finally, we present some preliminary results in Section 4. Also in Section 4, we demonstrate how wavelet networks may be applied to anomaly suppression in laser range images by fitting a multiresolution wavelet basis to the data in conjunction with the expectation-maximization (EM) algorithm. Here, the dyadic wavelet decomposition is used as an initialization for the wavelet network. The wavelet network is then used to refine the chosen wavelet basis.

2 Wavelet Networks

The wavelet network structure, proposed by Zhang and Benveniste [11.12], provides a link between the neural network and the wavelet decomposition. The 1 1/2 layer neural network is of the form

$$g(x) = \sum_{i=1}^{N} w_i \sigma(a_i^T x + b_i) \tag{1}$$

where $w_i, b_i \in \Re, a_i \in \Re^n$and $\sigma(\cdot)$ is a sigmoidal function. Neural networks of this form are often used in function approximation because, as shown in [13. 14], the neural network is a universal approximator. That is, given any continuous function $f(x)$ and any $\epsilon > 0$, there exists a neural network of the form of (1) such that $\|g(x) - f(x)\| < \epsilon$. This implies that if the network is properly trained (i.e. the values for w_i, a_i and b_i are correctly chosen and N is large enough), the neural network will be an arbitrarily good approximation of the function $f(x)$. Thus, the performance of the neural network is critically dependent on the choice of the values w_i, a_i and b_i. In practice, after a random initialization step, the neural network parameters are tuned via a back propagation procedure [15]. The wavelet network improves on neural networks by providing a more robust initialization procedure.

In the wavelet network, the sigmoid function of (1) is replaced by a wavelet function. Like neural networks, wavelet networks are universal approximators. The significant difference between neural networks and wavelet networks is that the initialization of a wavelet network is not random: rather, the wavelet network initialization follows directly from a commonly known transformation, namely, the discrete dyadic wavelet decomposition. To see this more clearly, recall the form of the (truncated) dyadic wavelet decomposition

$$g_d(x) = \sum_{s,r \in Q} \mathcal{W}(s,r)\alpha^{-s/2}\psi(\alpha^{-s}x - \beta r) \tag{2}$$

where $\psi : \Re \rightarrow \Re$ is a wavelet function that satisfies the frame property. In this formulation, the only parameters fitted from the data are the weights $\mathcal{W}(s,r)$. The dilation and translation parameters are predetermined and fixed. Often, the dilations and translations are constrained to a dyadic tree where the dilations are exponentially distributed and the translations are uniformly distributed. In contrast to a dyadic wavelet decomposition, in the wavelet network, following the initialization phase, *all* of the wavelet network parameters (w_i, t_i and d_i) are allowed to vary to fit the data.

More precisely, the wavelet network is of the form (see Figure 1) :

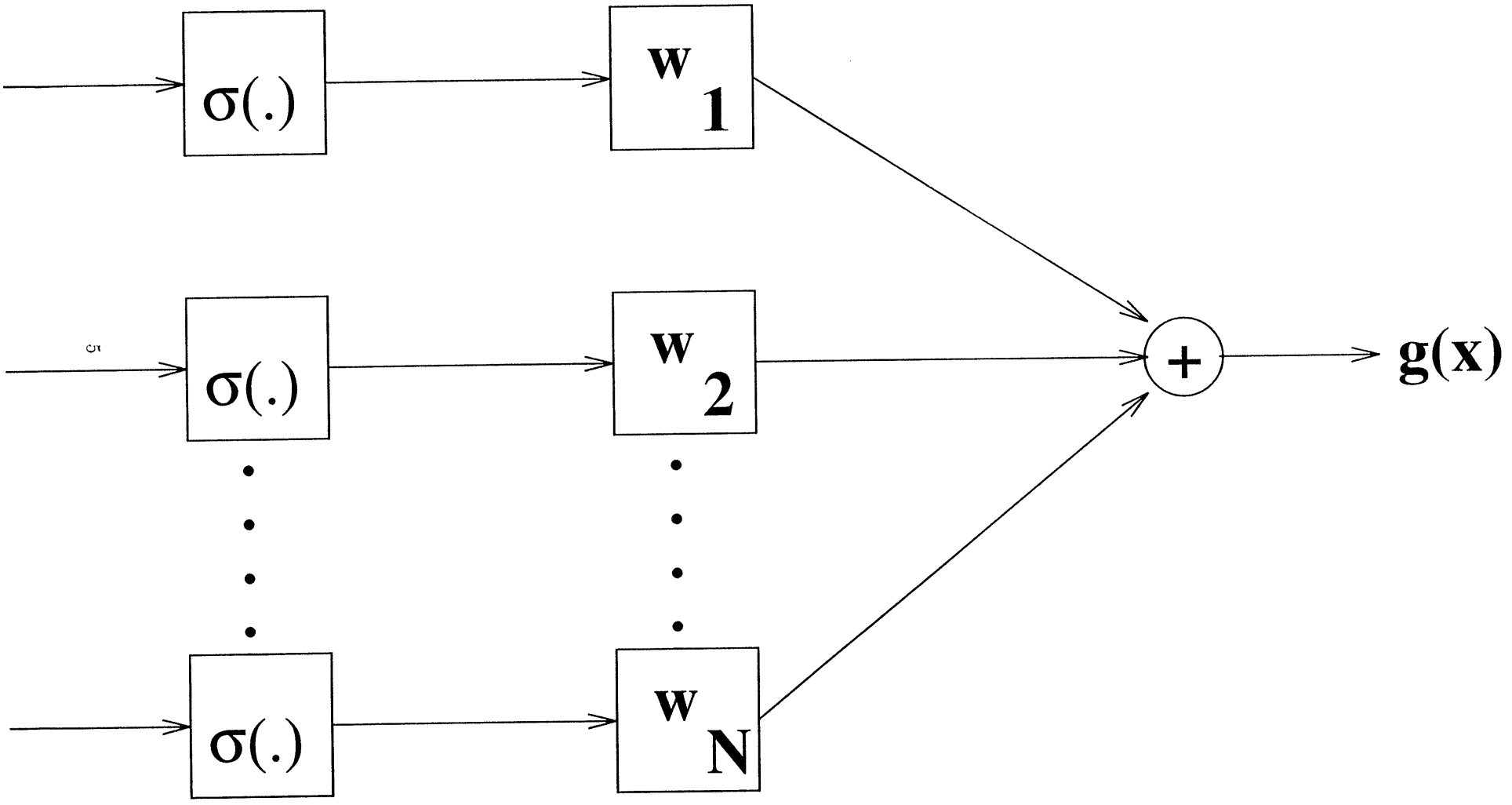
$$g(x) = \sum_{i=1}^{N} w_i \psi(D_i(x - t_i)) \tag{3}$$

where $w_i \in \Re$, $t_i \in \Re^2$ and $D_i = \text{diag}(d_i)$ with $d_i \in \Re^2$. The parameters $w_i, d_i,$ and t_i represent weight, dilation, and translation, respectively. Note that this is the same form as the 1-1/2 layer neural network in (1) with the sigmoid functions replaced by wavelet functions. Wavelet networks were originally developed in the context of function approximation. Given samples of a function $f(x)$, the wavelet network parameters ($w_i, d_i,$ and t_i) are chosen to minimize the cost

$$C = \|g(x) - f(x)\|_2^2 \tag{4}$$

The wavelet network is initialized using the dyadic wavelet decomposition (2). Then, the parameters, $w_i, d_i,$ and t_i are updated using a stochastic gradient algorithm to minimize (4).

Although developed in the context of function approximation, wavelet networks naturally fit in our proposed


σ(.)
σ(.)
σ(.)
w 1
w 2
w N
+
g(x)

geometric and multiscale object feature framework because of the geometric nature of the network parameters (i.e. translations, rotations, and dilations). Recall that the error criterion for wavelet networks given in (4) is the mean-square error. Since this is not a geometric error criterion, the wavelet network decomposition does not provide a geometric based object representation. Thus, we will combine wavelet networks with morphological feature extraction to obtain our feature vectors. This combination yields scale-to-scale isoloation of features. That is, detail of any given geometric scale will not corrupt the features at any other geometric scale.

3 Morphological Feature Extraction

As previously mentioned, morphological processing is commonly used to analyze the underlying shape or texture of an image. Further, to gain insight into the overall structure of the image, the image must be processed with a family of structuring elements. In this section, we describe some common types of morphological filtering and provide an interpretation of the family of processed images.

Morphological filtering on sets is defined in terms of the fundamental operations of dilation and erosion. The dilation and erosion of a set F by a structuring element S are given in (5) and (6), respectively.

$$F \oplus S = \bigcup_{s \in S} (F + s) \tag{5}$$

$$F \ominus S = \bigcap_{s \in (-S)} (F + s) \tag{6}$$

In (5) and (6), the translation and reflection operations are defined as

$$F + s = \{f + s | f \in F\} \tag{7}$$

$$-S = \{-s | s \in S\} \tag{8}$$

A morphological opening is then defined as an erosion followed by dilation :

$$O(F, S) = [F \ominus (S)] \oplus S \tag{9}$$

An opening removes sharp corners and breaks the set in two at narrow passes. A morphological closing is then defined as a dilation followed by an erosion :

$$C(F,S) = [F \oplus (S)] \ominus S \tag{10}$$

A closing fills in holes and covers over small protrusions. Thus, an opening may be thought of as smoothing from the inside of the set. On the other hand, a closing may be thought of as smoothing from the outside of the set. Set morphology is easily extended to function morphology by replacing the function to be processed, f, and the structuring function, s, with their respective umbras.

In this work, we use morphological processing, specifically openings, to obtain a geometric multiscale representation of binary images. That is, the object is decomposed into its canonical parts via morphological filtering.

4 Preliminary ATR Results

The wavelet network error criterion is mean-square error which is well suited to the function approximation application for which wavelet networks were developed. In applying wavelet networks in ATR and particularly for the development of a geometric feature vector, this error criterion leads to feature vectors (wavelet network parameters) that minimize mean-square error but do not convey geometric information. As a first step in including geometry in our feature vectors, we will use the technique known as granulometries to decompose the object into its canonical parts. Then, we will apply wavelet networks to each of these parts to obtain an object feature vector.

4.1 Morphological Filtering

In this section, we apply morphological filtering to decompose a simulated binary image of a tank into its canonical parts, which we assume to be of differing morphological scales. Specifically, we use granulometries to decompose the object and provide an interpretation of the change of area graphs produced by this procedure.

To incorporate this type of decomposition in our feature vector, we must determine the morphologically significant scales. One way to do this is to examine the $\mu_L(O(F, rS))$ where μ_L is Lebesgue measure (area) and rS is a structuring element scaled by r. Suppose that the object under consideration was entirely composed of scaled and shifted versions of the structuring element S which were non overlapping and of different scales. In this case, $\mu_L(O(F, rS))$ would be a monotonically decreasing series of steps. If $\mu_L(O(F, rS))$ is flat for $r_1 < r < r_2$, then we conclude that there are no pieces or subparts of F in the range of scales $r \in [r_1, r_2]$. If there is a discontinuity at $r = r_3$, we conclude that there is a piece of F at scale r_3. Of course, we will almost never encounter such a specialized structure. However, we can interpret $\mu_L(O(F, rS))$ for more realistic shapes in a similar manner. For example, if $\mu_L(O(F, rS))$ is relatively flat (i.e. has a slope of approximately 0), we conclude that there are no structures in F of the scale r. Similarly, abrupt changes in the slope of $\mu_L(O(F, rS))$ are significant since these correspond to changes in the underlying morphology of the object. In particular, step-like discontinuities in $\mu_L(O(F, rS))$ indicate that there is a subpart of F at scale r which is of the same shape as the structuring element S. Basically, the morphologically significant scales will be the scales for which the slope of $\mu_L(O(F, rS)) \approx 0$ and scales close to a step discontinuity.

To illustrate this consider the binary simulated tank, F in Figure 3. In Figure 2, $\mu_L(O(F, rS))$ is given as a function of r. The structuring element S is a square of side r. From this graph we conclude that the morphologically significant scales are $r = \{1, 6, 20, 22\}$. The output images (I_r) of the morphological opening at each of these scales is given in Figure 4 and the difference images (I_{r-s}) between these scales is given in Figure 5. Consider the set of images $\{I_{22}, I_{22-20}, I_{6-1}\}$. This set of images is a decomposition of the tank F into its canonical parts, namely, the gun, turret, and body.

Similarly, we can decompose a binary tank from model board data. This tank is taken from a model board simulated battle scene imaged using IR. The original image was threshold to yield the image shown in Figure 7 at $r = 1$. Again, the function $\mu_L(O(F, rS))$, given in Figure 6, has similar structure which indicates morphologically significant features. In this case, the structuring element S used was a rectangle of dimensions $3r \times 2r$. We conclude that the morphologically significant features are at scales $r = \{1, 4, 10\}$. Finally, the output at these significant scales is shown in Figure 7 and the difference between the scales is shown in Figure 8.

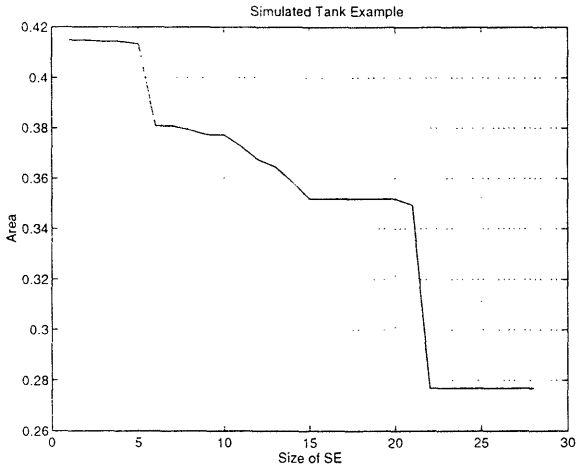


Figure 2: Area of morphologically processed imaged as a function of r.

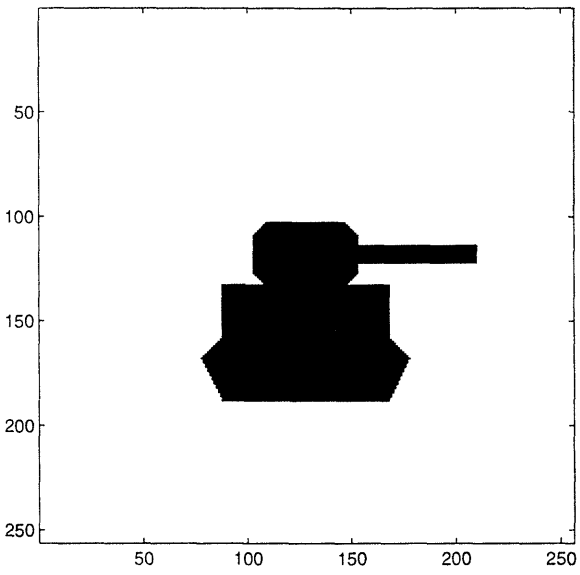


Figure 3: Simulated Tank.

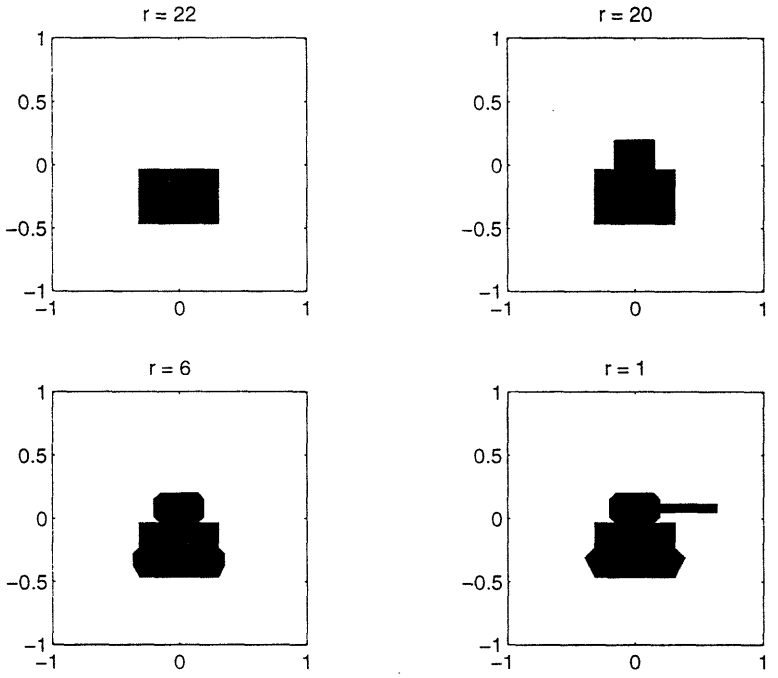


Figure 4: Output of morphologically filtering at significant scales for simulated tank.

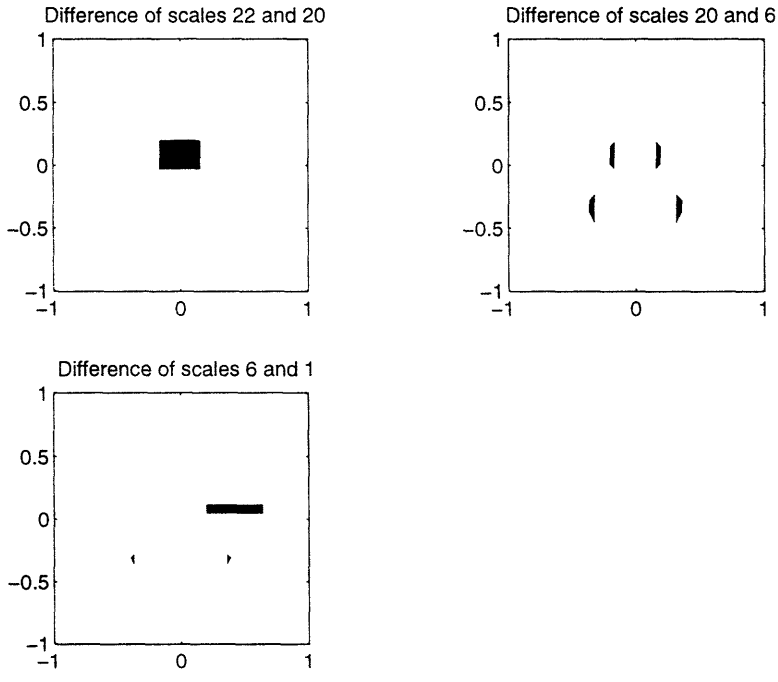


Figure 5: Difference between significant scales for simulated tank.

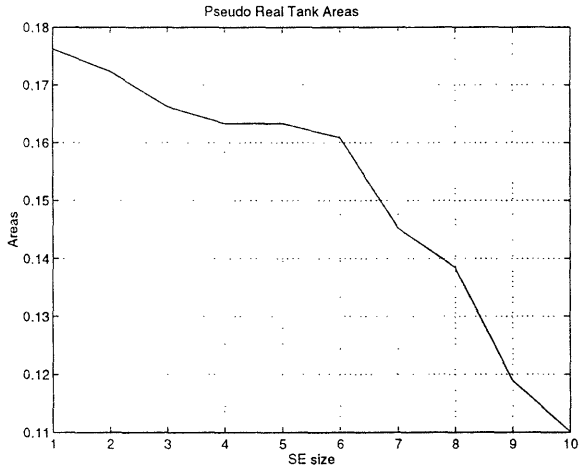


Figure 6: Area of morphologically processed imaged as a function of r.

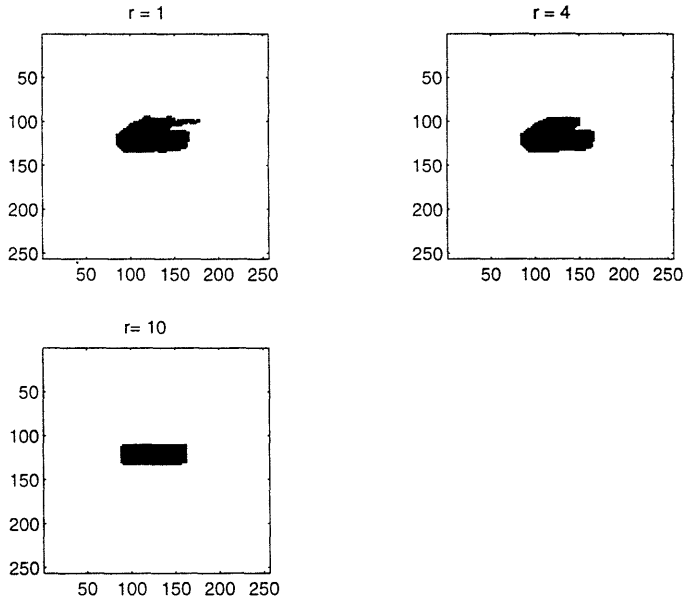


Figure 7: Output of morphologically filtering at significant scales for model board tank.

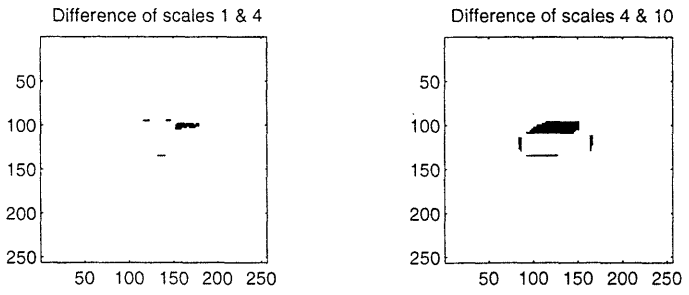


Figure 8: Difference between significant scales for model board tank.

4.2 Wavelet Network Feature Extraction

In this section, we obtain a wavelet network decomposition of each of the canonical pieces of the simulated tank from the previous section. The wavlet network is used in phase because it is robust to noise and variability in the underlying image. These are preliminary results in which we fit only one wavelon to each of the canonical pieces of the original image. In future work, we will examine methods to incorporate more wavelons.

Having obtained a morphological multiscale decomposition of the object, we may now apply the wavelet network. The wavelet network parameters will then be our feature vector. To do this, we fit one wavelon to each of the pieces of the simulated tank in the set of images $\{I_{22}, I_{22-20}, I_{6-1}\}$. The wavelet network parameters are given in Table 4.2 and the final wavelet network approximation is given in Figure 9. These parameters specify the size and relative position of each of the pieces of the simulated tank.

This is, of course, only a preliminary result meant to show the promise of wavelet networks in feature extraction. This coarse wavelet network decomposition compresses the image in a manner very similar to human perception. As subsequent wavelons are fit to each piece, the approximate image will more closely resemble the original. Note, however, even in the absence of more detailed features, this representation conveys the underlying structure of the

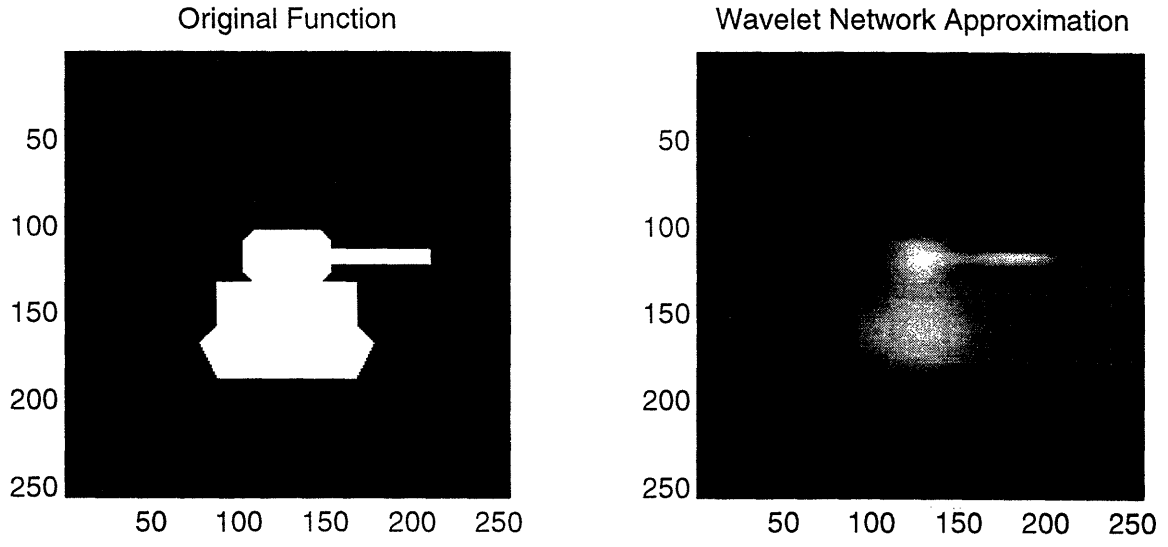


Figure 9: Wavelet network representation of simulated tank

original object. At this stage, this information might be used to determine pose for use in other object recognition schemes.

	Translation	Dilation	Weight
Wavelon #1	0.0000	2.7871	1.4172
	-0.2467	4.0289	
Wavelon #2	0.4220	3.9941	1.4085
	0.0859	25.3231	
Wavelon #3	0.0000	5.5595	1.4066
	0.0898	7.5658	

Table 1: Final Wavelet Network Parameters

4.3 A Laser Radar Example

Finally, we demonstrate how wavelet networks may be used to refine some chosen wavelet basis. In keeping with our proposed integrated approach to ATD/R, we consider an application of wavelet networks to laser range data, as described in [16]. In [16], the goal is anomaly suppression in laser range data where a multiresolution wavelet basis is fitted to the data in conjunction with the expectation-maximization (EM) algorithm. Here we show a simplified example with data similar to that used in [16], to highlight the benefits of using wavelet networks.

Consider the true signal or range profile shown in Figure 10. To accurately represent this using a dyadic wavelet decomposition where the dilations are exponentially distributed and the translations are uniformly distributed, one would need to include many scale levels. Exclusion of some of these scale levels yields a reconstruction that does not accurately represent the true signal. For example, consider the dyadic wavelet decomposition of the true range profile in Figure 10. Here we used the Haar wavelet basis, $\psi(d(x-t))$ with 16 wavelons where $\psi(x)$ defined as

$$\psi(x) = \begin{cases} 1 & \text{for } x \leq 256 \\ -1 & \text{for } 256 < x < 512 \end{cases} \tag{11}$$

Specifically, the set of dilation/translation pairs used are $\{(d,t)\} = \{(0,0),(2^s, 512r/2^s)\}$ for $s = \{0,1,2,3\}$ and $r = \{0\ldots., 2^s - 1\}$. This reconstruction does not capture the overall shape of the range profile and yields a nomalized mean square error of 0.5446. The normalized mean square error (NSRMSE) is defined as

$$\text{NSRMSE}(g) = \sqrt{\frac{\sum_{k=1}^{K}[f(x_k) - g(x_k)]^2}{\sum_{k=1}^{K}(f(x_k) - \bar{f})^2}} \tag{12}$$

where f is the function to be approximated, g is the wavelet network approximation, K is the number of observations and $\bar{f} = \frac{1}{K}\sum_{k=1}^{K} f(x_k)$. This error may be significantly reduced by using a wavelet network to refine the Haar wavelet basis. The wavelet network is initialized with the dyadic wavelet decomposition with the same number of wavelons as above. The resulting wavelet network decomposition is shown in Figure 10. This decomposition identifies the discontinuities in the true range profile much more accurately. In addition, the normalized mean square error has been reduced to 0.3022.

5 Conclusion

In conclusion, in this paper, we have described some preliminary steps towards developing geometric and multiscale features to be used in object recognition. Our current focus has been a two-phase approach of morphological decomposition followed by wavelet network approximation. The final wavelet network parameters may be thought of as a coarse scale approximation to the original object. In addition, we have demonstrated how wavelet networks

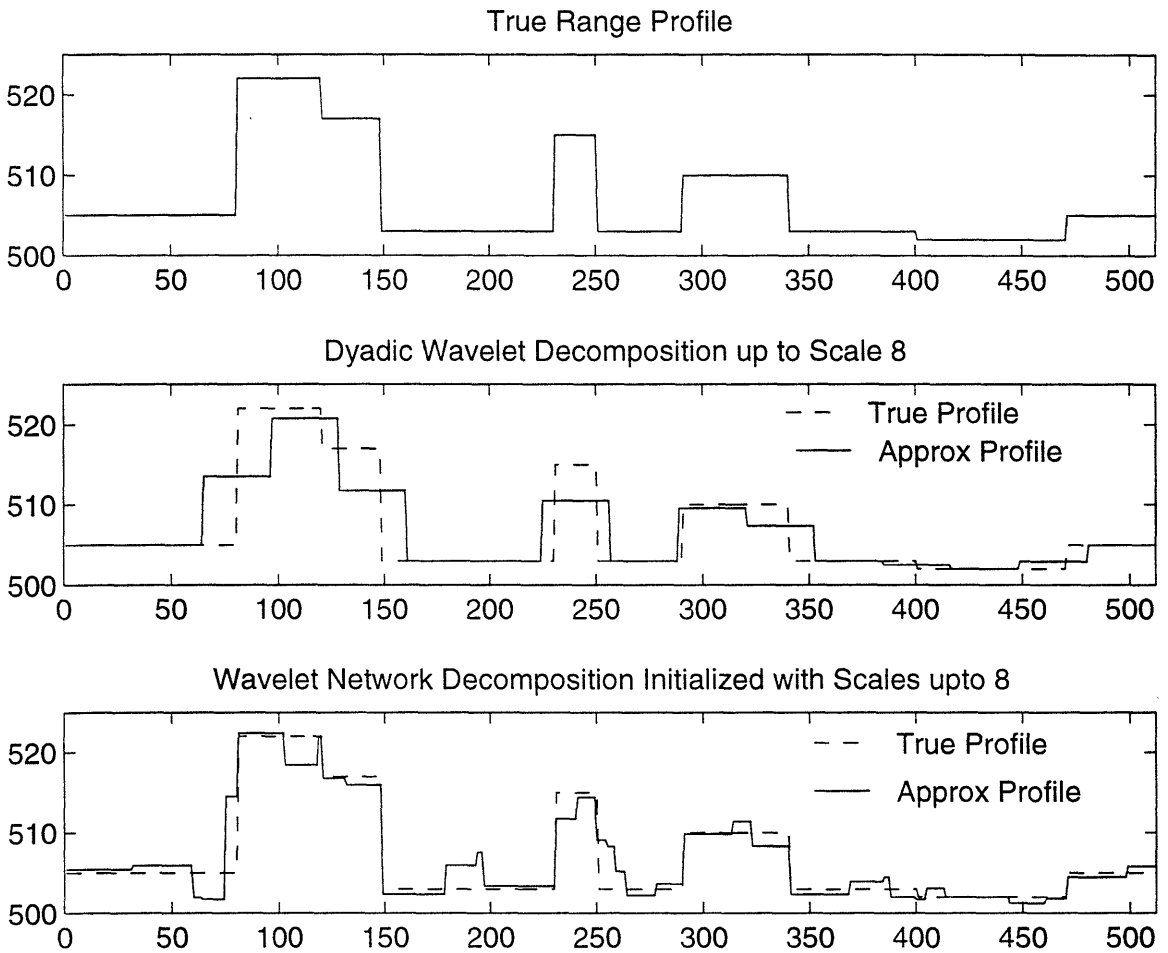


Figure 10: Application of wavelet networks to range profile estimation

may be used in laser radar range profile estimation. Here the wavelet network was used to refine the chosen wavelet basis. Future work in this area will include an investigation of other shape descriptors (e.g. pattern spectrum [17]) and developing a stochastic model framework.

References

[1] C. Giardina and E. Dougherty, *Morphological Methods in Image and SIgnal Processing.* Prentice Hall, 1988.

[2] J. Serra, *Image Analysis and Mathematical Morphology.* Academic Press, 1982.

[3] P. Maragos and R. Schafer, "Morphological filters-parts I and II," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. ASSP-35, no. 8, pp. 1153–1184, 1987.

[4] J. Verley and R. Delanoy, "Adaptive mathematical morphology for range imagery," *IEEE Transactions on Image Processing*, vol. 2, pp. 272–275, April 1993.

[5] T. Esselman and J. Verly, "Some applications of mathematical morphlogy to range imagery," in *Proceedings of ICASSP*, pp. 245–248, April 1987.

[6] A. Toet, "A morphological pyramidal image decomposition," *Pattern Recognition Letters*, vol. 9, pp. 255–261, May 1989.

[7] A. Toet, "Image fusion by a ratio of low-pass pyramid," *Pattern Recognition Letters*, vol. 9, pp. 245–253, May 1989.

[8] R. Haralick, X. Zhuang, C. Lin, and J. Lee, "The digital morpholoogical sampling theorem," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, pp. 2067–2090, December 1989.

[9] H. Heijmans and A. Toet, "Morphological sampling," *CVGIP:Image Understanding*, vol. 54, pp. 384–400, November 1991.

[10] P. Burt and E. Adelson, "The Laplacian pyramid as a compact image code," *IEEE Transactions on Communications*, vol. COM-31, April 1983.

[11] Q. Zhang and A. Benveniste, "Wavelet networks," *IEEE Transactions on Neural Networks*, vol. 3, November 1992.

[12] Q. Zhang, "Wavelet network: the radial structure and an efficient initialization procedure," *Technical Report of Linkoping University*, 1992.

[13] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of Control, signals and systems*, vol. 2, pp. 303–314, 1989.

[14] K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks*, vol. 2, pp. 359–366, 1989.

[15] R. Hecht-Nielsen, "Theory of the backpropagation neural network," in *Proceedings of IJCNN*, 1989.

[16] I. Fung, "Multiresolution laser radar range profiling with the expectation-maximization algorithm," Master's thesis, MIT, 1992.

[17] P. Maragos, "Pattern spectrum and mulitscale shape representation," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 11, no. 7, 1989.