Information systems based on neural network and wavelet methods with application to decision making, modeling and prediction tasks

D.A. Karras
University of Ioannina, Department of Informatics, Ioannina, Greece,
S.A. Karkanis
NCSR "Demokritos", Institute of Nuclear Technology, Ag. Paraskevi, Greece, and
B.G. Mertzios
Democritus University of Thrace, Department of Electrical and Computer Engineering, Xanthi, Greece

Introduction

Information processing of signals and images has become an essential part of contemporary scientific and developmental activity in the field of computer science and engineering. It is used in telecommunications, in diagnosis, in the transmission and analysis of satellite images, in medical imaging, in quality control and in time series forecasting among many important disciplines of computer aided engineering. The tasks where these information processing systems should be involved in order to be integral parts of such applications are decision making, prediction and modeling. For instance, medical diagnosis and quality control involve decision making (deciding, using signal and image information, whether a person suffers from an illness or a product should be rejected as useless), telecommunications and data transmission involve modeling of signals and images (e.g. for their efficient compression) and finally, time series forecasting involves prediction of signal fluctuations in time.

All of these applications involve the analysis and interpretation of complex time series in one (signals) or two dimensional form (images). Therefore, designing and building robust information systems that process signals and images becomes increasingly important in contemporary research and development activities. The desired properties of such systems include accurate analysis, efficient coding (modeling is a prerequisite), rapid transmission, and then reconstruction of the delicate oscillations or fluctuations as a function of time or space. These characteristics are essential since the information contained in signals and images is effectively present in the complicated

Kybernetes, Vol. 27 No. 3, 1998, pp. 224-236, © MCB University Press, 0368-492X

arabesques appearing in their representations. Thus, anomalies and discontinuities play an instrumental role in signal and image representation and characterization since they carry important information about fluctuations in time or space. The majority of the above mentioned applications involve in one way or another the detection and prediction of such anomalies. For instance, quality control involves defect recognition, medical diagnosis involves peak detection and modeling (NMR etc.), and finally, data transmission and time series forecasting involve modeling and prediction of abnormal fluctuations.

Therefore, the designing and building of robust information processing systems for decision making, modeling and prediction tasks to be used in the applications discussed above involve dealing with the problem of efficiently and accurately detecting and predicting signal and image anomalies in time or space respectively. The goal of this paper can be summarized as an attempt to offer a solution to this problem, suggesting a novel methodology. This methodology employs neural network and wavelet techniques. The rationale underlying the use of these methods is that wavelets offer a successful time-frequency signal and image representation for enhanced information localization (and thus detection of abnormal fluctuations), while neural networks offer a powerful approach to solve the classification and function approximation problems obviously involved in the decision making, modeling and prediction tasks under consideration. The novelty of our approach in the use of these methodologies is that we suggest the fusion of the information coming from the original signal or image input with the information coming from the wavelet transformed original input by using neural networks, so as to accomplish improved detection of abnormalities.

This novel methodology of building the information systems under consideration is applied to the design of a system for defect recognition from images and to a system for NMR signal modeling and prediction. Moreover, it can find applications in almost every information system processing signals and images for use in decision making, modeling and prediction tasks.

Defect recognition from images is becoming increasingly significant in a variety of applications since quality control plays a very important role in contemporary manufacturing of virtually every product. In addition, peak and abnormal fluctuation detection-prediction plays a critical role in signal characterization (for instance, in medical diagnosis) and time series forecasting. Despite the lot of interest, little work has been done in this field since these classification and function approximation problems, respectively, present many difficulties. The resurgence of interest, however, for neural network research has revealed the existence of powerful classifiers-nonlinear function approximators. In addition, the emergence of the 1-D and 2-D wavelet transform (Kolaczyk, 1994; Meyer, 1993) as a popular tool in signal and image processing offers the ability of robust feature extraction in signals and images. Combinations of both techniques have been used with success in various applications (Lee *et al.*, 1996). Therefore, it is worth attempting to investigate whether they can jointly offer a viable solution to the defect recognition and abnormal fluctuation modeling-prediction problem. To this end, concerning the design of information systems that process images, we propose a novel methodology in detecting defective areas in them by examining the discrimination abilities of their textural properties. Besides neural network classifiers and the 2-D wavelet transform, the tools utilized in such an analysis are co-occurrence matrices based textural feature extraction (Haralick *et al.*, 1973) and SVD analysis. Concerning the design of information systems that process signals, we propose a novel methodology, employing the 1-D wavelet transform and neural networks for function approximation, for abnormal fluctuation modeling-prediction.

The first problem, that is defect recognition from images, can be clearly viewed as an image segmentation one, where the image should be segmented in defective and non-defective areas only, unlike its conventional consideration. The classical segmentation problem, that is dividing an image into homogeneous regions, the discovery of a generally effective scheme, remains a challenge. To this end, many interesting techniques have been suggested so far, including spatial frequency techniques and relevant ones like texture clustering in the wavelet domain (Porter and Canagarajah, 1996). Most of these methodologies use very simple features like the energy of the wavelet channels or the variance of the wavelet coefficients (Unser, 1995). Our approach stems from this line of research. However, there is need for much more sophisticated feature extraction methods if one wants to solve the segmentation problem in its defect recognition incarnation, taking into account the high accuracy required. Following this reasoning we propose to incorporate in the research efforts the co-occurrence matrices analysis, since it offers a very precise tool for describing image characteristics and especially texture (Haralick *et al.*, 1973). It clearly provides second order information about pixel intensities when the majority of the other feature extraction techniques do not exploit it at all.

The second problem, that is abnormal fluctuation modeling-prediction, can be viewed as a function approximation task. Fusion of original signal data and the coefficients of their corresponding DWT (discrete wavelet transform) multi-resolution analysis for achieving improved solutions, is the novelty suggested. The optimal features in the wavelet domain utilized here, unlike the complex ones discussed above for defect recognition, for enhanced function approximation are simply the wavelet coefficients of the selected channels.

There are two main stages of the suggested novel information processing systems. Namely, optimal feature selection in the wavelet domain (optimal in terms of the information these features carry) and neural network based classification-function approximation. The viability of the concepts and methods employed in the proposed approach is illustrated in the experimental section of this paper, where it is clearly shown that, by achieving a 98.75 per cent defective area classification accuracy as well as an excellent reconstruction of an NMR signal, our methodology is very promising for use in the design of effective information systems for processing signals and images in a variety of applications.

The characteristics of the wavelet transform are summarized below. In the following sections a detailed description of the suggested methodology is offered. Then we show the promising results obtained and finally, we conclude the paper.

An introduction to the wavelet transform

Wavelets offer a general mathematical approach for hierarchical function decomposition. According to this transformation, a function, which can be a function representing an image, a curve, signal etc., can be described in terms of a coarse level in addition with details that range from broad to narrow scales. Wavelets offer a novel technique for computing the levels of detail present, under a framework that is based on a chain of *approximation* vector spaces $\{Vj \subset L^2(\Re^2), j \in Z\}$ and a *scaling* function ϕ such that the set of functions $\{2^{-j/2}\phi(2^{-j}t-k) k \in Z\}$ form an orthonormal basis for V_j. These two components introduce a mathematical framework presented by Mallat (1989) and are called multi-resolution analysis.

A multi-resolution analysis (MRA) scheme of $L^2(\Re^2)$ can be defined as a sequence of closed subspaces $\{V_j \subset L^2(\Re^2), j \in Z$ satisfying the following properties :

Containment : $V_j \subset V_{j-1} \subset L^2$; for all $j \in Z$.

Decrease : $\lim_{j \to \infty} V_j = 0$, i.e. $\bigcap_{j>N} V_j = \varnothing$, for all N .

Increase : $\lim_{j \to -\infty} V_j = L^2$, i.e. $\bigcup_{j<N} V_j = L^2$, for all N .

Dilation : $u(2t) \in V_{(j-1)} \Leftrightarrow u(t) \in V_j$.

Generator : There is a function $\phi \in V_0$ whose translation $\{\phi(t-k : k \in Z)\}$ forms a basis for V_0 .

By defining *complementary subspaces* $W_j = V_{j-1} - V_j$ so that $V_{j-1} = V_j + W_j$ then we can write, according to the "increase" property that

$$L^2(\Re^2) = \sum_{j \in Z} W_j \tag{1}$$

The subspaces W_j are called *wavelet subspaces* and contain the difference in signal information between the two spaces V_j and V_{j-1}. These sets contribute to a wavelet decomposition of L^2 according to (1). In Mallat (1989) it has been proved that a mother wavelet ψ can be created such that the set of functions $\{\psi(2^{-j}t-k) k \in Z)$ forms a basis for W_j. The W_j are mutually orthogonal and the set of scaled and dilated wavelets $\{2^{-j/2}\psi(2^{-j}t-k) j \in Z, k \in Z\}$ provides an orthonormal *wavelet basis* for $L^2(\Re^2)$. Approximating and detailed signals can be obtained by projecting the input signal to the corresponding (approximation or detailed) space. Practically the approximation and detail projection coefficients associated with V_j and W_j are computed from the approximation and detail coefficients at the next higher scale V_{j-1}, using a quadrature mirror filter (QMF) pair and a pyramidal subband coding scheme (Kocur *et al.,* 1996; Wickerhauser, 1994).

Optimal feature selection in the wavelet domain

The problem of texture discrimination, aiming at segmenting the defective areas in images, is considered in both the time and the wavelet domain, since it has been demonstrated that discrete wavelet transform (DWT) can lead to better texture modeling (Ryan *et al.*, 1996). On the other hand, the task of abnormal fluctuation modeling-prediction is considered in both the time and wavelet domain so as to exploit the signal representation capabilities of the wavelet transformation (Freeman, 1995). Besides, in this way we can better exploit the well known local information extraction properties of wavelet signal decomposition as well as the well known features of wavelet denoising procedures (Donoho and Johnstone, 1995). We use the popular 2-D and 1-D discrete wavelet transform schemes (Kolaczyk, 1994; Meyer, 1993) in order to obtain the wavelet analysis of the original images and signals containing defects and abnormal fluctuations respectively. It is expected that this kind of information considered in the wavelet domain should be smooth but owing to the well known time-frequency localization properties of the wavelet transform, the defective areas and the abnormal fluctuations, whose statistics vary from the ones of the image-signal background, should more or less clearly emerge in the foreground. We have experimented with the standard 2-D and 1-D wavelet transforms using nearly all the well known wavelet bases like Haar, Daubechies, Coiflet, Symmlet etc. as well as, in the case of defect recognition information processing system design, with Meyer's and Kolaczyk's 2-D wavelet transforms (Kolaczyk, 1994). However, and this is very interesting, only the 2-D and 1-D Haar wavelet one level transforms have exhibited the expected and desired properties. All the other orthonormal, continuous and compactly supported wavelet bases have smoothed the images and signals so much that the defective areas and abnormal fluctuations do not appear in the subbands. We have performed a one-level wavelet decomposition of the images and signals, thus resulting in four and two main wavelet channels respectively. Concerning the wavelet decomposition of the images to be handled by the proposed first information system, among the three channels 2, 3, 4 (frequency index) we have selected for further processing the one whose histogram presents the maximum variance. A lot of experimentation has shown that this is the channel corresponding to the most clear appearance of the defective areas. On the other hand, concerning the wavelet decomposition of the signals to be processed by the second information system, we have selected the second channel of the detailed wavelet coefficients. Regarding the first information system, the subsequent step in the proposed methodology is to raster scan both the image obtained from the selected wavelet channel and the original image with sliding windows of $M \times M$ and $2M \times 2M$ dimensions respectively. We have experimented with 256×256 images and we have found that $M = 8$ is a good candidate size for the sliding window. Correspondingly, regarding the second information system, the subsequent step is to scan both the signal obtained from the selected wavelet channel and the original signal with sliding windows of length M and $2M$ respectively. We have experimented with an NMR signal containing 1,024 samples. Thus, the two obtained wavelet channels contain 512 samples. We have experimentally found that $M = 16$ is the minimum window length yielding meaningful wavelet coefficients.

Concerning the second suggested novel information system for processing NMR signals no further feature extraction analysis takes place. We simply use as input to the signal modeling-prediction neural network subsequently employed in the proposed novel methodology, the feature vector obtained directly from the sliding windows in time and wavelet domains. Thus, 32 samples from each window scanning the original signal along with its corresponding 16 wavelet coefficients of the wavelet channel with frequency index 2, jointly form a feature vector of 48 components.

In the first information system under consideration, further feature extraction is conducted. For each sliding window we perform two types of analysis in order to obtain features optimal in terms of information content. First, we use the information that comes from the co-occurrence matrices (Haralick *et al.*, 1973). These matrices represent the spatial distribution and the dependence of the gray levels within a local area. Each (i,j)th entry of the matrices represents the probability of going from one pixel with gray level (i) to another with a gray level (j) under a predefined distance and angle. More matrices are formed for specific spatial distances and predefined angles. From these matrices, sets of statistical measures are computed (called feature vectors) for building different texture models. We have considered four angles, namely 0, 45, 90, 135 as well as a predefined distance of one pixel in the formation of the co-occurrence matrices. Therefore, we have formed four co-occurrence matrices. Owing to computational complexity issues regarding co-occurrence matrices analysis we have quantized the image obtained from the selected wavelet channel into 16 gray levels instead of the usual 256 levels, without adverse effects in defective area recognition accuracy. This procedure also renders the on-line implementation of the proposed system highly feasible. Among the 14 statistical measures, originally proposed by Haralick (1973), that are derived from each co-occurrence matrix, we have considered only four of them. Namely, angular second moment, correlation, inverse difference moment and entropy.

- ***Energy - Angular Second Moment*** $$f1 = \sum_i \sum_j p(i,j)^2$$
- ***Correlation*** $$f2 = \frac{\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i * j) p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

- ***Inverse Difference Moment*** $f3 = \sum_i \sum_j \frac{1}{1+(i-j)^2} p(i,j)$

- ***Entropy*** $f4 = - \sum_i \sum_j p(i,j) \log(p(i,j))$

We have found experimentally that these measures provide high discrimination accuracy that can be only marginally increased by adding more measures in the feature vector. Thus, using the above mentioned four co-occurrence matrices we have obtained 16 features describing spatial distribution in each 8×8 sliding window in the wavelet domain. In addition, we have formed another set of eight features for each such window by extracting the singular values of the matrix corresponding to this window. SVD analysis has recently been successfully related to invariant pattern recognition (Al-Shaykh and Doherty, 1996). Therefore, it is reasonable to expect that it provides a meaningful way for characterizing each sliding window, thus preserving first order information regarding this window, while, on the other hand, the co-occurence matrices analysis extracts second order information. Therefore, we have formed, for each sliding window in the image of the selected wavelet channel, a feature vector containing 24 features that uniquely characterizes it in the wavelet domain. In addition, we have formed another set of four features that uniquely characterize the corresponding 16×16 sliding window in the original 256×256 image. More specifically, we have first transformed the original image into one of equal dimensions containing instead of each pixel value its corresponding probability of existence within the image. This new image is raster scanned by 16×16 sliding windows. For each such window we obtain a set of four features by calculating the above four mentioned statistical measures. Finally, these 28 component feature vectors feed the neural classifier of the next stage of the suggested methodology.

Information processing using neural networks

In both information systems we employ a neural network technique for either classification or function approximation. Concerning the suggested novel information system for processing images with defective areas, associated with decision making tasks in quality control, we subsequently use a neural classifier. After obtaining information about the structure and other characteristics of each image, utilizing the above depicted methodology, we employ a supervised neural network architecture of the multilayer feed forward type (MLPs), trained with the online back propagation error algorithm, having as a goal to decide whether a texture region belongs to a defective part or not. The inputs to the network are the 28 features described above. The best network architecture that has been tested in our experiments is the 28-35-35-1. The desired outputs during training are determined by the corresponding sliding window location. More specifically, if a sliding window belongs to a defective area the desired output of the network is one, otherwise it is zero. We have defined, during MLP training phase, that a sliding window belongs to a defective area if any of the pixels in the 4×4 central window inside the original 8×8 corresponding sliding window belongs to the defect. The reasoning underlying this definition is that the decision about whether a window belongs to a defective area or not should come from a large neighborhood information, thus preserving the 2-D structure of the problem and not from information associated with only one pixel (e.g the central pixel). In addition and probably more significantly, by defining the two classes in such a way, we can obtain many more training patterns for the class corresponding to the defective area, since defects, normally, cover only a small area of the original image. It is important for the effective neural network classifier learning to have enough training patterns for each one of the two classes but, on the other hand, to preserve as much as possible the a priori probability distribution of the problem. We have experimentally found that a proportion of 1:3 for the training patterns belonging to defective and non-defective areas respectively is very good for achieving both goals.

Correspondingly, regarding the second suggested novel information system for processing NMR signals, which normally contain abnormal fluctuations, associated with modeling and prediction tasks in signal coding and time series forecasting, we next employ a function approximation neural network. It is again a supervised neural network architecture of the MLP type, trained with the online back propagation error algorithm, having as a goal to predict from the $M = 32$ samples belonging to each sliding window the corresponding next two signal values. The inputs to the network are the $32+16 = 48$ features above described. The best network architecture that has been tested in our experiments is the 48-35-35-2. The desired outputs during training are for each sliding window its corresponding two next signal values. The training set has been formed by such (48,2) pairs randomly selected from the set of all pairs corresponding to the original signal. Therefore, this task cannot be characterized as a purely extrapolation one but it includes both prediction and interpolation concepts in a joined fashion.

Results and discussion

The efficiency of our approach in designing information systems effective in processing images and signals is illustrated by demonstrating the performance of the two novel systems described in the previous sections.

Decision making in automated inspection images

The first one, concerning decision making in automated inspection images, is tested in the textile image shown in Figure 1, which contains a very thin and long defect in its upper side as well as some smaller defects elsewhere. This image is 256×256, while the four wavelet channels obtained by applying the 2-D Haar wavelet transform are 128×128. These wavelet channels are shown in Figure 2. There exist 14,641 sliding windows of 8×8 size in the selected wavelet

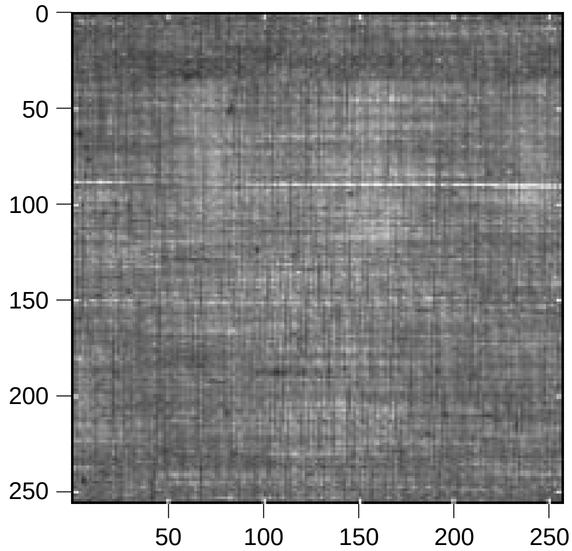

Figure 1.
Textile image with a defect

channel 3, shown in Figure 3. The neural network has been trained with a training set containing 1,009 patterns extracted from these sliding windows as described above. Out of the 1,009 patterns 280 belong to the long and thin defective area of the upper side only, while the rest belong to the class of non-defective areas. The learning rate coefficient was 0.3 while the momentum one was 0.4. The neural network has been tested on all the 14,641 patterns coming from the sliding windows of the third wavelet channel. The results obtained by

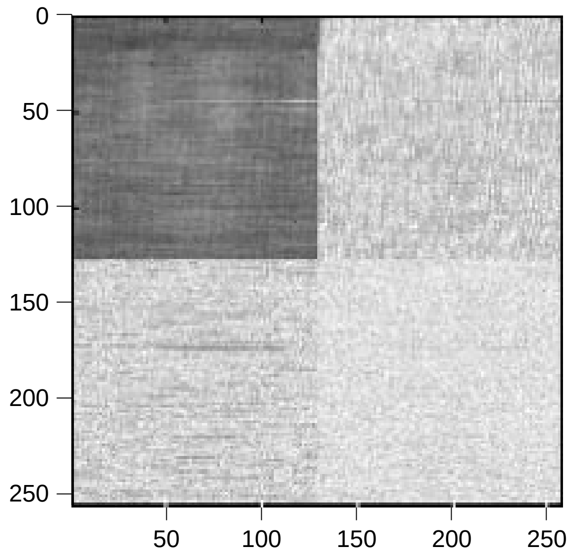

Figure 2.
Wavelet transformed original image

utilizing the complete set of 28 features (16 come from the statistical measures in the wavelet domain, eight from the SVD and four from the statistical measures in the original image) are shown in Figure 5. Figure 4 illustrates the results obtained by omitting the four statistical measures corresponding to the original image. Note that the network in both cases was able to generalize and also find some other minor defects, while another network of the same type, trained with the 64 pixel values of the sliding windows, under exactly the same conditions, was able to find only the long and thin defect. However, Figure 5 exhibits the superiority of the suggested methodology over the approach of

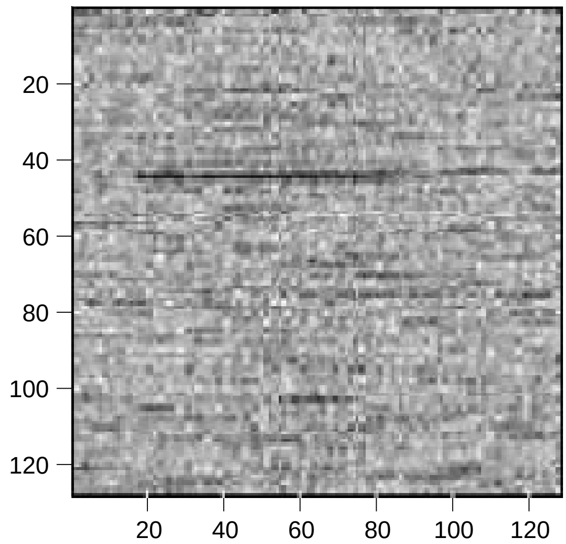

Figure 3.
Wavelet channel chosen

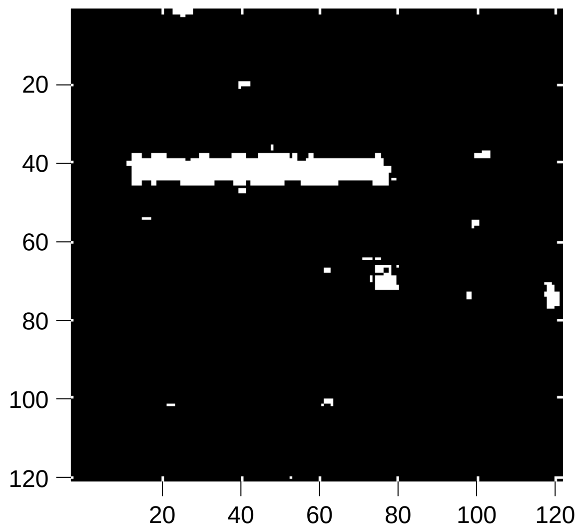

Figure 4.
Resulted image – white regions represent defects

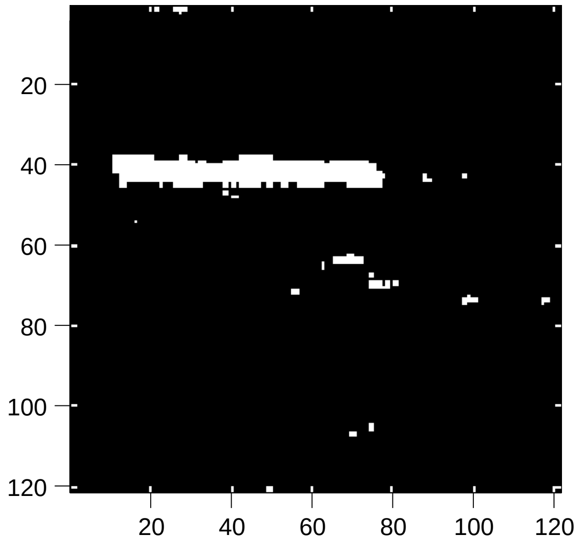

Figure 5.
Resulted image – white regions represent defects

using only wavelet domain information. This fact demonstrates the efficiency of our feature extraction methodology. Finally, in terms of classification accuracy we have achieved an overall 98.75 per cent regarding Figure 5 and 98.48 per cent regarding Figure 4.

Modeling and prediction tests

The second system, concerning modeling and prediction tasks in signal coding and time series forecasting, is tested in the NMR signal shown in Figure 6. This signal contains 1,024 samples and by following the methodology depicted in previous sections, similar to the one developed for defect recognition, we can construct a set of 991 patterns of (48,2) pairs if one exploits wavelet domain information and (32,2) pairs otherwise. The training set comprised 800 patterns randomly selected out of the 991. The neural system is again tested in the set of the 991 patterns yielding the results shown in Figure 7 and 8 respectively. Note the superiority in accuracy in the signal representation, shown by reconstructing the signal from only its predicted values, obtained by employing the proposed methodology in Figure 7. Finally, the prediction error is 0.02 in the average per predicted value.

Conclusions

We have proposed a novel methodology for designing information systems based on wavelet and neural network techniques, suitable for processing signals and images for decision making, modeling and prediction tasks. The excellent results obtained by using two such novel systems for defect recognition in quality control and for NMR signal modeling and prediction clearly demonstrate that our methodology deserves further evaluation in building information systems.

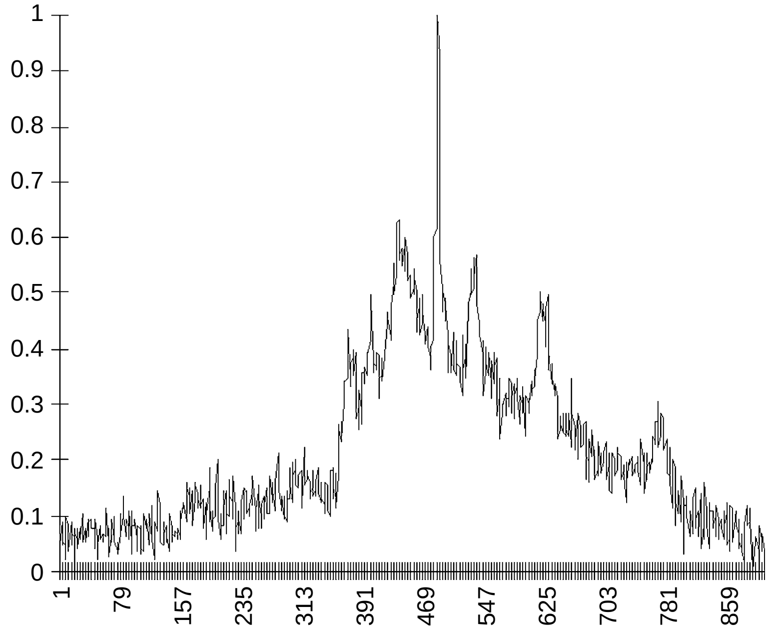

Figure 6.
NMR signal

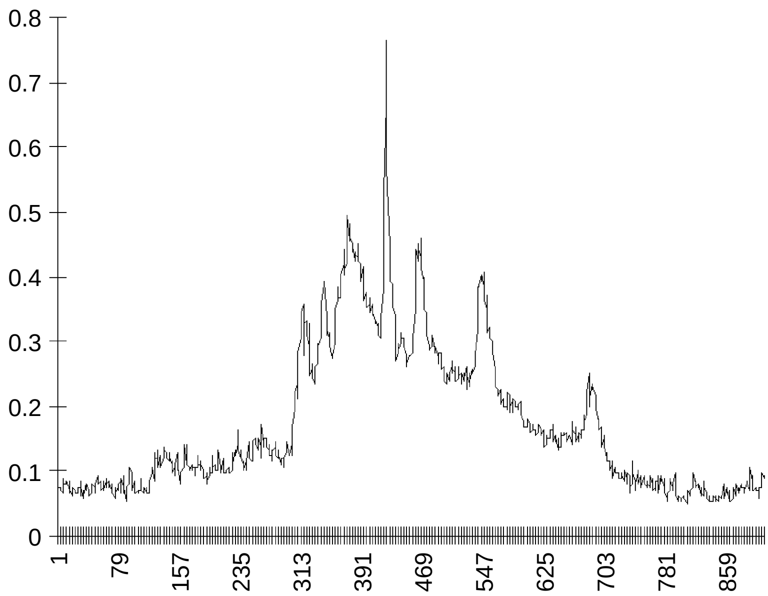

Figure 7.
NN + wavelet
reconstructed signal

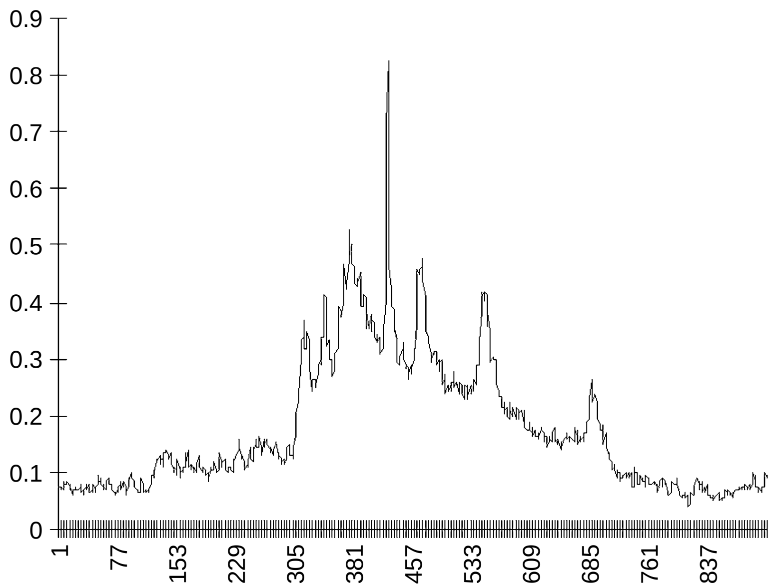

Figure 8.
NN reconstructed signal

References and further reading

Al-Shaykh, O.K. and Doherty, J.E. (1996), "Invariant image analysis based on radon transform and SVD", *IEEE Trans. Circuits and Systems,* Vol. 43 No. 2, pp. 123-33.

Antonini, M., Barlaud, M., Mathieu, P. and Daubechies, I. (1992), "Image coding using wavelet transform", *IEEE trans Image Processing*, Vol.1, pp. 205-20.

Donoho, D.L. and Johnstone, I.M. (1995), "Ideal time-frequency denoising", Technical Report, Dept. of Statistics, Stanford University.

Freeman, M.O. (1995), "Wavelet signal representations with important advantages", *Photonics New,* pp. 8-14.

Haralick, R.M., Shanmugam, K. and Dinstein, I. (1973), "Textural features for image classification", *IEEE Trans. Systems, Man and Cybernetics,* Vol. SMC-3, No. 6, pp. 610-21.

Kocur, C.M., Rogers, S.K. *et. al.* (1996), "Using neural networks to select wavelet features for breast cancer", *IEEE Eng. in Medicine and Biology*, Vol. 15 No. 3, pp. 95-102.

Kolaczyk, E. (1994), "WVD solution of inverse problems", Doctoral Dissertation, Stanford University, Dept. of Statistics.

Lee, C.S. *et. al.* (1996), "Feature extraction algorithm based on adaptive wavelet packet for surface defect classification", *ICIP*-96, 16-19 Sept., Lausanne, Switzerland.

Mallat, S.G. (1989), "A theory for multi-resolution signal decomposition: the wavelet representation", *IEEE Trans. on Pattern Analysis and Machine Intelligence,* Vol. 11 No. 7, pp. 674-93.

Meyer, Y. (1993), *Wavelets: Algorithms and Applications,* SIAM, Philadelphia.

Porter, R. and Canagarajah, N. (1996), "A robust automatic clustering scheme for image segmentation using wavelets", *IEEE Trans. on Image Processing*, Vol. 5 No. 4, pp. 662- 5.

Ryan, T.W., Sanders, D., Fisher, H.D. and Iverson, A.E. (1996), "Image compression by texture modeling in the wavelet domain", *IEEE trans. Image Processing*, Vol. 5 No. 1, pp. 26-36.

Unser, M. (1995), "Texture classification and segmentation using wavelet frames", *IEEE trans. Image Processing,* Vol. 4 No. 11, pp. 1549-60.

Wickerhauser, M.V. (1994), *Adapted Wavelet Analysis from Theory to Software,* IEEE Press.