International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 9, No. 1 (2011) 111–130

DOI: 10.1142/S0219691311003967

World Scientific
www.worldscientific.com

PYRAMIDAL HYBRID APPROACH: WAVELET NETWORK WITH OLS ALGORITHM-BASED IMAGE CLASSIFICATION

OLFA JEMAI*

Higher Institute of Computer Sciences, Road El Jorf
Km 22.5 - 4119 Medenine, Tunisia
and
Research Group on Intelligent Machines
National Engineering School of Sfax (ENIS)
BP 1173, Sfax, 3038, Tunisia
olfa.jemai@ieee.org

MOURAD ZAIED[†], CHOKRI BEN AMAR[‡]
and MOHAMED ADEL ALIMI[§]

Research Group on Intelligent Machines
National Engineering School of Sfax (ENIS)
BP 1173, Sfax, 3038, Tunisia
[†]*mourad.zaied@ieee.org*
[‡]*Chokri.benamar@ieee.org*
[§]*adel.alimi@ieee.org*

Received 13 October 2009
Revised 10 January 2010

Taking advantage of both the scaling property of wavelets and the high learning ability of neural networks, wavelet networks have recently emerged as a powerful tool in many applications in the field of signal processing such as data compression, function approximation as well as image recognition and classification. A novel wavelet network-based method for image classification is presented in this paper. The method combines the Orthogonal Least Squares algorithm (OLS) with the Pyramidal Beta Wavelet Network architecture (PBWN). First, the structure of the Pyramidal Beta Wavelet Network is proposed and the OLS method is used to design it by presetting the widths of the hidden units in PBWN. Then, to enhance the performance of the obtained PBWN, a novel learning algorithm based on orthogonal least squares and frames theory is proposed, in which we use OLS to select the hidden nodes. In the simulation part, the proposed method is employed to classify colour images. Comparisons with some typical wavelet networks are presented and discussed. Simulations also show that the PBWN-orthogonal least squares (PBWN-OLS) algorithm, which combines PBWN with the OLS algorithm, results in better performance for colour image classification.

Keywords: Wavelet network; beta function; OLS algorithm; frames; classification.

AMS Subject Classification: 22E46, 53C35, 57S20

*Corresponding author.

1. Introduction

Image classification is an important problem which is particularly difficult for traditional machine learning algorithms mainly because of the high number of input variables that may describe images (i.e. pixels). Indeed, with a high number of variables, these methods tend to produce very unstable models with low generalization. Computing times can also be detrimental in such extreme conditions.

To handle this high dimensionality, image classification systems usually rely on a pre-processing step, specific to the particular problem and application field. This step aims at extracting a reduced set of interesting features from the initially huge number of pixels. This reduced set is then used as new input variables for traditional learning algorithms, possibly tuned to the specific application. The limitation of this approach is clear: when considering a new problem or application field, it is necessary to manually adapt the pre-processing step by taking into account the specific characteristics of the new application. But, at the same time, recent advances in automatic learning have produced new methods that are able to handle more and more complex problems without requiring any *a priori* information about the application. These methods are increasingly competitive with methods specifically tailored for these fields.

To remedy these problems, a solution based on Wavelets Neural Networks (RBF neural networks with wavelets as transfer functions) is proposed. Wavelets are also excellent approximators and sensors of signals. Their time-frequency analysis capability always makes them an effective and innovating tool. In addition, their remarkable results in the domain of face recognition and classification (e.g., the Gabor wavelet in the EBGM method) encourage their integration in such a hybrid system.

In this paper, we are going to introduce a new training method for Wavelets Neural Networks (WNN) that represents a signal as a linear combination of wavelets with different dilations and translations that permit the direct calculation of connection weights of the network. The Wavelets Neural Network will be used to assess this algorithm in the field of images classification.

A wavelet network has more advantages than common networks, e.g., faster convergence, avoiding local minima and easy determination and adaptation of structure. Generally, there are two kinds of wavelet networks. One is to take the wavelet network as a special RBF network.[39] Training of this kind of wavelet networks is to some extent similar to that of RBF networks. The other kind of wavelet network is constructed from the inverse discrete wavelet transform.[29,36] With the development of wavelet networks, much work is devoted to the algorithms of training of these kinds of networks. Error back-propagation may be the most popular algorithm in the learning of wavelet networks. In the course of training, error back-propagation is often combined with Orthogonal Least Square-Backward Elimination,[24,39] which is used in the selection of network structures. In this paper, a new training algorithm for the wavelet networks is presented.

Inspired by the similarity between discrete wavelet transform and neural networks, Zhang and Benveniste[37] proposed the wavelet network in 1992. They constructed a special feed-forward neural network supported by wavelet theory. After constructing a wavelet library, which consists of a set of, scaled and translated wavelets (regressors), some network pruning algorithms were employed to select the most significant regressors.[38]

The main purpose behind the use of the OLS algorithm[8] is to find these regressors, which provide the most significant contribution to approximation error reduction. This has been widely accepted as an effective method for regressor selection in RBF networks.[9,18] Thus, we are going first to incorporate OLS algorithm into the wavelet network, and then exploit its potential in the classification process.

During the past decades wavelet analysis has become a powerful tool for multiresolution analysis. Important applications can be found in various fields, ranging from remote sensing to biomedical imaging and query by content in large image data bases. Intuitively, wavelet analysis is an ideal approach to the classification of images. Much previous works have been proposed to solve the image classification problem.[27,39,40] An algorithm of the back-propagation type has been derived for adjusting the parameters of the wavelet neural network.[3] Applications of wavelet neural network in the medical field include classification of coronary artery diseases,[1–3,30] characteristics of heart valve prostheses, interpretation of Doppler signals of the heart valve diseases, transmitting bio-signals and ECG segment classification.[3] However, to date, wavelet neural network based on a novel learning algorithm using direct calculations of weight connections involves a relatively new approach.

The rest of this paper is organized as follows. We first briefly review the structure of wavelet networks and then introduce the modified OLS algorithm for network pruning in Sec. 2. In Sec. 3, experiments are conducted to demonstrate the effectiveness of the optimized wavelet network, and point out their advantages from the viewpoint of image classification. Finally, we conclude with specifying the features of wavelet neural network in the last section.

2. Theoretical Background

2.1. *Wavelet and wavelet network*

Analyzing a signal from its corresponding graph does not give us access to all the information it contains. It is often necessary to transform it, that is, to give it another representation which clearly shows its features. The baron Jean Baptiste Joseph Fourier suggested that all functions must be able to express themselves in a simple way as a sum of sinus. In *the analytic theory of the heat*, Fourier gets the equations to the partial derivatives describing the transfers of heat, and thus developing them in infinite sum of trigonometric functions.

As an advanced alternative to the classical Fourier analysis, wavelets have been successfully used in many aspects of signal processing, such as de-noising, compressing, edge detection, and so on. The fundamental idea behind wavelets is to process data at different scales or resolutions. In such a way, wavelets provide a time-scale presentation of a sequence of input signal.[33]

Wavelet, generally named mother wavelet, must satisfy the properties of translation and dilation which can generate other wavelets $\psi_{a,b}(t), a > 0$ and b real.

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right). \tag{2.1}$$

The wavelet networks result from the combination of wavelets and neural networks.[28,29] First, continuous wavelet transform of function f is defined as the scalar product of f and the mother wavelet ψ.[21]

$$\omega(a,b) = \frac{1}{\sqrt{a}}\int f(x)\psi\left(\frac{x-b}{a}\right)dx. \tag{2.2}$$

The reconstruction of the function f from its transform is expressed by the following equation:

$$f(x) = \frac{1}{c_\psi}\int_R \int_{R_+} \omega(a,b)\frac{1}{\sqrt{a}}\psi\left(\frac{x-b}{a}\right)da\,db. \tag{2.3}$$

This equation expresses a function in terms of a sum of all dilations and all translations of the mother wavelet. If we have a finished number N_ω of wavelets $\psi_{a,b}$ obtained from the mother wavelet ψ, Eq. (2.4) will be considered as an approximation of an inverse transform.

$$f(x) \approx \sum_{j=1}^{N_\omega} c_j\psi_j(x). \tag{2.4}$$

This can also be considered as the decomposition of a function in a weighted sum of wavelets, where each weight c_j is proportional to $\omega(a_j, b_j)$. This establishes the idea of wavelet networks.[17,25]

The wavelet networks have been introduced by Zhang and Benveniste as a combination of the RBF neurons network and the decomposition in wavelets. Figure 1 displays the structure of the wavelet network of Zhang. The multilayered networks allow the representation of a nonlinear function by training while comparing their inputs and their outputs. This training is made while representing a nonlinear function by a combination of activation functions. The sigmoid function is often used as an activation one.

Zhang and Benveniste[37] replaced the sigmoid function by an admissible wavelet and they reached these results:

- The wavelet networks preserve the property of universal approximation of the RBF networks.
- A direct link exists between the weights of the network ω_i and the wavelets coefficients.
- A good approximation can be reached with a wavelet network of a small size.

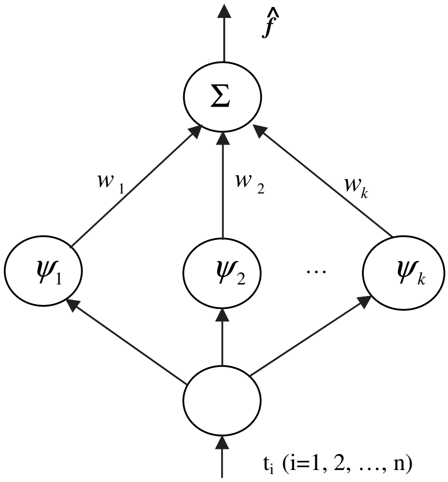

Fig. 1. Structure of wavelet network.

2.2. *The frames and the wavelet networks*

2.2.1. *The discrete wavelet transform (DWT)*

To analyze a discrete signal, it is known that the representation $\psi_{a,b}$ of Eq. (2.1) is very redundant.[15] Therefore, to get a finite number of wavelets, we discretize the dilation parameter (a) and the translation parameter (b) of the mother wavelet.[20]

$$\omega(a,b) = \frac{1}{\sqrt{a}} \sum f(x)\psi\left(\frac{x-b}{a}\right)dx. \tag{2.5}$$

To reconstitute the signal, the double integral in Eq. (2.3) is replaced by a double sum

$$f(t) = \frac{1}{c_\psi} \sum\sum \omega(a,b)\frac{1}{\sqrt{|a|}}\psi\left(\frac{t-b}{a}\right). \tag{2.6}$$

The inverse wavelet transform can be obtained only when the used wavelets form an orthogonal basis. For a general case (nonorthogonal basis), concepts like frame and dual frame must be introduced.

2.2.2. *The frames*

If we replace the double sum of Eq. (2.6) by a unique one, we can write the function f as:

$$f(t) = \sum_i \omega_i \psi_i(t). \tag{2.7}$$

According to Daubechies,[15] this relation is valid only if the wavelets used constitute an orthogonal basis. Generally, we can write f in term of its wavelet coefficients and a family of wavelets is said to be dual frame.

Assuming that $\psi \in L^2(R)$ a wavelet, S a sampling on a time-frequency grid, and $B_\psi = \{\psi_{a,b} \,|\, (a,b) \in S\}$ a discrete family of wavelets, B_ψ constitute a "wavelets frame" if $A > 0$ and $B < \infty$ exist, as for all $f \in L^2(R)$ if:

$$A\|f\|^2 \leq \sum_{(a,b)\in S} |\langle \psi_{a,b}, f\rangle|^2 \leq B\|f\|^2, \tag{2.8}$$

where A and B are called the limits of the frame.

When a discrete wavelets family forms a frame, it provides a complete representation without lossing any function f of L^2.[15] To provide more details, we introduce other terms: B_ψ is said orthogonal basis if for all $\psi_i, \psi_j \in B_\psi$.

$$\langle \psi_i, \psi_j\rangle = \delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j. \end{cases} \tag{2.9}$$

A frame is said basis if for all f belonging to L^2 the linear combination $f(t) = \sum_i \omega_i \psi_i(t)$ is unique. When a family of wavelets is at the same time orthogonal and basis it is said to be orthogonal basis.

In general, a frame is not an orthogonal basis (only the condition $A = B = 1$ gives an orthogonal basis). Also, it provides a redundant representation of the function $f(t)$. The A/B report is called a redundancy factor report. When a frame is redundant, the wavelet coefficients of the same neighborhood are intercorrelated. A better definition of the details and the fine structures results from the above stated function in the time-frequency representation.[36]

For an orthogonal basis all function f can be written in a unique manner:

$$f(t) = \sum_{(a,b)\in S} \omega(a,b)\psi\left(\frac{t-b}{a}\right) = \sum_{(a,b)\in S} \langle \psi_{a,b}, f\rangle \psi_{a,b}(t). \tag{2.10}$$

For other values of A and B this representation remains valid, but is no more B_ψ an orthogonal basis and the representation of f as a linear combination of wavelets is no more unique. Therefore, the term "dual frame" is introduced.[18] In these cases, f is written according to the dual frame $\tilde{B}_\psi = \{\tilde{\psi}_{a,b} \,|\, (a,b) \in S\}$.

$$f(t) = \sum_{(a,b)\in S} \langle \tilde{\psi}_{a,b}, f\rangle \psi_{a,b}(t) = \sum_{(a,b)\in S} \langle \psi_{a,b}, f\rangle \tilde{\psi}_{a,b}(t). \tag{2.11}$$

If the function ψ is the analyzing wavelet, the coefficients of wavelet are calculated by the scalar product of this wavelet by the function to be analyzed. The dual wavelet is used for the reconstruction (the inverse is true as well). In Fig. 2, we give an example of a wavelet and its dual.

- ***How to discretize the continuous wavelet transform to get frames?***
  To get a frame, the coefficients a and b must be discretized in the following manner[15]: $a = a_0^m, b = nb_0a_0^m$ with $a_0 > 1$ and $b_0 > 0$.

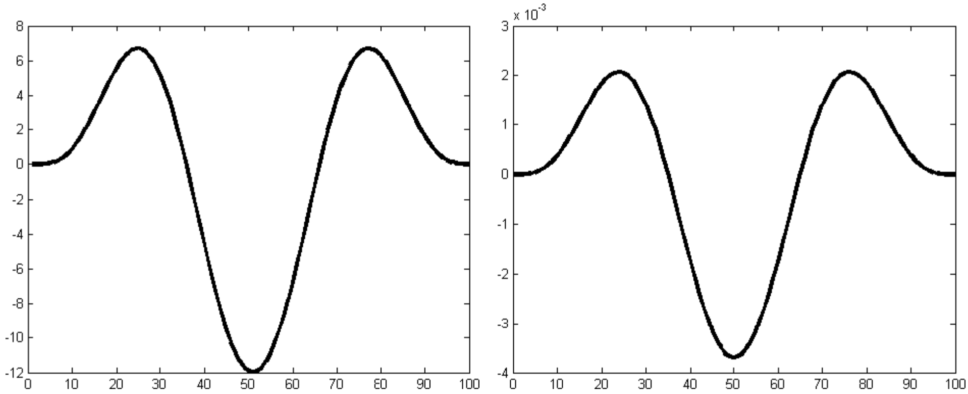

Fig. 2. A wavelet and its dual.

So for a signal including a_0^j points, we calculate only the coefficients:

$$\omega_{m,n}(f) = a_0^{-m/2} \sum \psi(a_0^{-m}t - nb_0)f(t), \quad m = 1, \ldots, j, \;\; n = 1, \ldots, a_0^{j-m}. \tag{2.12}$$

- We notice that we sampled more finely to the highest frequencies rather than to the lowest frequencies.
- For $a_0 = 2, b_0 = 1$ the sampling is said dyadic. The expression of the mother wavelet is given by:

$$\psi_{m,n}(t) = 2^{\frac{-m}{2}}\psi(2^{-m}t - n). \tag{2.13}$$

2.2.3. *Direct calculation of weights*

As we have seen in Sec. 2.2.2, a family of wavelet functions is not necessarily orthogonal. Therefore, for a given family Ψ of wavelets it is not possible to calculate a wavelet coefficient ω_i directly by a simple projection of the wavelet ψ_{m_i,n_i} onto the considered function. Thus, it was proposed[13,14] to use Eq. (2.14) to find the optimal coefficients ω_i for each fixed wavelet

$$E = \min_{m_i,n_i,\omega_i \text{ for all } i} \left\| f - \sum_{i=1}^{N} \omega_i \psi_{m_i,n_i} \right\|_2^2. \tag{2.14}$$

Equation (2.14) says that the ω_i,m_i and n_i are optimized, i.e. translation, dilation of each wavelet are chosen in such a way that the image f is optimally approximated by the weighted sum of wavelets ψ_{m_i,n_i}. Because optimization is a slow process, we suggest a direct calculation in the case of a finite wavelet family. The correct coefficients ω_i are computed by projecting f on the dual wavelets $\tilde{\psi}_{m_i,n_i}$.

The wavelet $\tilde{\psi}_{m_i,n_i}$ is a dual wavelet if

$$\langle \psi_{m_i,n_i}, \tilde{\psi}_{m_i,n_i} \rangle = \delta_{i,j} \tag{2.15}$$

and we find $\tilde{\psi}_{m_i,n_i}$ to be

$$\tilde{\psi}_{m_i,n_i} = \sum_j (\Psi)^{-1}_{i,j} \psi_{m_j,n_j} \tag{2.16}$$

where $(\Psi)_{i,j} = \langle \psi_i, \psi_j \rangle$.

2.3. *Orthogonal Least Squares (OLS) algorithm*

The orthogonal least squares (OLS) algorithm is an efficient implementation of the forward selection method for subset model selection. The ability to find good subset parameters with only a linearly increasing computational requirement makes this method attractive for practical implementations.

Let us represent these nonlinear predictors that have the linear in parameter structure as a linear regression model:

$$y = Xh + e \tag{2.17}$$

where y is the desired signal vector, X is the information matrix of size NK, h is the parameter vector of the model, and e the error vector of approximating y by Xh. The column vectors y and e contain N elements, that is, there are N data samples and N values of error.

The original X matrix has K columns. To create a parsimonious model which has R parameters, we are actually trying to pick R columns from the input matrix X to form a subset input matrix *Xs*. The OLS algorithm selects columns from the input matrix sequentially. At each selection, all the unused columns are studied to determine how each column will contribute to fit the desired vector y with the current subset *Xs*. The column that provides the best combination with *Xs* to model y will be picked to form the new *Xs*. The above procedure is repeated until the number of columns in *Xs* is equal to K. The selection procedure is made very efficient by employing orthogonalization schemes such as the Gram–Schmidt or the Householder transformation.[12] The Gram–Schmidt algorithm is a well-known standard numerical method that can be employed for orthogonalization of a basis function.[19] The details of the algorithm can be found in Chen *et al.*[7,8]

3. Proposed Approach

To define a wavelet network, we first take a family of n wavelets $\Psi = \{\psi_1, \ldots, \psi_n\}$ with different parameters of scaling and translation (generated by distributing the parameters on a dyadic grid) that can be chosen arbitrarily at this point.

According to Zhang, all function f belonging to the $L^2(R)$ space can be represented, with an arbitrary precision, by a wavelet network.

Given a classification or regression problem, the architecture of a wavelet network is exactly specified by the number of particular wavelets required. In this research, a library of wavelets will be selected to be the candidate hidden neurons (wavelets), and then the optimal architecture is constructed by pruning the hidden neurons.

3.1. *Training step*

3.1.1. *Wavelet network initialization*

A wavelet library, W, which consists of discretely dilated and translated versions of a given wavelet ψ, should be constructed according to the training data set. The wavelet family is:

$$W = \{\psi(a_0^{-m}x - nb_0) : m \in S_a, n \in S_{b(m)}\}. \tag{3.1}$$

In Eq. (3.1), $a_0, b_0 > 0$ are two scalar constants defining the discretization step sizes for dilation and translation. a_0 is typically dyadic. S_b and S_a are finite sets related to the size of the input data domain D.

The maximum scale level of the wavelet network should be determined first and consequently all the possible translation values will be calculated at each scale level.

3.1.2. *Network pruning algorithm*

After the construction of a wavelet library, the number of wavelets, N_s, may be decided manually or automatically. Therefore, in the N-candidate wavelet library, the goal is to find the best N_s wavelet regressors which minimize the error between output vectors and expected vectors.

Orthogonal Least Square algorithm[8] has been widely used in pruning RBF networks and wavelet networks as well.[32] It combines the orthogonal transform with the forward regression procedure to select regressors from a large candidate regressor set. The advantage of employing OLS is that the responses of the hidden layer neurons are decorrelated so that the contribution of individual candidate neurons to the approximation error reduction can be evaluated independently.

In the forward orthogonalization procedure, only one column of regressor matrix is orthogonalized at the ith iteration. The ith column is made orthogonal to each of the $i-1$ previously orthogonalized columns and the operation for $i = 2, \ldots, N$ is repeated.

The wavelets corresponding to the first K columns will be selected to construct the wavelet network. Figure 3 shows the wavelets library and some selected wavelets with the application of OLS algorithm that will be used in the hidden layer of the wavelet neural network. Those wavelets used as activation functions of the hidden neurones form a frame. To get an approximation of the learning image as an output of the network, we proceed as follows:

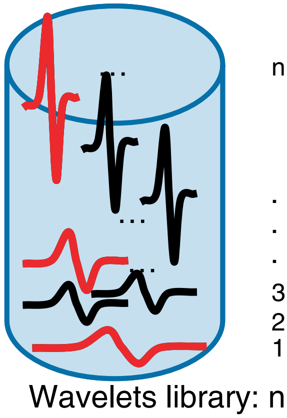

Fig. 3. The K selected wavelets by OLS algorithm (wavelets in red). (Color online.)

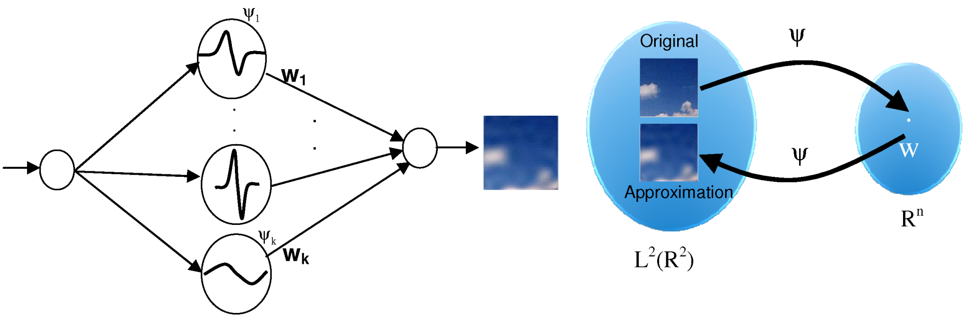

Fig. 4. Principle of OLS wavelet selection.

- We project the learning image (I) on the dual frame $(\tilde{\psi}_i)$ to calculate the connection weights (ω_i).

$$\omega_i = \langle I, \tilde{\psi}_i \rangle. \tag{3.2}$$

- We compute the output of the network $(\tilde{I})$ by a linear combination of the selected wavelets (ψ_i) and these connection weights (ω_i).

$$\tilde{I} = \sum_i \omega_i \psi_i \tag{3.3}$$

These steps are illustrated in Fig. 4.

3.1.3. *Pyramidal network training algorithm*

Pyramidal image models employ several images obtained at different stages of the learning algorithm. Let $f(x, y)$ be the original image. A pyramid image is a set of images $f_k(x, y), k = 1, \ldots, p$ which design the number of pyramid levels. The pyramid is formed according to these steps:

Proposed learning algorithm

***Step* 1:** Starting the learning by preparing a library of candidate wavelet functions to be selected as activation functions of the wavelet network. This step consists in the following:

(1) Choosing the mother wavelet covering all the support of the signal I to be analyzed.
(2) Building a library formed by the wavelets of the discrete wavelet transform using a dyadic sampling of the continuous wavelet transform.
(3) Setting as a stop learning condition an error Emin between the input and the output network or a number of neurons in the hidden layer of the network.

***Step* 2:** Applying OLS algorithm to select the optimized wavelets used in the hidden layer.

***Step* 3:** Fixing a definite number K of wavelets which will join the network at each stage of the pyramid structure in the training process.

***Step* 4:** At a given stage i of the pyramid, a wavelet network formed by K hidden neurons is built, and the network is computed to calculate the connection weights and finally the output of the network $\tilde{I}_i = \sum_{j=1}^{k} \omega_{ij}\psi_{ij}$ is calculated. The difference between the initial image and the approximation will be calculated, this difference will be considered as an original image in the next stage of the pyramid: $I_{i+1} = I_i - \tilde{I}_i$

***Step* 5:** The global approximation that is the sum of all approximations at each stage of the pyramid is calculated $\tilde{I} = \sum_{i=1}^{n} \tilde{I}_i$ with n is the number of stage of the pyramid. If the error Emin, or the number of the hidden neurons is reached then it is the end of learning.

Otherwise the next k wavelets of the library are selected and we go back to Step 4.

Figure 5 illustrates the pyramidal training algorithm of beta wavelet network.

3.1.4. *Algorithm flowchart*

The detailed steps of the proposed learning algorithm have been described in the last section. We can summarize this training algorithm according to the following flowchart as illustrated in Fig. 6.

3.2. *Overview of the classification process*

We propose in this paper a complete solution of image classification based on wavelet networks trained by PBWNN-OLS algorithm. The solution which we present proceeds, at first, by the approximation of the learning benchmark by a 2D

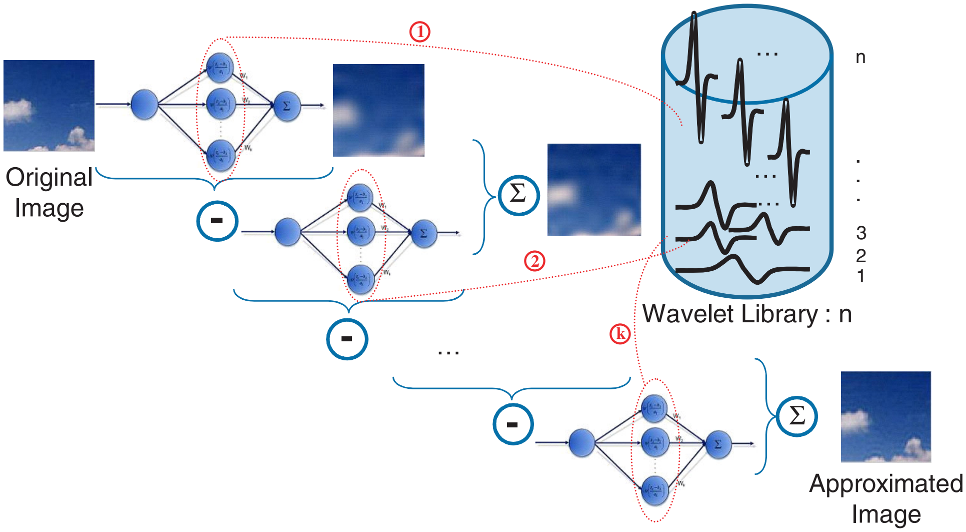

Fig. 5. Schema of pyramidal learning.

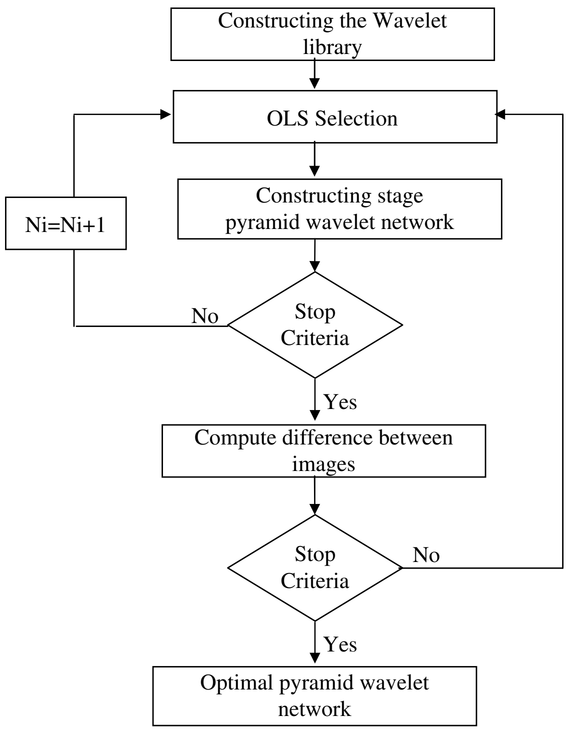

Fig. 6. Algorithm flowchart.

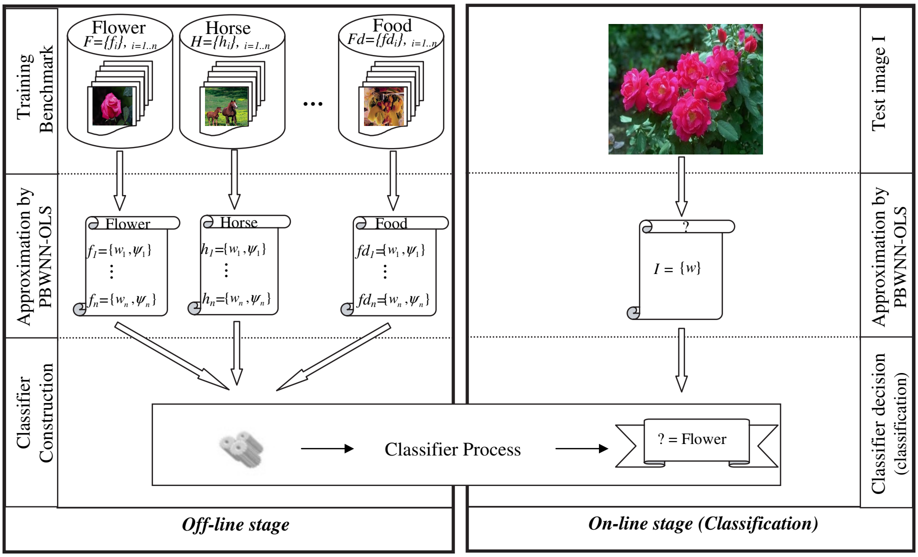

Fig. 7. Complete framework of image classification by pyramidal wavelet networks.

wavelet networks to produce the compact functional parameters of each network (constituted by wavelets and their coefficients). In classification stage (on-line), the test image is projected on the wavelet networks of the learning images and new weights (or coefficients) specific to this image are computed. The family of wavelets remains then unchanged (that of the learning images). Finally, we compare the coefficients of the learning images with the coefficients of the test image by computing the distances ($\min(distance(w_{ij}, w))$ where $i = 1, \ldots, c, j = 1, \ldots, n$, c is the number of classes and n the number of learning images in each class). The pipeline of all these stages is illustrated in Fig. 7.

3.3. *Definition and basic properties of the Beta function*

To experiment our approach, we use the Beta wavelets which is based on the Beta function defined by[34,39]:

$$\beta(x, p, q, x_0, x_1) = \begin{cases} \left(\dfrac{x - x_0}{x_c - x_0}\right)^p \left(\dfrac{x_1 - x}{x_1 - x_c}\right)^q & \text{if } x \in [x_0, x_1], \\ 0 & \text{otherwise,} \end{cases} \tag{3.4}$$

where $p, q, x_0 < x_1 \in R$ and $x_c = \frac{px_1 + qx_0}{p+q}$. The Beta function possesses the following properties:

$$\beta(x_0) = \beta(x_1) = 0 \quad \text{and} \quad \beta(x_c) = 1, \tag{3.5}$$

$$\frac{d\beta(x)}{dx} = \frac{px_1 + qx_0 - (p+q)x}{(x - x_0)(x_1 - x)}\beta(x), \tag{3.6}$$

$$\left.\frac{d\beta(x)}{dx}\right|_{x=x_c} = \left.\frac{d\beta(x)}{dx}\right|_{x=x_0} = \left.\frac{d\beta(x)}{dx}\right|_{x=x_1} = 0, \tag{3.7}$$

$$\frac{p}{q} = \frac{x_c - x_0}{x_1 - x_c}, \tag{3.8}$$

$$\frac{d^2\beta(x)}{dx} = \beta(x)A(x), \tag{3.9}$$

with

$$A(x) = \frac{1}{(x_1 - x)(x - x_0)}\left[\frac{1}{(x_1 - x)} - \frac{1}{x - x_0 - (p+q)(x+1) + px_1 + px_0}\right].$$

We demonstrate in Refs. 5, 28 and 34 that the derivatives of the Beta function are admissible wavelets. We can get different wavelets when we modify the values of the x_0, x_1, q and p parameters. Figure 8 shows examples of Beta wavelets.

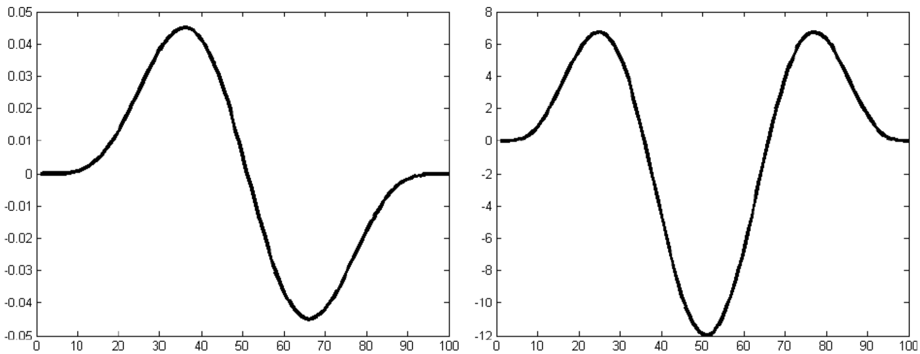

Fig. 8. Beta 1 and Beta 2 wavelets (first and second derivative of the Beta function).

We can obtain a 2D wavelet by multiplying two 1D wavelets. The 2D wavelet has four parameters: one for dilation (c), two parameters of translation (s_x, s_y) and one to rotate the wavelet (θ).

$$\beta(x, y) = \beta(x) \times \beta(y). \tag{3.10}$$

Figure 9 illustrates examples for the 2D Beta wavelets.

4. Experimental Results

Image classification attempts to classify images into semantic categories using low-level image features, and therefore, bridging the gap between high-level semantics and low-level features. The categorization of images into classes can be helpful both for semantic organizations of digital libraries and for automatic annotations of images.

The classification of natural imagery is quite hard in general since real images from the same semantic class may have large variations and images from different semantic classes may share a common background.

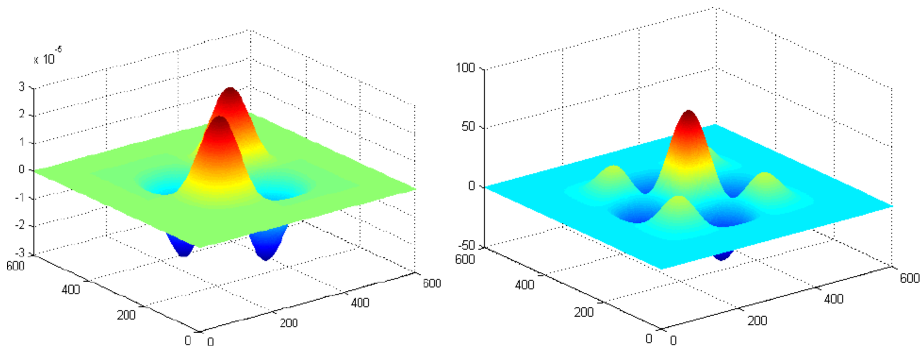

Fig. 9. 2D Beta function wavelets (first and second derivative of 2D Beta function).

In order to evaluate the proposed method in different conditions, various collections of images are used. We have varied the number of image classes, thus, we have used 2-image classes, 3-image classes and 10-image classes (for some samples, see Fig. 10). These images contain a wide range of content (scenery, animals, objects, etc.), colors and lighting conditions.

Therefore, we randomly shuffled the images in each class for such database and took several images as the training set and the test set. After repeating this procedure three times to ensure fairness, we obtained three sets of training data and test data. In Table 1, we give the number of training and test images sets for each database.

The average percentage of correctness classification of the three test sets is summarized in Table 2 for the different databases.

Table 1. Number of training and testing images for each database.

	Training set	Testing set
Base 1	62 images	62 images
Base 2	157 images	157 images
Base 3	147 images	147 images
Base 4	231 images	231 images
Base 5	500 images	500 images

Table 2. Correctness classification on three data sets.

	BWNN	PBWNN-OLS (our approach)
Base 1	98.39%	100%
Base 2	82.80%	90.44%
Base 3	83.67%	91.15%
Base 4	67.53%	80.08%
Base 5	60.20%	71.20%

(a) Base 1: Sample images from the 2-classes database (grass, sky).

[scale=.97](b) Base 2: Sample images from the 2-classes database (binocular, rifle).

(c) Base 3: Sample images from the 2-classes database (flower, butterfly).

(d) Base 4: Sample images from the 3-classes database (binocular, rifle, baseball-glove).

(e) Base 1: Sample images from the 10-classes database Wang (buses, dinosaurs, horses, flowers, mountains, food, elephants, monuments, Africa, beach).

Fig. 10. Sample images from various classes.

Table 3. Time consumption for the processing steps.

Processing stage	Time consumption BWNN	Time consumption PBWNN-OLS
OLS selector	1 h 30 mn	1 h 30 mn
Training	20 mn	12 s
Classification	2 mn	3 s

The examination of computing times is very important in the evaluation of a natural image classification algorithm. Therefore, the PBWNN-OLS training algorithm is more efficient than the classical BWNN used in Refs. 29 and 36 and computationally less expensive. Table 3 compares CPU-times spent to compute the different stages of the learning algorithm for one image, on a 1.66 GHz Intel Core 2 Duo with a 2 GB cache. This table shows us that computing the classification step with our approach is more than forty times as fast as computing it with the classical approach.

Based on the experiment results in Tables 2 and 3, both the classification ability and the learning speed of the network are remarkably superior to the classical wavelet network models.

4.1. *Comparison of the proposed classifier to popular methods in the literature*

We have carried out experiments on the SIMPLIcity database[31] which contains 1,000 images extracted from the well known commercial COREL database.[a] The database contains ten clusters representing semantic generalized meaningful categories such as "African people", "beaches", "buildings", "buses", "dinosaurs", "elephants", "flowers", "food", "horses" and "mountains". Needless to say, that the categories are extremly heterogeneous in terms of signal contents, as illustrated in Fig. 7. There are 100 images per cluster. Table 4 compares classification rates obtained with four different approaches:

- *KNN*(d) is the k nearest neighbors approach: to classify an image, we compute the distance between this image and every image of a database, the classes of the k closest images determine the class. We set k to 5, which gives the best average results.[23]
- *HMM*(T) is the approach based on hidden Markov models where T is the parameter which determines the number of states.[23]
- *GM*(d) gives results obtained when characterizing each image class by the generalized median, i.e. the string which minimizes its distance to every string in the class, and classifying images with respect to the class of its closest generalized median.[23]

[a]The SIMPLIcity database can be downloaded on the James Z. Wang website at http://wang.ist.psu.edu/jwang/test1.tar.

Table 4. Comparison of different classifiers.

Classification model	Classification rate
HMM(1)	63.2%
HMM(2)	68.1%
HMM(5)	70%
HMM(10)	70.2%
HMM(20)	70.1%
HMM(50)	70.6%
HMM(100)	70.3%
KNN(d_H)	64.3%
KNN(d_{Hw})	67.6%
GM(d_H)	65.9%
GM(d_{Hw})	66.4%
DT	64.5%
BWNN	60.2%
PBWNN-OLS (our approach)	**71.2**%

- *DT* is the classification model called Decision Trees applied in Ref. 6. A decision tree is a class discriminator that recursively partitions the training set until each partition consists entirely or dominantly of examples from one class.

The experimental results reported here seem very promising and the proposed approach outperforms the other methods.

5. Conclusion

In this paper, wavelet network has been analyzed and successfully applied to image classification, so it has not only effectively improved the classification performance of the network but also has much faster learning capacity than the existing wavelet network architecture. First, we have employed the OLS algorithm to optimize the structure of wavelet network, so we can reach the goal with a minimum number of hidden neurons. Then, we have utilized the pyramid wavelet network structure for the image classification process to increase the speed of the learning algorithm. The experimental results have shown the efficiency of OLS algorithm in network pruning. Based on the experimental results, the novel pyramidal wavelet network model has greatly improved both the learning speed and the classification performance compared with the existing wavelet network and other models in classification process.

Acknowledgments

The authors would like to acknowledge the financial support of this work by grants from the General Direction of Scientific Research and Technological Renovation (DGRSRT), Tunisia, under the ARUB program 01/UR/11/02.

References

1. E. Avci and I. Turkoglu, Modelling of tunnel diode by adaptive-network-based fuzzy inference system, *Int. J. Computa. Intell.* **1**(1) (2003) 231–233.
2. E. Avci, I. Turkoglu and M. Poyraz, A new approach based on scalogram for automatic target recognition with x-band Doppler radar, *Asian J. Inform. Technol.* **4**(1) (2005) 133–140.
3. E. Avci, I. Turkoglu and M. Poyraz, Intelligent target recognition based on wavelet packet neural network, *Experts Syst. Appl.* **29**(1) (2005) 175–182.
4. C. Ben Amar and O. Jemai, Wavelet networks approach-based image compression, *ICGST Int. J. Graphics Vision Image Process.* **SI1** (2005) 37–45.
5. C. Ben Amar, M. Zaied and M. A. Alimi, Beta wavelets synthesis and application to lossy image compression, *J. Adv. Eng. Software* **36**(7) (2005) 459–474.
6. A. Boucher, T.-H. Dang and T.-L. Le, Classification vs. recherche d'information: Vers une caractérisation des bases d'images, 12èmes Rencontres de la Société Francophone de Classification (SFC), Montréal, Canada (2005).
7. S. Chen, S. A. Billings and W. Luo, Orthogonal least squares methods and their application to nonlinear system identification, *Int. J. Contr.* **50**(5) (1989) 1873–1896.
8. S. Chen, C. F. Cowan and P. M. Grant, Orthogonal least squares learning algorithm for radial basis function networks, *IEEE Trans. Neural Networks* **2**(2) (1991) 302–309.
9. S. Chen, E. S. Chng and K. Alkadhimi, Regularized orthogonal least squares algorithm for constructing radial basis function networks, *Int. J. Contr.* **64**(5) (1996) 829–837.
10. J. L. Cheng, Wavelet neural networks with a hybrid learning approach, *J. Inform. Sci. Eng.* **22** (2006) 1367–1387.
11. E. S. Chng, S. Chen and B. Mulgrew, Efficient computational schemes for the orthogonal least squares algorithm, *IEEE Trans. Signal Process.* **43**(1) (1995) 373–376.
12. E. S. Chng, H. Yang and S. Bos, Orthogonal least squares learning algorithm with local adaptation process for the radial basis function network, *IEEE Trans. Signal Process. Lett.* **3**(8) (1996) 253–255.
13. J. Daugman, Complete discrete 2D Gabor transform by neural networks for image analysis and compression, *IEEE Trans. Acoustics Speech Signal Process.* **36** (1988) 1169–1179.
14. I. Daubechies, The wavelet transform, time-frequency localization and signal analysis, *IEEE Trans. Inform. Theory* **36** (1990) 961–1005.
15. I. Daubechies, *Ten Lectures on Wavelets* (Society for Industrial and Applied Mathematics (SIAM), 1992).
16. M. Ehler, Compacity supported multivariate, pairs of dual wavelet frames obtained by convolution, *Int. J. Wavelets Multiresolut. Inf. Process.* **6**(2) (2008) 183–208.
17. C. Foucher and G. Vaucher, Compression d'images et réseaux de neurones, Revue Valgo, *Valgo* **2001** (2001) No. 01-02, 10 pp.; La revue en ligne de l'Association des Connexionnistes en Thése (ACTH) (ISSN 1625-9661).
18. J. B. Gomm and D. L. Yu, Selecting radial basis function network centers with recursive orthogonal least squares training, *IEEE Trans. Neural Networks* **11** (2000) 306–314.
19. C. L. Lin, J. F. Wang, C. Y. Chen, C. W. Chen and C. W. Yen, Improving the generalization performance of RBF neural networks using a linear regression technique, *Expert Sys. Appl.* **36** (2009) 12049–12053.
20. H.-G. Li, Y. Li, C.-P. Sun and X.-L. Li, A new discrete transform fast algorithm with rational dilation factor, *Int. J. Wavelets Multiresolut. Inf. Process.* **6**(4) (2008) 609–616.

21. S. S. Luengar, E. C. Cho and V. Phoha Vir, *Foundations of Wavelets Networks and Applications* (Chapman and Hall/CRC, 2000).
22. S. G. Mallat , A theory for multiresolution signal decomposition: The wavelet representation, *IEEE Trans. Pattern Anal. Mach. Intel.* **11**(7) (1989) 674–693.
23. M. Mouret, C. Solnon and C. Wolf, Classification of images based on hidden markov models, in *IEEE Workshop on Content Based Multimedia Indexing*, Chania, Crete (2009), pp. 169–174.
24. M. Othmani, W. Bellil, C. Ben Amar and A. M. Alimi, A new structure and training procedure for multi-mother wavelet networks, *Int. J. Wavelets Multiresolut. Inf. Process.* **8**(1) (2010) 149–175.
25. Y. Oussar, Réseaux d'ondelettes et réseaux de neurones pour la modélisation statique et dynamique de processus, Thèse de doctorat, Université Pierre et Marie Curie (1998).
26. S. Said, B. Ben Amor, M. Zaied, C. Ben Amar and M. Daoudi, Fast and efficient 3D face recognition using wavelet networks, in *16th IEEE International Conference on Image Processing*, Cairo, Egypt (2009), pp. 4153–4156.
27. A. Sengur, I. Turkoglu and M. Cevdet Ince, Wavelet packet neural networks for texture classification, *Expert Syst. Appl.* **32** (2007) 527–533.
28. H. Szu, B. Telfer and S. Kadambe, Neural network adaptive wavelets for signal representation and classification, *Opt. Eng.* **31** (1992) 1907–1961.
29. H. Szu, B. Telfer and J. Garcia, Wavelet transforms and neural networks for compression and recognition, *Neural Networks* **9** (1996) 695–708.
30. I. Turkoglu, A. Arslan and E. Ilkay, An intelligent system for diagnosis of the heart valve diseases with wavelet packet neural networks, *Comput. Biol. Med.* **33** (2003) 319–331.
31. J. Z. Wang, J. Li and G. Wiederhold, SIMPLIcity: Semantics-sensitive integrated matching for picture libraries, *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(9) (2001) 947–963.
32. A. L. Wei, G. M. Zeng, G. H. Huang, J. Liang and X. D. Li, Modeling of a permeate flux of cross-flow membrane filtration of colloidal suspensions: A wavelet network approach. *Int. J. Environ. Sci. Tech.* **6**(3) (2009) 395–406.
33. R. Yan and R. X. Gao, Base wavelet selection for bearing vibration signal analysis, *Int. J. Wavelets Multiresolut. Inf. Process.* **7**(4) (2009) 411–426.
34. M. Zaied, C. Ben Amar and M. A. Alimi, Award a new wavelet based beta function, in *Proc. International Conference on Signal, System and Design*, *SSD03*, Vol. 1, Tunisia (2003), pp. 185–191.
35. M. Zaied, C. Ben Amar and A. M. Alimi, Beta wavelet networks for face recognition, *J. Decision Systems JDS* **14**(1–2) (2005) 109–122.
36. M. Zaied, O. Jemai and C. Ben Amar, Training of the Beta wavelet networks by the frames theory: Application to face recognition, in *Proc. International Workshops on Image Processing Theory, Tools and Applications*, Tunisia (2008), pp. 165–170.
37. Q. Zhang and A. Benveniste, Wavelet networks, *IEEE Trans. Neural Networks* **3** (1992) 889–898.
38. Q. Zhang, Using wavelet network in nonparametric estimation, *IEEE Trans. Neural Networks* **8**(2) (1997) 227–236.
39. Z. Zhang, Learning algorithm of wavelet network based on sampling theory, *Neurocomput.* **71** (2007) 244–269.
40. Z. Zhang, Iterative algorithm of wavelet network learning from non uniform data, *Neurocomput.* **72** (2009) 2979–2999.