International Journal of Wavelets, Multiresolution and Information Processing

A NOVEL APPROACH FOR FACE RECOGNITION BASED ON FAST LEARNING ALGORITHM AND WAVELET NETWORK THEORY

MOURAD ZAIED*

National Engineering School of Gabes
Rue Omar Ibn-Elkhattab 6029, Gabes, Tunisia
and
REsearch Group on Intelligent Machines, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia
mourad.zaied@ieee.org

SALWA SAID†, OLFA JEMAI‡
and CHOKRI BEN AMAR§

REsearch Group on Intelligent Machines, National Engineering School of Sfax
†*salwa.said@ieee.org*,‡*olfa.jemai@ieee.org*,§*chokri.benamar@ieee.org*

Received (Day Month Year)
Revised (Day Month Year)
Communicated by (xxxxxxxxxx)

This paper presents a new appoach of face recognition based on wavelet network using 2D fast wavelet transform and multiresolution analysis. This approach is divided in two stages: the training stage and the recognition stage. The first consists to approximate every training face image by a wavelet network. The second consists in recognition of a new test image by comparing it to all the training faces, the distances between this test face and all images from the training set are calculated in order to identify the searched person. The usual training algorithms presents some disadvantages when the weights of the wavelet network are computed by applying the back propagation algorithm or by direct solution which requires computing an inversion of matrix, this computation may be intensive when the learning data is too large. We present in this paper our solutions to overcome these limitations. We propose a novel learning algorithm based on the 2D Fast Wavelet Transform. Furthermore, we have increased the performances of our algorithm by introducing the Levenberg-Marquardt method to optimize the learning functions and using the Beta wavelet which has at both an analytical expression and wavelet filter bank. Extensive empirical experiments are performed to compare the proposed method with others approaches as PCA, LDA, EBGM and RBF neural network using the ORL and FERET benchmarks.

Keywords: Wavelet Network; Fast Wavelet Transform; Learning Algorithm; Face Recog-

*Corresponding author. Address: ENIG, Av. Omar Ibn El Khattab-Zerig-6029, Gabes, Tunisia. Tel: 216-98-241151

nition.

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

1. Introduction

Over the past few decades, biometry has been widely applied to several applications such as access control, border control systems and so on. Fingerprint and irisbased identity recognition showed good performances; however they require a user's cooperation, who finds them intrusive. Consequently, the current trend goes to biometrics which can be collected on the move, as the face under different lighting conditions, facial expressions and changes in head pose.

Ksantini & all propose a novel Bayesian logistic discriminant (BLD) model that avoids the small sample size problem by using a sparsity-promoting Gaussian prior over the unknown parameters or weights. Moreover, novel subclass and multinomial versions of the model are proposed to address the problems posed by nonlinearly separable classes and to perform polychotomous classification.[24] Sanqiang & all presents a new Constrained Profile Model (CPM), in cooperation with Flexible Shape Model (FSM) to form an efficient localization framework. Through Gabor feature constrained local alignment, the proposed method not only avoids local minima in landmark localization, but also circumvents the exhaustive global optimization.[42] Jian Yang & all presents the concept of color space normalization(CSN)and two CSN techniques,i.e., the within-color-component normalization technique and the across-color-component normalization technique, for enhancing the discriminating power of color spaces for face recognition.[37] Haifeng Hu & all propose a nonlinear face recognition technique based on neighborhood preserving discriminant analysis (NPDA). The kernel trick is adopted to allow the efficient computation of local Fisher discriminant in high-dimensional feature space. Moreover, a direct solution for obtaining the optimal feature vectors in feature space is presented which can preserve the most discriminative information.[14] Dattatray & all presents a new pattern recognition framework for face recognition based on the combination of Radon and wavelet transforms, which is invariant to variations in facial expression, and illumination. It is also robust to zero mean white noise. The technique computes Radon projections in different orientations and captures the directional features of face images. Further, the wavelet transform applied on Radon space provides multiresolution features of the facial images. Being the line integral, Radon transform improves the low-frequency components that are useful in face recognition. For classification, the nearest neighbor classifier has been used.[15]

Neural Networks have been supplied with an access to the frequency analysis with the introduction of wavelets. A wavelet network has more advantages than common networks, e.g., faster convergence, avoiding local minimum, easy decision and adaptation of structure.[7,38,39] Usually, there are two kinds of wavelet networks, which are constructed from different ideas.[9] One is to take the wavelet network as a special RBF network.[30,40] Moreover the training of this kind of wavelet networks is to some extent similar to that of RBF networks. The other kind of wavelet networks is constructed from the wavelet theory.

With the development of wavelet networks, many works would be devoted to the learning algorithms of wavelet networks. The back-propagation error may be the most popular algorithm in the learning of wavelet networks.[29,30,31,5] In the course of training, the error back-propagation is often combined with Orthogonal Least Square-Backward Elimination,[27,29,30,31] which is used in the selection of network structures. At the same time, many algorithms are proposed to initialize the weights,[35] accelerate the convergence [38,18] and adjust the structures of wavelet networks [38] when the error back-propagation is applied to the training algorithms. Other algorithms besides the error back-propagation are also proposed to train the wavelet networks, e.g., Kalman filter,[31] genetic algorithms [6,20,33] and immune algorithms.[26] These above mentioned algorithms mostly stem from the typical neural networks, so they seldomly utilize the excellent properties of wavelets, fully, in the frequency domain though they have accelerated convergence, avoided local minimum and overcome overfitting to some extent.

In this paper, we propose a new algorithm for the learning of wavelet networks in multiresolution domain (MRA) by application of 2D Fast wavelet transform. This paper is divided into four parts. The first part briefly reviews the theory of wavelet networks, the multiresolution analysis and the Fast wavelet transform theory in 2D case. The second part presents the proposed learning algorithm of wavelet networks using 2D Fast Wavelet Transform. In the third part we present steps of recognition stage. In the last part, experiments are conducted to demonstrate the efficiency of the proposed approach by using the ORL and the FERET datasets. To better evaluate the performance of our approach, we have compare it with other methods of face recognition. We finish the paper by the conclusion and the future work.

2. Review of wavelet transform and wavelet network in 2D case

2.1. *The Continuous wavelet transform (CWT)*

The continuous wavelet transform, for a given function $f(x, y) \in L^2(R)$, is given by:

$$\omega(a, b_x, b_y, \theta) = \frac{1}{\sqrt{|a|}} \int\int f(x, y)\psi_\theta(\frac{x - b_x}{a}, \frac{y - b_y}{a})dxdy =< \psi_{a,b_x,b_y,\theta}, f > \tag{2.1}$$

All the analyzing wavelets are generated from one mother wavelet by varying the dilation parameter (a), the translation parameters (b_x, b_y) and the rotation parameter (θ).

The corresponding inverse wavelet transform that rebuilds the function f from its coefficients of wavelets is:

$$f(x,y) = \frac{1}{C_\psi}\int_0^\infty \frac{da}{a^3}\int\int db_x db_y \int_0^{2\Pi} d\theta\, \omega(a,b_x,b_y,\theta)\frac{1}{\sqrt{|a|}}\psi_\theta(\frac{x-b_x}{a},\frac{y-b_y}{a})dxdy \quad (2.2)$$

The coefficient C_ψ must be different to zero and different to ∞ so that this inverse transform exists. This condition, always known as the admissibility condition, must be verified at the begining of any wavelet transform.
Knowing that C_ψ is always positive, the admissibility condition can be written as follow:

$$0 < c_\psi = 2\pi^2 \int^\infty \frac{dw}{\|w\|}\int_0^{2\pi} d\theta \left\|\hat{\psi}(w\cos\theta, w\sin\theta)\right\|^2 < \infty \quad (2.3)$$

2.2. *The discrete wavelet transform (DWT)*

The CWT has the drawbacks of redundancy and impracticability with digital computers. As parameters (a, b_x, b_y) take continuous values, the resulting CWT is a very redundant representation, and impracticability is the result of redundancy. Therefore, the scale and shift parameters are evaluated on a discrete grid of time-scale plane leading to a discrete set of continuous basis functions.[13,4,25] The discretization is performed by setting [36]: $a = a_0^m, b_x = nb_{x_0}a_0^m, b_y = nb_{y_0}a_0^m$ with $a_0 > 1$, $b_{x_0}, b_{y_0} > 0$ and $\theta = l\theta_0$, $\theta_0 > 0$ with $l \in N$. and So for analyzing a 2D signal, we use only the family wavelets:

$$\psi_l({a_0^{-m}}x - n_x b_{x_0}, {a_0^{-m}}y - n_y b_{y_0}) \quad \text{with } m = 1,...,j; n_x = 1,...,{a_0^{j-m}}; n_y = 1,...,a_0^{j-m} \quad (2.4)$$

In the case of DWT, the analyzing wavelets are characterized by three new parameters: the position parameter n, the scale parameter m ($1 \le m \le j$, j represent the number of scales) and the rotation parameter. The analyzing formula can be written:

$$w_{m,n_x,n_y,l} = {a_0^{-m/2}}\sum\sum \psi_l({a_0^{-m}}x - n_x b_{x_0}, {a_0^{-m}}y - n_y b_{y_0})f(x,y) \quad (2.5)$$

And the inverse wavelet transform is:

$$f(x,y) = \sum w_{m,n_x,n_y,l}\psi_l({a_0^{-m}}x - n_x b_{x_0}, {a_0^{-m}}y - n_y b_{y_0}) \quad (2.6)$$

The most widely used form of such discretization with $a_0 = 2$, $b_{x_0} = 1, b_{y_0} = 1$ on a dyadic time-scale grid. Such a wavelet transform is described as the standard DWT.

The selection of the wavelet is made in such a way that basis function set $\psi_{m,n_x,n_y,l}$ constitute an orthonormal basis. Several such wavelet bases have been reported in literature to evaluate a 2D signal using the summation of finite basis over index m, n_x, n_y and l with finite DWT coefficients with almost no error. All these wavelets can be derived with an arbitrary resolution and with finite DWT coefficients.

2.3. *The mutiresolution analysis (MRA)*

Multi-resolution analysis (MRA) is often used for signal representations and signal processing because it can represent signals at the split resolution and scale space. In multi-resolution analysis, a signal is viewed at various levels of approximations or resolutions. A complicated signal is divided into several simpler signals by applying MRA, and each signal is considered separately. From the viewpoint of signal spaces, the multiresolution analysis consists on, firstly, a scaling function $\phi(x) \in L^2(R)$ which constitutes an orthonormal basis by varying its position on a given scale (j). The functions of every scale generate an approximation of a given signal f to analyze. Secondly, additional functions, i.e. wavelet functions, are then used to encode the difference in information between adjacent approximations.[8,4]
With the one-dimensional scaling function ϕ and corresponding wavelet ψ, we can design separable two-dimensional scaling and wavelet functions. The separable wavelets are viewed as tensor products of one-dimensional wavelets and scaling functions. If $\psi(x)$ is the one-dimensional wavelet associated with one-dimensional scaling function $\phi(x)$, then three 2-D wavelets associated with three sub-images are given as [34]:

$$\psi^H(x,y) = \phi(x)\psi(y) \tag{2.7}$$

$$\psi^V(x,y) = \psi(x)\phi(y) \tag{2.8}$$

$$\psi^D(x,y) = \psi(x)\psi(y) \tag{2.9}$$

$$\phi(x,y) = \phi(x)\phi(y) \tag{2.10}$$

where the symbols H, V and D stand for the directional wavelet coefficients. The transformation basis functions are defined as

$$\phi_{m,n_x,n_y,}(x,y) = 2^{(-m/2)}\phi(2^{-m}x - n_x, 2^{-m}y - n_y) \tag{2.11}$$

$$\psi_{m,n_x,n_y,}(x,y) = 2^{(-m/2)}\psi(2^{-m}x - n_x, 2^{-m}y - n_y), i = \{H,V,D\} \tag{2.12}$$

Thus, the resulting two-dimensional wavelet transform can be used in an image $f(x,y)$ of size $M * N$

$$V_{m,n_x,n_y,} = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\phi_{m,n_x,n_y,}(x,y) \tag{2.13}$$

$$W^i_{m,n_x,n_y,} = \frac{1}{\sqrt{MN}} \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} f(x,y)\psi^i_{m,n_x,n_y,}(x,y), i = \{H,V,D\} \tag{2.14}$$

The implementation of a MRA of an image produces three detailed sub-images (HL, LH, HH), which contain respectively the detail coefficients (W^H,W^V, W^D) corresponding to three different directional-orientations (Horizontal, Vertical and Diagonal) and a lower resolution sub-image LL which contain the approximation coefficients V. The analyzing process can be iterated in a similar manner on the LL channel to provide multilevel decomposition. Multilevel decomposition hierarchy of an image is illustrated in figure 1:

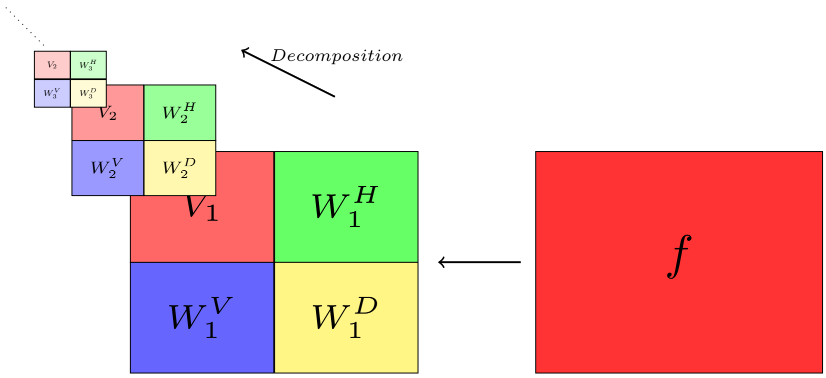

Fig. 1. Multilevel decomposition hierarchy of an image with 2-D MRA

The image f can be rebuilted by exploiting all the details and the approximation of the last scale:

$$f(x,y) = \sum V_{m,n_x,n_y,}\phi_{m,n_x,n_y,} + \sum\sum W^i_{m,n_x,n_y,}\psi^i_{m,n_x,n_y,}(x,y), i = \{H,V,D\} \quad (2.15)$$

In general, when the $\phi_{m,n_x,n_y,}$ constitutes a non orthogonal basis, it is necessary to compute first the duals basis composed by the functions $\tilde{\phi}_{m,n_x,n_y,}$ of the scaling functions $\phi_{m,n_x,n_y,}$ to be able to calculate the $V_{m,n_x,n_y,}$ coefficients. A dual basis of a set of functions can be computed by [22]:

$$\tilde{\phi}_j = \sum_{i=1}^{N} (\Phi)^{-1}_{j,i}\phi_i \quad \text{with } \Phi_{j,i} =< \phi_j, \phi_i > \quad (2.16)$$

The approximation of f at the scale m and the positions n_x, n_y is gotten by:

$$V_{m,n_x,n_y,} =< f, \tilde{\phi}_{m,n_x,n_y,} > \quad (2.17)$$

With the same manner we can compute the dual family of wavelets and compute the detail coefficients:

$$W^i_{m,n_x,n_y,} =< f, \tilde{\psi}^i_{m,n_x,n_y,} >, i = \{H,V,D\} \quad (2.18)$$

The following figures represent the decomposition steps of signal f using dual wavelets and scaling functions ($\tilde{\psi}^i_{m,n_x,n_y,}, \tilde{\phi}_{m,n_x,n_y,}$) and the reconstitution steps with primal wavelets and scaling functions $(\psi^i_{m,n_x,n_y,}, \phi_{m,n_x,n_y,})$:

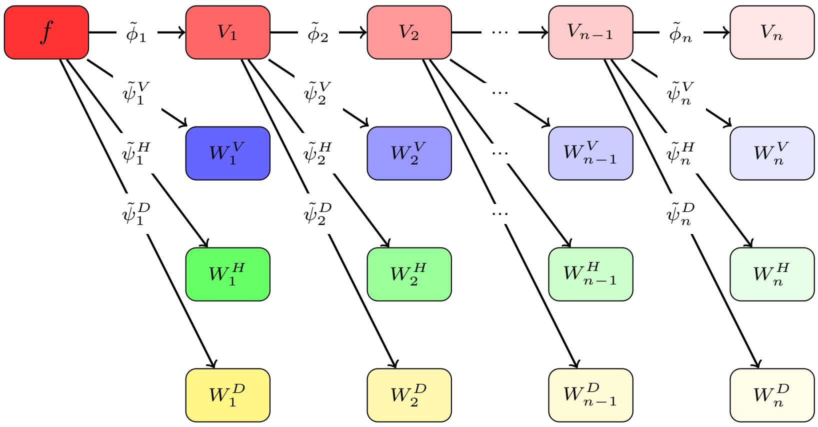

Fig. 2. 2D MRA Analysis of an image f using dual wavelets and scaling functions

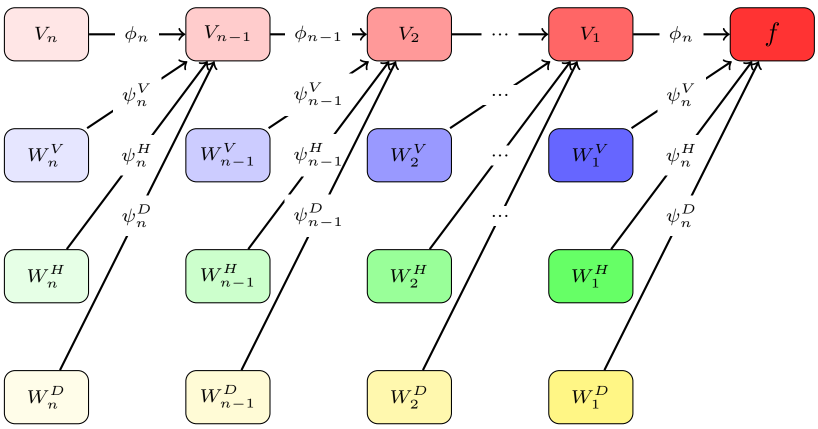

Fig. 3. 2D MRA reconstruction of an image f using primal wavelets and scaling functions

2.4. *Wavelet Network Theory*

In 1992, Zhang and Benveniste have introduced a new theory called "Wavelet networks" using a combination of artificial neural networks based on radial basis function and wavelet decomposition. In addition, they have explained how a Wavelet networks can be generated and showed how they can be used for pattern matching.

Zhang showed that Wavelet networks are able to handle nonlinear regression of moderately large input dimension with sparse training data. Moreover, he replaced the transfer function by an admissible wavelet and he demonstrated that a wavelet network preserve the property of universal approximation of the RBF networks.[29] Firstly, wavelet network is defined by pondering a set of wavelets dilated and translated from one mother wavelet with weight values to approximate a given signal f. In the 2D version, the rotation of mother wavelet is added as a third parameter. The output of the network is given by the following equation which represents an approximaton of equation (2.2) when we use finit number of wavelets.

$$\tilde{f} = \sum_{i=1}^{n} \omega_i \psi_i \tag{2.19}$$

The corresponding architecture is presented in figure 4, and an example of 2D neurone from the wavelet network is shown in the figure 5.

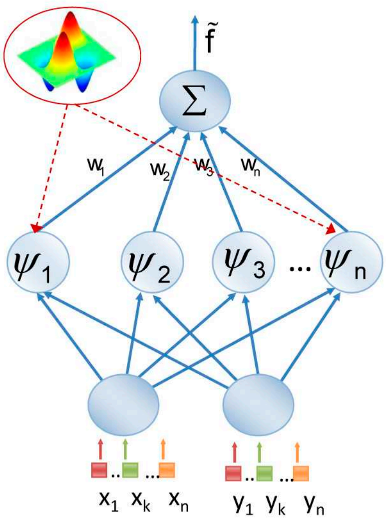

Fig. 4. Wavelet network.

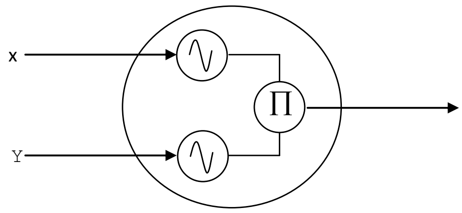

Fig. 5. 2D Neuron based on 1D wavelet functions.

This architecture can be extended by adding some dilated and translated versions of the scaling function of the corresponding used wavelet in the hidden layer of the network. In this case, the approximation of the signal is:

$$\tilde{f} = \sum_{i=1}^{p} \omega_i^k \psi_i^k + \sum_{j=1}^{q} \nu_j \phi_j, \; k = \{H, V, D\} \tag{2.20}$$

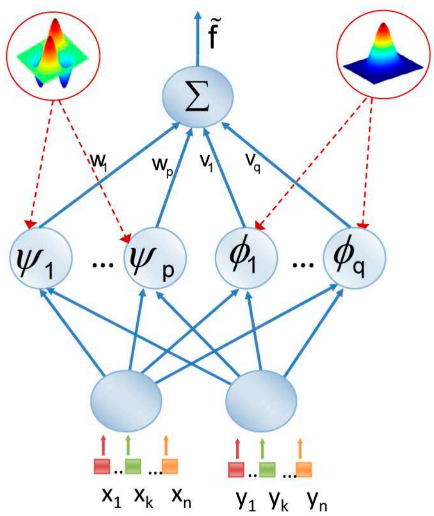

Fig. 6. Wavelet network using scaling function.

To compute the output weight connections of the wavelet network, many authors [13,41] proposed to use the technique of the projection of the signal f on the dual basis of the wavelets and the scaling functions of the hidden layer. This learning technique gives exact values of the weights, but it has a major shortcoming when we determinate the weights of the hidden layer to the output layer, this lead to compute the inversion of a matrix Φ. The direct solution causes an intensive computation as the matrix is too large.

In the section 3, we will propose a rapid method to calculate the connection weights of a dyadic wavelet network. Then, it will be generalized to any kind of wavelet network. Our method is based on the Fast Wavelet Transform (FWT), in what follows, we will review some of the relevant FWT theories, and then we will detail our proposed learning algorithm.

2.5. *The 2D Fast Wavelet Transform (FWT)*

The FWT aims to give a simple way to handle the multiresolution analysis, in other word with the FWT we want to compute the approximation coefficients and the detail coefficients with other technique, simpler and easier than those which are based on projection on the dual basis.
In order to fasten the calculation of the approximation and the detail coefficients, rapid algorithms for decomposition and reconstitution using filter bank are invented. Known by FWT, those algorithms decrease enormously the consumed time of decomposition and reconstitution steps.
The four subbands (HL, LH, HH, LL) computed by applying an analyse with wavalets and scaling functions in previous, are arise from separable applications of vertical and horizontal filter bank. The filters h and g shown in Fig 7 are one-dimensional low pass filter and high pass filter, respectively. To obtain the next coarse level of wavelet coefficients, the subband LL alone is further decomposed and critically sampled using similar filter bank shown in Fig 7. These results in two-level wavelet decomposition. Similarly, to obtain further decomposition, the second LL will be used. This process continues until some final scale is reached. The decomposed image can be reconstructed (i.e., inverse DWT) using a reconstruction or synthesis filter bank shown in Fig 8.[2] This figure present the principle of the FWT of an image.[2]

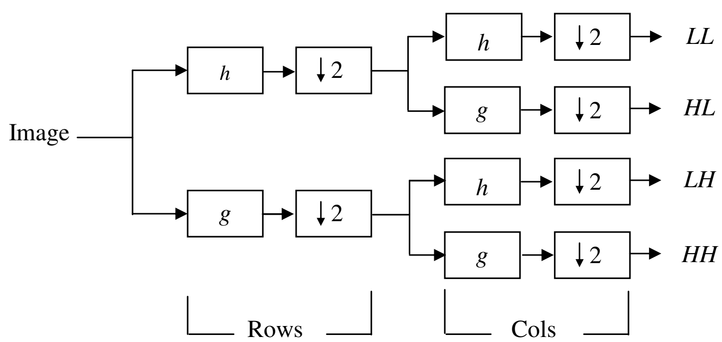

Fig. 7. Rapid decomposition of an image with filter bank.

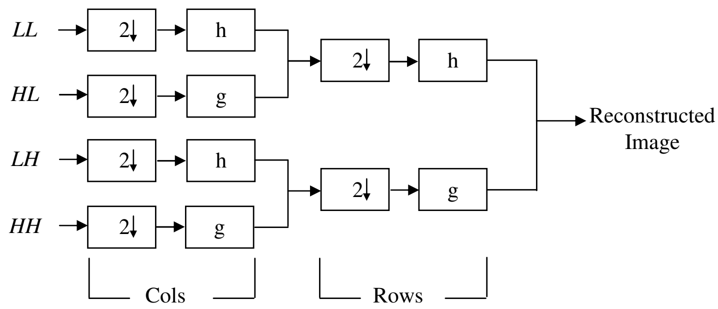

Fig. 8. Rapid reconstruction of an image with filter bank.

3. Learning Wavelet Network using 2DFWT

In this study, we used the FWT to achieve the decomposition process of a given signal, and the inverse wavelet transform (Eq. 15) to reconstruct it.

In this section we will show how we can learn a wavelet network only using the FWT technique. Firstly, the proposed learning algorithm are presented. Secondly, we detail how we create the library of wavelets and scaling functions. Third, we presents the calculation of the weights. Fourth, the optimization of the function nodes and finally Computing the corresponding weight of the optimized function.

3.1. *Proposed learning algorithm*

Step 1: Building a library of candidate wavelets and scaling functions.[40,17,11,12] This step Includes the following items:

(1) Choose the mother wavelet covering all the image to approximate.
(2) Build a library that contains wavelets and scaling functions to be used.

Let's nominate those functions g_l with $l = 1 \cdots M \times N$.
Step 2: An error E between the signal f and the output of the network is used as a stop learning condition,.
Step 3:Set a signal s equal to the signal f to be learned and initialize the output network to $\tilde{f} = 0$.
Step 4: Apply a series of FWT to the signal s using the dual set $(\tilde{h},\tilde{g})$ filters in order to calculate the coefficients α_l corresponding to the g_l functions of the library. s can be written :

$$s = \sum_{l=1}^{M\times N} \alpha_l g_l \tag{3.1}$$

Step 5: Compute the contribution $\alpha_l g_l$ of all the activation functions in the library to reconstruct the signal s.
Step 6: Select from the library the function g_k(k is the number of the selected function) that contributes more in the reconstruction of the signal s.

Step 7: Optimize the parameters of the g_k function (dilation, position and rotation) to better approximate the signal s and improve the accuracy of the output network.
Step 8: The gotten optimized function g'_k doesn't belong any more to the library, we compute the corresponding weight α'_k(detailed in next section).
Step 9: Add the function g'_k to the hidden layer of the network that approximate f and set its corresponding weight to α'_k. The output of the network is $\tilde{f} = \tilde{f} + \alpha'_k g'_k$
Step 10: Optimizing the wavelet network architecture: The function g'_k not necessarily forms a basis with the previous added functions to the hidden layer. In this case g'_k forms a frame with those functions, a process to reduce the hidden nodes number and up to date the weights of the network is introduced (detailed in section 3.6).
Step 11:Compute the residual signal $s = f - \tilde{f}$ and return to the Step 4 if you didn't come to the end of the learning.

3.2. *Creation of the library of wavelets and scaling functions*

A sampling on a dyadic grid of dilation and translation parameters is proceded to the mother wavelet and the corresponding scaling functions, in order to construct the library of candidate functions to join to our wavelet network.

The content of this library depends on the level of MRA proceded: the first level contains an equal number of scaling functions, vertical wavelets, horizontal wavelets and diagonal wavelets. The total number of those functions is $M * N$ (which is the size of the signal). In the second level, the scaling functions of the first scale are replaced by functions more dilated (twice than the precedent scale). We have in the library $\frac{3}{4}(M * N)$ of wavelets of the first level, $\frac{3}{4} * \frac{(M*N)}{4}$ of wavelets of the second level and $\frac{1}{4} * \frac{(M*N)}{4}$ of scaling functions of the second level. We can stop the sampling at any scale k with $k \leq m$(eq.4) but we must complete the library by the corresponding scaling functions of the last scale (Fig. 9).

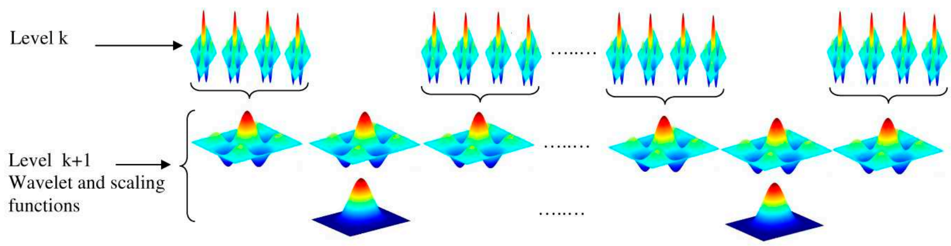

Fig. 9. The library wavelets and scaling functions after k MRA scales.

3.3. *Calculation of the weights*

To compute the weights corresponding to the activation functions that belong to the library, we analyze k (k is the number of the scales) times the signal f with the dual filters ($\tilde{h}$,$\tilde{g}$) of the scaling function $\tilde{\phi}$ and the wavelet $\tilde{\psi}$. The result is a sum of signals W_j^i corresponding to the weights of the wavelets of the scales j ($j = 1 \cdots k$) and orientations i ($i = H, V, D$) and the weights V^k of the scaling functions of the scale k.

3.4. *Optimization of the function nodes*

To increase the efficiency of the network, we optimize the selected function g_k of the step 6 by the search of his translations, dilation and rotation, which provide the best approximation of the residual signal s. Knowing that the optimization is easier when the analytical expression of the function to optimize is known, we opted for the Beta wavelets and scaling functions as activation functions. These functions have the distinction of having, at the same time, analytical expressions and wavelet filter bank.[4] This characteristic motivates the use of these functions in our algorithm. In addition, they have given very interesting results in many works.[4,32,40,17] More details about Beta wavelets and their caracteristics are given in.[4,30,40,41] An exemple of beta wavelets and the associated scaling function is represented in figure 10.

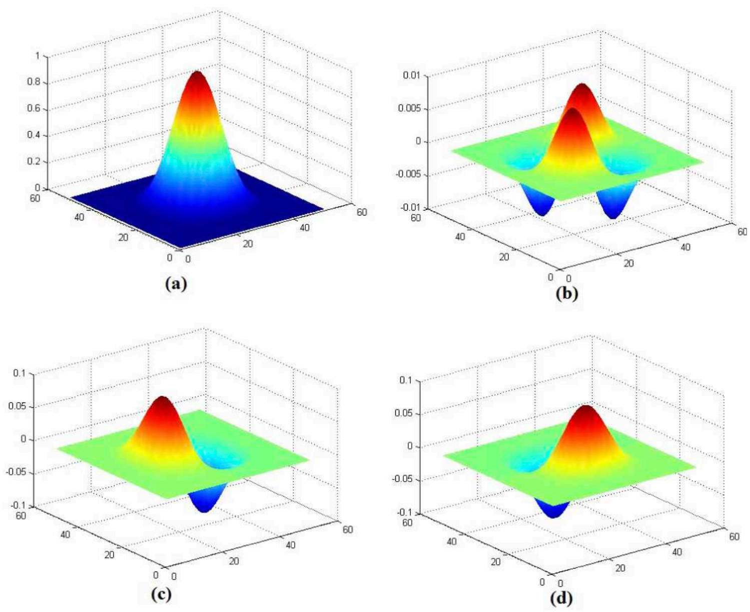

Fig. 10.2D Beta wavelets and associated scaling function(a):ϕ,(b):ψ^D,(c):ψ^H,(d):ψ^V.

Several optimization methods exist. In this work, we choose the method of

Levenberg-Marquardt. We optimize the selected function g_k. Indeed, the parameters of this function are used as initialization of this algorithm. The result is a function g'_k that performs better the output of the network and approximates the best the residual signal s.

At a given iteration k, to get the activation function g'_k, we optimize the dilation parameter a , the position parameters b_x and b_y and the rotation parameter θ of the function g_k by minimizing the function:

$$E_k = \min_{a,b_x,b_y,\theta}(\sqrt{(f - \tilde{f} - \alpha_k g_k)^2}) \tag{3.2}$$

3.5. *Computing the corresponding weight of the optimized function*

When we select and optimize a function from the library of the wavelet and the scaling functions to use it in the hidden wavelet network layer, we compute its corresponding weight by the flowing way:

(1) We have

$$s = \alpha_1 g_1 + \cdots + \alpha_k g_k + \cdots + \alpha_L g_L \tag{3.3}$$

(2) We decompose by FWT the signal g'_k

$$g'_k = \gamma_1 g_1 + \cdots + \gamma_k g_k + \cdots + \gamma_L g_L \tag{3.4}$$

(3) Because $\gamma_k \neq 0$ we can write (Proof. Details are provided in Appendix)

$$g_k = \frac{1}{\gamma_k} g'_k - \frac{\gamma_1}{\gamma_k} g_1 - \cdots - \frac{\gamma_L}{\gamma_k} g_L \tag{3.5}$$

In the equation (3.3) we replace g_k by the equation (3.5)

$$\begin{aligned} s &= \alpha_1 g_1 + \cdots + \alpha_k (\frac{1}{\gamma_k} g'_k - \frac{\gamma_1}{\gamma_k} g_1 - \cdots - \\ &\quad \frac{\gamma_{k-1}}{\gamma_k} g_{k-1} - \frac{\gamma_{k+1}}{\gamma_k} g_{k+1} - \cdots - \frac{\gamma_L}{\gamma_k} g_L) \\ &\quad + \cdots + \alpha_L g_L \\ &= (\alpha_1 - \alpha_k \frac{\gamma_1}{\gamma_k}) g_1 + \cdots + \frac{\alpha_k}{\gamma_k} g'_k + \cdots + \\ &\quad (\alpha_L - \alpha_k \frac{\gamma_L}{\gamma_k}) g_L \end{aligned} \tag{3.6}$$

We deduce the weight α'_k of the activation function g'_k

$$\alpha'_k = \frac{\alpha_k}{\gamma_k} \tag{3.7}$$

3.6. *Optimizing the wavelet network architecture*

The purpose of this section is to deduce the number of functions in the hidden layer of the network. The idea is that after optimizing the parameters of the selected function, the result g'_k is not necessarily that it forms a base with those of the hidden layer. In this case,the added activation function g'_k to the hidden layer forms a frame with the previous added ones, it can be expressed by a linear combination of those existing functions. This process reduces enormously and compresses the network size. Let's suppose that we are at the second iteration of the learning and that the first two added functions are linearly independent, the output of the network is then:

$$f = \alpha'_1 g'_1 + \alpha'_2 g'_2 \tag{3.8}$$

At the third iteration, the approximation of the signal f is given by

$$f = \alpha'_1 g'_1 + \alpha'_2 g'_2 + \alpha'_3 g'_3 \tag{3.9}$$

Assuming that in this iteration, the functions are not linearly independent (the added function depends on the first tow ones), f can be written:

$$\begin{aligned} f &= \alpha'_1 g'_1 + \alpha'_2 g'_2 + \alpha'_3(\vartheta_1 g'_1 + \vartheta_2 g'_2) \\ &= (\alpha'_1 + \alpha'_3\vartheta_1) g'_1 + (\alpha'_2 + \alpha'_3\vartheta_2) g'_2 \end{aligned} \tag{3.10}$$

we can notice that the function g'_3 was not used as an actication function in the hidden layer, but to update the connection weights of the previous iteration. the α'_1 is updated to $(\alpha'_1 + \alpha'_3\vartheta_1)$ and the α'_2 is updated to $(\alpha'_2 + \alpha'_3\vartheta_2)$.
This example clearly shows that a selected function at a given iteration, will be used as a kernel function when it forms a base with the functions of the hidden layer previously selected. Otherwise, it will update the connection weights. It is easy to verify that the function g'_k forms a basis with the kernel functions of the network. It is just enough that its verifies the condition $< \tilde{f}, g'_k > \neq 0$. The method was proposed fistly by Kruger in his article [23] and it is more detailed in.[40]

4. Face recognition approach

In the state of the art, 2D face recognition techniques are sensetive to pose and luminance variations. A number of approaches have been proposed to explicitly handle pose and luminance variations.[21] In this work we present our solution to overcome these limitations by: Projecting the image in the wavelet space which is less sensetive to luminance changes and exploiting the morphological caracteristics of the transfer functions (wavelet and scaling functions). In fact, a wavelet network approximating a given face can also approximate the same face with another pose by varying the translatin, dilation and rotation parameters. In this section we present our contribution in the recognition stage. For each probe image p to recognize, we modify the wavelet networks aleady stored in the training stage by optimizing the kernel functions parameters, in order to approximate this test image. Our method is composed by the folowing steps:

Step 1: The parameters (a, b_x, b_y, θ) of the transfer functions of every training wavelet network (α, g) are modified in order to approximate the image p. Those optimizations are made by using the Levenberg Marquardt method. This results on the creation of new wavelet networks (α, g')
Step 2: The α are no more the corresponding weights of the g' functions, so the new α' connections weights of the obtained networks are recomputed by the 2DFWT technique.
Step 3: Every network (α', g') is compared to its origin (α, g) by computing the Euclidian distances.
Step 4: We compare then these distances. Generally, the smallest one concerns the searched person.

5. Experimental results

In this section we present the results of our face recognition experiment using the ORL face database of the Cambridge AT&T Laboratories and the FERET ones. To evaluate better the performances of our approach FBWN2D, we compared it to other methods of face recognition (neural networks (RBF), PCA, LDA, EBGM).

5.1. *Experiments on ORL face database*

This database contains 10 different images of each of 40 distinct subjects taken between April 1992 and April 1994 at Olivetti Research Laboratory, Cambridge, UK. For some subjects, the images were taken at different times, varying the lighting, facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). All the images were taken against a dark homogeneous background with the subjects in an upright, frontal position (with tolerance for some side movement). The images are in '.pgm' format and of dimension 92 x112 (width x height) 8-bit gray levels (Fig. 11).

In the next figure we present the results using the ORL dataset, we can see that our approach gives the better recognition rate in comparison with other techniques as PCA, RBF, EBGM and LDA which gives the less performances.

5.2. *Experiments on FERET face database*

For more evaluation, we used the standard FERET data set including the data partitions (subsets) for recognition tests. The gallery consists of 14 126 images and there are four subsets of probe images (fb, fc, dup1, and dup2) that are compared to the gallery images in recognition stage.[1] Examples of images from FERET dataset are presented in figure 13.

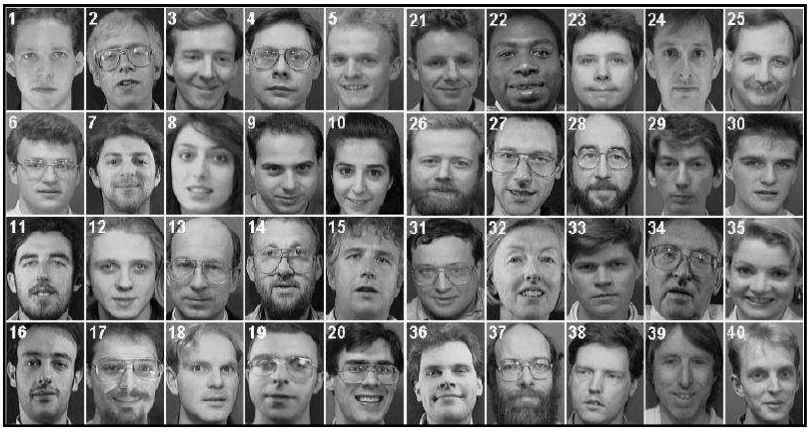

Fig. 11. Subjects of the ORL dataset.

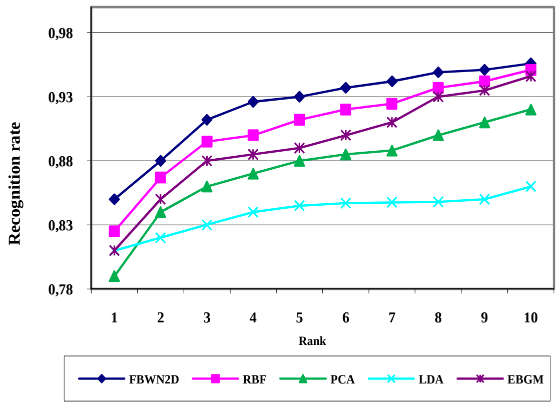

Fig. 12. Recognition rate of the ORL database.

Fig. 13. Examples of images from FERET Dataset.

Every subject from this dataset have many sessions, by varying the expressions, luminance conditions and pose. I the next table we present the number of images of each subset.

Table 1. Description of the different subset from FERET Dataset.

Subset Name	Size of Training Subset	Size of Test Subset
fb	1195	1196
fc	194	1196
duplicate I	722	1196
duplicate II	234	864

The next figures shows the recognition rates of our approach in comparaison with other appoachs sited above, using the four subsets from the FERET dataset. Our approach gives for all subsets the best rates, specially in the rank 1 which is he most secure. For tha others ranks, we have less security, more rate of false acceptance and better rates of recognitions.

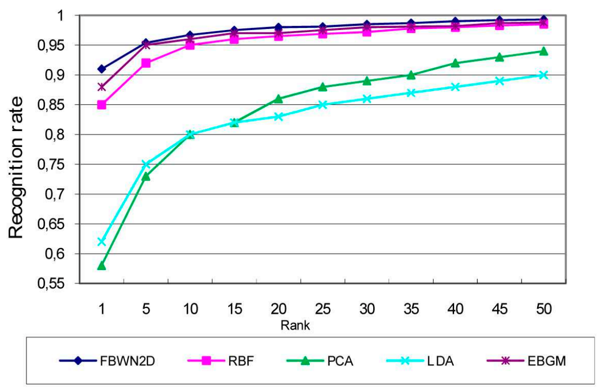

Fig. 14. Recognition rate of the fb set.

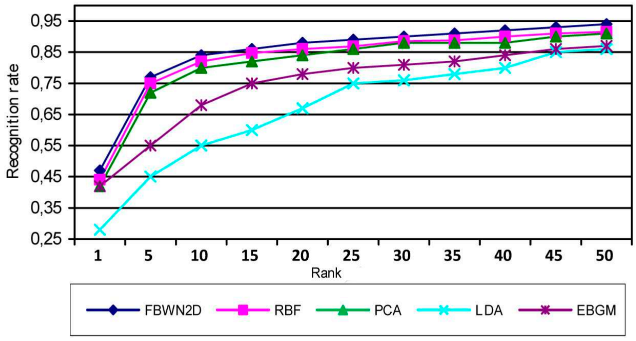

Fig. 15. Recognition rate of the fc set.

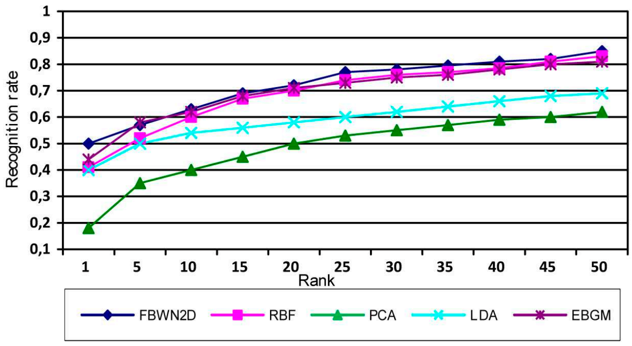

Fig. 16. Recognition rate of the duplicate I set.

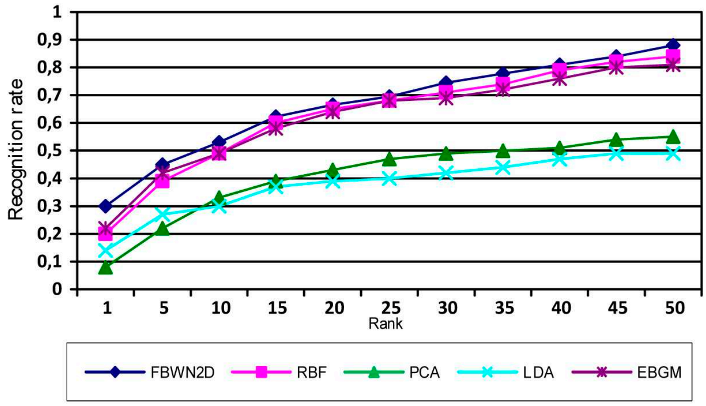

Fig. 17. Recognition rate of the duplicate II set.

6. Conclusion

The approach used in this work to recognize faces by numeric vision is based on the Beta wavelets networks. An algorithm of training of these networks based on the 2D fast wavelet transform and multiresolution analysis theory has been proposed and has been implemented. Our method of face recognition consists in two main stages: The training stage which has as object to optimize a wavelets network for every training face using the proposed algorithm. The second is the recognition stage. It's based on the capacity of stored training networks to approximate the face to recognize. Experiments on two dataset (ORL and FERET) are done to evaluate the efficiency of our proposed approach. The performances of the Beta wavelets networks used for face recognition are clear and the results obtained are promoters. The robustness and the rapidity of the proposed training algorithm that are based on 2D fast wavelet transform theory increased these performances. We are looking actually to extend this learning algorithm to 3D case and see its performance in 3D face recognition.

Acknowledgment

The authors would like to acknowledge the financial support of this work by grants from General Direction of Scientific Research (DGRST), Tunisia, under the ARUB program

Appendix A. Proof of $\gamma_k \neq 0$

proof. Assume $\{g_k\}_k$ is a bases,then we can represent f with a unique manner as

$$f = \alpha_1 g_1 + \cdots + \alpha_k g_k + \cdots + \alpha_L g_L \tag{A.1}$$

Since the g'_k is the optimized of g_k, based on (24), it is also obvious that f can be represented as

$$f = \alpha'_1 g_1 + \cdots + \alpha'_{k-1} g_{k-1} + \alpha'_k g'_k + \alpha'_{k+1} g_{k+1} + \cdots + \alpha'_L g_L \tag{A.2}$$

since $\alpha_k g_k$ contribute the most in the reconstruction of the signal f, so this means that $\alpha_k \neq 0$
let suppose that $\gamma_k = 0$. Substituting formula (24) into (A.2), it yields

$$f = \alpha'_1 g_1 + \cdots + \alpha'_{k-1} g_{k-1} + \alpha'_k(\gamma_1 g_1 + \cdots + \gamma_k g_k + \cdots + \gamma_L g_L) + \alpha'_{k+1} g_{k+1} + \cdots + \alpha'_L g_L \qquad (A.3)$$

since $\gamma_k g_k = 0$, formula (A.3) is represented as

$$f = (\alpha'_1 + \alpha'_k \gamma_1) g_1 + \cdots + (\alpha'_{k-1} + \alpha'_k \gamma_{k-1}) g_{k-1} + (\alpha'_{k+1} + \alpha'_k \gamma_{k+1}) g_{k+1} + \cdots + (\alpha'_L + \alpha'_k \gamma_L) g_L \qquad (A.4)$$

It follows from (A.1) and (A.4) that $\alpha_k = 0$, so this is contradictory to what is mentioned above that $\alpha_k g_k$ contribute the most in the reconstruction of the signal f which implies that $\alpha_k \neq 0$. Hence, the supposition is false which yields to conclude that $\gamma_k \neq 0$

References

1. V. Akalin , Face recognition using eigenfaces and neurak networks, Master of science. The graduate school of natural and applied sciences. The Middle East technical university. Ankara, Turkey December 2003.
2. S. Arivazhagan, S. Deivalakshmi, K. Kannan, B.N. Gajbhiye, C. Muralidhar, Sijo N. Lukose, M.P. Subramanian, *Multi-resolution system for artifact removal and edge enhancement in computerized tomography images*, Pattern Recognition Letters 28 (2007) 1769-1780
3. W. Bellil ,C. Ben Amar and A.M. Alimi, *Beta wavelet networks for function approximation*, 7th International Conference on Adaptative and Natural Computing Algorithms ICANNGA'05, Coimbra, Portugal, March 21-23, 2005, Riberio et al. (eds), Springer Computer Science, SpringerWien NewYork, vol. 1, pp 18-21.
4. C. Ben Amar, M. Zaied and M. A. Alimi,*Beta wavelets. Synthesis and application to lossy image compression, Journal of Advances in Engineering Software*, Elsevier Edition, Vol. 36, N7, PP 459 - 474. 2005.
5. C. Ben Amar and O. Jemai *Wavelet Networks Approach for Image Compression". ICGST: The International Congress for global Science and Technology*, GVIP International Journal on Graphics, Vision and Image Processing, GVIP Special Issue on Image Compression, pp. 15- 23, 2007.
6. L. ChangGyoon, K. Kangchui, K. Eungkon,*Modeling for an adaptive wavelet network parameter learning using genetic algorithms*, in: Proceedings of the Fifteenth IASTED International Conference on Modeling and Simulation, CA, USA, 2004, pp. 55-59.
7. L. Chunguang, W. Peng, E. Javier,*Complex-valued wavelet network*, J. Comput. System Sci. 67 (3) (2003) 623-632.
8. I. Daubechie,*Ten Lectures on Wavelets*, CBMS-NSF Regional Conferences Series in Applied Mathematics, SIAM (Capital City Press, Montpellier, Vermont), 1992.

9. C. Daniel, O. Umit, *Wavelet neural networks: a design perspective*, in: IEEE International Symposium on Intelligent Control-Proceedings, 1994, pp. 376-381.
10. M. Ehler, *Compacity supported multivariate*, pairs of dual wavelet frames obtained by convolution, Int. J. Wavelets Multiresolut. Inf. Process. 6(2) (2008) 183-208.
11. R. Ejbali, M. Zaied, and Chokri Ben Amar, *Intelligent approach to train wavelet networks for Recognition System of Arabic Words*, International Joint Conference on Knowledge Discovery KDIR'10, 25-28 October, Valencia-Spain.
12. R. Ejbali, M. Zaied, and C. Ben Amar, *Wavelet network for recognition system of arabic word*, International Journal of Speech Technology, vol. 13, pp. 163-174, 2010.
13. C. Hu, R. Feris and M. Turk, *Real-time View-based Face Alignment using Active Wavelet Networks*, Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp 215, 2003.
14. H. Hu, P. Zhang, Z. Ma, *Direct kernel neighborhood discriminant analysis for face recognition*, Pattern Recognition Letters 30 (2009) 902-907.
15. D. V. Jadhav, R S. Holambe, *Feature extraction using Radonand wavelet transforms with application to face recognition*, Neurocomputing 72 (2009) 1951-1959
16. O. Jemai, M. Zaied, C. Ben Amar and A.M. Alimi, *Pyramidal Hybrid Approach: Wavelet Network with OLS Algorithm Based-Image Classification*, International Journal of Wavelets, Mutiresolution and Information Processing, Vol. 9, No. 1, pp. 1-20, (2011)
17. O. Jemai, M. Zaied, C. Ben Amar and A. M. Alimi, *FBWN: an architecture of Fast Beta Wavelet Networks for Image Classification*, 2010 IEEE World Congress on Computational Intelligence (IEEE WCCI 2010), the 2010 International Joint Conference on Neural Networks (IJCNN 2010), p. 1953-1960, July, 18-23, 2010 - CCIB, Barcelona, Spain.
18. X. Jinhua, W.C. Ho. Daniel, *A constructive algorithm for wavelet neural networks*, First International Conference on Natural Computation, Changsha, China, 2005, pp. 730-739.
19. Z. Jun, G.G. Walter, Y. Miao, *Wavelet neural networks for function learning*, IEEE Trans. Signal Process. 43 (6) (1995) 1485-1496.
20. C. Kai, H. Yoshio, *An adaptive hybrid wavelet neural network and its application, in: Proceedings of the 2004 IEEE International Conference on Robotics and Biomimetics*, IEEE ROBIO 2004, Shenyang, China, 2004, pp. 73-86.
21. W. C. Kao, M. C. Hsu, Y. Y. Yang, *Local contrast enhancement and adaptive feature extraction for illumination-invariant face recognition*, Pattern Recognition.
22. V. Kruger, *Gabor wavelet networks for object representation*, Ph. D thesis, 2001.
23. V. Kruger and G. Sommer, *Affine real-time face tracking using gabor wavelet networks*, IEEE Computer Society Pattern Recognition, Barcelona, Spain, Sept. 3-8, 2000.
24. R. Ksantini, B. Boufama, D. Ziou, B. Colin, *A novel Bayesian logistic discriminant model: An application to face recognition*, Pattern Recognition (2009), doi:10.1016/j.patcog.2009.08.021.
25. 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.
26. H. Min, C. Baotong, *A novel learning algorithm for wavelet neural networks,*

in: First International Conference on Natural Computation, ICNC, Changsha, China, 2005, pp. 1-7.

27. M. Othmani, W. Bellil, C. Ben Amar and A. M. Alimi, *A New Structure and Training Procedure for Multi-Mother Wavelet Networks*, nt. J. Wavelets Multiresolut. Inf. Process., IJWMIP, vol. 8, issue 1, 2010, pp. 149-175.
28. M. Othmani, W. Bellil and C. Ben Amar,*3D Object Modelling using Multi-mother Wavelet Network*, International Conference on Neural Computation, Madeira Portugal, October 5-7, 2009.
29. Z. Qianhua, A. Benvenist,*Wavelet network*, IEEE Trans. Signal Process. 13 (6) (1992) 889-898.
30. Z. Qinghua,*Using wavelet network in non parameters estimation*, IEEE Trans. Neural Networks 8 (1997) 227-236.
31. S. Qingmei, G. Yongchao,*A stepwise updating algorithm for multiresolution wavelet neural networks, in: Proceedings of the International Conference on Wavelet Analysis and Its Applications (WAA)*, Chongqing, China, 2003, pp. 633-638.
32. S. Said, B. Ben Amor, M. Zaied, C. Ben Amar and M. Daoudi,*Fast and efficient 3D face recognition using wavelet networks*, 16th IEEE International Conference on Image Processing. pp. 4153-4156, 2009.
33. K. Seong-Joo, K. Taek-Yong, S. Jae-yong, J. Hong-Tae,*Design of the scaling-wavelet neural network using genetic algorithm*, in: Proceedings of the International Joint Conference on Neural Networks, Honolulu, HI, 2002, pp. 2174-2179.
34. C.Y. Wen, J. K. Chen,*Multi-resolution image fusion technique and its application to forensic science*, Forensic Science International 140 (2004) 217-232.
35. O. Yacine, D. Gerard,*Initialization by selection for wavelet network training*, Neurocomputing 34 (2002) 131-143.
36. 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.
37. J. Yang, Ch.Liu, L.Zhang,*Color space normalization: Enhancing the discriminating power of color spaces for face recognition*, Pattern Recognition (2009),doi:10.1016/j.patcog.2009.11.014.
38. B. Yevqeniy, L. Nataliya, P. Iryna, V. Olena,*An adaptive learning algorithm for a wavelet neural network*, Expert Systems 22 (5) (2005) 235-240.
39. L. Ying, D. Zhidong, Z. Xi,*Approximated non-tensor pre-wavelets neural network based on compound training algorithm*, J. Inform. Comput. Sci. 2 (4) (2005) 715-722
40. M. Zaied, O. Jemai and C. Ben Amar,*Training of the Beta wavelet networks by the frames theory: Application to face recognition*, The International Workshops on Image Processing Theory, Tools and Applications, IPTA'08, p. 165-170, Tunisia November 2008.
41. M. ZAIED, C. BEN AMAR and M. A. ALIMI,*Beta Wavelet Networks for Face Recognition, Journal of Decision Systems*, Lavoisier 2005 Edition, Vol. 14, N. 1-2/2005, pp. 109-122, 2005.
42. S. Zhao, Y. Gao, B. Zhang,*Gabor feature constrained statistical model for efficient landmark localization and face recognition*, Pattern Recognition Letters 30 (2009) 922-930.